

Article

Predicting Remaining Useful Life of Rolling Bearings Based on Reliable Degradation Indicator and Temporal Convolution Network with the Quantile Regression

Qiaoping Tian ¹  and Honglei Wang ^{1,2,*}¹ School of Management, Guizhou University, Huaxi District, Guiyang 550025, China; tqp616732695@163.com² Key Laboratory of "Internet +" Collaborative Intelligent Manufacturing in Guizhou Province, Guiyang 550025, China

* Correspondence: hlwang@gzu.edu.cn

Abstract: High precision and multi information prediction results of bearing remaining useful life (RUL) can effectively describe the uncertainty of bearing health state and operation state. Aiming at the problem of feature efficient extraction and RUL prediction during rolling bearings operation degradation process, through data reduction and key features mining analysis, a new feature vector based on time-frequency domain joint feature is found to describe the bearings degradation process more comprehensively. In order to keep the effective information without increasing the scale of neural network, a joint feature compression calculation method based on redefined degradation indicator (DI) was proposed to determine the input data set. By combining the temporal convolution network with the quantile regression (TCNQR) algorithm, the probability density forecasting at any time is achieved based on kernel density estimation (KDE) for the conditional distribution of predicted values. The experimental results show that the proposed method can obtain the point prediction results with smaller errors. Compared with the existing quantile regression of long short-term memory network (LSTMQR), the proposed method can construct more accurate prediction interval and probability density curve, which can effectively quantify the uncertainty of bearing running state.

Keywords: smart manufacturing; remaining useful life prediction; reliability; features compression and computing; quantile regression



Citation: Tian, Q.; Wang, H. Predicting Remaining Useful Life of Rolling Bearings Based on Reliable Degradation Indicator and Temporal Convolution Network with the Quantile Regression. *Appl. Sci.* **2021**, *11*, 4773. <https://doi.org/10.3390/app11114773>

Academic Editor: Tat-Hean Gan and David Mba

Received: 21 April 2021

Accepted: 21 May 2021

Published: 23 May 2021

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



Copyright: © 2021 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

With the development of sensor technology, signal data shows explosive growth. Under such circumstances, effective information extraction and information value analysis become the key links for the efficient use of big data to promote the development of manufacturing industry. In Smart Manufacturing technology, accurate remaining useful life (RUL) prediction provides decision guidance for the formulation of appropriate preventive maintenance and replacement strategies. On the premise of ensuring the operational reliability of mechanical equipment, it can avoid resource waste due to over frequency maintenance or other serious consequences caused by mechanical parts failure. The efficiency of prediction and maintenance decision making will be improved through the mining and analysis of key features of the bearings operation status monitoring data. Therefore, the prediction of RUL of mechanical equipment parts (such as gears and bearings) has been paid more and more attention by scholars. As an essential part of rotating machinery, the performance degradation or failure of rolling bearings directly affects the performance and operation reliability of mechanical equipment. Prognostics Health Management (PHM) of rolling bearings is a process of monitoring and predicting current or future degradation based on historical operational degradation data, designed to assist personnel in making reasonable maintenance decisions to prevent or avoid bearing failure. At present, most sensor-based PHM technologies are mainly based on data-driven statistical models that de-

termine rules in a probabilistic manner [1,2] and artificial intelligence algorithms that rely on machine learning tools [3]. Some experts and scholars have put forward the rolling bearing RUL prediction method based on nonlinear degradation and model [4,5].

At the same time, under the background of rapid development of technology, mechanical equipment status monitoring data shows a trend of massive growth, which provides information basis for real-time acquisition and advance prediction of equipment operation status information. However, when a significant measurement error is introduced due to the complex and changeable operating environment of the equipment, environmental interference, abnormal sensor or system disturbance, some isolated singularities deviating from expected values may occur in the original monitoring data. During the operation of the equipment, due to the occurrence of faults or defects, the monitoring data will also appear abnormal. Obviously, such data contains important information, which cannot be processed in the same way as noise data, anomalous sensor data, and anomalous data caused by environmental interference. Therefore, in obtaining the running state of the equipment, noise processing, and key features mining analysis become the essential work of predicting the operating state of the material and carrying out PHM tasks.

Because of the problem that the high dimension original monitoring data often contain different types of noise signals, it is necessary to extract sufficient information from the contaminated original monitoring data and restore the pure data to achieve the effectiveness of degradation indicators. In recent years, key features mining analysis technology has made rapid development in the field of deep learning. Many degradation indicators construction methods and neural networks have achieved satisfactory performance in different scenarios, vibration signal feature extraction technology based on wavelet transform [6,7], wavelet packet decomposition technology [8] has been frequently used in health indicator construction task, and has been proved to be effective. Based on the idea of deep learning, as an unsupervised learning method, autoencoder, is widely used in dimensionality reduction and information retrieval tasks in the field of PHM because of its robust feature extraction and generalization ability [9–11]. For the past few years, some experts and scholars have proposed to use autoencoder, or improved version of autoencoder, such as stacked sparse autoencoder [12], stacked denoising autoencoder [13,14], convolutional autoencoder [15], and to reconstruct sensor readings and automatically extract degradation characteristics to obtain unsupervised degradation indicator values to identify the severity of equipment degradation. The long short-term memory network (LSTM) [16], convolutional neural network (CNN) [17], deep convolution neural network (DCNN) [18], Convolution and LSTM Hybrid Deep Neural Networks [19], recursive gating unit (Gru) neural network [20], and recurrent convolutional neural network [21] are combined to realize the task of PHM.

The above research work provides a useful reference for the effective learning of the degradation state characteristics of bearing operation. In particular, literature [22] realized the effective prediction of bearings RUL based on deep autoencoder and deep neural networks. On this basis, we will focus on the dimension, purity and efficiency of data. In practical engineering applications, the actual operation data and forecast model parameters of the bearing have a substantial uncertainty, so the traditional point prediction results will inevitably have errors, which is difficult to reflect the uncertainty of bearings degradation state. In the establishment of maintenance strategy based on reliability, if able to quantify the uncertainty of prediction results, they can run for maintenance decision-making and risk assessment to provide more abundant information. The probability prediction offers an effective way to quantitatively balance the risk, which can better describe the possible fluctuation range, uncertainty, and risk of the RUL in the process of future operation degradation, so it has more research value.

For the probability density prediction, most studies adopt nonparametric modelings, such as quantile regression (QR) [23], kernel density estimation(KDE) [24], etc. Which can directly calculate the distribution function or quantile. Taylor [25] introduced a more flexible model in 2000, namely, the quantile regression neural network (QRNN); how-

ever, the shallow structure of QRNN lacks enough ability to simulate the complex time characteristics of time series model. With the continuous development of deep learning, some deep learning models have been gradually applied to the research of time series data. On this basis, considering the correlation between RUL and time, the previous time information can be associated with the current task, the LSTM network model [26] that can learn long-term dependence relationship is constructed. Compared with LSTM and other networks, CNN has a natural advantage in large-scale parallel data processing. Temporal convolution network (TCN) [27] is an improvement of one-dimensional CNN for time series problems, literature [27] shows that TCN has the advantages of faster prediction speed and higher accuracy in most scenarios. At present, there is little research on TCN in the field of engineering reliability, and there is no report on the research of fusion of TCN and neural network in the probability density prediction of RUL.

This study mainly focuses on the problems such as the current running state of the motor rolling element bearings and RUL, which cannot truly represent the abnormal monitoring data caused by various reasons during the equipment condition monitoring process. In this paper, a degradation indicator (DI) construction method based on data reduction and key features mining analysis is proposed. The QR was combined with TCN to improve the performance of the prediction model. Firstly, time domain and frequency domain features are compressed by stacked denoising autoencoder (SDAE). Through redefinition and combination of correlation, redundancy, and monotonicity, the sensitivity measurement standard Std_{DI} of DI is further proposed. The optimal features are selected as optimal DI set Opt_{DI} . Then bearings DI in set Opt_{DI} and its corresponding RUL are taken as input and output variables of temporal convolution network quantile regression (TCNQR) model. The relationship between input variables and response variables under different quantiles is obtained, based on the above results, the probability density function of the predicted bearing RUL is obtained by KDE, which provides more abundant information than traditional point prediction results.

The main contributions of this paper are summarized as follows:

- (1) The original features are compressed and reconstructed based on the SDAE compression method to obtain low-dimensional representative features and retain sufficient information on the premise of not increasing the size of the neural network.
- (2) The redundancy, correlation, and monotonicity measures were incorporated into the DI selection criteria. The sensitivity standard of DI is redefined to further reduce the dimension of input feature variables on the premise of ensuring the efficiency of DI set.
- (3) Combining TCN with QR, a probability density prediction method based on TCNQR is proposed to obtain the predicted value of bearings RUL probability density, obtain more comprehensive and effective information of bearings degradation state, and further reflect the uncertainty of bearings RUL, so as to guide the equipment maintenance decision of actual production and manufacturing activities and avoid large errors and economic losses.

The rest of the paper is organized as follows: Section 2 introduces the basic theory of SAE and QR. The detailed implementation process of the construction of bearings DI set and the TCNQR prediction model are presented in Section 3. The performance of the proposed method was verified using the motor rolling element bearings datasets from Xi'an Jiaotong University (Shaanxi, China) and compared with other methods in Section 4. Finally, conclusions are drawn in Section 5.

2. Basic Theory

2.1. Basic Theory of SAE

Equipment condition monitoring data often has noise signals in the data due to the working environment or operating conditions, which leads to “impure” data. To realize the accurate analysis of the data, we need to “clean” the data, extract the effective features, and restore the pure data. In practical application, principal component analysis (PCA), as a classical method to improve the signal-to-noise ratio (SNR) and linear dimension reduction, is limited in subspace feature extraction [28], and the determination of principal component weight is often subjective, so it is not suitable for the understanding of nonlinear features. However, Hinton [29] in 2006 proposed a low dimensional expression method of learning high-dimensional features by training deep “autoencoder” network, which is essentially a signal compression model based on neural network. Autoencoder(AE) compression embodies its advantages in feature compression and provides a new direction for the research in this field. The simplified network structure of shallow AE is shown in Figure 1.

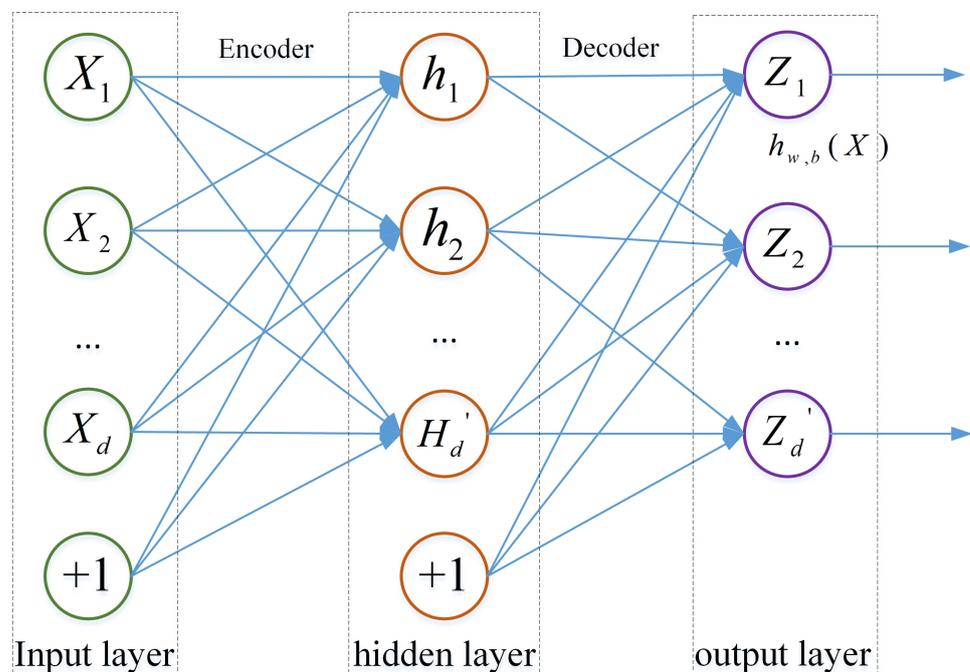


Figure 1. Simplified network structure of autoencoder (AE).

