





An Ensemble Learning and RUL Prediction Method Based on Bearings Degradation Indicator Construction

Qiaoping Tian 🕩 and Honglei Wang *

School of Management, Guizhou University, Huaxi District, Guiyang 550025, China; tqp616732695@163.com * Correspondence: hlwang@gzu.edu.cn

Received: 5 November 2019; Accepted: 28 December 2019; Published: 2 January 2020



Abstract: The prediction of the remaining life of a bearing plays a vital role in reducing the accident-related maintenance costs of machinery and in improving the reliability of machinery and equipment. To predict bearing remaining useful life (RUL), the abilities of statistical characteristics to reflect the bearing degradation state differ, and the single prediction model has low generalization ability and a poor prediction effect. An ensemble robust prediction method is proposed here to predict bearing RUL based on the construction of a bearing degradation indicator set: the initial bearing degradation indicator subsets were constructed using the Fast Correlation-Based Filter with Approximate Markov Blankets (FCBF-AMB) and Maximal Information Coefficient (MIC) selection methods. Through the cross-operation of the obtained subsets, we obtained a set of robust degradation indicators. These selected degradation indicators were fed into the long short-term memory (LSTM) neural network prediction model enhanced by the AdaBoost algorithm. We found through calculation that the average prediction accuracy of the proposed method is 91.40%, 92.04%, and 93.25% at 2100, 2250, and 2400 rpm, respectively. Compared with other methods, the proposed method improves the prediction accuracy by 1.8% to 14.87% at most. Therefore, the method proposed in this paper is more accurate than the other methods in terms of RUL prediction.

Keywords: remaining useful life prediction; three-stage feature selection; degradation indicator; LSTM-AdaBoost prediction model

1. Introduction

Rolling bearings are one of the key components supporting rotating shafts in rotating mechanical equipment. Bearing failure is often considered one of the most common causes of mechanical equipment failure [1,2]. Bearing reliability is critical for the reliability, durability, and efficiency of mechanical equipment. Any accidental failure of a bearing may cause have various negative effects [3,4] ranging from production downtime to casualties or even catastrophic environmental pollution. To address these issues, online detection of bearing health is urgently required to effectively enhance the safety of mechanical equipment operation [5–7], predict bearing remaining useful life (RUL), and to implement an action plan to prevent catastrophic events and extend the bearing life cycle [7]. Advances in bearing RUL prediction technology have provided increasingly powerful technical support for intelligent bearing RUL prediction and health management [8–10]. In the past few decades, the research has achieved theoretical results that have been widely applied. Most bearing RUL prediction methods apply a model-based or data-driven approach [11,12].

The model-based method mainly relies on an accurate mathematical model of bearing degradation, but bearing degradation is a complicated and difficult problem [13]. The data-driven approach uses data mining and artificial intelligence [14] to explore the potential relationship between current bearing

state data and RUL. Data-driven methods have become a promising approach. Data-driven analysis methods can be used as objective and rational tools to understand the data and make decisions [15,16]. For example, in the field of life sciences [17], the data-driven approach was used to conduct diagnosis [18,19]. In the field of engineering applications, the data-driven method has been used to obtain information on road lighting infrastructure. Based on the feature selection technology of supervised and unsupervised filters, the dimension of feature space was reduced to classify and identify lamps, which was ultimately used to evaluate and optimize the performance of the street lighting at night [20]. Some experts used ensemble data-driven statistical models to map comparative shallow landslide susceptibility to obtain the relationship between heavy rain and shallow landslides [21].

The deep learning approach [22] has advantages for bearing RUL prediction [6], providing new opportunities for this research field [10,23]. A typical deep learning framework consists of four phases: data acquisition and processing, feature extraction and calculation, learning model building, and prediction. In today's big data era, the premise of accurate bearing RUL prediction is to extract as much effective information as possible from massive amounts of monitoring data [24]. However, the data are increasingly complicated and high-dimensional. The irrelevant and redundant features in these high-dimensional data increase the complexity of the learning model, and can even reduce the prediction accuracy, which creates the problem known as "dimensional disaster" [25]. In the feature extraction and calculation stage, deep learning has some shortcomings: the time-domain features are less able to reflect the bearing degradation process details, the frequency-domain features are not sensitive to medium-term bearing degradation, and the time frequency-domain characteristics of wavelets can cause information loss. These three problems usually lead to information redundancy and increases the neural network nodes, which in turn leads to difficult training and over-fitting of the mode [6]. In this process, the traditional algorithm mainly finds a set of features with high contribution rate. Some authors [26–28] defined different feature types based on the contribution of features to degradation information (DI). In this case, feature correlation is a measure of the degradation-stage-related information.

A feature that does not contain information about the bearing DI is considered insignificant and therefore unnecessary for the prediction task. Removing such features can improve the prediction model and speed up the learning algorithm. Conversely, the relevant features are those that can reflect bearing DI. To minimize the prediction error, it may not be necessary to select all relevant features, but instead only select the feature subset with the highest contribution rate and the strongest prediction ability. Feature subsets with these properties may not be unique due to redundancy effects. Redundancy is usually measured by feature correlation; if the values of two features are relevant, then they are redundant.

With the feature selection method, the representative feature subset is selected, the features with a high contribution rate and sensitivity that are favorable for prediction are retained, and the complete set is replaced to construct and train the learning model. Experts and scholars have studied this field, especially based on artificial intelligence and statistical methods, using feature compression methods or similar monotonic methods [6,10]. The optimal feature selection method should not only reduce data dimensions, but also eliminate redundant and irrelevant features. Therefore, considering correlations in feature selection plays a crucial role in reducing data dimensions [29]. However, in the construction of feature subsets, only relying on a single correlation or sensitivity measurement method will bias the calculation results to some extent, which will reduce the robustness of the feature subsets. Therefore, we aimed to use the three-stage feature selection method to extract sensitive features and construct a bearing degradation indicator set. Based on two different feature extraction methods, the initial subsets of bearing degradation indicators were constructed, and the cross-operation of these subsets was applied to obtain the robust set that can fully reflect bearing degradation information.

In the deep learning model establishment and prediction stage, scholars introduced a bearing RUL prediction method based on a recurrent neural network (RNN). However, for practical problems,

gradient disappearance occurs. Hochreiter et al. [30] proposed a long short-term memory (LSTM) model in 1997 to overcome the RNN problem of gradient disappearance.

Some experts have proposed using the LSTM neural network to predict bearing RUL based on the bearing degradation bottleneck feature, waveform entropy (WFE) indicator, time factor, or based on the deep feature representation method [26,31–33]. Compared with the previous artificial intelligence algorithms, the predictive ability of the LSTM significantly improved. The above research used a single artificial intelligence algorithm to predict bearing RUL; however, the single artificial intelligence algorithm has weaker generalization and low robustness. The bearing RUL cannot be predicted well outside the sample. To address this problem, we wanted to enhance the LSTM neural network prediction model using the AdaBoost algorithm.

To overcome the aforementioned shortcomings, we propose an ensemble robust prediction method to predict bearing RUL based on the construction of a bearing degradation indicator set.

The main contributions of this paper are summarized as follows:

- To reveal the state of bearing degradation more fully, we integrated the selected high contribution rate and sensitive features to form a more representative and robust feature set, defined as the bearing degradation indicator set.
- (2) To ensure the robustness of the constructed set of bearing degradation indicators, a new framework for three-stage feature selection is proposed for bearing RUL prediction, which more comprehensively considers the correlation between features and bearing degradation state.
- (3) The AdaBoost algorithm is proposed to enhance the prediction ability, the prediction accuracy, and the generalization ability of the LSTM prediction model.

