



A remaining useful life prediction method for bearing based on deep neural networks

Hua Ding^{*}, Liangliang Yang, Zeyin Cheng, Zhaojian Yang

College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan 030024, China
Shanxi Key Laboratory of Fully Mechanized Coal Mining Equipment, Taiyuan 030024, China

ARTICLE INFO

Keywords:

Remaining useful life prediction
Stratified sampling
3 sigma criterion
DCNN
Generalization ability
Deep learning

ABSTRACT

With the purpose of improving the prediction accuracy and generalization ability of remaining useful life (RUL) prediction models, this paper proposes a new method to predict the RUL of bearings based on the convolutional neural network (CNN). First, the 3 sigma criterion is applied to denoise the original data and remove gross errors. Subsequently, the frequency features are obtained from the original data by the fast Fourier transform (FFT), and the root mean square is employed as the tracking metric. Then, stratified sampling, which differs from the traditional time series data partitioning method, is applied to data partitioning to completely learn the data features. A deep convolutional neural network (DCNN) model without a pooling layer, which consists of three convolutional layers and two fully connected layers, is constructed to avoid feature loss. Finally, the NASA IMS dataset is utilized to assess the preprocessing method, DCNN accuracy and generalization ability. The experimental results show the effectiveness of the proposed method.

1. Introduction

Remaining useful life (RUL) prediction of mechanical equipment is extremely important for industrial manufacturing. Accurate RUL prediction or reliability analysis can reduce the maintenance downtime of mechanical equipment, improve productivity and reduce production costs [1–3]. More importantly, accurate RUL prediction or reliability analysis can ensure the adoption of appropriate maintenance decisions for workers [4], avoid casualties and property losses caused by mechanical equipment damage [5], and improve production safety. Therefore, RUL prediction or reliability analysis of mechanical equipment is a significant component for improving production safety, and the development and improvement of the accuracy of RUL prediction techniques is urgent [6,7].

At present, there are some commonly used denoising methods for data preprocessing including median filtering denoising [8], hard threshold denoising [9], soft threshold denoising [10], and singular value decomposition (SVD) [11]. These methods are widely used in data processing of various research, such as digital image denoising [12], brain electrical signal denoising [13], mining steel rope damage signal processing [14], and testing data of bridge denoising [15].

Generally, existing methods for RUL prediction of mechanical

equipment can be grouped into two categories: model-driven approaches [16] and data-driven approaches [17]. Model-driven approaches employ a qualitative analysis of monitored equipment and require certain expertise to construct the running state characteristics of equipment [18]. For complex working environments, relying on expertise to analyze RUL is difficult, causes low accuracy and does not allow online updates [19]. Common model-driven methods include the Erying model [20], particle filter [21] and Weibull distribution [22]. Data-driven approaches do not require prior professional knowledge. The relationship between the current monitoring data and the operating state of the equipment is established if sufficient complete historical operation data are available [23], and online monitoring data are employed to evaluate the RUL of mechanical equipment. The monitoring data can accurately reflect the state of a device, reduce interference from human factors, and improve the prediction accuracy. Early popular algorithms include the support vector machine (SVM) [24], relevance vector machine (RVM) [25], hidden Markov model (HMM) [26], Gaussian process regression [27], and Gaussian mixture model [28]. Some researchers use artificial neural networks and multilayer perceptrons for RUL prediction of equipment and obtain great results [29–31].

With the rapid development of computer technology in recent years,

* Corresponding author at: College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan 030024, China.
E-mail address: dinghua2002@163.com (H. Ding).

deep learning has achieved remarkable progress in target recognition [32,33], image processing [34], computer vision [35,36], etc., which are mainly attributed to its powerful information mining and feature extraction capabilities. The special structure and remarkable functions of deep learning have attracted increasing attention. Some researchers have attempted to apply deep learning to regression analysis. Polson et al. [37] employed a deep learning model to predict traffic flow and proved that the deep learning model can capture nonlinear temporal and spatial features. Huang et al. [38] proposed the use of the deep convolutional neural network (CNN)-long short-term memory (DCNN-LSTM) model to predict particulate matter.

Deep learning can represent unforeseen data relation, establish complex nonlinear map relationships, be adapted to new problems relatively easily and auto extract features without human intervention, which lead to powerful learning ability [39–41]. Thus, deep learning methods have also been introduced by some researchers to predict the RUL of mechanical equipment. The feed-forward neural network (DBN-FNN) [42], LSTM based on the encoder-decoder (LSTM-ED) [43], ConvLSTM [44], stacked sparse autoencoder (SSAE) [45] and an improved neural network based on an adjacent difference neural network [46] have been employed to predict the RUL of machine components, which can obtain more ideal results than other methods listed in the research and demonstrate that deep learning methods are helpful for the improvement of RUL prediction. Bearings are key components and frequent causes of operation failures and have been widely used in modern machines [47]. Precise RUL prediction for bearings is necessary and significant. The deep belief network (DBN), which is the label for the rapid development of the DNN, has been combined with a particle filter to obtain the expected RUL prediction values of bearings [48]. With the purpose of improving the RUL prediction values, the recurrent neural network (RNN) [49], the AE combined with bidirectional gated recurrent unit (bi-GRU) [50], etc., have been applied to RUL prediction of bearings. Other deep learning methods for RUL prediction also have been proposed by researchers. Due to its extensive applications and strong ability for feature extraction, CNN can obtain excellent results for RUL prediction. Yoo et al. [51] developed a time-frequency image construction health index (HI) and predicted the RUL using CNN. Ren et al. [52] employed a new feature extraction of a spectrum-principal-energy-vector method, which is based on DCNN, to predict the RUL of bearings.