The AE is a neural network that reconstructs the input signal from the target expression. The purpose is to obtain a dimension reduction feature expression $H = \{h_{(1)}, h_{(2)}, h_{(3)}, \dots\}$ of the data through training based on the input unlabeled data $X = \{x_{(1)}, x_{(2)}, x_{(3)}, \dots\}$. The AE encodes the input vector X to the compression feature of the hidden layer through activation function mapping to express H :

$$H = f_{(\theta)}(x) = s(Wx + b) \tag{1}$$

The feature expression H of the hidden layer is reconstructed by mapping and decoding as follows:

$$\hat{x} = g_{\theta'}(H) = s(W'H + b') \tag{2}$$

where, $\theta = (W, b)$ and $\theta' = (W', b')$ are coding model parameters and decoding model weight parameters respectively; W and W' are encoding weight matrix and decoding

weight matrix respectively; b and b' are bias vectors; s function is the activation function, and the general expression is $s(u) = \text{sigmoid}(u) = 1/(1 + \exp(-u))$.

The reconstruction result \hat{x} cannot reproduce the input x completely and accurately. Our goal is to find the minimum reconstruction error of parameters θ and θ' , and to use loss function to represent the training effect and to minimize the loss function. At this time, the common characteristics of the input information x and the reconstruction information \hat{x} are extracted to the maximum. Generally, there are square error loss function and cross entropy loss function, which are respectively expressed as:

$$L(x, \hat{x}) = \frac{1}{n} \left\| x - \hat{x} \right\|^2 \quad (3)$$

$$L(x, \hat{x}) = - \sum_{i=1}^n [x_i \lg(\hat{x}_i) + (1 - x_i) \lg(1 - \hat{x}_i)] \quad (4)$$

The optimization function is expressed as:

$$(\theta^*, \theta^{*'}) = \operatorname{argmin} \frac{1}{n} \sum_{i=1}^n L(x_i, \hat{x}_i) = \operatorname{argmin} \frac{1}{n} \sum_{i=1}^n L(x_i, g_{\theta'}[f_{\theta}(x_i)]) \quad (5)$$

Considering that there will be noise information in the actual operation process of bearings, this paper uses the de-noising autoencoder (DAE) to add noise to the original features, that is, the original input vector x is added with noise to get \tilde{x} , the unit is randomly selected according to a certain proportion and forced to be set to 0, and then the de-noised data \tilde{x} is trained for encoding and decoding. At this time, the loss function is consistent with the traditional AE loss function, and the optimization function is as follows:

$$(\theta^*, \theta^{*'}) = \operatorname{argmin} \frac{1}{n} \sum_{i=1}^n L(x_i, \hat{x}_i) = \operatorname{argmin} \frac{1}{n} \sum_{i=1}^n L(x_i, g_{\theta'}[f_{\theta}(\tilde{x}_i)]) \quad (6)$$

The cost function is:

$$J_{DAE}(W, b) = \frac{1}{m} \sum_{i=1}^m \left(\frac{1}{2} \left\| h_{W,b}(\tilde{x}_i) - H_i \right\|^2 \right) \quad (7)$$

where $h_{W,b}(\tilde{x}_i)$ is the activation value of neurons in the hidden layer.

The combination of DAE is stacked into a deep learning hierarchical structure, that is, multiple DAE are cascaded to complete the task of feature extraction layer by layer, and representative features with lower dimension are obtained. Its essence is to take the damaged information with noise as the input signal, so that the reconstructed signal has a certain robustness to the noise in the input signal. The stacked denoising autoencoder (SDAE) takes the hidden layer output of each layer as the input of the next layer, and obtains the parameter sum of each layer. In this paper, Hinton's layer by layer greedy learning algorithm is used to construct a SDAE network. The main idea of the algorithm is to train only one layer of the network at a time, that is, to train a DAE with only one hidden layer at a time. When the DAE of this layer reaches the optimization, the next DAE will be trained.

Build the three-layer SDAE model, the input layer nodes is n , hidden layer nodes is m , the original input data x is de-noised according to the proportion coefficient λ to get \tilde{x} . According to Equations (1)–(6), the hidden layer H of the first layer is obtained, and the hidden layer output of the first layer network is obtained according to formula (7) of the minimized cost function, that is, the model parameters $\theta_1 = (W, b)$ are obtained. According to the layer by layer greedy training algorithm, the hidden layer outputs of the second and third layer networks are obtained in turn.

The model parameters W and b are obtained by updating the weight once in each iteration using the gradient descent method. The process is as follows:

- (1) Let $\Delta W_1 = 0, \Delta b_1 = 0$
- (2) Calculate $\nabla_{w(l)}L(x, \hat{x}), \nabla_{b(l)}L(x, \hat{x})$
- (3) $\Delta W_l = \Delta W_l + \nabla_{w(l)}L(x, \hat{x})$
- (4) $\Delta b_l = \Delta b_l + \nabla_{b(l)}L(x, \hat{x})$
- (5) $W_l = W_l - \alpha(\frac{1}{m}\Delta W_l), b_l = b_l - \alpha(\frac{1}{m}\Delta b_l)$, where α is the learning rate.

2.2. Basic Theory of Quantile Regression

The essence of quantile regression (QR) [30] is to estimate the different conditional quantiles of response variables by taking different values of τ , so as to obtain the regression prediction model under all quantiles. The calculation formula is:

$$Q_Y(\tau|X) = \alpha_0(\tau) + \alpha_1(\tau)X_1 + \alpha_2(\tau)X_2 + \dots + \alpha_k(\tau)X_k = X \cdot \alpha(\tau) \tag{8}$$

where, $\alpha(\tau) = [\alpha_0(\tau), \alpha_1(\tau), \dots, \alpha_k(\tau)]'$ is the model parameter related to quantile τ , the appropriate $\alpha(\tau)$ needs to be solved by optimizing the following formula :

$$\min \sum_{i=1}^N \rho_{\tau}(Y_i - X_i\alpha(\tau)) = \min \sum_{i|Y_i \geq X_i\alpha(\tau)} \tau(Y_i - X_i\alpha(\tau)) + \min \sum_{i|Y_i < X_i\alpha(\tau)} (\tau - 1)(Y_i - X_i\alpha(\tau)) \tag{9}$$

where, N is the number of samples, ρ is the optimization function.

$$\rho_{\tau}(u) = \begin{cases} \tau u, u \geq 0, \\ (\tau - 1)u, u < 0. \end{cases} \tag{10}$$

After obtaining the estimated value of $\alpha(\tau)$, the estimated value of the response variable under τ quantile can be obtained according to Equation (8).

3. Methodology

To formulate a reasonable and effective maintenance plan, it is necessary to take the RUL as the reference. Bearing as the spare part of mechanical equipment, its RUL prediction is mainly divided into the following four steps:

- (1) bearing operation state data acquisition;
- (2) Representative feature extraction of bearing running state information;
- (3) Predictive modeling;
- (4) RUL prediction.

Figure 2 is the data-driven based RUL prediction.

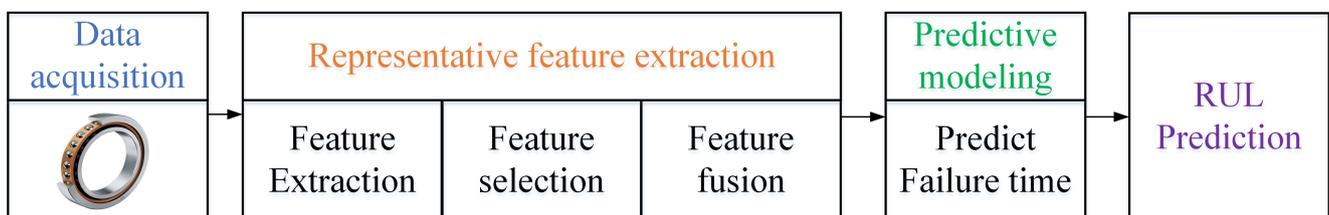


Figure 2. Data-driven based RUL prediction.

Based on the original bearings vibration signal data, different features are extracted. Data reduction and key features analysis techniques are used to drive the construction of DI set. that is, the SDAE model is used to reduce the dimension of time-domain and frequency-domain features, the initial representative feature set F_{SDAE} are extracted from the time-domain and frequency-domain features based on the SDAE compression method; the features in F_{SDAE} are de-redundant, and the de-redundant feature set $DEMIC_{FF}$ is obtained

based on maximum information coefficient (MIC) method. The optimal DI set Opt_{DI} which can reveal the degradation state of the bearing is constructed based on the sensitivity standard Std_{DI} proposed in this paper. Then, the elements in DI set Opt_{DI} are input into the TCNQR network to predict the RUL probability density of the bearing. Finally, the prediction results are compared and analyzed. The structure of the bearings RUL prediction model proposed is shown in Figure 3.

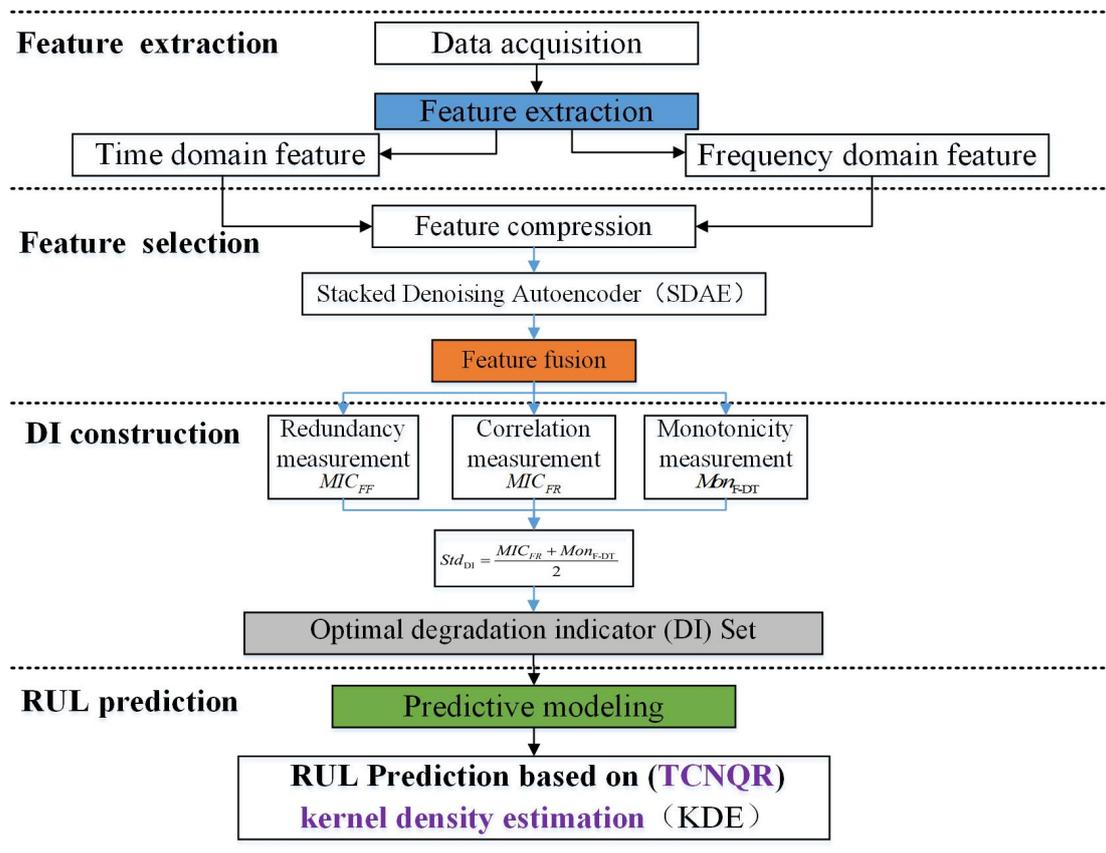


Figure 3. The structure of the bearings RUL prediction model proposed.

3.1. Feature Extraction

The time-domain feature represents the change of bearings vibration signal with the passage of time. Although the time-domain features of bearings vibration signal cannot provide enough degradation information to predict the bearings RUL, it describes the attenuation trend of bearings and reveals whether the running state of bearings is stable. For example, according to the amplitude change of vibration signal and other time-domain features, the bearing can be preliminarily judged whether damage occurs, but different time domain features have different ability to characterize the health status of bearing operation, so it is particularly important to select appropriate time-domain features in bearing running condition monitoring. Frequency variation can display the noise information of bearing vibration signal, and the influence of noise signal can be eliminated by processing the frequency domain features of bearings vibration signal. Therefore, the time and frequency domain characteristics of bearings running signal are selected as the initial research object.

3.2. Feature Compression

In order to select efficient features, this paper adopts the SDAE to compress the features in the time domain and frequency domain.

The SDAE network is used to realize feature compression. Due to the large number of time-domain and frequency-domain features, it is difficult to combine different time-

domain and frequency-domain features to predict the bearing RUL. However, the SDAE network can realize feature compression, ensure the similarity between the decoding results and the input data, and restore the original features to the maximum extent. In the training stage of SDAE network, the input and output are both time domain and frequency domain feature vectors extracted from the original training set. We put the features compressed by SDAE into feature subset F_{SDAE} as the basis for the construction of the optimal DI set Opt_{DI} .