The rest of the paper is organized as follows: Section 2 introduces the basic LSTM prediction model theory and two kinds of feature selection methods. Section 3 presents the detailed implementation process of this three-stage feature selection method that was applied to the construction of a bearing degradation indicator set and an improved LSTM-AdaBoost prediction model. The performance of the proposed method was verified using the XJTU-SY bearing datasets from Xi'an Jiaotong University (XJTU, Xi'an, China) and compared with other methods in Section 4. Finally, conclusions are drawn in Section 5.

2. Basic Theory and Algorithm

Based on the three main problems experienced: the feature correlation measurement standard in the feature extraction and calculation process, the computational complexity in the predictive modelling process, and the generalization ability of the prediction model, this section lists the relevant theories that can be used to solve these problems.

First, the initial reference degradation indicator subset F^* was screened by the fast correlation-based filter (FCBF) solution and approximate Markov Blanket to construct an initial subset of reference degradation feature indicators that can characterize the bearing degradation process. Secondly, the maximum information coefficient (MIC) was used to measure the correlation between features and features, as well as the correlation between features and bearings degradation state, to construct the initial reference degradation indicator subset F_{F-R} with maximum correlation between features and real RUL and the subset F_{F-F} with minimum redundancy between features. Thirdly, cross-operation was adopted for the initial reference bearing degradation indicator subsets to reduce the computation load, shorten the training time of prediction model, and reduce the computational complexity of the prediction modelling process. The reason for choosing different correlation measurement methods to construct the bearing degradation indicator subsets was to avoid the single correlation measurement method being affected by outliers in the data set, resulting in bias of the constructed bearing degradation indicator set and affecting the prediction accuracy.

The results obtained using different correlation measurement methods were cross-operated to retain the effective indicators to the maximum extent. Finally, the AdaBoost algorithm was used to

enhance the prediction model of LSTM neural network, and multiple weak predictors were assembled into a strong predictor to predict the bearing remaining useful life.

2.1. LSTM

Recurrent neural network (RNN) is a type of neural network dedicated to processing time-series data samples. Each layer of its output is output to the next layer and to a hidden state, which is used by the current layer when processing the next sample, as shown in Figure 1. Module M of the RNN reads the input $x^{(t)}$ and obtains the output $h^{(t)}$. Circulation is used to complete the transfer to the next step of information from the current step.



Figure 1. Recurrent neural network (RNN) expansion structure.

The above chain network structure reveals that RNN is essentially sequence-dependent. However, in practical applications, problems of gradient disappearance and gradient explosion occur. To solve these problems faced by RNN, Hochreiter et al. [30] constructed a LSTM architecture that involves a memory cell. This model resembles a standard RNN with a hidden layer. Each repeating module has a simple tanh layer. The LSTM has the same structure, but the only difference is that the structure inside each module is different, each node in the ordinary hidden layer is replaced by a storage units. The specific structure [34] of the model is shown in Figure 2. This structure ensures the RNN model has the long short-term memory in the form of weights and ephemeral activations.



Figure 2. The architecture of a long short-term memory (LSTM) memory cell.

 $x^{(t)}$ is the input vector at the current time, $h^{(t-1)}$ is the hidden layer state value of the previous time(t-1), and the memory unit is the memory of the neuron state, which is used to record the current time state. The forget gate in the LSTM decides what information is retained or forgotten. The forgetting gate is calculated by the sigmoid function. The input gate decides whether to update

the state of the LSTM using the current input; the output gate decides whether to pass on the hidden state to the next iteration.

$$\begin{split} g^{(t)} &= & \tanh\left(W^{g^X}x^{(t)} + W^{g^h}h^{(t-1)} + b_g\right), \\ i^{(t)} &= & \sigma\left(W^{i^X}x^{(t)} + W^{i^h}h^{(t-1)} + b_i\right), \\ f^{(t)} &= & \sigma\left(W^{f^X}x^{(t)} + W^{f^h}h^{(t-1)} + b_f\right), \\ o^{(t)} &= & \sigma\left(W^{o^X}x^{(t)} + W^{o^h}h^{(t-1)} + b_o\right), \\ s^{(t)} &= & g^{(t)} \times i^{(t)} + s^{(t-1)} \times f^{(t)}, \\ h^{(t)} &= & \tanh\left(s^{(t)}\right) \times o^{(t)}. \end{split}$$

where *W* and *b* values are the layer weights and biases, respectively; σ and *tanh* represent the sigmoid activation function and hyperbolic tangent activation function, respectively; $x^{(t)}$ and $h^{(t-1)}$ are the input layer and hidden layer at time *t*, respectively; $g^{(t)}$, $i^{(t)}$, $f^{(t)}$, and $o^{(t)}$ are the output values of the input node, the input gate, the forget gate, and the output gate, respectively; and $s^{(t)}$ is an internal state at the current time.

2.2. Feature Selection

To reduce the computational burden and improve the prediction accuracy, it is necessary to select the sensitive features of the bearing degradation indicators that clearly represent the bearing degradation state information, and eliminate the irrelevant or redundant features that are useless or even affect the prediction accuracy of bearing RUL [35]. In this paper, we propose a three-stage feature selection method based on FCBF-AMB and MIC, which reduces feature redundancy and reduces feature data dimension based on the bearing degradation indicator subsets fusion method.

2.2.1. FCBF Feature Selection Method and Markov Blanket

The fast correlation-based filter (FCBF) solution feature selection method is based on the idea of significance and adopts the backward sequential search strategy to find the feature subset quickly and effectively. Symmetrical uncertainty (SU) was used as a correlation metric to select symmetrical features and remove redundant features [36].

Calculate the symmetric uncertainty of each feature:

$$SU(f_i, R) = 2\left[\frac{IG(R|f_i)}{H(R) + H(f_i)}\right]$$

where H(R) and $H(f_i)$ represent the information entropy of the real RUL value R and feature f_i , respectively [37]; $IG(R|f_i)$ represents the information gain (*IG*) and measures the reduction in uncertainty about the real RUL value R given the value of feature f_i .

Given a threshold value λ , if $SU(f_i, R) \ge \lambda$, f_i is a strongly correlated feature for the real RUL value R, it should be retained or deleted otherwise.

In this paper, symmetric uncertainty SU in the FCBF feature selection method is adopted as the metric standard to approximate the Markov Blanket. We applied approximate Markov blanket [25,36] to identify and delete redundant features. Feature redundancy can be determined using the Markov blankets [38] concept. Formally, it is defined as:

Definition 1 (Markov Blankets). In the feature set F, for a given feature $f_i \in F$, if there is $f_i \perp \{F - M_i - \{f_i\}, R\} \mid M_i$, the feature subset $M_i \subset F(f_i \notin M_i)$ is the Markov Blanket of f_i .

In the above definition, \perp denotes independent and $|M_i|$ denotes conditional on M_i . In other words, the Markov Blanket condition in the definition states, where the feature set F is divided into three mutually exclusive parts: feature f_i , feature subset M_i , and feature subset $F - M_i - \{f_i\}$. These three subsets have no intersection, and the union is the feature set F. If feature subset M_i is given, the feature f_i is independent of the feature subset $F - M_i - \{f_i\}$ and the real RUL value R.

Definition 2 (Approximate Markov Blankets). For the two features f_i and $f_j(i \neq j)$, the condition of f_i being the approximate Markov blanket of f_j is: $SU(f_i, R) > SU(f_j, R)$ and $SU(f_i, R) < SU(f_i, f_j)$.

The approximate Markov Blanket (AMB) is computed by comparing the correlation between feature f_i and feature f_j , and the *SU* value of f_i and the real RUL value *R*. If the correlation *SU* between different features is large, then f_j is an AMB.