The literature shows that deep learning can achieve excellent results in RUL prediction by combining different neural networks, and the goal of improving RUL prediction value is still in progress. However, researchers seldom verify the models' generalization ability. The generalization ability is an important index for evaluating the prediction performance of a model [53]. Instead, the prediction accuracy of deep neural networks has always been the focus, and most researchers attempt to improve accuracy using different methods. Thus, an ideal model not only achieves excellent prediction accuracy but also has great generalization ability; that is, the model can achieve the same prediction accuracy for different datasets without adjusting the parameters and structure.

By considering data preprocessing and prediction models, this paper proposes a new prediction method based on CNN theory to improve the prediction accuracy and enhance the generalization of the model. First, the 3 sigma criterion is introduced to preprocess the data and eliminate the gross monitored errors. Subsequently, frequency features are obtained by the fast Fourier transform (FFT), health tracking is constructed, and a training set and testing set are built by stratified sampling based on time- and frequency-domain features. To improve the learning of the data characteristics and avoid features loss, a DCNN model without the pooling layer is constructed, where the first three layers are convolutional layers, and the fourth layer is a fully connected layer that integrates the features after convolution. The final layer is a fully connected layer, which is also known as the output layer. RNN, LSTM and Window-CNN (W-CNN) models are compared with the proposed DCNN

model; the results show that the proposed method can achieve high prediction accuracy and generalization ability. The main contributions of this paper are summarized as follows:

1. For data preprocessing, the 3 sigma criterion is employed to remove gross monitored errors based on the characteristics of the data distribution and data acquisition, which avoids the influence of data errors and improves the prediction accuracy.
2. Based on classification, stratified sampling is employed to divide the dataset and ensure that the training set includes the operation data for each stage of the monitored device. By using the local feature perception ability of the convolution core in the CNN convolution layer and adjusting the size of the convolution core, a DCNN prediction model without a pooling layer is constructed to enhance the feature learning ability of the model.
3. By taking advantage of data preprocessing and feature learning, a method for RUL prediction for bearings is proposed based on DCNN, which can improve the prediction accuracy and generalization ability.

The remainder of this paper is organized as follows: Section 2 introduces the CNN infrastructure. Section 3 explains the model construction and related theories proposed in this paper. Section 4 presents the experimental results and analysis. Section 5 concludes the paper and discusses future works.

2. CNN structure and principle

As one of the main deep learning frameworks, CNN has been extensively applied in image recognition, natural language processing and other fields mainly due to its powerful feature extraction ability. A CNN consists of convolutional layers, pooling layers and fully connected layers, as shown in Fig. 1.

Convolution operations for the input vectors are completed by moving convolution kernels in the convolutional layer. An activation function is adopted to achieve a nonlinear fit. Currently, many types of activation functions, including the sigmoid function, tanh function, and rectified linear unit (ReLU), are employed. When the features are complicated or the differences between two features are not obvious, the sigmoid function is employed. When the differences between two features are obvious, the tanh function, which expands the features in the process of circulation and is easily distinguished, is employed. The input data must be normalized before using either of the activation functions; otherwise, the data tend to change gently after the activation function, and the gradient vanishes. With the ReLU function, the input data do not require normalization, which avoids the vanishing gradient and forms the sparsity of the network to achieve better feature extraction.

After the feature mapping of the input vectors by the convolutional layer, the main features are extracted by the pooling layer, which further reduces the number of parameters and completes the dimension reduction using the method of mean pooling or max pooling. After sampling, a new feature map is generated to improve the results and prevent overfitting. The fully connected layer connects and organizes the highly abstract features that are formed by convolution, learns the nonlinear combination of features in a simple manner, and passes the processing results to the output layer.

3. Proposed method

Based on the strong feature extraction and learning ability of the CNN, this paper proposes a regression prediction method for the RUL prediction for bearings.

According to the distribution characteristics and acquisition methods of the monitoring data, the gross monitored errors in the original data are removed using the 3 sigma criterion. After denoising, the frequency features are obtained from time-domain data by the FFT, and the RMS is

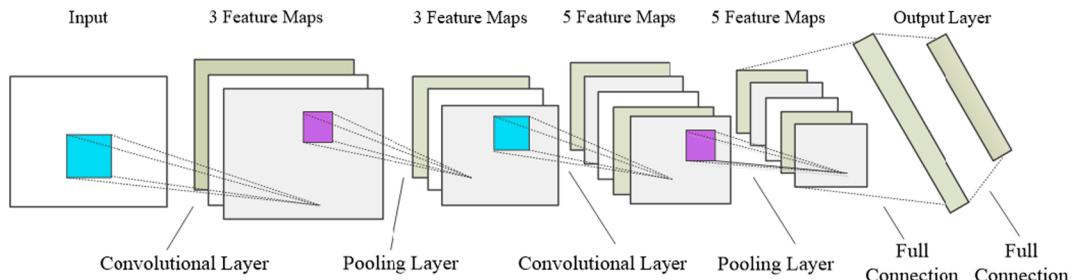


Fig. 1. CNN structure schemes.