3.3. Feature Fusion

Through the SDAE compression features, to a certain extent, can improve the efficiency of features, and the characteristics of advantage can be found, but because of the existence of redundancy, the optimal feature attributes together is not necessarily the optimal and sensitive feature subset. The sensitive features that can represent the running state of the bearing should be monotonically related to the degradation process of the bearing and have low redundancy with other features, since irrelevant and redundant features will affect the prediction efficiency and accuracy, eliminating those features that cannot provide sufficient fault information has become the main task in the feature selection stage.

For the measurement of correlation, most studies use the method of Person linear correlation coefficient [31,32] to measure the correlation, so as to select the features with higher contribution rate. However, the nonlinear relationship between variables is often ignored, resulting in the unreliability of measurement results, because the complex relationship between general variables cannot be modeled by a single function. The MIC [33] solves the above problems very well. The MIC is suitable for measuring the linear and nonlinear relationship between variables in the measurement data, and mining the non functional dependency relationship between variables.

Based on these goals, MIC is used to measure the relationship between features and degradation trend and the correlation between features.

The MIC is calculated mainly by mutual information (MI) and meshing method. MI is used to measure the degree of correlation between variables. 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:

$$MI(A, B) = \sum_{a \in A} \sum_{b \in B} p(a, b) \log \frac{p(a, b)}{p(a)p(b)} \quad (11)$$

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 :

$$M(D)_{x,y} = \frac{MI^*(D, x, y)}{\log \min\{x, y\}} \quad (12)$$

MIC is defined as:

$$MIC(D) = \max_{xy < B(n)} \{M(D)_{x,y}\} \quad (13)$$

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, $\omega(1) \leq B(n) \leq o(n^{1-\epsilon})$, $0 < \epsilon < 1$. In this paper, the MIC is used to measure the correlation between features and degradation trend as well as features and features. In essence, it is a normalized maximum MI with the value interval of $[0, 1]$.

Definition 1 (The correlation-MIC_{FR}). Suppose feature set $F = \{f_{1t}, f_{2t}, \dots, f_{mt}; R_{1t}, R_{2t}, \dots, R_{mt}\}$, m is the feature number, t is the time series, R_{mt} is the RUL corresponding to the feature f_m at time t . The correlation between any feature f_i and the RUL R_i is defined as MIC_{F_i, R_i} and denoted as MIC_{FR} , the larger the value of MIC_{FR} is, then f_i is the strong correlation feature. On the contrary, f_i is a weak correlation feature; If the value of MIC_{FR} is 0, f_i is an independent feature.

Definition 2 (The redundancy-MIC_{FF}). The redundancy (a kind of correlation) between any feature f_i and feature f_j is defined as MIC_{F_i, F_j} , denotation as MIC_{FF} , the higher the value of MIC_{FF} , the stronger the substitutability between f_i and f_j , namely the stronger the redundancy; the value of MIC_{FF} is 0, indicating that f_i and f_j are independent of each other.

Calculating the redundancy MIC_{F_i, F_j} between features in feature subset F_{SDAE} , the MIC – FF matrix will be obtained that can measure the correlation between features. Finding the minimum values for each column in the MIC – FF matrix and combining these minimum values into a set $m_{FF} = \{m_{FF0}, m_{FF1}, \dots, m_{FFn}\}$, where each column corresponds to one feature, and there are n columns in this matrix. The maximum value will be as the FF – threshold. Counting the number of elements in each column that are less than the threshold value, combining the numbers into a set $N_{FF} = \{Num_{FF0}, Num_{FF1}, \dots, Num_{FFn}\}$, sorting the numbers to find the median. If the number of values is better than the median, the features corresponding to this column will be the elements of the De-redundant feature subset ${}_{DE}MIC_{FF}$. Similarly, the correlation MIC_{FR} is obtained for the elements in subset ${}_{DE}MIC_{FF}$, so as to obtain the degree of influence of those features on the degradation state after the redundancy is removed.

Definition 3 (The monotonicity-Mon_{F-DT}). In order to measure the performance of the feature more comprehensively, monotonicity is used as one of the measures. We will measure the monotonicity of features and degradation tendency, denoted as Mon_{F-DT} .

$Mon_{F-DT} = \left| \frac{Fs > 0}{L-1} - \frac{Fs < 0}{L-1} \right|$, Fs is the different feature sequence, L is the sample length of the full life cycle, and when $Mon_{F-DT} = 1$, the feature and degradation tendency are completely monotonous.

Definition 4 (The Standard-Std_{DI}). Both feature measures MIC_{FR} and Mon_{F-DT} are limited within the range of $[0, 1]$. They are positively correlated with the performance of candidate features and are suitable as feature selection measures. The linear combination of the above two indexes is taken as the selection standard of bearing degradation index, denoted as Std_{DI} .

$Std_{DI} = \frac{MIC_{FR} + Mon_{F-DT}}{2}$, based on the MIC_{FR} value and Mon_{F-DT} value of all features in feature subset ${}_{DE}MIC_{FF}$, the Std_{DI} values of all features in feature subset ${}_{DE}MIC_{FF}$ are obtained. If $Std_{DI} > 0.5$, the feature will be selected into the optimal DI set Opt_{DI} .

Optimal DI Set Construction

A subset of F_{SDAE} is composed of low-dimensional feature expression based on SDAE feature compression method. We will further de-redundant the features in the subset F_{SDAE} .

Based on the MIC feature selection method [34], a de-redundancy feature subset ${}_{DE}MIC_{FF}$ can be obtained by de-redundancy operation.

After the de-redundant feature subset ${}_{DE}MIC_{FF}$ is obtained, we will calculate the correlation and monotonicity of the features in the ${}_{DE}MIC_{FF}$ subset, and measure the sensitivity of the DI based on the proposed Std_{DI} standard, which ultimately constitutes the optimal DI set Opt_{DI} . The steps for obtaining the set Opt_{DI} are shown in Algorithm 1.

Algorithm 1 Optimal DI construction method.

Input: The de-redundant feature subset ${}_{DE}MIC_{FF}$, feature subset ${}_{DE}MIC_{FF} = \{f_1, f_2, \dots, f_m, R\}$, real RUL value R , m is the number of features in the subset ${}_{DE}MIC_{FF}$.

Output: optimal DI set Opt_{DI} .

- 1: **for** $f_i \in {}_{DE}MIC_{FF}$ **do**
- 2: Calculate MIC_{FR} values and Mon_{F-DT} values separately, obtaining the $MIC - Mon$ matrix,
- 3: **for** Every value in every row of the $MIC - Mon$ matrix **do**
- 4: Calculate the Std_{DI} value based on $Std_{DI} = \frac{MIC_{FR} + Mon_{F-DT}}{2}$,
- 5: **if** $Std_{DI} > 0.5$ **then**,
- 6: Select the features in subset ${}_{DE}MIC_{FF}$ to form the optimal DI set Opt_{DI} .
- 7: **end if**
- 8: **end for**
- 9: **end for**

3.4. Quantile Regression of Temporal Convolution Network (TCNQR)

Based on the results of literature [27], TCN has faster prediction speed and higher prediction accuracy. In this paper, quantile regression (QR) was combined with TCN, and the RUL probability density prediction of bearings was realized based on the optimal DI construction method.

The optimization problem of quantile regression (QR) will be introduced into the TCN to realize the bearing RUL probability density prediction. By optimizing the following objective function, the parameters of the TCNQR model are estimated.

$$Y_i = g(X_i, W, b) \quad (14)$$

X_i is the input data constructed by the above DI construction method. The RUL prediction can be expressed as $Y_i^p = g(X_i, \hat{W}, \hat{b})$, \hat{W} and \hat{b} are the optimal estimate of the weight parameter and the bias parameter.

The essence of TCN is the optimization of CNN for time series problems. By adjusting parameters such as convolution kernel, number of convolution layers and expansion coefficient, the global perception of sequence data with specified length is realized, it can be expressed as:

$$F(s) = (X * f)(s) = \sum_{i=0}^{o-1} f(i) \cdot X_{s-d \cdot i} \quad (15)$$

where, X is the input time series; $*$ is the convolution operation; o is the size of the convolution kernel; d is the dilation factor; D is the expansion coefficient.

Further, in order to simplify the calculation of parameters and accelerate the convergence speed of the algorithm, TCN rewrote the weight parameter W as a joint representation of modulus m and direction V through weight normalization [27], as shown in Equation (16).

$$W = \frac{m}{\|V\|} V \quad (16)$$

To avoid too much information loss in the process of feature extraction, TCN replaces the convolutional layer with the residual block, and the input data passes through 1×1 convolution processing, reaches the specified dimension, it is added with the feature data extracted by the dilated causal convolution [27] to serve as the final output of this layer. As shown in the Figure 4, the two dilated causal convolution layers and their corresponding weight normalization, spatial dropout [35] (which added after each dilated

convolution for regularization), activation function and residual connection link, and finally formed a new residual block, which is the basic unit of deep TCN.

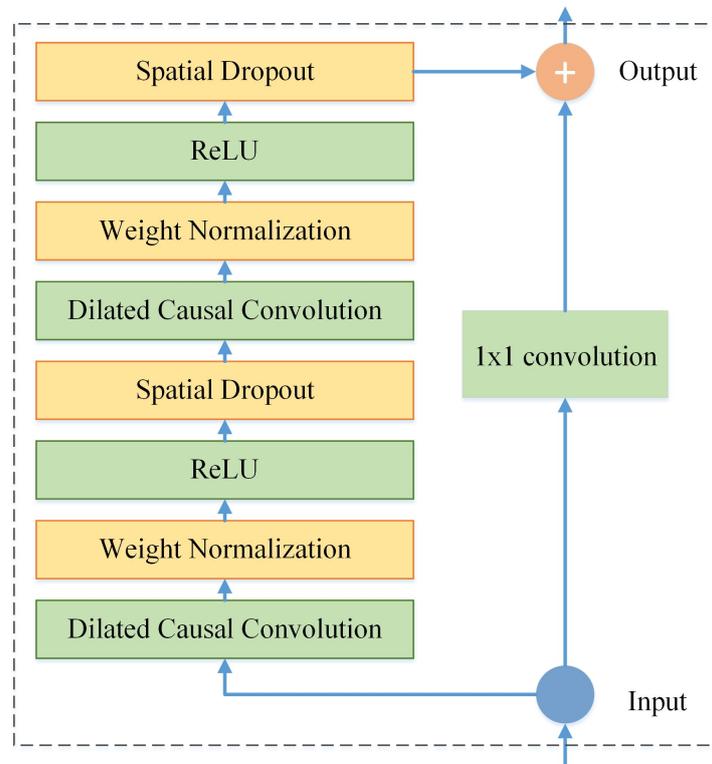


Figure 4. Schematic diagram of residual blocks.

In this paper, combining the QR theory and TCN algorithm, Quantile Regression of temporal convolution network (TCNQR) algorithm is proposed for bearings RUL prediction problem. The QR loss function is as follows:

$$\begin{aligned}
 Loss = \min_{W,b} \sum_{i=1}^N \rho_{\tau}(Y_i - Y_i^p) &= \min_{W,b} \sum_{i|Y_i \geq g(X_i, \hat{W}, \hat{b})} \tau |Y_i - g(X_i, \hat{W}, \hat{b})| \\
 &+ \min_{W,b} \sum_{i|Y_i < g(X_i, \hat{W}, \hat{b})} (1 - \tau) |Y_i - g(X_i, \hat{W}, \hat{b})| \quad (17)
 \end{aligned}$$

The value range of τ is 0 to 1. Through constant optimization and adjustment of W and b , equation (17) can be minimized. In TCNQR, equation (14) is rewritten as:

$$\hat{Q}_Y(\tau | X) = g(X, \hat{W}(\tau), \hat{b}(\tau)) \quad (18)$$

where $\hat{Q}_Y(\tau | X)$ is the optimal estimate of Y when input data is X in the case of determining quantile τ .

3.5. Kernel Density Estimation (KDE)

Kernel density estimation (KDE) [36] is a method to estimate the probability density function of unknown variables from the data itself [37]. Taking the above conditional quantile as the input value of KDE and selecting the appropriate window width for probability density prediction, the following quantile function can be obtained. $Z_i = \hat{Q}_Y(\tau | X)$, $i = (1, \dots, n)$, n is the number of quantiles.

Assuming that Z_1, Z_2, \dots, Z_n is an independent quantile function derived from the estimated probability distribution, then its kernel density is estimated as: $f_h(Z) = \frac{1}{nh} \sum_{i=1}^n K(\frac{Z-Z_i}{h})$, h is the window width, and $K(\cdot)$ is the kernel function.