Definition 3 (Predominant feature). A feature f_i is a predominant feature of F if it does not have any approximate Markov blanket in F. Predominant features are not removed at any stage.

The process of using the AMB feature selection method to find and delete redundant features is as follows:

FCBF consists of two stages: obtaining the subset of relevant features and selecting the predominant features from the subset. A relevant feature f_i is predominant if no other relevant feature f_j exists, such that f_j is an AMB for f_i . The feature subset composed of all predominant features is the initial bearing degradation indicator subset F^* , which represents the degradation state of bearings.

2.2.2. Maximum Information Coefficient (MIC)

Reshef et al. [39] proposed the MIC theory and solution method, focusing on the linear and nonlinear metric relationships between variables, and further exploring the non-function dependencies between variables through this metric relationship. The MIC mainly uses mutual information as an indicator of the degree of correlation between variables and meshing methods are used for calculation.

Given variable $A = \{a_i, i = 1, 2, \dots n\}$ and variable $B = \{b_i, i = 1, 2, \dots n\}$, where *n* is the number of samples, the mutual information (*MI*) is defined as follows:

$$MI(A,B) = \sum_{a \in A} \sum_{b \in B} p(a,b) \log \frac{p(a,b)}{p(a)p(b)},$$

where P(a, b) is the joint probability density of *A* and *B*, and P(a) and P(b) are the boundary probability densities of *A* and *B*, respectively.

Suppose set $D = \{(a_i, b_i), i = 1, 2, \dots, n\}$ is a set of finite ordered pairs. It defines a division *G*, which is used to divide the value range of variable *A* into *x* segments and divide the value range of variable *B* into *y* segments. *G* is a grid with a size of $x \times y$. Calculate MI(A, B) within each grid partition obtained, since the same grid can be divided several ways. The maximum value of MI(A, B) under different division methods is chosen as the *MI* value of a division *G*.

The maximum mutual information formula of *D* under a division is defined as $MI^*(D, x, y) = \max MI(D|G)$, where D|G denotes data *D* are divided by *G*. The maximum information coefficient (*MIC*) uses *MI* to indicate the quality of the grid; a feature matrix is formed by maximum normalized *MI* values under different divisions. The feature matrix is defined as $M(D)_{x,y}$ and the formula is $M(D)_{x,y} = \frac{MI^*(D,x,y)}{\log \min\{x,y\}}$.

MIC is defined as: $MIC(D) = \max_{xy < B(n)} \{M(D)_{x,y}\}$, where *n* is the sample size of the sample and B(n) is a function of sample size and represents the upper limit of the grid $x \times y$. Generally, $\omega(1) \le B(n) \le o(n^{1-\varepsilon}), 0 < \varepsilon < 1$. We set $B(n) = n^{0.6}$ in the experiment [39].

Suppose feature set $F = \{f_1, f_2, \dots, f_m, R\}$. The number of features is *m* and the real RUL values are *R*. *MIC* is used to define the correlation between feature f_i and real RUL values *R* as $MIC(f_i, R)$. Similarly, *MIC* is used to define the correlation between feature f_i and feature f_j as $MIC(f_i, f_j)$. We prefer to select larger $MIC(f_i, R)$ and smaller $MIC(f_i, f_j)$ features to form a set of bearing degradation indicators.

To reduce the dimension of the bearing degradation indicator set feature data, we propose the following three-stage bearing degradation indicator set construction framework based on feature subsets fusion method.

3. Methodology

3.1. Proposed Degradation Indicator Set

The structure of the proposed bearing RUL prediction model is shown in Figure 3. The original data used for bearing RUL prediction include the bearing vibration signal. First, different features are extracted from the vibration signal data, including time-domain features and frequency-domain features. Secondly, a three-stage feature selection method is used to extract and reduce the sensitive features of the feature data to construct the indicator set for bearing degradation. Then, the most sensitive features selected in the degradation indicator set are input into LSTM-AdaBoost for RUL prediction.

This section describes the procedure for construction of the proposed bearing degradation indicator set. As shown in Figure 3, the procedure is mainly composed of three stages: feature extraction, selection of sensitive features, and construction of the degradation indicator set.



Figure 3. A flowchart of the proposed bearing degradation indicator set.

The characteristics of several subsets in a given data set can produce predictive models with similar performance, but the predictive power may be different. According to the algorithm's search strategy or sample bias, some features can be selected [40]. In general, features extracted by different feature extraction methods with similar performance are highly correlated. We assumed that the relevant features are separately calculated and extracted to ensure their independence in the search process.

FCBF is a selection algorithm that uses correlation fast filter features. The feature ranking method is adopted to delete irrelevant or weakly correlated features. This method has low time complexity,

but it cannot remove redundant features. Some experts and scholars addressed this problem by using an approximate Markov blanket de-redundancy method based on the FCBF feature selection result using the MIC as the measurement standard [25].

In this study, the FCBF-AMB feature selection method [40] was used to construct the initial bearings degradation indicator subset. To ensure the independence of the initial bearing degradation indicator subsets, we used FCBF-AMB and the MIC algorithm to extract features and form different initial bearings degradation indicator subsets, respectively. First, the subset of relevant features is obtained and arranged in descending order to identify and delete the weak correlations with irrelevant features, and to add strong correlation features to the initial feature set F' according to their SU with respect to the real RUL. The predominant feature f_i is selected from the feature set F and placed into subset F'. Next, let feature f_i become the first feature in this subset F'. By definition, feature f_i is a predominant feature for each of the remaining relevant features f_j . Check whether f_i is an AMB for f_j . If so, f_j is removed from the subset F'. Then, repeat the process until no predominant features remain in the feature set F'. Construct the initial bearings degradation indicator subset F^* . The details are outlined in Algorithm 1.

Algorithm 1 FCBF-AMB feature selection method.

Input: Original feature set $F = \{f_1, f_2, \dots, f_m, R\}$, real RUL values R, SU threshold value λ . **Output:** Initial bearing degradation indicator subset F^* .

1: for $f_i \in F$ do Calculate the symmetric uncertainty $SU(f_i, R)$ between the features and real RUL values R, 2: if $SU(f_i, R) > \lambda$ then, 3: Add feature f_i to feature subset F' and rank it in descending order, 4: end if 5: 6: end for 7: for $f_i \in F'$ do for $f_j \in F' \setminus \{f_i\}$ do 8: **if** f_i is an AMB for f_i , **then** 9: Add f_i to F^* , break; 10: end if 11: end for 12: Remove predominant feature f_i from F'. 13: 14: end for

Steps 1–6 include the process for removing the irrelevant and weakly correlated features using the symmetric uncertainty feature ordering method, to finally obtain a feature subset F' with strong correlation with the bearing degradation state. F' contains many redundant features that will be deleted in the approximate Markov blanket method in steps 7–14. The predominant feature f_i is selected from feature subset F' and deleted; predominant feature f_i is added to the initial bearing degradation indicator subset. The above process is repeated until feature subset F' is an empty subset.

After the first stage of processing obtains a smaller subset of bearing degradation indicators, two subsets of bearing degradation indicator based on the *MIC* feature selection method are constructed in the second stage.

Calculate the correlation $MIC(f_i, f_j)$ between features, the correlation $MIC(f_i, R)$ between features and real RUL values. MIC - FF refers to the matrix that can measure the correlation between features, whereas MIC - FR refers to the matrix that can measure the correlation between features and real RUL values.