employed as the tracking metric to describe the trend in the bearing vibration. The features are normalized to avoid the influence of data fluctuation, and the training set and testing set are constructed based on the time and frequency features by stratified sampling. The training set is employed to train the prediction model. If the loss function is minimal after training, the trained model is saved. If the loss function is not minimal, back propagation (BP) is employed to tune the weights \mathbf{W} and bias \mathbf{b} of the convoluted and fully connected layers in reverse until the loss function reaches the minimum. The trained model is tested using testing data. If the index value is optimal (illustrated in Section 3.2.2), the predicted RUL is obtained. If the index value is not optimal, the training times, kernel size and depth of the hidden layer are adjusted to obtain the optimal index value. A flowchart of the method is shown in Fig. 2.

3.1. Data preprocessing

1) DATA DENOISING

The 3 sigma criterion can be employed to eliminate gross errors from the measurement data by assuming that the data obey a Gaussian distribution and multiple measurements have equal precision; that is, the data distribution is almost entirely concentrated in the $(\mu - 3\sigma, \mu + 3\sigma)$ interval, and the proportion of data beyond this interval is 0.27%. The data beyond the interval are regarded as noise and are eliminated in the processing step. The expression is presented as

$$P(\mu - 3\sigma < x < \mu + 3\sigma) = 0.9973 \quad (1)$$

where μ represents the mean, and σ represents the standard deviation [54].

`density.default(x = x, bw = 0.9, kernel = c("gaussian"))`

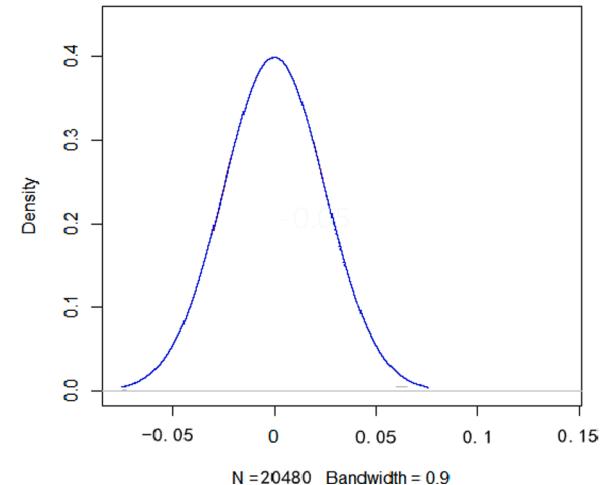


Fig. 3. Verification of raw data distribution.

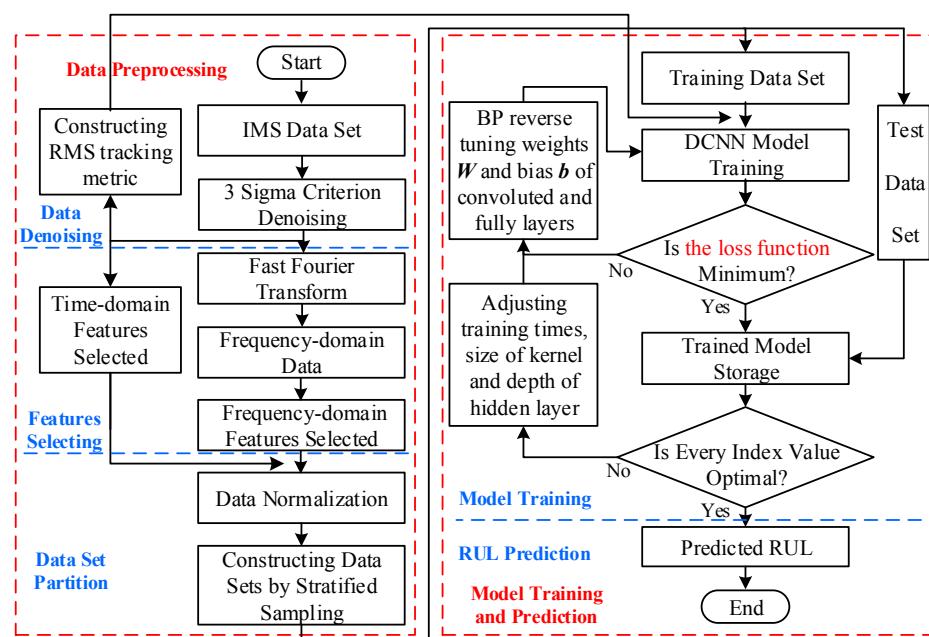


Fig. 2. Flowchart of the proposed method.

The monitoring data of each time point in the dataset conform to a Gaussian distribution, as shown in Fig. 3. Data sampling is performed many times with the same conditions, in accordance with the requirements for using the 3 sigma criterion denoising method.