For the selection of kernel function, there are Gaussian kernel function, matrix kernel function and trigonometric kernel function. In this paper, Gaussian kernel function is selected, and the expression is: $K(u) = \frac{1}{\sqrt{2\pi}} \exp[-\frac{u^2}{2}]$, the value range of h is 1.8 – 2.0.

3.6. Prediction Accuracy Measures

In this paper, the mean absolute error (MAE) and root mean square error (RMSE) are used as the evaluation indexes of deterministic point prediction results, and the prediction interval coverage (PICP) and Mean prediction interval width (MPIW) are used as the reliability and clarity evaluation indexes of probabilistic prediction results [38].

- (1) Mean absolute error (MAE)

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - Y_i^P|$$

- (2) Root mean square error (RMSE)

$$RMSE = \frac{\sum_{i=1}^n (Y_i - Y_i^P)^2}{\sum_{i=1}^n Y_i^2}$$

n is the number of prediction points, Y_i is the real RUL of the sample, Y_i^P is the predicted value.

- (3) Reliability

Prediction interval coverage probability (PICP) is usually used to evaluate the accuracy of prediction interval. It is composed of upper bound and lower bound of coverage target value.

$$PICP = \frac{1}{n} \sum_{i=1}^n C_i$$

A larger $PICP$ value means that more target values Y_i^P fall within the constructed prediction interval. n is the number of prediction points, C_i is the coverage of the prediction interval, $C_i = 1$, if $Y_i^P \in [l_i, u_i]$, otherwise, $C_i = 0$. u_i and l_i are the upper and lower bounds of the target value Y_i^P respectively, and the optimal value of $PICP$ is 100%, which means that all the target values Y_i^P fall within the prediction interval, that is, the coverage rate is 100%.

- (4) Clarity

Mean prediction interval width (MPIW) shows the average width of prediction interval [39].

$$MPIW = \frac{1}{nD} \sum_{i=1}^n (u_i - l_i)$$

where D is the difference between the maximum value and the minimum value of the target values Y_i^P , and the target with D is used to normalize the average width. The bandwidth of the prediction interval should be as small as possible, which determines the amount of information in the prediction interval. Therefore, the width of prediction interval can be used as a standard to compare the prediction results.

4. Experiment and Analysis

The bearings run-to-failure data acquired from accelerated degradation tests were used to demonstrate the effectiveness of the proposed method. The proposed method was compared with other methods.

4.1. Data Description

The bearings testbed is shown in Figure 5. Detailed information about the platform and experiments can be found in [40].

As tabulated in Table 1, 15 rolling element bearings were tested under three different operating conditions. The first two bearings and the last bearing in every operating condition were regarded as a training set and the others were used as a testing set.

The change trend of the original time domain and frequency domain characteristics with the monitoring time is shown in Figure 6. The amplitude of vibration signal varies with time, which indicates that vibration signal plays an important role in the evaluation of bearings performance degradation.

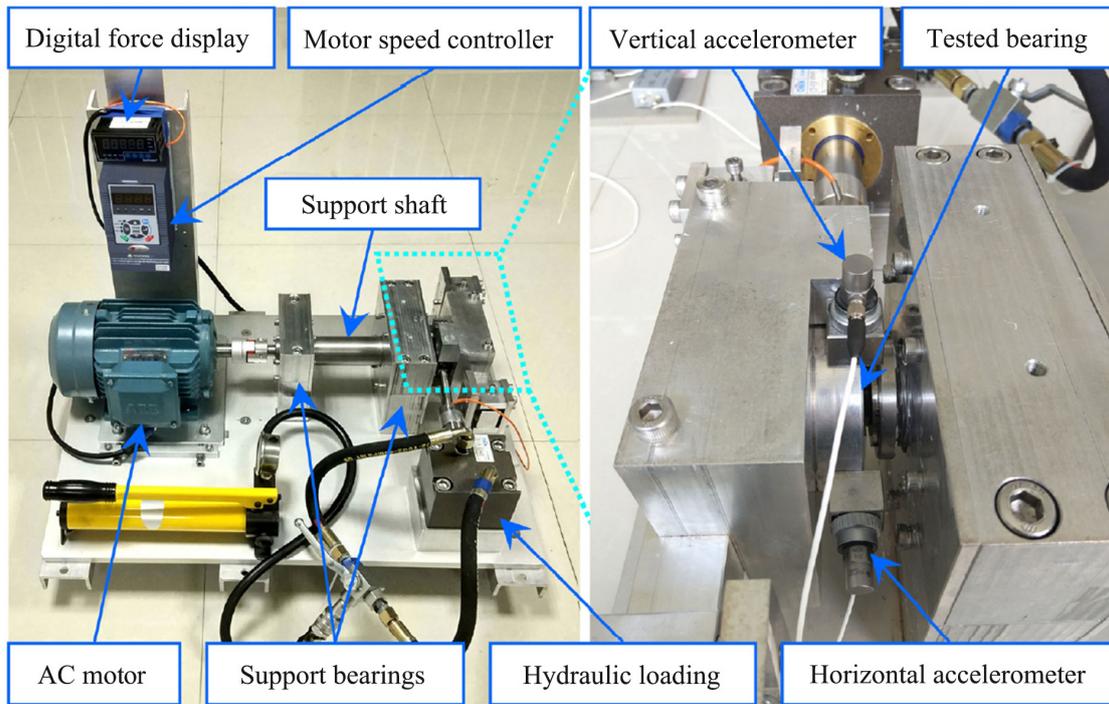


Figure 5. Bearing testbed.

Table 1. Operating Conditions of the Tested Bearings.

| Operating Condition | Rotating Speed (rpm) | Radial Force (kN) | Bearings Dataset |
|---------------------|----------------------|-------------------|------------------|
| Condition A | 2100 | 12 | BearingA_1 |
| | | | BearingA_2 |
| | | | BearingA_3 |
| | | | BearingA_4 |
| | | | BearingA_5 |
| Condition B | 2250 | 11 | BearingB_1 |
| | | | BearingB_2 |
| | | | BearingB_3 |
| | | | BearingB_4 |
| | | | BearingB_5 |
| Condition C | 2400 | 10 | BearingC_1 |
| | | | BearingC_2 |
| | | | BearingC_3 |
| | | | BearingC_4 |
| | | | BearingC_5 |

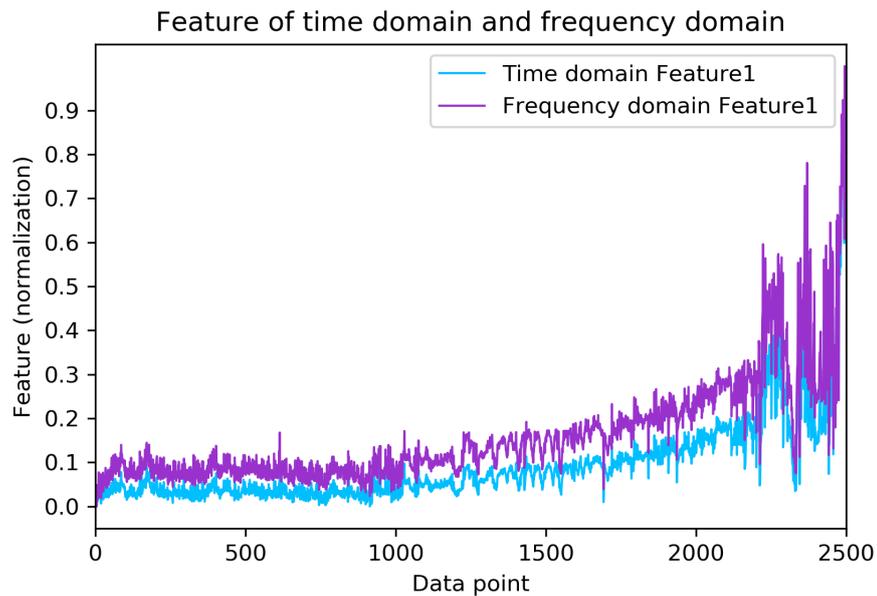


Figure 6. The change trend of time domain and frequency domain features with monitoring time.

4.2. Experiment

4.2.1. Data Preprocessing

The original vibration signal data usually contains rich degradation information. In order to effectively characterize the degradation state, it is necessary to process the vibration signal properly, such as feature extraction and transformation.

In the bearing vibration signal data, in order to avoid information loss, we extract multiple features from time domain and frequency domain to form the initial feature set. In addition, to improve the convergence speed and prediction accuracy of the prediction model, all characteristics are normalized and set between [0,1].

When bearing failure occurs in mechanical equipment, both time-domain and frequency-domain signals will change. Vibration signal data has many time-domain and frequency-domain characteristics, and different features have different representational capabilities for bearing health state. Some time-domain features even have almost no representational capabilities. Therefore, we need to select the appropriate time domain and frequency domain characteristics to achieve efficient prediction of bearings RUL.

The initial features are extracted from the original vibration signals, the time-domain and frequency-domain features in both horizontal and vertical directions are calculated, and the total number of features in both directions is 50, the time-domain feature and frequency-domain features were calculated using the feature parameters listed in Table 2 [41]. $F_0 - F_{11}$ and $F_{25} - F_{36}$ are the time-domain features in horizontal direction and vertical direction, respectively; $F_{12} - F_{24}$ and $F_{37} - F_{49}$ are the frequency-domain features in horizontal direction and vertical direction, respectively. The research content of this paper includes 24 kinds of time-domain features and 26 kinds of frequency-domain features.

As the detailed characteristic parameters in Table 2: The time domain characteristic parameters F_0 , $F_2 - F_4$, and F_{11} reflect the amplitude and energy of vibration in time domain; F_1 and $F_5 - F_{10}$ reflect the time series distribution of the signal. The frequency domain characteristic parameter F_{12} reflects the vibration energy of bearing in frequency domain; $F_{13} - F_{15}$, F_{17} , and $F_{21} - F_{24}$ represent the concentration of the spectrum; F_{16} and $F_{18} - F_{20}$ reflect the change of dominant frequency band position.

Table 2. The feature parameters.

| Feature (Horizontal) | Feature (Vertical) | Time-Domain Feature Parameters | Feature (Horizontal) | Feature (Vertical) | Frequency-Domain Feature Parameters |
|----------------------|--------------------|---|----------------------|--------------------|---|
| F0 | F25 | $f_0 = \frac{\sum_{n=1}^N x(n)}{N}$ | F12 | F37 | $f_{12} = \frac{\sum_{k=1}^K s(k)}{K}$ |
| F1 | F26 | $f_1 = \sqrt{\frac{\sum_{n=1}^N (x(n)-f_0)^2}{N-1}}$ | F13 | F38 | $f_{13} = \frac{\sum_{k=1}^K (s(k)-f_{12})^2}{K-1}$ |
| F2 | F27 | $f_2 = \left(\frac{\sum_{n=1}^N \sqrt{ x(n) }}{N} \right)^2$ | F14 | F39 | $f_{14} = \frac{\sum_{k=1}^K (s(k)-f_{12})^3}{K(\sqrt{f_{13}})^3}$ |
| F3 | F28 | $f_3 = \sqrt{\frac{\sum_{n=1}^N (x(n))^2}{N}}$ | F15 | F40 | $f_{15} = \frac{\sum_{k=1}^K (s(k)-f_{12})^4}{K(f_{13})^2}$ |
| F4 | F29 | $f_4 = \max x(n) $ | F16 | F41 | $f_{16} = \frac{\sum_{k=1}^K \tilde{f}_k s(k)}{\sum_{k=1}^K s(k)}$ |
| F5 | F30 | $f_5 = \frac{\sum_{n=1}^N (x(n)-f_1)^3}{(N-1)f_2^3}$ | F17 | F42 | $f_{17} = \sqrt{\frac{\sum_{k=1}^K (\tilde{f}_k - f_{16})^2 s(k)}{K}}$ |
| F6 | F31 | $f_6 = \frac{\sum_{n=1}^N (x(n)-f_0)^4}{(N-1)f_1^4}$ | F18 | F43 | $f_{18} = \sqrt{\frac{\sum_{k=1}^K \tilde{f}_k^2 s(k)}{\sum_{k=1}^K s(k)}}$ |
| F7 | F32 | $f_7 = \frac{f_4}{f_3}$ | F19 | F44 | $f_{19} = \sqrt{\frac{\sum_{k=1}^K \tilde{f}_k^4 s(k)}{\sum_{k=1}^K \tilde{f}_k^2 s(k)}}$ |
| F8 | F33 | $f_8 = \frac{f_4}{f_2}$ | F20 | F45 | $f_{20} = \frac{\sum_{k=1}^K \tilde{f}_k^2 s(k)}{\sqrt{\sum_{k=1}^K s(k) \sum_{k=1}^K \tilde{f}_k^4 s(k)}}$ |
| F9 | F34 | $f_9 = \frac{f_3}{\frac{1}{N} \sum_{n=1}^N x(n) }$ | F21 | F46 | $f_{21} = \frac{f_{17}}{f_{16}}$ |
| F10 | F35 | $f_{10} = \frac{f_4}{\frac{1}{N} \sum_{n=1}^N x(n) }$ | F22 | F47 | $f_{22} = \frac{\sum_{k=1}^K (\tilde{f}_k - f_{16})^3 s(k)}{K f_{17}^3}$ |
| F11 | F36 | $f_{11} = \sum_{n=1}^N x(n) ^2$ | F23 | F48 | $f_{23} = \frac{\sum_{k=1}^K (\tilde{f}_k - f_{16})^4 s(k)}{K f_{17}^4}$ |
| | | | F24 | F49 | $f_{24} = \frac{\sum_{k=1}^K (\tilde{f}_k - f_{16})^{\frac{1}{2}} s(k)}{K \sqrt{f_{17}}}$ |

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 $k = 1, 2, \dots, K$, K is the number of spectral lines. \tilde{f}_k is the frequency value of the k -th spectral line.