We find the minimum values for each column in the MIC - FF matrix and combine these minimum values into a set $min_{FF} = \{min_{FF0}, min_{FF1}, \dots, min_{FF24}\}$, where each column corresponds to one feature, and there are 25 columns in this matrix. We find the maximum value as the FF - threshold. Then, we count the number of elements in each column that are less than the threshold value, and combine the numbers into a set $Num_{FF} = \{Num_{FF0}, Num_{FF1}, \dots, Num_{FF24}\}$, and sort

the numbers to find the median. If the number of values is greater than the median, the features corresponding to this column are weakly correlated, the more likely to be selected, they will be the elements of the bearings degradation indicator subset F_{F-F} as strong irrelevance features.

Similarly, the $MIC(f_i, R)$ values are sorted in descending order to find the median of the FR – *threshold* value in the MIC - FR matrix. Then, the values greater than the threshold, will be the elements of the bearing degradation indicator subset F_{F-R} as strongly relevant features. In the process of feature extraction for the bearing training sets, we found that the median value of $MIC(f_i, R)$ was 0.5 and the maximum value of min_{FFi} in the set min_{FF} was 0.1. We set the features' maximum value FR – *threshold* to 0.5 and the minimum value FF – *threshold* to 0.1.

The steps for obtaining the bearing degradation indicator subsets are shown in Algorithm 2.

Algorithm 2 MIC feature selection method.

Input: Original data set *D*, original feature set $F = \{f_1, f_2, \dots, f_m, R\}$, real RUL value *R*. **Output:** Initial bearing degradation indicator subset F_{F-F} , subset F_{F-R} . 1: for $f_i \in F$ do, Calculate maximum information coefficient $MIC(f_i, f_j)$, obtaining the MIC - FF matrix, 2: for Every value in every column of the *MIC* – *FF* matrix do, 3: Find the minimum values and obtain the set $min_{FF} = \{min_{FF0}, min_{FF1}, \dots, min_{FF24}\},\$ $4 \cdot$ end for 5: 6: end for 7: **for** Every min_{FFi} in set min_{FF} **do**, Find the maximum values in in set min_{FF} as the FF - threshold, 8: 9: end for 10: for Every column in MIC - FF matrix do, Count the number of elements in each column that are less than the FF - threshold, obtain the set $Num_{FF} = \{Num_{FF0}, Num_{FF1}, \cdots, Num_{FF24}\},\$ 12: end for 13: **for** Num_{FFi} in set Num_{FF} **do**, 14: Find the median number Num_{med} in set Num_{FF} , 15: if *Num_{FFi}* >*Num_{med}* then, Select the features corresponding to the feature columns and form the feature subset F_{F-F} , 16: end if 17: 18: end for 19: for $f_i \in F$ do Calculate maximum information coefficient $MIC(f_i, R)$, obtaining the MIC - FR matrix, 20: **for** Every value in every row of the *MIC* – *FR* matrix **do** 21: Rank the values and find the median value FR_{med} as the FR – threshold, 22: 23: if $MIC(f_i, R) > FR_{med}$ then, Select the features to form a subset F_{F-R} . 24: 25. end if end for 26: 27: end for

The third stage is called the feature subsets fusion method. The bearing degradation indicator subset F^* constructed in the first stage based on FCBF-AMB, and the bearing degradation indicator subset F_{F-F} and subset F_{F-R} constructed in the second stage, are cross-operated to construct the optimal indicator set F_{opt} , which characterizes the bearings degradation state.

In the above stages, three subsets of bearing degradation indicators are obtained. Subset F^* is an initial subset of degradation indicator with strong correlation and low redundancy. Subset F_{F-F} is a strongly uncorrelated subset composed of features with low redundancy. Subset F_{F-R} is a strongly correlated subset that consists of features that have strong correlations with failure modes.

3.2. LSTM-AdaBoost Ensemble Learning and Prediction Model

After constructing the LSTM neural network model mentioned in Section 2.1, the prediction ability of the model did not meet the requirements for robust prediction. AdaBoost is an iterative algorithm that was originally mainly used in classification problems, and it is sensitive to abnormal features. We considered using the AdaBoost algorithm to enhance the LSTM network model and achieve robust prediction.

Suppose we want to make the m-step ahead prediction for a time-series. The iterative prediction strategy is implemented in this paper, which can be expressed as: $\hat{x}_{t+m} = f(x_t, x_{t-1}, \dots, x_{t-(p-1)})$, where \hat{x} is the predicted value, x_t is the actual value in period t, and p denotes the lag orders.

In this study, the AdaBoost algorithm was introduced to integrate a set of LSTM predictors. The proposed LSTM-AdaBoost ensemble learning approach consists of seven steps as shown in Algorithm 3.

Algorithm 3 LSTM-Adaboost Algorithm.

Input: Training data set: $S = \{(x_{t_1}, \hat{x}_{t_1}), (x_{t_2}, \hat{x}_{t_2}), \dots, (x_{t_N}, \hat{x}_{t_N})\}$, LSTM weak predictor. **Output:** Strong predictor P(x)

- 1: Initialize the weight vector. The weight distribution of the training data is initialized to: $W = (\frac{1}{N}, \frac{1}{N}, \cdots, \frac{1}{N}), k = 1, 2, \cdots, K,$
- 2: Suppose the weight distribution is W_k , the prediction error of the predictor P_k on the training data set is calculated by $\varepsilon_k^i = |P_k(x_i) y_i| / E_k$, where $E_k = \sup_i (|P_k(x_i) y_i|)$, the output interval of the function is [0, 1],
- 3: Calculate the total error of training sample sets: $\varepsilon_k = \sum_{i=1}^n W_k^i \varepsilon_{k'}^i$
- 4: Calculate the weights of the current predictor $a_k = \frac{1}{2} \ln \left(\frac{1}{\beta_i} \right)$, $\beta_i = \frac{\varepsilon_k}{1 \varepsilon_k}$,
- 5: Update the distribution of weights of training datasets as follows: $W_{k+1}^i = \frac{W_k^i \beta_k^{-\epsilon_i}}{Z(k)}$, where $Z(k) = \sum_{k=1}^{N} W^i$

$$\sum_{i=1}^{k} W_{k'}^{i}$$

- 6: Repeat steps 1–5 until all the LSTM predictors are obtained. Record the connection weight of the LSTM predictors $W = (w_1, w_2, \dots w_K)$, where $w_i = \frac{a_i}{\sum_{i=1}^K a_i}$,
- 7: Build the final predictor and integrate the above trained predictors according to the connection weights to obtain the final strong predictor.

$$P(x) = w_1 P_1(x) + w_2 P_2(x) + \cdots + w_K P_K(x).$$

Through the LSTM-AdaBoost ensemble learning approach, multiple weak predictors are integrated into a strong predictor and the features of the degradation indicator set are predicted by the strong predictor. Finally, the prediction results are ensembled to obtain the remaining useful life of the bearings in the next moment. The main steps are as follows:

First, the feature extraction method proposed above is used to construct the indicator set of bearings degradation F_{opt} . Next, for each feature of the bearing degradation indicator set F_{opt} , the LSTM-AdaBoost ensemble learning approach is adopted to obtain the predicted remaining useful life value $\hat{f}_{i,(t+1)}$ corresponding to each feature at moment t + 1.

Finally, the above prediction results are ensembled to obtain the predicted value $R\hat{U}L_{t+1}$ of useful bearing life at t + 1. That is $R\hat{U}L_{t+1} = \sum_{i=1}^{n} \hat{f}_{i,t+1}$.

4. Experiment and Analysis

The run-to-failure data acquired from accelerated degradation tests of rolling element bearings were used to demonstrate the effectiveness of the proposed prediction approach. The proposed approach was compared with other two features selected methods.