2) TIME AND FREQUENCY FEATURE SELECTION

The frequency-domain data were obtained from the original time-domain by FFT; the features are shown in Table 1.

3) CONSTRUCTION OF THE TRACKING METRIC

The tracking metric is the basis for the lifetime analysis, including stability and accuracy, and can completely reflect the vibration trend and reliability of mechanical equipment. The RMS, which can reflect the changing trend of the whole dataset, has reasonable stability, satisfies the requirements for a tracking metric, and therefore is chosen as the health tracking metric. The RMS of each group describes the vibration of the dataset at the corresponding time point, and the RMS of the whole dataset reflects the change trend of the amplitude. RMS is expressed as

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (2)$$

where x_i represents the monitoring data of the i th point.

4) NORMALIZATION

The monitoring data include the amplitude of each point in the bearing operation [55]. The positive and negative signals in the data represent the direction of the amplitude, and the normalized data vary in [0,1]. The denoising data are normalized to avoid excessively large or small values.

$$x_{inorm} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (3)$$

where x_i represents the monitoring data of the i th point, and x_{min} and x_{max} represent the minimum value and maximum value at the same time point after the data are denoised.

5) DATASET PARTITION

The monitoring data of the last two moments are inconsistent with the actual trend, and the error is large. To avoid the impact of this error, this part of the data is eliminated in the experiment. Stratified sampling is proposed to construct a training set and testing set for the prediction model to learn the characteristics of each stage in the process of bearing operation. The complete historical data of the bearing operation, which consist of time and frequency features (illustrated in the 2nd point of part A in Section 3), are divided into groups according to the time series. Each group contains 5 data points and is extracted according to the ratio of 4:1; that is, one data point is extracted from every four data points and

Table 1
Selected features.

Domain	Features	Expression
Time	Kurtosis	$Kurt = E\left[\left(\frac{x - \mu}{\delta}\right)^4\right]$
	Skewness	$Skew = E\left[\left(\frac{x - \mu}{\delta}\right)^3\right]$
	Peak-to-Peak	$P - P = max(x_i) - min(x_i)$
	Variance	$Var = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2$
Frequency	Spectral Skewness	$Skew = \sum_{i=1}^N \left(\frac{f_i - \bar{f}}{\sigma}\right)^3 S(f_i)$
	Spectral Kurtosis	$Kurt = \sum_{i=1}^N \left(\frac{f_i - \bar{f}}{\sigma}\right)^4 S(f_i)$

is then sequentially extracted until the end. The extracted data are employed as the testing set, and the remaining data constitute the training set, as shown in Fig. 4.

3.2. Prediction model construction

1) STRUCTURE OF PREDICTION MODEL

This paper constructs a DCNN prediction model based on TensorFlow. ReLU is applied as the activation function to accurately fit the data and make predictions; the expression is shown in formula (4). Dropout is applied in DCNN forward propagation to avoid overfitting. Dropout causes some randomly selected neurons in the network layer to be inactivated each time such that these neurons do not participate in the operation, which improves the generalization ability of the model. Here, dropout is set to 0.7, which means 30% of the neurons are inactivated randomly during the operation. The BP algorithm is applied during the BP update of the network, and the loss function, which is applied to evaluate the accuracy of the model, acts as an index for the reverse optimization; the expression is shown in formula (5). The Adam optimization function is utilized to improve the results. The main idea of the Adam function is to adjust the learning rate of each parameter by using the first-order moment estimation and second-order moment estimation of the gradient. After the bias correction, the learning rate is concentrated in a certain range to stabilize the parameter. Using the Adam algorithm to optimize the model includes four steps: calculating the gradient of each parameter, as show in formula (6); calculating the velocity update amount and the modified velocity update amount, as show in formulas (7) and (8), which can dynamically adjust the velocity depending on updated parameters and the gradient accumulation square; calculating the gradient accumulation square and the modified gradient accumulation square, as show in formulas (9) and (10), the velocity update amount and gradient accumulation square are the key technologies of Adam; and updating the parameters, as show in formula (11), where η is the step length factor, which depends on the velocity update amount and gradient accumulation square. The testing set is input into the trained model to predict the RUL and analyze the reliability. The structure and parameters of the model are shown in Table 2.

$$f(x) = max(0, x) \quad (4)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (5)$$

$$d\omega = \frac{\partial L(\omega)}{\omega} \quad (6)$$

where ω represents the parameter, $d\omega$ is the first moment of ω , and $L(\omega)$ is the function of ω .

$$v = \beta_1 v + (1 - \beta_1) d\omega \quad (7)$$

where β_1 is the exponential decay rate of the first-order moment estimation, and v is the velocity update amount.

$$\dot{v} = \frac{v}{1 - \beta_1 t} \quad (8)$$

where β_1^t represents the exponential decay rate of the first moment estimation at the current moment, and \dot{v} is the modified velocity update amount.

$$Sd\omega = \beta_2 Sd\omega + (1 - \beta_2)d\omega^2 \quad (9)$$

where β_2 is the exponential decay rate of the second moment estimation, $Sd\omega$ is the gradient accumulation square, and $d\omega^2$ is the second moment of ω .