4.2.2. Construction of Bearing Optimal Degradation Indicator Set

In this paper, the SDAE feature compression method is used to compress and extract low dimensional features from 24 time-domain features and 26 frequency-domain features in horizontal and vertical directions. In this experiment, different output features are used to test. Before the original time-domain and frequency-domain features are input into the SDAE network, the input data are normalized to ensure the effectiveness of the compression [22]. It can be seen from Figure 7 that the decoding error decreases with the increase of the number of features. However, when the number of features is 35, the average decoding error of bearings from three working conditions tends to be flat. To retain decoding information as much as possible, the first 35 features are selected, that is, the network inputs 50-dimensional time domain and frequency domain vectors, and outputs 35-dimensional compressed time domain and frequency domain features. It is called the SDAE compression feature set F_{SDAE} .

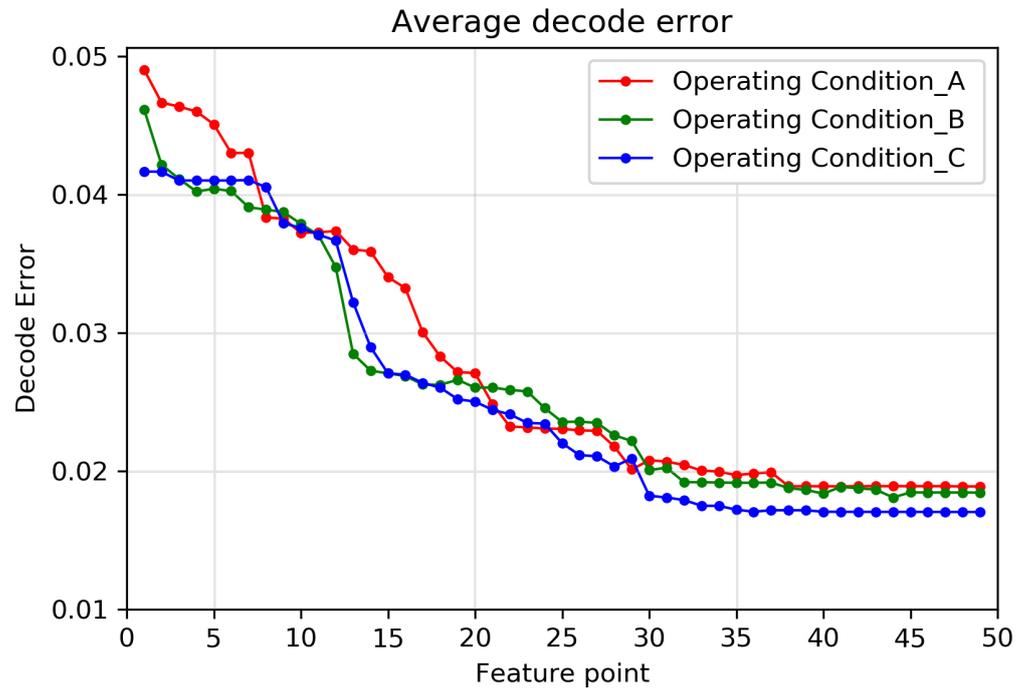


Figure 7. Visualization of training process-SDAE.

Through SDAE compression, 35 features are first extracted and a feature set F_{SDAE} is constructed. Then the redundancy of 35 features under three working conditions are measured by using the MIC method. Taking the bearings of working condition A as an example, the MIC between features and features MIC_{FF} will be calculated to construct de-redundant feature subset $DEMIC_{FF}$. Finally, we got 22 features as shown in the Figure 8, Figure 8 shows the average redundancy value of 22 features in the de-redundant feature subset $DEMIC_{FF}$.

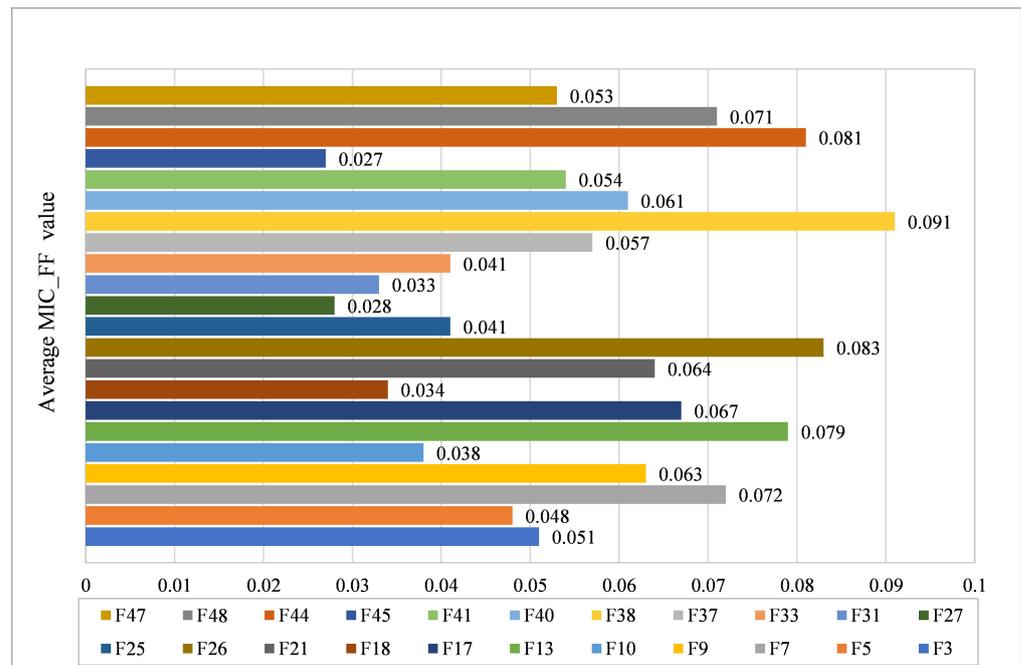


Figure 8. The average MIC_{FF} values of the features in the subset $DEMIC_{FF}$.

At last, the correlation and monotonicity of elements in feature subset $DEMIC_{FF}$ are calculated, the correlation between features and degradation trend MIC_{FR} , and the monotonicity Mon_{F-DT} between features and degradation trend obtained by monotonicity measurement method Mon_{F-DT} . In order to get the optimal feature set Opt_{DI} , the index Std_{DI} is used to select features, and 0.5 is taken as the threshold, as shown in Figure 9. Features higher than 0.5 will become members of the optimal feature set Opt_{DI} , and features lower than 0.5 will not be retained. Therefore, there are 14 features in the final optimal feature set Opt_{DI} . The MIC_{FR} value, Mon_{F-DT} value, and Std_{DI} value of the features in the final optimal feature set Opt_{DI} are shown in Figure 10.

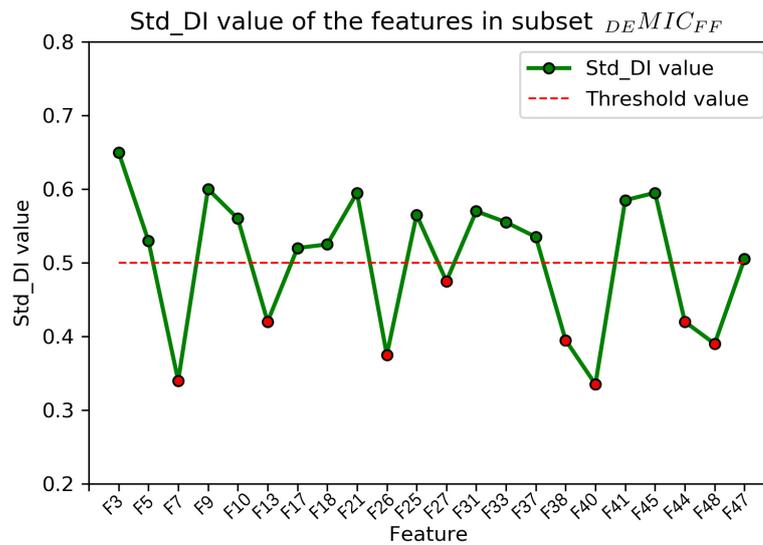


Figure 9. The Std_{DI} calculation results of subset $DEMIC_{FF}$.

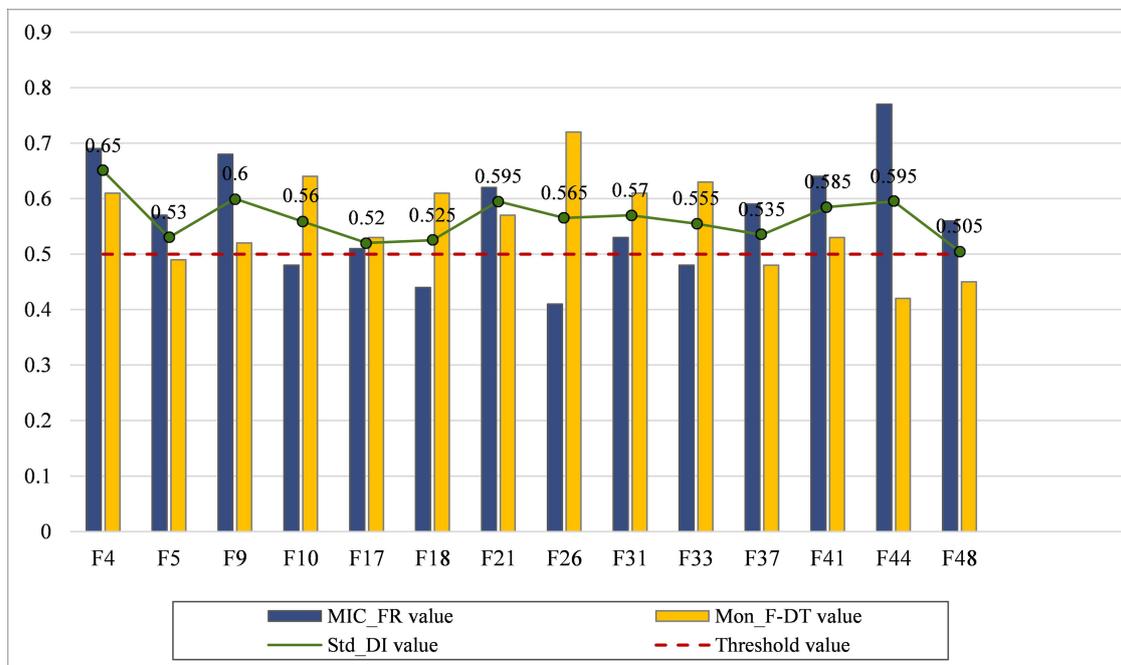
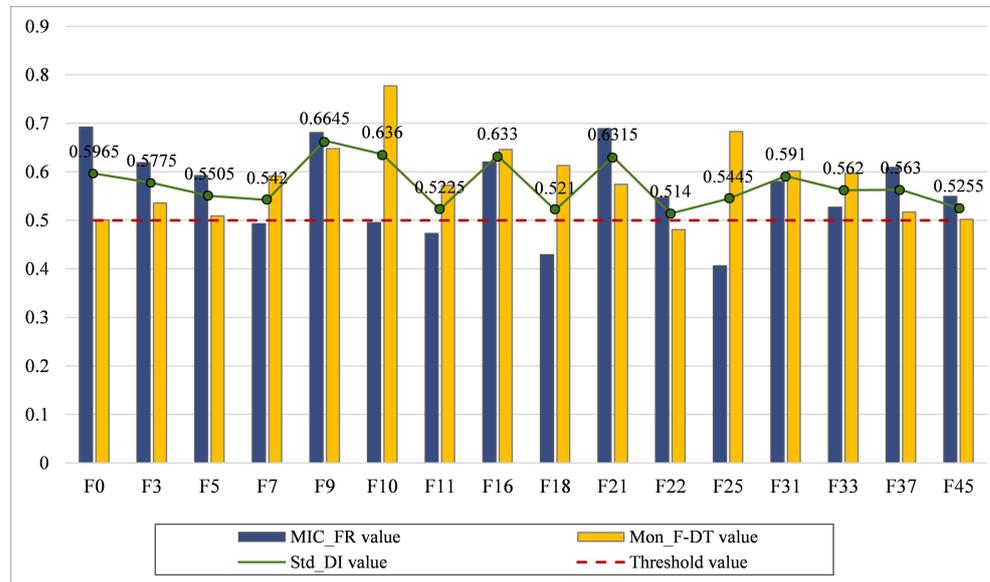