4.1. Data Description

The bearings testbed is shown in Figure 4. These faults occurred accidentally in accelerated degradation experiments. XJTU-SY bearing datasets were provided by the Institute of Design Science and Basic Component at Xi'an Jiaotong University (XJTU), Shaanxi, China, and the Changxing Sumyoung Technology Co., Ltd. (SY), Zhejiang, China. The data sets contained complete run-to-failure data of 15 rolling element bearings that were acquired by conducting many accelerated degradation experiments. This testbed was designed to conduct the accelerated degradation tests of rolling element bearings under different operating conditions (different radial forces and rotating speeds). The tested bearings were type LDK UER204.

This platform can conduct accelerated degradation tests of bearings to provide real experimental data that characterize the degradation of bearings during their whole operating life.

To acquire the run-to-failure data of the tested bearings, two type PCB 352C33 accelerometers were horizontally and vertically mounted on the bearing to monitor its vibration. The sampling frequency was set to 25.6 kHz. As shown in Figure 5, a total of 32,768 data points (i.e., 1.28 s) were recorded for each sampling, and the sampling period was 1 min. Detailed information about the platform and experiments can be found in [41].

As tabulated in Table 1, 15 rolling element bearings were tested under three different operating conditions. Among them, the first two bearings in every operating condition were regarded as a training set and the others were used as a testing set. Figure 6 shows the vibration signal of test bearing 1-1 during its whole life cycle. The amplitude of vibration signal increases with time, which indicates that vibration signal plays an important role in bearing performance degradation assessment.



Figure 4. Bearing testbed.



Figure 5. Sampling setting for vibration signals.

Operating Condition	Rotating Speed (rpm)	Radial Force (KN)	Bearings Dataset
Condition A	2100	12	Bearing1_1 Bearing1_2 Bearing1_3 Bearing1_4 Bearing1_5
Condition B	2250	11	Bearing2_1 Bearing2_2 Bearing2_3 Bearing2_4 Bearing2_5
Condition C	2400	10	Bearing3_1 Bearing3_2 Bearing3_3 Bearing3_4 Bearing3_5

Table 1. Operating Conditions of the Tested Bearings.



Figure 6. Typical vertical vibration signals.

4.2. Experiment

4.2.1. Data Preprocessing and Feature Extraction

Because the vibration signal collected by the sensor contains important degradation information, appropriate transformation of the vibration signal can reflect the degradation state of the bearings. To avoid information loss, multiple features in the time and frequency domains are extracted to form feature set for selection. In addition, to accelerate the convergence of the prediction model and improve the prediction accuracy, all the features are normalized. The data preprocessing details are as shown in Algorithm 4.

Algorithm 4 Data preprocessing

Input: Data sample S = {s₁, s₂, ..., s_n}, n is the number of samples.
Output: Original feature set F after data preprocessing.
1: for s_iinS do
2: Calculate each feature in the time and frequency domains, the calculated features are normalized and set between [0,1] to form the original feature set F = {f₁, f₂, ..., f_m}, and m is the number of features.

3: end for

When the bearings in mechanical equipment fail, the amplitude and probability distribution of the time-domain signal change. Signal frequency components, energy of different frequency components, and the position of the main energy spectrum of the spectrum change, which can effectively characterize the state of bearing health, provide the information about the noise in the bearing vibration signal [6,42]. Some features are useless, so choosing the appropriate time-domain and frequency-domain features is the key to effectively predicting the bearing RUL. To obtain more DI and fully reflect the running state of bearings, the feature parameters in the time and frequency-domain are comprehensively used here.

Each of these vibration signals is processed to extract 12 time-domain features, such as mean, variance, and kurtosis. A total of 13 frequency-domain features characterize the degradation of bearing performance, as shown in Table 2. In this study, the time-domain feature and frequency-domain features were calculated using the feature parameters listed in Table 2 [43].

Feature	Time-Domain Feature Parameters	Fosturo	Frequency-Domain Feature Parameters
Teature	Time-Domain reature ratameters	reature	
F0	$f_0 = \frac{\sum_{n=1}^{N} x(n)}{N}$	F12	$f_{12} = \frac{\sum\limits_{k=1}^{\infty} s(k)}{K}$
F1	$f_1 = \sqrt{\frac{\sum_{n=1}^{N} (x(n) - f_0)^2}{N - 1}}$	F13	$f_{13} = \frac{\sum_{k=1}^{K} (s(k) - f_{12})^2}{K - 1}$
F2	$f_2 = \left(rac{\sum_{n=1}^N \sqrt{ x(n) }}{N} ight)^2$	F14	$f_{14} = \frac{\sum_{k=1}^{K} (s(k) - f_{12})^3}{K \left(\sqrt{f_{13}}\right)^3}$
F3	$f_3 = \sqrt{\frac{\sum_{n=1}^{N} (x(n))^2}{N}}$	F15	$f_{15} = \frac{\sum_{k=1}^{K} (s(k) - f_{12})^4}{\frac{K}{(f_{13})^2}}$
F4	$f_4 = \max x(n) $	F16	$f_{16} = \frac{\sum\limits_{k=1}^{\widehat{\Sigma}} \tilde{f}_k s(k)}{\sum\limits_{k=1}^{K} s(k)}$
F5	$f_5 = \frac{\sum_{n=1}^{N} (x(n) - f_1)^3}{(N-1)f_2^3}$	F17	$f_{17} = \sqrt{\frac{\sum_{k=1}^{K} \left(\tilde{f}_{k} - f_{16}\right)^{2} s(k)}{K}}$
F6	$f_6 = \frac{\sum_{n=1}^{N} (x(n) - f_0)^4}{(N-1)f_1^4}$	F18	$f_{18} = \sqrt{rac{\sum_{k=1}^{K} \hat{f}_{k}^{2} s(k)}{\sum_{k=1}^{K} s(k)}}$
F7	$f_7=rac{f_4}{f_3}$	F19	$f_{19} = \sqrt{\frac{\sum_{k=1}^{K} \tilde{j}_{k}^{4} s(k)}{\sum_{k=1}^{K} \tilde{j}_{k}^{2} s(k)}}$
F8	$f_8 = \frac{f_4}{f_2}$	F20	$f_{20} = \frac{\sum_{k=1}^{K} \tilde{f}_{k}^{2} \tilde{s}(k)}{\sqrt{\sum_{k=1}^{K} s(k) \sum_{k=1}^{K} \tilde{f}_{k}^{4} s(k)}}$
F9	$f_9 = rac{f_3}{rac{1}{N}\sum_{n=1}^{N} x(n) }$	F21	$f_{21} = \frac{f_{17}}{f_{16}}$
F10	$f_{10} = rac{f_4}{rac{1}{N}\sum_{n=1}^{N} x(n) }$	F22	$f_{22} = \frac{\sum_{k=1}^{K} \left(\tilde{f}_k - f_{16} \right)^3 s(k)}{K f_{17}^3}$
F11	$f_{11} = \sum_{n=1}^{N} \left x\left(n \right) \right ^2$	F23	$f_{23} = \frac{\sum_{k=1}^{K} \left(\tilde{f}_k - f_{16} \right)^4 s(k)}{K f_{17}^4}$
		F24	$f_{24} = rac{\sum_{k=1}^{K} \left(\tilde{f}_k - f_{16} ight)^{rac{1}{2}} s(k)}{K \sqrt{f_{17}}}$
	where $x(n)$ is the time-domain signal series, for $n = 1, 2, \dots, N$, N is the number of each sample points.		where $s(k)$ is the frequency-domain signal series, for $k = 1, 2, \dots, K$, <i>K</i> is the number of spectral lines.
	- •		J_k is the frequency value of the k-th spectral line.