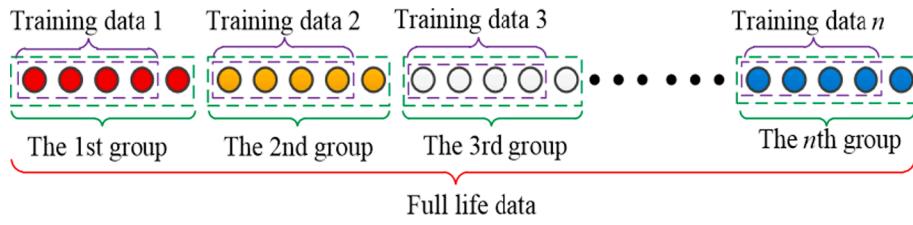


Fig. 4. Stratified sampling.

Table 2
Structure of the DCNN.

Layer (Type)	Kernel Size	Strides	Output Connection
Input layer	None	None	1st convolutional layer
1st convolutional layer	6 × 1	6	2nd convolutional layer
2nd convolutional layer	1 × 1	1	3rd convolutional layer
3rd convolutional layer	1 × 1	1	1st fully connected layer
1st fully connected layer	None	None	2nd fully connected layer

$$Sd\omega^t = \frac{Sd\omega}{1 - \beta_2^t} \quad (10)$$

where β_2^t is the exponential decay rate of the second moment estimation at the current moment, and $Sd\omega$ is the modified gradient accumulation square.

$$\omega = \omega - \frac{\eta}{\sqrt{Sd\omega} + v} \quad (11)$$

The input features are transformed into a one-dimensional vector, and from the top of the vector to the bottom of the vector, each of the six elements corresponds to six features (shown in Table 1) at the same time point such that the first group represents the first time point, etc., as shown in Fig. 5. The kernel size in the first convolutional layer is 6×1 and the stride is 6, which means that the kernel simultaneously computes six features for each operation, as shown in Fig. 6. Two convolutional layers with 1×1 kernels follow to enhance the ability of feature learning. The 1st fully connected layer integrates the features after

convolution, and the 2nd fully connected layer computes the predicted results. ReLU is chosen as the activation function.

2) MODEL EVALUATION

The model is evaluated by the root-mean-square error (RMSE), goodness of fit (R^2), adjusted goodness of fit (adjusted R^2), scoring function, mean absolute percentage error (MAPE) and relative accuracy (RA) [56].

RMSE is an index that measures the difference between the predicted value and the actual value. As RMSE approaches 0, the predicted result approaches the real value, and the prediction becomes more accurate. The expression is

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (12)$$

where y_i represents the actual value, and \hat{y}_i represents the predicted value.

R^2 is an excellent criterion that is applied in statistics to evaluate the fit of regression models. The value R^2 generally varies within [0, 1]. The closer the value to 1, the better the fit of the model. The value of the function is close to 0 when the model fits poorly. The expression is

$$R^2 = 1 - \frac{\sum_{i=0}^n (\hat{y}_i - y_i)^2}{\sum_{i=0}^n (\bar{y} - y_i)^2} \quad (13)$$

where \bar{y} represents the mean, and n represents the number of samples in the data.

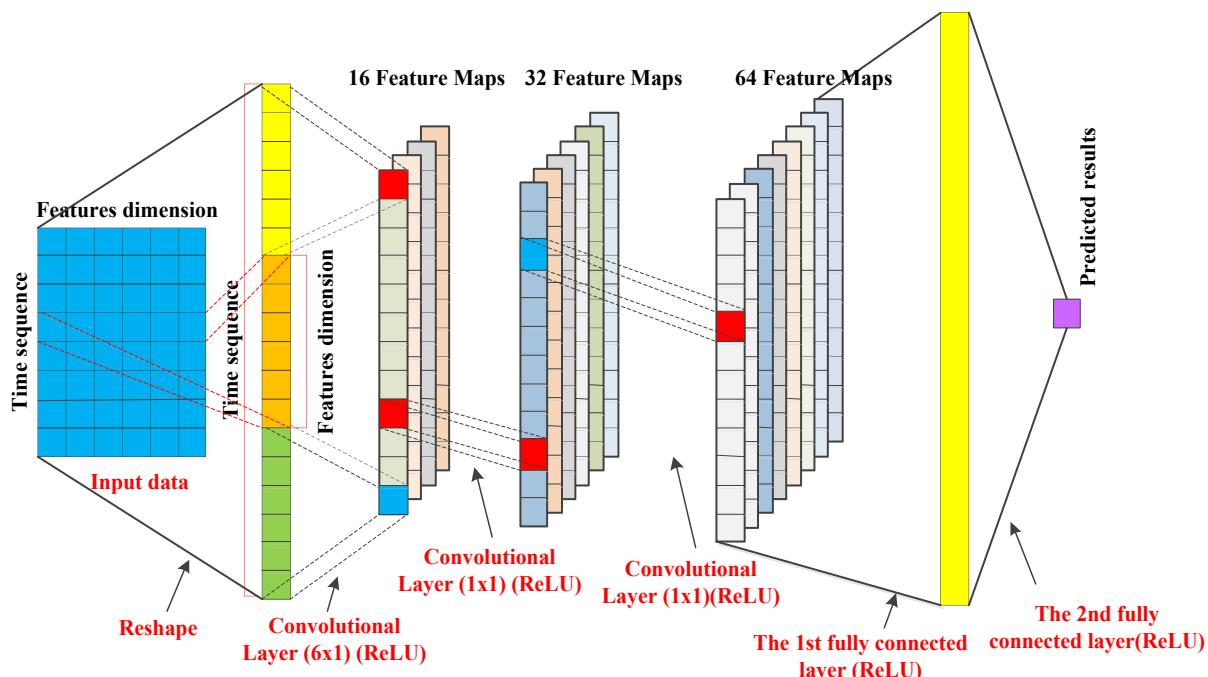


Fig. 5. Structure of DCNN.