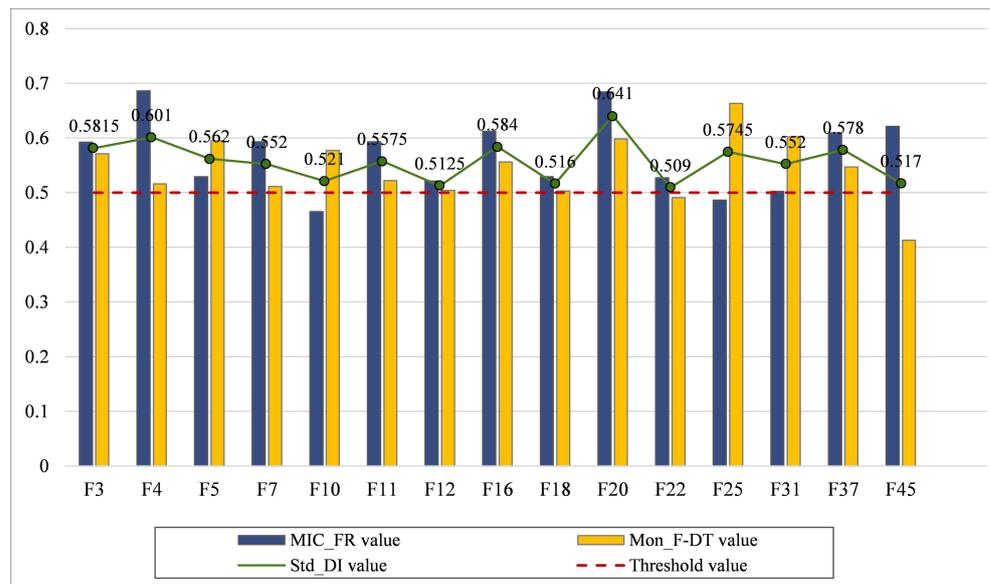
Figure 10. Features in the final optimal feature set Opt_{DI} under working condition A.

In the same process, the bearings degradation indicator set in working condition B and working condition C are constructed with the same method and the specific parameters of the features in the constructed optimal degradation indicator set are shown in Figure 11, in

which Figure 11a represents the composition of the degradation indicator set under working condition B, and Figure 11b represents the composition of the degradation indicator set under working condition C.



(a)



(b)

Figure 11. Features in the final optimal feature set Opt_{DI} . (a) Features in the final optimal feature set Opt_{DI} under working condition B, (b) Features in the final optimal feature set Opt_{DI} under working condition C.

4.2.3. Train Prediction Model

After obtaining the set Opt_{DI} , the prediction model is trained. In the prediction stage, the feature vectors of the test data set are input into the trained TCNQR network, and the predicted RUL values will be output. The prediction results will be evaluated based on the prediction accuracy measures in Section 3.6.

In the TCNQR network, the number of residual blocks and the size of the convolution kernel will affect the computation amount and computation speed of the neural network, and the appropriate dropout rate can effectively solve the problem of neural network

over fitting, the network is built by using Keras deep learning framework, the activation function is the default Relu function. Using Adam optimization algorithm, we design independent adaptive learning rates for different parameters. The number of iterations is 100, the quantile interval is 0.01, then a TCNQR network was constructed. The optimal structure and parameters of the TCNQR model under each quantile are determined by training samples, based on the trained model, the quantiles of the predicted bearing at all quantiles are obtained, which are substituted into the Gaussian kernel density estimation function to estimate the probability density curve of the predicted bearing RUL.

4.3. Results and Analysis

4.3.1. Comparison of Point Prediction Results

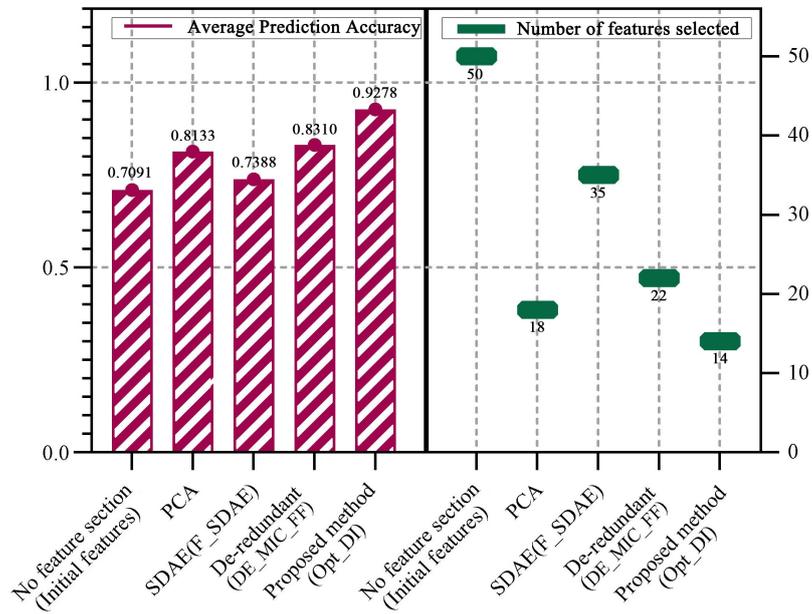
To reflect the advantages of constructing bearing DI set by the method proposed in this paper, different DI construction methods were applied to the bearings under three different working conditions. Each bearing in the test set was run 10 times; we obtained the average prediction accuracy of 3 bearings under the every operating condition. Figure 12 (a) – (c) depict the average prediction accuracy and the number of features in the DI set of operating condition A, operating condition B, and operating condition C, respectively. As shown in Figure 12, the proposed method extracts fewer features than the other feature selection methods, and has relatively high accuracy. This is mainly because the proposed method measures redundancy, correlation and monotonicity on the basis of dimension reduction, which ensures the sensitivity robustness of the degradation index set based on dimension reduction features.

The median of TCNQR prediction result can be regarded as estimating the RUL of bearing as a certain value or at a certain point, in order to further quantify the effectiveness of the prediction results, we take MAE and RMSE as the evaluation indexes of the prediction results, the results were compared with the predicted results of TCNQR. We predicted the bearings under different operating conditions in the test set 10 times, and calculated the MAE and RMSE of the three bearings under each operating condition, compared the RUL prediction results based on the PCA construction method, SDAE construction method, de-redundant construction method and the DI construction method proposed in Table 3.

According to the comparison results in Table 3, it is found that the proposed method has obvious advantages in the error of RUL deterministic prediction results under these three different working conditions, on the one hand, the feature selection method we proposed in the same prediction model has a lower error, which proves the effectiveness of the DI construction method proposed in this paper. On the other hand, under the same DI construction method, TCNQR prediction model has better prediction performance. The results show that the proposed method achieves the goal of optimizing the prediction model from the perspective of data analysis.

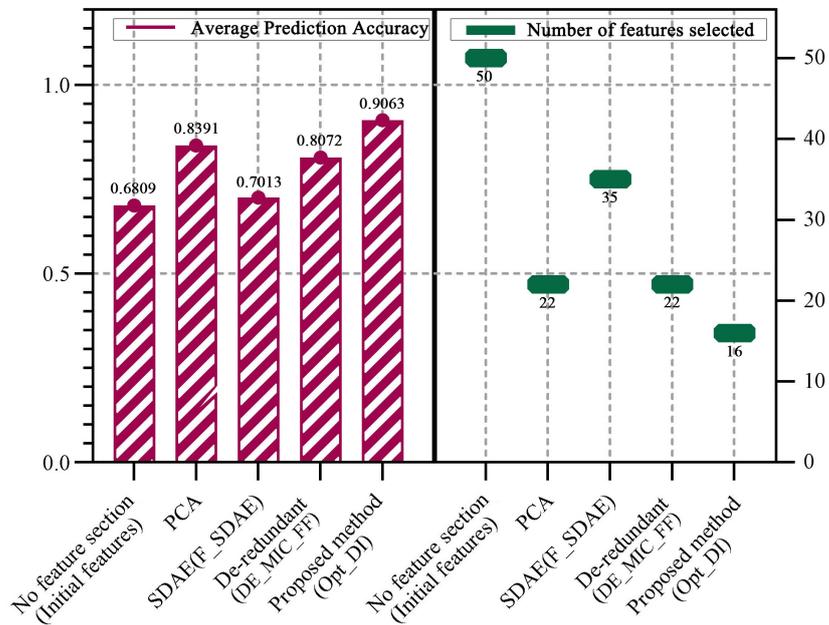
Taking the bearing_3 of different conditions as case, taking the median of density function as the result of deterministic point prediction, and the influence of DI construction on the prediction results is considered. In order to increase the readability of the comparison graph, three sets of degradation indicators with fewer features were selected to compare the predicted results with the real RUL. The prediction results of TCNQR prediction model are shown in the Figure 13. Figure 13a–c are the deterministic prediction results of three different bearing_3 under working condition A, condition B, and condition C, respectively. It can be seen that the prediction results of the method proposed in this paper is closest to the real RUL value, this is more clearly shows the effectiveness of the DI method proposed in this paper.

Average Prediction Accuracy and the Number of features selected under Condition A



(a)

Average Prediction Accuracy and the Number of features selected under Condition B



(b)

Figure 12. Cont.

Average Prediction Accuracy and the Number of features selected under Condition C

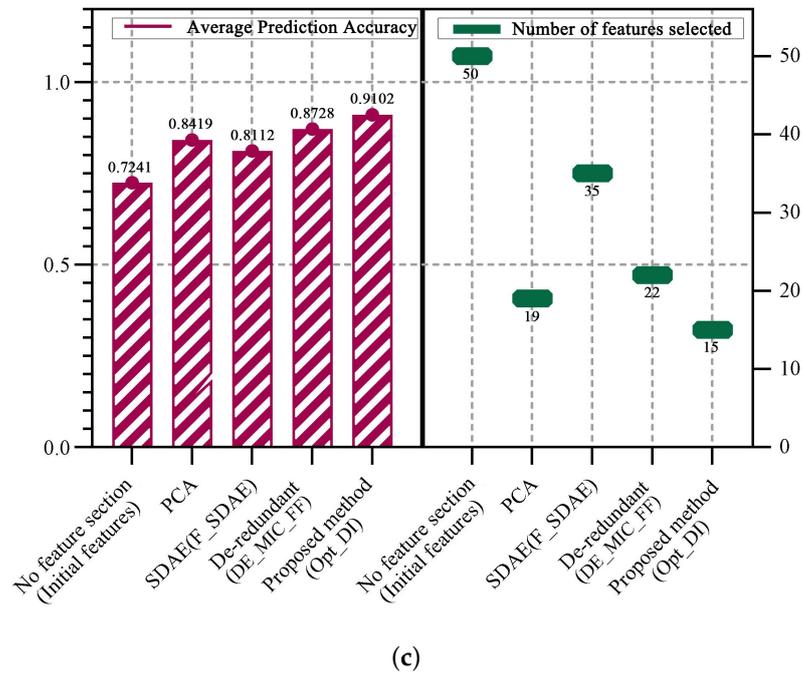


Figure 12. Average prediction accuracy and number of features selected under three operating conditions. (a) average prediction accuracy and the number of features selected under condition A, (b) average prediction accuracy and the number of features selected under condition B, (c) average prediction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

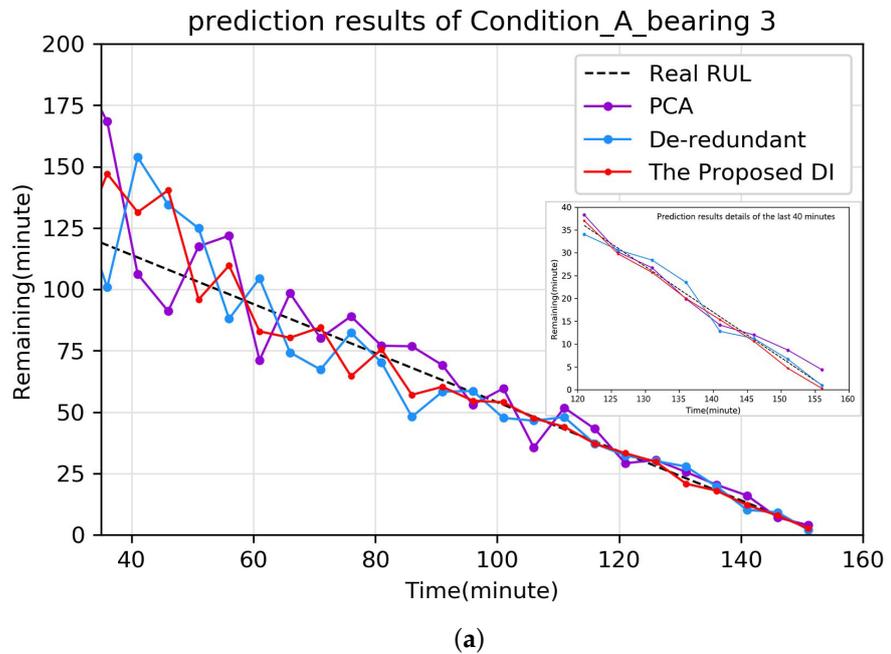


Figure 13. Cont.