Table 2. The feature parameters.

4.2.2. Construction of Bearing Degradation Indicator Set

The features mentioned above represent bearings degradation from the different perspective. However, if all these features have been taken as input parameters to the model, then it may result into model over-fitting. Thus, before using these features as input parameters to the model, it is desirable to select the most sensitive features from the feature set and remove the less indicative features to improve the model accuracy [44].

In this paper, the three-stage feature selection method is used to select the sensitive features that can characterize bearings degradation state, which is used to construct the bearings degradation indicator set. Taking operating condition A as an example, the construction process of degradation indicator set is described in detail as follows. Figure 7 shows the sensitive features extracted by the first-stage FCBF feature extraction method. The reference value symmetric uncertainty *SU* is sorted in descending order, the threshold λ given in this paper is 0.1, i.e., the feature with the SU value greater than 0.1 is selected to be placed in the feature subset *F*', then, an AMB de-redundancy method is used to de-redundant the features in feature subset *F*', construct the feature subset in Figure 8 as the initial bearings degradation indicator subset *F**.



Figure 7. The sensitive features were extracted and sorted in descending order in the feature subset F'.



Figure 8. Initial bearing degradation indicator subset F^* .

In the second stage, the *MIC* method mentioned above is used to measure the correlation between features and failure modes, features and features, and construct a strong subset F_{F-R} with strong correlations between features and failure modes, and a strongly uncorrelated subset F_{F-F} consisting of less redundant features. The two subsets of bearing degradation indicators are shown in Figure 9a,b.



Figure 9. Initial bearings degradation indicator (a) subset F_{F-R} and (b) subset F_{F-F} .

In the third stage, the above three bearing degradation indicator subsets, F^* , F_{F-R} , and F_{F-F} , are cross-operated based on the fusion method to obtain a strong correlation and low redundancy optimal degradation indicator set F_{opt} . The final bearings degradation indicator set consists of eight features shown in Figure 10, which will be applied to the bearing remaining useful life prediction as the degradation indicators of the bearing.



Figure 10. Optimal degradation indicator set *F*_{opt} of operating condition A.

According to the proposed feature selection method, features selected by the proposed method are shown in Figure 11.



Figure 11. Features selected by the proposed method.

4.2.3. Train Prediction Model

After obtaining the optimal degradation indicator set revealing the state of bearing degradation, the prediction model is trained. The model is trained by LSTM network, and the AdaBoost algorithm is used to optimize the LSTM prediction model to form a strong predictor. The input of the model is the degradation indicators of the optimal set, the output is the RUL of the bearings. After the training process, the trained model is used to predict the bearing RUL.

4.3. Results and Analysis

To reflect the advantages of constructing the bearing degradation indicator set by using the three-stage feature selection method proposed in this paper, different bearings selection methods were applied to the bearings under three different working conditions to construct the bearing degradation indicator set. Each bearing in the test set was run 10 times; we obtained the average prediction accuracy of three bearings under the same operating conditions. Figure 12a–c depict the average prediction accuracy and feature selection of operating condition A, operating condition B, and operating condition C, respectively. There were three bearings in each condition. As shown in Figure 12, the proposed method extracts fewer features than the other two feature selection methods, and has relatively high accuracy. This is mainly because the proposed method conducts cross-operation on different subsets of degradation indicators, ensuring the robustness of the set of degradation indicators based on reducing feature dimensions.



Figure 12. Average prediction accuracy and number of features selected under three operating conditions. (a) average prediction accuracy and the number of features selected under condition A, (b) average prediction accuracy and the number of features selected under condition B, (c) average prediction accuracy and the number of features selected under condition C.

The feature selection method proposed in this paper is based on the correlation and redundancy measurement, aiming to reduce the complexity of the model and ensure the sensitivity and high contribution rate of the features. mRMR is a method based on the correlation and redundancy measurement. Principal component analysis (PCA) is a dimensionality reduction method that also has a significant effect on reducing the data dimension, and FCBF+Markov Blanket has also been applied [36]. Therefore, these methods were compared with the proposed method in this study. Based on the bearings degradation indicator set constructed above, we used the LSTM neural network and LSTM-AdaBoost ensemble algorithm to predict the RUL of bearing 1_3, bearing 2_3, and bearing 3_3 under three operating conditions.

Figure 13a,b reveal the results of predicting the RUL of bearing 1_3 using different feature selection methods and prediction models. The figure shows that in the early stage of prediction, the prediction result of all the three methods deviate considerably. The prediction results of the non-feature selection method deviates more from the real RUL value, indicating that feature selection is necessary for RUL prediction. The proposed method is the first to fit near the real curve. By comparing the prediction effects of different prediction models, we conclude that AdaBoost algorithm plays a role in improving the prediction accuracy. The prediction results of the other two bearings are shown in Figure 13c–f. The prediction results of the RUL of these two bearings are similar to those of bearing 1_3, further demonstrating that the method proposed in this paper is more robust and produces a better prediction effect under different operating conditions and degradation stages. The effectiveness of this method has been proven.



(e) RUL Prediction results from LSTM.

(**f**) RUL Prediction results from LSTM-AdaBoost.

Figure 13. RUL prediction resultsusing three different features selection methods for three bearings.

The accuracy of model prediction is measured using the mean square error (MSE). Table 3 shows the comparison results of the MSE between no feature selection, PCA, mRMR method, FCBF + Markov Blanket method, and the proposed method. We predicted the bearings under different operating conditions in the test set 10 times, and calculated the mean value of the MSE of the three bearings under each operating condition to represent the predicted results under different operating conditions. To more clearly determine the prediction effects of different methods, taking bearing 1_3 as an example, we selected the prediction results for 12 moments of the bearing, which was the average values obtained by the prediction model running 10 times. We compared the different methods and list the absolute error= |Predicted RUL - Real RUL|. Table 4 provides the details of the predicted results.

In Tables 3 and 4, the MSE and absolute error of non-feature selection method were the largest because irrelevant and redundant data existed in the original data set, even noise. If no feature selection process or filtering process is applied, some outliers also contribute to the prediction model, which leads to significant deviation of the prediction model and reduces the prediction accuracy.

This further proves that feature selection is critical during bearing RUL prediction process. The two tables show that PCA and mRMR do not perform well either because, at the initial stage of bearing degradation, the prediction model learning the degradation characteristics is not complete, resulting in a large error. When the prediction model learns the degradation characteristics, it is affected by the low indicators or weak contribution rate of degradation characteristics to some extent. The prediction error of the method based on FCBF + Markov Blanket is higher than our proposed method. This shows that the more comprehensive the comprehensive features, the more accurate the prediction results. Compared with the results of the two prediction models, the error is significantly reduced for the proposed model, which proves that the LSTM-AdaBoost ensemble prediction method provides improved prediction accuracy.

The proposed method can approximate the real RUL curve of bearings because the method is more robust and avoids the single feature selection method, which may lead to feature sensitivity bias and make the contribution rate of some features larger or smaller under certain measurement standards. The AdaBoost algorithm is a robust lifting algorithm that further guarantees the generalization ability of the prediction model. The experimental results showed that the proposed method has good practical value for bearing RUL prediction.