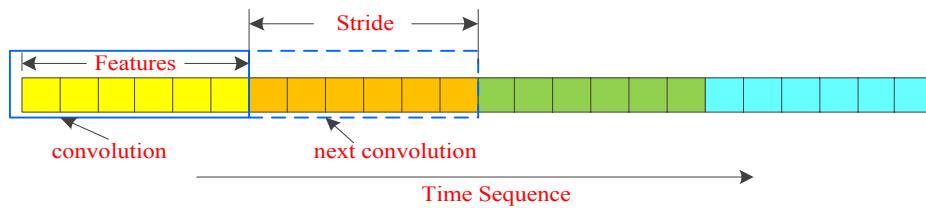


Fig. 6. Operation of the 1st convolution layer.

According to formula (13), when the number of samples is relatively large, R^2 increases, so the goodness of fit may not accurately represent the performance of the evaluation model, especially in the case of large sample sizes. To overcome this problem, an adjusted goodness of fit is proposed to avoid the influence of the number of data samples on the model evaluation. The expression is

$$\text{Adjusted_}R^2 = 1 - (1 - R^2) * \frac{n - 1}{n - p - 1} \quad (14)$$

where n represents the number of samples in the data, and p represents the number of features in the data.

The scoring function applies a large penalty to overestimated values and a small penalty to underestimated values, such that the closer the score to 0, the more accurate the results, expressed as

$$\text{score} = \begin{cases} \sum_{i=1}^N \left(e^{\frac{d_i}{13}} - 1 \right) & \text{if } d_i < 0 \\ \sum_{i=1}^N \left(e^{\frac{d_i}{10}} - 1 \right) & \text{if } d_i \geq 0 \end{cases} \quad (15)$$

$$d_i = \widehat{\text{RUL}}_i - \text{RUL}_i$$

where $\widehat{\text{RUL}}_i$ represents the predicted RUL of the i th time point, and RUL_i represents the actual RUL of the i th time point.

MAPE considers not only the error between the predicted value and the actual value, but also the proportion of the error to the actual value as follows:

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - \widehat{y}_i|}{y_i} \quad (16)$$

RA is defined as a measure of error in RUL prediction relative to the actual RUL, such that the closer the value of RA to 1, the better the prediction effect [55]:

$$\text{RA} = 1 - \frac{|y_i - \widehat{y}_i|}{y_i} \quad (17)$$

Generally, to ensure adequate work safety, the end of life of equipment, which determines the value of RUL, is normally set to a value that is less than its failure set point. The end of life of equipment is always affected by future production requirements, maintenance time, trend projection and extrapolation, etc. [56]

4. Experiment

4.1. DataSet description

The bearing dataset employed in the experiment to form a contrast and improve the validation is obtained from NASA IMS. There are three testing datasets in the IMS file, each of which contains operational data from monitoring four bearings. The first dataset contains the monitoring results for each bearing in the horizontal and vertical directions, and the second and third datasets contain the monitoring results of the bearing in only one direction. During the monitoring process, the bearings rotate

at a constant rate of 2000 r/min driven by an alternating current motor. Data are collected every 10 min [57] at a frequency of 20.48 kHz [58]. Each dataset records the signals from the normal operation of the bearing to the end of its life.

This study investigated the second IMS dataset, which contains monitoring information for the operation of four bearings. Each bearing is equipped with a single sensor that records the vibration of 20,480 points each time [57,59]. The RMS trend of each bearing is shown in Fig. 7.

The abscissa represents the time point of bearing operation; the interval between each time point is 10 min; and the ordinate is the RMS value that corresponds to different time points. The analysis shows that the degradation trend of each bearing is different. With the increase in the operation time, the RMS value of bearing 1 changes the most. At the 600–800 time point, the RMS value of bearing 1 suddenly changes. After reaching the 800 time point, the RMS value of bearing 1 increases sharply. Compared with bearing 1, the RMS values of the other three bearings change slowly; that is, they degrade smoothly, which shows a different degradation trend of the four bearings.