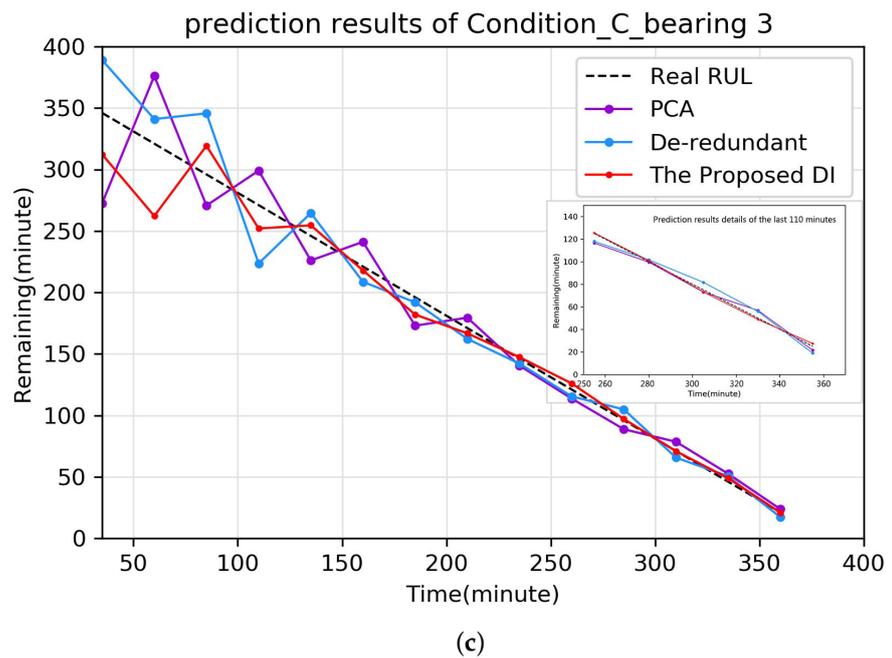
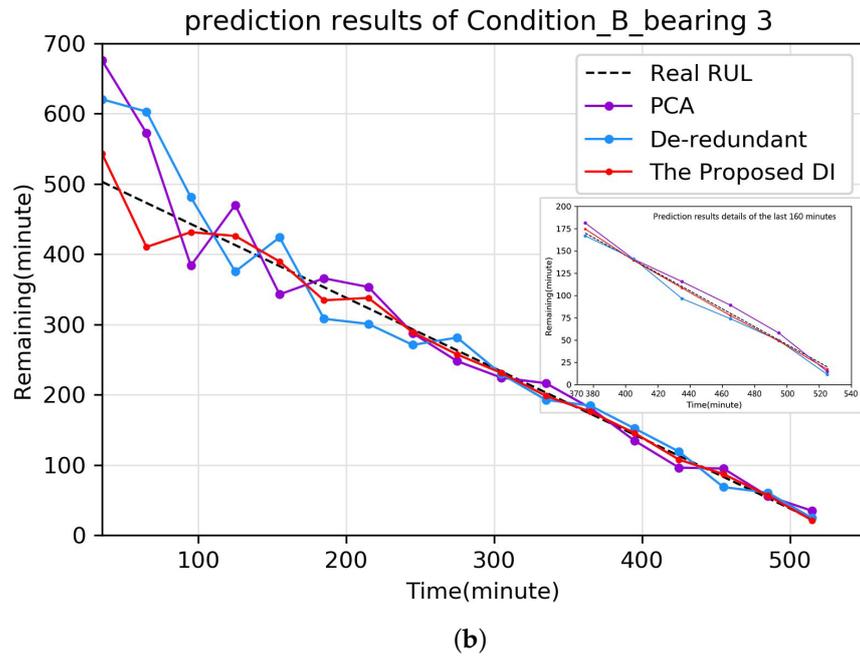


Figure 13. Deterministic point prediction results based on different DI method. (a) deterministic point prediction results of bearing 3 under working condition A, (b) deterministic point prediction results of bearing 3 under working condition B, (c) deterministic point prediction results of bearing 3 under working condition C.

Table 3. Comparison of prediction accuracy at motor speeds of 2100 rpm, 2250 rpm and 2400 rpm using different DI set.

| Working Condition | DI Set | The Prediction Results | | | |
|----------------------|---------------------|------------------------|---------------|-------------|--------------|
| | | MAE (LSTMQR) | RMSE (LSTMQR) | MAE (TCNQR) | RMSE (TCNQR) |
| Condition A 2100 rpm | Initial feature set | 2.928 | 3.207 | 2.801 | 3.034 |
| | PCA feature set | 2.407 | 2.683 | 2.473 | 2.650 |
| | F_{SDAE} | 1.706 | 1.976 | 1.391 | 1.589 |
| | $DEMIC_{FF}$ | 1.411 | 1.694 | 1.317 | 1.533 |
| | Opt_{DI} | 0.729 | 0.903 | 0.584 | 0.783 |
| Condition B 2250 rpm | Initial feature set | 2.643 | 2.916 | 2.509 | 2.733 |
| | PCA feature set | 2.296 | 2.602 | 2.217 | 2.406 |
| | F_{SDAE} | 1.781 | 1.940 | 1.445 | 1.690 |
| | $DEMIC_{FF}$ | 1.803 | 2.117 | 1.283 | 1.466 |
| | Opt_{DI} | 0.657 | 0.851 | 0.579 | 0.747 |
| Condition C 2400 rpm | Initial feature set | 2.727 | 2.985 | 2.648 | 2.811 |
| | PCA feature set | 2.279 | 2.549 | 2.335 | 2.560 |
| | F_{SDAE} | 1.502 | 1.737 | 1.277 | 1.492 |
| | $DEMIC_{FF}$ | 1.412 | 1.689 | 1.182 | 1.369 |
| | Opt_{DI} | 0.701 | 0.884 | 0.458 | 0.619 |

4.3.2. Comparison of KDE Prediction Results

To verify that the method proposed in this paper can more effectively predict the random uncertainty characteristics of the RUL, based on the above contents, the probability density prediction results of bearing A_4 are obtained by TCNQR. Taking three operation moments (75 min, 90 min and 105 min) of bearing A_4 as an example, the probability density curve of bearing RUL is obtained. The real RUL values corresponding to the above there operation moments are 47 min, 32 min and 17 min. The prediction results are shown in the Figure 14. Figure 14d–f are the KDE prediction results of 75 min, 90 min and 105 min, respectively.

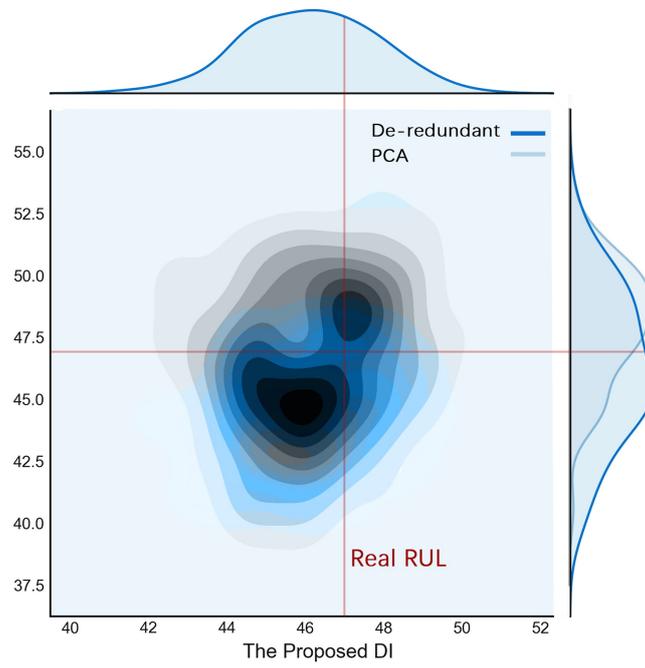
As can be seen from the above figure, the density of the predicted value of the proposed method is concentrated in a range smaller than the real value (at three different moments), which meets the demand for early warning in practical engineering applications. Compared with PCA method, the prediction curve of De-redundant method is more concentrated and closer to the real RUL value. Of course, the peak value of the probability density curve of the RUL prediction result of the proposed method is the closest to the real RUL value, indicating that the probability density forecasting method of bearing RUL based on the TCNQR model and KDE is better than other methods.

In order to verify the effectiveness of the proposed method, the RUL interval prediction of three bearings under different working conditions was carried out, the PICP value and MPIW value of each model at the confidence level of 90% were recorded, and the results were compared in Table 4.

Table 4. Reliability and clarity comparison of prediction results at three different moments.

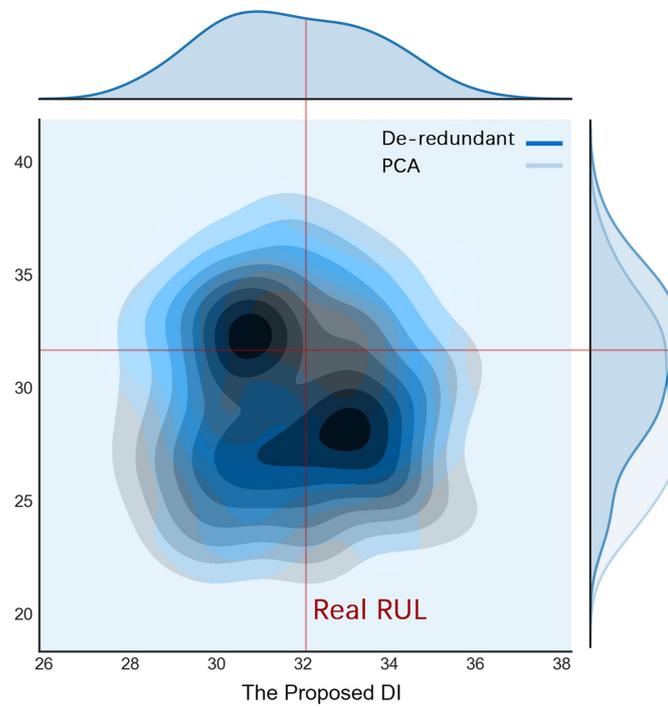
| Moment | Evaluation Indexes | DI Construction Method | | | | | |
|---------|--------------------|------------------------|--------------------|------------------------------|----------------------------|-----------------------------|----------------------------|
| | | PCA-DI Set (LSTMQR) | PCA-DI Set (TCNQR) | F_{SDAE} - DI Set (LSTMQR) | F_{SDAE} -DI Set (TCNQR) | Opt_{DI} -DI Set (LSTMQR) | Opt_{DI} -DI Set (TCNQR) |
| 30 min | PICP | 84.93% | 87.17% | 90.71% | 91.49% | 93.43% | 95.16% |
| | MPIW | 64.22% | 63.98% | 63.06% | 60.72% | 60.17% | 58.44% |
| 60 min | PICP | 87.12% | 89.74% | 90.13% | 92.35% | 93.39% | 95.02% |
| | MPIW | 62.91% | 62.13% | 59.28% | 58.90% | 56.13% | 53.06% |
| 100 min | PICP | 87.03% | 90.19% | 89.28% | 93.04% | 95.87% | 96.25% |
| | MPIW | 62.15% | 59.74% | 59.87% | 55.39% | 58.62% | 41.06% |

KDE Prediction results based on different DI method (75 minutes)



(a)

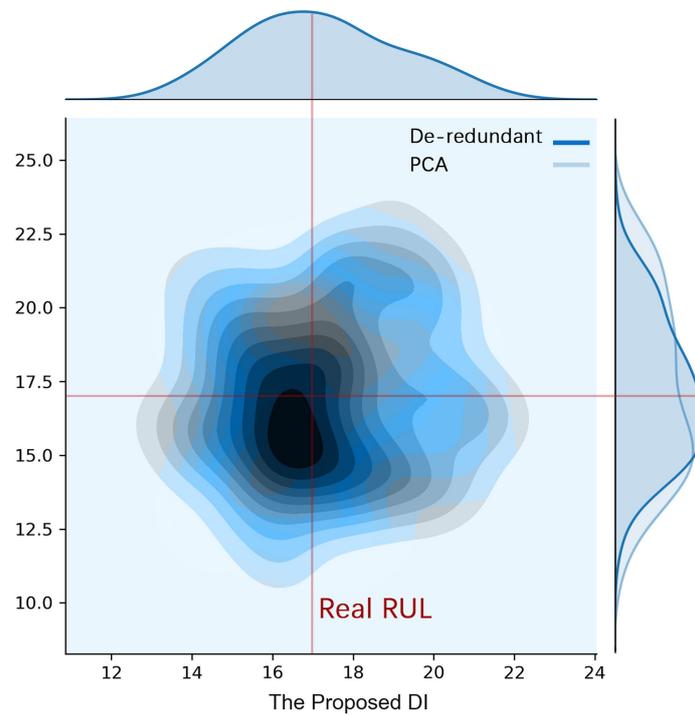
KDE Prediction results based on different DI method (90 minutes)



(b)

Figure 14. Cont.