Operating Condition	Feature Selection Method	Prediction Results (MSE)					
o pontang contanton		Prediction Results (MSE) LSTM LSTM-AdaBoost 471.28 317.43 359.44 244.06 221.64 162.90 82.96 67.09 19.68 10.02 322.81 263.26 206.92 141.32 186.46 127.24 46.62 41.18 13.29 7.06					
	No feature selection	471.28	317.43				
	PCA	359.44	244.06				
Condition A 2100 rpm	mRMR	221.64	162.90				
_	FCBF + Markov Blanket	82.96	67.09				
	Proposed method	19.68	10.02				
	No feature selection	322.81	263.26				
	PCA	206.92	141.32				
Condition B 2250 rpm	mRMR	186.46	127.24				
	FCBF + Markov Blanket	46.62	41.18				
	Proposed method	13.29	7.06				
	No feature selection	419.36	179.02				
	PCA	143.52	97.21				
Condition A 2400 rpm	mRMR	193.77	102.06				
_	FCBF + Markov Blanket	64.27	28.64				
	Proposed method	21.49	15.06				

Table 3. Comparison of prediction accuracy of LSTM and LSTM-AdaBoost at motor speeds of 2100, 2250, and 2400 rpm using different feature selection methods.

Moment Real RUL		No Feature Selection					PCA				mRMR				FCBF + Markov Blanket				Proposed Method			
		LST	LSTM		LSTM-AdaBoost		LSTM		LSTM-AdaBoost		LSTM		LSTM-AdaBoost		LSTM		LSTM-AdaBoost		LSTM		LSTM-AdaBoost	
(Min) (Min	()	Predicted	Absolute	Predicted	Absolute	Predicted	Absolut	e Predicted	l Absolute	Predicted	l Absolute	Predicted	l Absolute	e Predicted	l Absolute	Predicted	Absolut	e Predicted	Absolute	e Predicted	Absolute	
	(M1n)	RUL	Error	RUL	Error	RUL	Error	RUL	Error	RUL	Error	RUL	Error	RUL	Error	RUL	Error	RUL	Error	RUL	Error	
45	113	102.43	10.57	124.7	11.7	122.27	9.27	119.3	6.3	99.35	13.65	118.92	5.92	119	6	110.32	2.68	107	6	116.61	3.61	
55	103	113.7	10.7	108.23	5.23	98.3	4.7	98.11	4.89	95.87	7.13	111.4	8.4	99.3	3.7	105.25	2.25	105	2	101	2	
65	93	101.09	8.09	99.64	6.64	105	12	97.04	4.04	88.3	4.7	98.18	5.18	89.6	3.4	95.77	2.77	95.9	2.9	95.28	2.28	
75	83	92.14	9.14	89.96	6.96	86.6	3.6	78.15	4.85	79.17	3.83	90.96	7.96	90.2	7.2	87.13	4.13	86.5	3.5	81.32	1.68	
85	73	68.22	4.78	67.8	5.2	79.1	6.1	69.32	3.68	77	4	75.82	2.82	76.31	3.31	74.29	1.29	75.13	2.13	72.91	0.09	
95	63	67.04	4.04	58.39	4.61	69.4	6.4	65.51	2.51	60.4	2.6	59.79	3.21	61.4	1.6	64.6	1.6	64.02	1.02	63.29	0.29	
105	53	42.93	10.07	56.77	3.77	55.8	2.8	54.91	1.91	47.2	5.8	57.03	4.03	56	3	54.94	1.94	54.72	1.72	54.41	1.41	
115	43	37.91	5.09	46.47	3.47	47	4	46.76	3.76	40.8	2.2	45.19	2.19	42.11	0.89	45.79	2.79	42.09	0.91	43.55	0.55	
125	33	36.25	3.25	38.06	5.06	30.58	2.42	31.77	1.23	30.2	2.8	35.27	2.27	36.5	3.5	34.18	1.18	35.2	2.2	34.29	1.29	
135	23	28.03	5.03	26.48	3.48	24.9	1.9	25.73	2.73	24.6	1.6	22.39	0.61	26.17	3.17	22.16	0.84	24.8	1.8	23.53	0.53	
145	13	15.23	2.23	15.5	2.5	15.02	2.02	14.7	1.7	14.37	1.37	15.91	2.91	15.41	2.41	12.4	0.6	14.33	1.33	13.6	0.6	
155	3	4.29	1.29	3.73	0.73	4.02	1.02	3.79	0.79	3.77	0.77	3.72	0.72	3.68	0.68	2.49	0.51	2.63	0.37	2.87	0.13	

Table 4. Comparison of prediction results of bearing 1_3 using different methods.

5. Conclusions

The bearing RUL prediction accuracy largely depends on the performance of the degradation indicator set. This paper proposes an ensemble learning method to improve the prediction accuracy of bearing RUL. We mainly studied the feature extraction phase and prediction modelling phase in the process of bearing RUL prediction. For the feature extraction phase, a three-stage feature selection method was proposed to construct the bearing degradation indicator set; for the prediction modelling phase, AdaBoost algorithm was used to enhance the LSTM neural network. Finally, the features of the bearing degradation indicators set were input into the LSTM-AdaBoost prediction model for ensemble learning and robust prediction. Through experimental verification, the proposed method was applied to the XJTU-SY bearing datasets, and the method was compared with different feature selection methods and different prediction modelling methods. The results showed that the method can effectively predict bearing RUL.

Author Contributions: H.W. designed research and conceptualization; Q.T. performed research and verified the proposed method; H.W. and Q.T. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by National Natural Science Foundation of China under grant 71962004.

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

References

- Boskoski, P.; Gasperin, M.; Petelin, D. Bearing fault prognostics based on signal complexity and Gaussian process models. In Proceedings of the 2012 IEEE Conference on Prognostics and Health Management (PHM), Denver, CO, USA, 18–21 June 2012.
- 2. Rasmussen, C.E. Gaussian Processes in Machine Learning. In *Advanced Lectures on Machine Learning*; Bousquet, O., von Luxburg, U., Ratsch, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2004; Volume 3176.
- 3. Amirhossein, G.; Lee, H.H. Probabilistic frequency-domain discrete wavelet transform for better detection of bearing faults in induction motors. *Neurocomputing* **2016**, *188*, 206–216.
- 4. Liu, J.; Wang, W.; Golnaraghi, F. An enhanced diagnostic scheme for bearing con- dition monitoring. *IEEE Trans. Instrum. Meas.* **2010**, *59*, 309–321.
- 5. Sikorska, J.Z.; Hodkiewicz, M.; Ma, L. Prognostic modelling options for remaining useful life estimation by industry. *Mech. Syst. Signal Process.* **2011**, *25*, 1803–1836. [CrossRef]
- 6. Guo, L.; Li, N.P.; Jia, F.; Lei, Y.G. A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing* **2017**, *240*, 98–109. [CrossRef]
- 7. Wang, D.; Shen, C.Q. An equivalent cyclic energy indicator for bearing performance degradation assessment. *J. Vib. Control* **2016**, *22*, 2380–2388. [CrossRef]
- 8. Guo, L.; Lei, Y.G.; Li, N.P.; Yan, T. Machinery health indicator construction based on convolutional neural networks considering trend burr. *Neurocomputing* **2018**, *292*, 142–150. [CrossRef]
- 9. Loutas, T.H.; Roulias, D.; Georgoulas, G. Remaining Useful Life Estimation in Rolling Bearings Utilizing Data-Driven Probabilistic E-Support Vectors Regression. *IEEE Trans. Reliab.* **2013**, *62*, 821–832. [CrossRef]
- 10. Ren, L.; Sun, Y.Q.; Cui, J.; Zhang, L. Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. *J. Manuf. Syst.* **2018**, *48*, 71–77. [CrossRef]
- 11. Sutrisno, E.; Oh, H.; Vasan, A.S.S.; Pecht, M. Estimation of remaining useful life of ball bearings using data driven methodologies. In Proceedings of the 2012 IEEE Conference on Prognostics and Health Management, Denver, CO, USA, 18–21 June 2012; pp. 1–7.
- 12. Caesarendra, W.; Widodo, A.; Thom, P.H.; Yang, B.S. Combined Probability Approach and Indirect Data-Driven Method for Bearing Degradation Prognostics. *IEEE Trans. Reliab.* **2011**, *60*, 14–20. [CrossRef]
- 13. Si, X.S.; W, W.B.; Hu, C.H.; Zhou, D.H. emaining useful life estimation—A review on the statistical data driven approaches. *Eur. J. Oper. Res.* **2011**, *213*, 1–14. [CrossRef]
- 14. Liu, F.; Shen, C.Q.; He, Q.B.; Zhang, A. Wayside bearing fault diagnosis based on a data-driven Doppler effect eliminator and transient model analysis. *Sensors* **2014**, *14*, 8096–8125. [CrossRef] [PubMed]