4.2. Experimental results and analysis

1) DATA PREPROCESSING RESULTS

The 3 sigma denoising criterion is adopted to eliminate gross monitored errors in each group of monitoring data. After denoising, the frequency-domain signal is obtained by an FFT. Next, the time and frequency features are selected, and the RMS is obtained for each time point of the dataset and employed as the basis for measuring the bearing amplitude. The RUL of a bearing is predicted according to the historical operation data of the bearing by observing the current trend of the time and frequency features. The preprocessing results of the first bearing are shown in Fig. 8. The dataset used here consists of 984 sets of data—namely, 984 time points. The results are shown by the abscissa

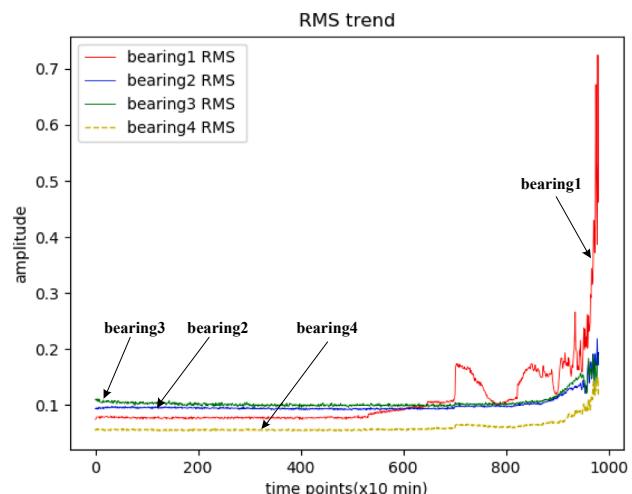


Fig. 7. RMS trend of each bearing.

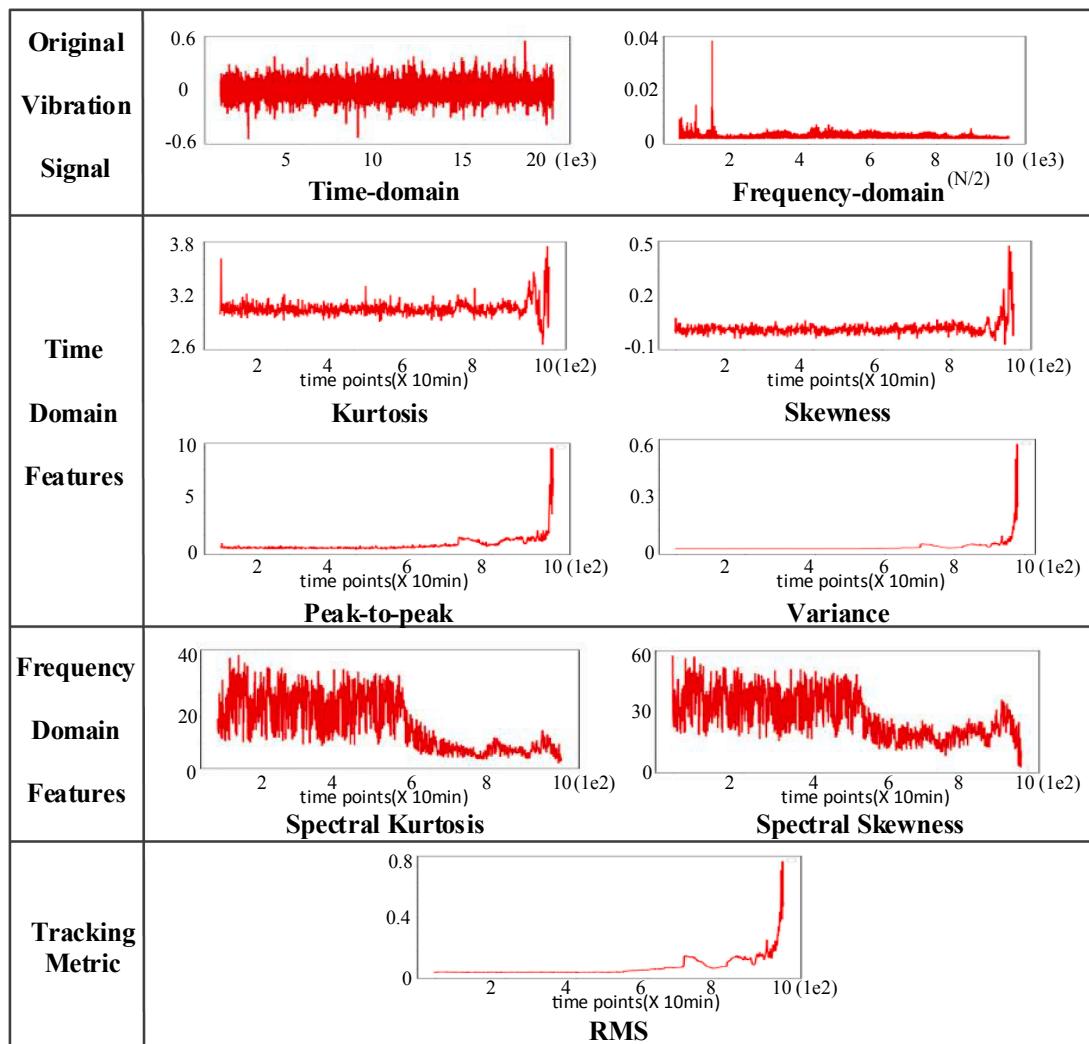


Fig. 8. Preprocessing results.

coordinate in Fig. 8.