KDE Prediction results based on different DI method (105 minutes)



(c)

Figure 14. KDE prediction results at different operating moment. (a) deterministic KDE prediction results at the moment of 75 min, (b) deterministic KDE prediction results at the moment of 90 min, (c) deterministic KDE prediction results at the moment of 105 min.

Table 4 shows the prediction results at different time points, among which the RUL prediction results at three different moments are shown. Sometimes the $PICP$ value of PCA method and F_{SDAE} method is less than 90%, it is less than the given confidence degree, which indicates that the method can not guarantee the accuracy of the prediction results. In terms of accuracy, the $PICP$ values of de-redundant method and DI method proposed in this paper are all greater than 90% at most times, especially the DI method proposed in this paper has higher prediction accuracy. From the perspective of $MPIW$ value analysis, the $MPIW$ value of PCA method and de-redundant method are both larger than the DI construction method proposed in this paper. It can be seen that the method proposed in this paper can obtain a relatively narrow prediction range on the premise of meeting the accuracy. Through the horizontal comparison between LSTMQR and TCNQR prediction model, it can be seen that the $PICP$ value of TCNQR is larger than LSTMQR, and the $MPIW$ value of TCNQR is significantly smaller, which further shows that TCNQR has better prediction performance in probability density prediction of time series problems.

5. Conclusions

The performance of rolling bearings DI set will affect the prediction accuracy of the bearing RUL to a large extent. In this paper, the SDAE feature compression method is introduced to compress complex multi-dimensional degradation information. The optimal feature set Opt_{DI} is obtained based on the redefined bearing DI standard Std_{DI} . The influence of DI on bearing RUL prediction is further quantified. Combining TCN with QR, the TCNQR method based on kernel density estimation is used to realize the bearing RUL prediction. Through experiment validation, the proposed method was compared with different DI construction methods and prediction models. The results showed that the method could not only improve the prediction accuracy, but also obtain the probability density function

of the RUL of bearings operation degradation at any time. Through the method proposed in this paper, a higher prediction accuracy and a narrower prediction range can be obtained, and an optimal decision can be made for employees when making maintenance plans, thus reducing resource waste and production downtime and other consequences caused by bearing failure due to over frequency maintenance.

Author Contributions: H.W. designed research and conceptualization; Q.T. performed research and verify 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, in part by the Key Laboratory of "Internet +" Collaborative Intelligent Manufacturing in Guizhou Provence under Grant [2016]5103.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. Son, K.L.; Fouladirad, M.; Barros, A.; Levrat, E.; Lung, B. Remaining useful life estimation based on stochastic deterioration models: A comparative study. *Reliab. Eng. Syst. Saf.* **2013**, *112*, 165–175. [[CrossRef](#)]
2. Si, X.S.; Wang, B.; Hu, C.H.; Zhou, D.H. Remaining useful life estimation—A review on the statistical data driven approaches. *Eur. J. Oper. Res.* **2011**, *213*, 1–14. [[CrossRef](#)]
3. Li, X.; Ding, Q.; Sun, J.Q. Remaining useful life estimation in prognostics using deep convolution neural networks. *Reliab. Eng. Syst. Saf.* **2018**, *172*, 1–11. [[CrossRef](#)]
4. Lei, Y.; Li, N.; Jia, F.; Lin, J.; Xing, S. A Nonlinear Degradation Model Based Method for Remaining Useful Life Prediction of Rolling Element Bearings. In Proceedings of the 2015 Prognostics and System Health Management Conference, Beijing, China, 21–23 October 2015; pp. 1–8. [[CrossRef](#)]
5. Lei, Y.; Li, N.; Gontarz, S.; Lin, J.; Radkowski, S.; Dybala, J. A Model-Based Method for Remaining Useful Life Prediction of Machinery. *IEEE Trans. Reliab.* **2017**, *65*, 1314–1326. [[CrossRef](#)]
6. Qian, Y.; Yan, R. Remaining Useful Life Prediction of Rolling Bearings Using an Enhanced Particle Filter. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 2696–2707. [[CrossRef](#)]
7. Li, H.; Liu, T.; Wu, X.; Chen, Q. Enhanced Frequency Band Entropy Method for Fault Feature Extraction of Rolling Element Bearings. *IEEE Trans. Ind. Inform.* **2020**, *16*, 5780–5791. [[CrossRef](#)]
8. Yen, G.G.; Lin, K.C. Wavelet packet feature extraction for vibration monitoring. *IEEE Trans. Ind. Electr.* **2000**, *47*, 650–667. [[CrossRef](#)]
9. Liu, H.; Zhou, J.; Zheng, Y.; Wei, J.; Zhang, Y. Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders. *ISA Trans.* **2018**, *77*, 167–178. [[CrossRef](#)] [[PubMed](#)]
10. Deng, Z.; Li, Y.; Zhu, H.; Huang, K.; Tang, Z.; Wang, Z. Sparse stacked autoencoder network for complex system monitoring with industrial applications. *Chaos Solitons Fractals* **2020**, *137*. [[CrossRef](#)]
11. Yu, W.; Kim, I.Y.; Mechefske, C. An improved similarity-based prognostic algorithm for RUL estimation using an RNN autoencoder scheme. *Reliab. Eng. Syst. Saf.* **2020**. [[CrossRef](#)]
12. Chen, R.X.; Chen, S.Y.; He, M.; He, D.; Tang, B.P. Rolling bearing fault severity identification using deep sparse auto-encoder network with noise added sample expansion. *J. Risk Reliab.* **2017**, *231*. [[CrossRef](#)]
13. Vincent, P.; Larochelle, H.; Bengio, Y.; Manzagol, P.A. Extracting and Composing Robust Features with Denoising Autoencoders. In Proceedings of the Machine Learning, Proceedings of the Twenty-Fifth International Conference (ICML 2008), Helsinki, Finland, 5–9 June 2008.
14. Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.A. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J. Mach. Learn. Res.* **2010**, *11*, 3371–3408.
15. Kaji, M.; Parvizian, J.; Venn, H. Constructing a Reliable Health Indicator for Bearings Using Convolutional Autoencoder and Continuous Wavelet Transform. *Appl. Sci.* **2020**, *10*, 8948. [[CrossRef](#)]
16. Malhotra, P.; Vishnu, T.V.; Ramakrishnan, A.; Anand, G.; Vig, L.; Agarwal, P.; Shroff, G. Multi-sensor prognostics using an unsupervised health index based on LSTM encoder-decoder. *arXiv* **2016**, arXiv:1608.06154.
17. Yoo, Y.J.; Baek, J.G. A Novel Image Feature for the Remaining Useful Lifetime Prediction of Bearings Based on Continuous Wavelet Transform and Convolutional Neural Network. *Appl. Sci.* **2018**, *8*, 1102. [[CrossRef](#)]
18. Ren, L.; Sun, Y.; Wang, H.; Zhang, L. Prediction of Bearing Remaining Useful Life With Deep Convolution Neural Network. *IEEE Access* **2018**, 13041–13049. [[CrossRef](#)]

19. Kong, Z.; Cui, Y.; Xia, Z.; Lv, H. Convolution and Long Short-Term Memory Hybrid Deep Neural Networks for Remaining Useful Life Prognostics. *Appl. Sci.* **2019**, *9*, 4156. [[CrossRef](#)]
20. Yu, J.; Xu, Y.; Liu, K. Planetary gear fault diagnosis using stacked denoising autoencoder and gated recurrent unit neural network under noisy environment and time-varying rotational speed conditions. *Meas. Sci. Technol.* **2019**, *30*. [[CrossRef](#)]
21. Wang, B.; Lei, Y.; Yan, T.; Li, N.; Guo, L. Recurrent convolutional neural network: A new framework for remaining useful life prediction of machinery. *Neurocomputing* **2020**, *379*, 117–129. [[CrossRef](#)]
22. Ren, L.; Sun, Y.; 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](#)]
23. Koenker, R.; Bassett, J.G. Regression quantiles. *Econometrica* **1978**, *46*, 33–50. [[CrossRef](#)]
24. Lv, P.; Yu, Y.W.; Fan, Y.Y.; Tang, X.F.; Tong, X.R. Layer-constrained variational autoencoding kernel density estimation model for anomaly detection. *Knowl.-Based Syst.* **2020**, *196*. [[CrossRef](#)]
25. Taylor, J.W. A quantile regression neural network approach to estimating the conditional density of multi-period returns. *J. Forecast.* **2000**, *19*, 299–311. [[CrossRef](#)]
26. Tian, C.; Ma, J.; Zhang, C.H.; Zhan, P.P. A Deep Neural Network Model for Short-Term Load Forecast Based on Long Short-Term Memory Network and Convolutional Neural Network. *Energies* **2018**, *11*, 3493. [[CrossRef](#)]
27. Bai, S.; Kolter, J.Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv* **2018**, arXiv:1803.01271.
28. Varon, C.; Alzate, C.; Suykens, J.A.K. Noise Level Estimation for Model Selection in Kernel PCA Denoising. *IEEE Trans. Neural Netw. Learn. Syst.* **2015**, *26*, 2650–2663. [[CrossRef](#)]
29. Hinton, G.; Salakhutdinov, R.R. Reducing the Dimensionality of data with neural networks. *Science* **2006**, *313*, 504–507. [[CrossRef](#)] [[PubMed](#)]
30. Hallock, K.F.; Koenker, R.W. Quantile Regression. *J. Econ. Perspect.* **2001**, *15*, 143–156.
31. Mao, W.; He, J.; Zuo, M.J. Predicting Remaining Useful Life of Rolling Bearings Based on Deep Feature Representation and Transfer Learning. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 1594–1608. [[CrossRef](#)]
32. Li, X.; Zhang, W.; Ding, Q. Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction. *Reliab. Eng. Syst. Saf.* **2019**, *182*, 208–218. [[CrossRef](#)]
33. Reshef, D.N.; Reshef, Y.A.; Finucan, H.K.; Grossman, S.R. Detecting Novel Associations in Large Data Sets. *Science* **2011**, *334*, 1518–1524. [[CrossRef](#)]
34. 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.* **2018**, *8*, 1–17. [[CrossRef](#)]
35. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.
36. Epanechnikov, V.A. Nonparametric estimation of a multidimensional probability density. *Theory Probab. Appl.* **1969**, *14*, 153–158. [[CrossRef](#)]
37. Sheather, S.J.; Jones, M.C. A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation. *J. R. Stat. Soc. Series B Method.* **1991**, *53*, 683–690. [[CrossRef](#)]
38. Wan, C.; Xu, Z.; Pinson, P.; Dong, Z.Y.; Wong, K.P. Optimal Prediction Intervals of Wind Power Generation. *IEEE Trans. Power Syst.* **2014**, *39*, 1166–1174. [[CrossRef](#)]
39. Khosravi, A.; Nahavandi, S.; Creighton, D.; Atiya, A.F. Comprehensive Review of Neural Network-Based Prediction Intervals and New Advances. *IEEE Trans. Neural Netw.* **2011**, *22*, 1341–1356. [[CrossRef](#)]
40. Wang, B.; Lei, Y.; Li, N.; Li, N. A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings. *IEEE Trans. Reliab.* **2020**, *69*, 401–412. [[CrossRef](#)]
41. 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](#)]