- Krush, M.T.; Agnihotri, R.; Trainor, K.J. A Contingency Model of Marketing Dashboards and Their Influence on Marketing Strategy Implementation Speed and Market Information Management Capability. *Eur. J. Mark.* 2016, *50*, 2077–2102. [CrossRef]
- Wilson, R.D. Using Clickstream Data to Enhance Business-to-Business Web Site Performance. J. Bus. Ind. Mark. 2010, 25, 177–187. [CrossRef]
- 17. Laukens, K.; Eyckmans, M.; Neuter, N.D. Preparing students for the data-driven life science era through a real-world viral infection case. *J. Biol. Educ.* **2019**, 1–10. [CrossRef]
- 18. Irit, H.; Deeb, D.; Naim, S.; Elad, Y.T. Can internet search engine queries be used to diagnose diabetes? Analysis of archival search data. *Acta Diabetol.* **2019**, *56*, 1149–1154.
- 19. Fu, C.; Chang, W.J.; Liu, W.Y.; Yang, S.L. Data-driven group decision making for diagnosis of thyroid nodule. *Sciece China Inf. Sci.* **2019**, *62*, 1–23. [CrossRef]
- 20. Kumar, S.; Deshpande, A.; Ho, S.S.; Ku, J.S.; Sarma, S.E. Urban Street Lighting Infrastructure Monitoring using a Mobile Sensor Platform. *IEEE Sens. J.* **2016**, *16*, 4981–4994. [CrossRef]
- 21. Dou, J.; Yunus, A.P.; Xu, Y.R.; Zhu, Z.F.; Chen, C.W. Torrential rainfall-triggered shallow landslide characteristics and susceptibility assessment using ensemble data-driven models in the Dongjiang Reservoir Watershed, China. *Nat. Hazards* **2019**, *97*, 579–609. [CrossRef]
- 22. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature 2015, 521, 436. [CrossRef]
- 23. Ren, L.; Cui, J.; Sun, Y.Q.; Cheng, X.J. Multi-bearing remaining useful life collaborative prediction: A deep learning approach. *J. Manuf. Syst.* **2017**, *43*, 248–256. [CrossRef]
- 24. Hu, C.H.; Pei, H.; Wang, Z.Q.; Si, X.S. A new remaining useful life estimation method for equipment subjected to intervention of imperfect maintenance activities. *Chin. J. Aeronaut.* **2018**, *31*, 514–528. [CrossRef]
- 25. Sun, G.L.; Li, J.B.; Dai, J.; Song, Z.C. Feature selection for IoT based on maximal information coefficient. *Future Gener. Comput. Syst.* **2018**, *89*, 606–616. [CrossRef]
- 26. Wang, F.T.; Liu, X.F.; Deng, G. Remaining Life Prediction Method for Rolling Bearing Based on the Long Short-Term Memory Network. *Neural Process. Lett.* **2019**, *50*, 2437–2454 [CrossRef]
- 27. Zhao, L.; Wang, X. A deep feature optimization fusion method for extracting bearing degradation features. *IEEE Access* **2018**, *6*, 19640–19653. [CrossRef]
- 28. Wang, D.; Tsui, K.; Miao, Q. Prognostics and Health Management: A Review of Vibration Based Bearing and Gear Health Indicators. *IEEE Access* **2018**, *6*, 665–676. [CrossRef]
- 29. Tang, X.H.; Wang, J.C.; Lu, J.G.; Liu, G.K. Improving bearing fault diagnosis using maximum information coefficient based feature selection. *Appl. Sci. Basel* **2018**, *8*, 1–17. [CrossRef]
- 30. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. Neural Comput. 1997, 9, 1735–1780. [CrossRef]
- Tang, G.; Zhou, Y.G.; Wang, H.Q.; Li, G.Z. Prediction of bearing performance degradation with bottleneck feature based on LSTM network. In Proceedings of the 2018 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Houston, TX, USA, 14–17 May 2018; pp. 1–6.
- 32. Zhang, B.; Zhang, S.H.; Li, W.H. Bearing performance degradation assessment using long short-term memory recurrent network. *Comput. Ind.* 2019, 106, 14–29. [CrossRef]
- 33. Mao, W.T.; He, J.L.; Tang, J.M.; Li, Y. Predicting remaining useful life of rolling bearings based on deep feature representation and long short-term memory neural network. *Adv. Mech. Eng.* **2019**, *10*, 1–18. [CrossRef]
- Zhang, Y.Z.; Xiong, R.; He, H.W.; Pecht, M.G. Long short-term memory recurrent neural network for remaining useful life prediction of lithium-ion batteries. *IEEE Trans. Veh. Technol.* 2018, 67, 5695–5705. [CrossRef]
- 35. Lei, Y.G.; Li, N.P.; Guo, L.; Li, N.B. Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mech. Syst. Signal Process.* **2018**, *104*, 799–834. [CrossRef]
- 36. Yu, L.; Liu, H. Eficient Feature Selection Via Analysis of Relevance and Redundancy. *J. Mach. Learn. Res.* **2004**, *5*, 1205–1224.
- 37. Yang, Y.M.; Pedersen, J.O. A comparative study on feature selection in text categorization. In Proceedings of the 14th International Conference on Machine Learning, Nashville, TN, USA, 8–12 July 1997; pp. 412–420.
- 38. Koller, D.; Sahami, M. Toward optimal feature selection. In Proceedings of the 13th International Conference on Machine Learning, Bari, Italy, 3–6 July 1996; pp. 284–292.
- Reshef, D.N.; Reshef, Y.A.; Finuca, H.K.; Grossman, S.R. Detecting Novel Associations in Large Data Sets. Science 2011, 334, 1518–1524. [CrossRef] [PubMed]

- 40. Miguel, G.H.; Francisco, G.V.; Belen, M.B.; Moreno-Vega, J.M. High-dimensional feature selection via feature grouping: A variable neighborhood search approach. *Inf. Sci.* **2016**, *326*, 102–118.
- 41. Wang, B.; Lei, Y.G.; Li, N.P.; Li, N.B. A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings. *IEEE Trans. Reliab.* **2018**, 1–12. [CrossRef]
- 42. Yu, J.B. Adaptive hidden Markov model-based online learning framework for bearing faulty detection and performance degradation monitoring. *Mech. Syst. Signal Process.* **2017**, *83*, 149–162. [CrossRef]
- 43. Lei, Y.G.; He, Z.J.; Zi, Y.Y. A new approach to intelligent fault diagnosis of rotating machinery. *Expert Syst. Appl.* **2008**, *35*, 1593–1600. [CrossRef]
- 44. Kundu, P.; Chopra, S.; Lad, B.K. Multiple failure behaviors identification and remaining useful life prediction of ball bearings. *J. Intell. Manuf.* **2017**, *7*, 1–13. [CrossRef]



 \odot 2020 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 (http://creativecommons.org/licenses/by/4.0/).