2) COMPARISON MODELS

The RNN, LSTM, W-CNN and DCNN with pooling layers (normal DCNN, N-DCNN) are chosen as the comparison models. To make the comparison more convincing and effective, the optimal prediction results of each model are selected for comparison; thus, the division proportion of each model dataset is different.

a. RNN

To ensure comparison fairness, the RNN prediction model is constructed based on the TensorFlow framework. The data are divided into a training set and a testing set in an 8:2 ratio, and four RNN layers are constructed. The last layer is the fully connected layer. ReLU is applied as the activation function, the mean squared error (MSE) is regarded as the loss function, and the Adam optimization algorithm is adopted. The neurons of each layer are 33, 17, 6 and 1, and the dropout is 0.7.

b. LSTM

An LSTM prediction model with three hidden layers is constructed based on the TensorFlow framework. The dataset is divided into a training set and a testing set according to a 7:3 ratio. The last layer

applies ReLU as the activation function. MSE is applied as the loss function, and the Adam optimization algorithm is adopted. LSTM consists of four layers; the neurons in each layer are 35, 24, 11 and 1, and the dropout is 0.7.

c. WINDOW-CNN

The W-CNN network is constructed based on the TensorFlow framework as a comparison model for RUL prediction. The denoised data are input into the time window to establish a sequential relationship among the data. The time window sliding step size is set to 1, and the overlap ratio is 2/3. The dataset is divided into a training set and testing set according to a 7:3 ratio. The activation function of each convolutional layer is ReLU, and the Adam optimization algorithm is adopted. The structure of W-CNN is described as follows: the first layer is the convolutional layer (size: 3×1 , stride: 2×1), the 2nd layer is the pooling layer (size: 2×1 , stride: 2×1), the 3rd is the convolutional layer (size: 2×1 , stride: 2×1), and the 4th layer is the pooling layer (size: 2×1 , stride: 1×1).

d. NORMAL DCNN

The number of convolutional layers and the size of convolutional kernels for a normal DCNN network are the same as in the proposed DCNN, but the pooling layer follows each convolutional layer in normal

DCNN.

e. DCNN

For the DCNN proposed in this paper, the dataset is divided into a training set and a testing set based on the ratio 8:2.

3) MODEL VERIFICATION

The monitoring data of the first bearing in the second group of the IMS dataset are selected as the model training data. Three experiments are performed to validate the effectiveness of the proposed method. Experiment 1 evaluates the method of data denoising and the prediction model proposed in this paper, Experiment 2 evaluates the generalization ability of the proposed model, and Experiment 3 examines the validity of stratified sampling, where each experiment records the best prediction results.

a. Experiment 1

For the running data of the first bearing, the predictions of five models (RNN, LSTM, W-CNN, normal DCNN and DCNN) are compared using different data preprocessing methods without denoising and with sliding median denoising, hard threshold denoising, soft threshold denoising, SVD and 3 sigma criterion denoising to determine the impact of denoising methods on prediction accuracy.

1. Data without denoising

According to the qualitative analysis shown in Fig. 9, the prediction curves of the RNN, LSTM and normal DCNN models considerably differ from the true curves, and the fit of the W-CNN and DCNN models is ideal. In terms of the 95% confidence interval and predicted error, the minimum fluctuation occurs with DCNN. Combined with the RMS

change trend of bearing 1 in Fig. 7, the sudden change in RMS occurs in the later period of operation, which causes an increase in the bearing operation instability, a gradual worsening of the fitting of the RNN and LSTM prediction curve and actual curve, an increase in the absolute value of error, and a lesser effect on the change in the DCNN prediction curve by the change in the bearing instability.

The quantitative analysis in Table 3 shows that the optimal values for each evaluation index are $\text{RMSE} = 0.01445$, $R^2 = 0.95846$, $\text{Adjusted } R^2 = 0.95812$, $\text{Score} = 0.23231$, $\text{MAPE} = 0.04378$, and $\text{RA} = 0.96622$, which are achieved by the DCNN model except for RMSE. In particular, the R^2 and $\text{Adjusted } R^2$ values clearly demonstrate the advantages of the proposed method compared with the other models.

The comprehensive qualitative and quantitative analyses indicate that when the data are not denoised, the predicted values of the DCNN model are closest to the true values.

2. Sliding median denoising

The qualitative analysis of the prediction results in Fig. 10 shows clear differences between the predicted curve and the actual curve of the RNN, LSTM and normal DCNN models. In the later stage of bearing operation, the RMS value suddenly changes; that is, the instability increases, as shown in Fig. 7. Thus, the areas of the 95% confidence interval and predicted error are larger. The prediction curves of W-CNN and DCNN are better but the 95% confidence interval and predicted error from DCNN occur in smaller ranges. DCNN can still obtain more accurate prediction values when the bearing operates unstably in the later stage.

The quantitative analysis in Table 4 indicates that the optimal values for each evaluation index are $\text{RMSE} = 0.01360$, $R^2 = 0.96455$, $\text{Adjusted } R^2 = 0.96419$, $\text{Score} = 0.08439$, $\text{MAPE} = 0.03647$, and $\text{RA} = 0.98353$. Compared with the evaluation index values of the W-CNN model, the evaluation index values of the DCNN model are close to the ideal values (illustrated in Section 3.2.2).

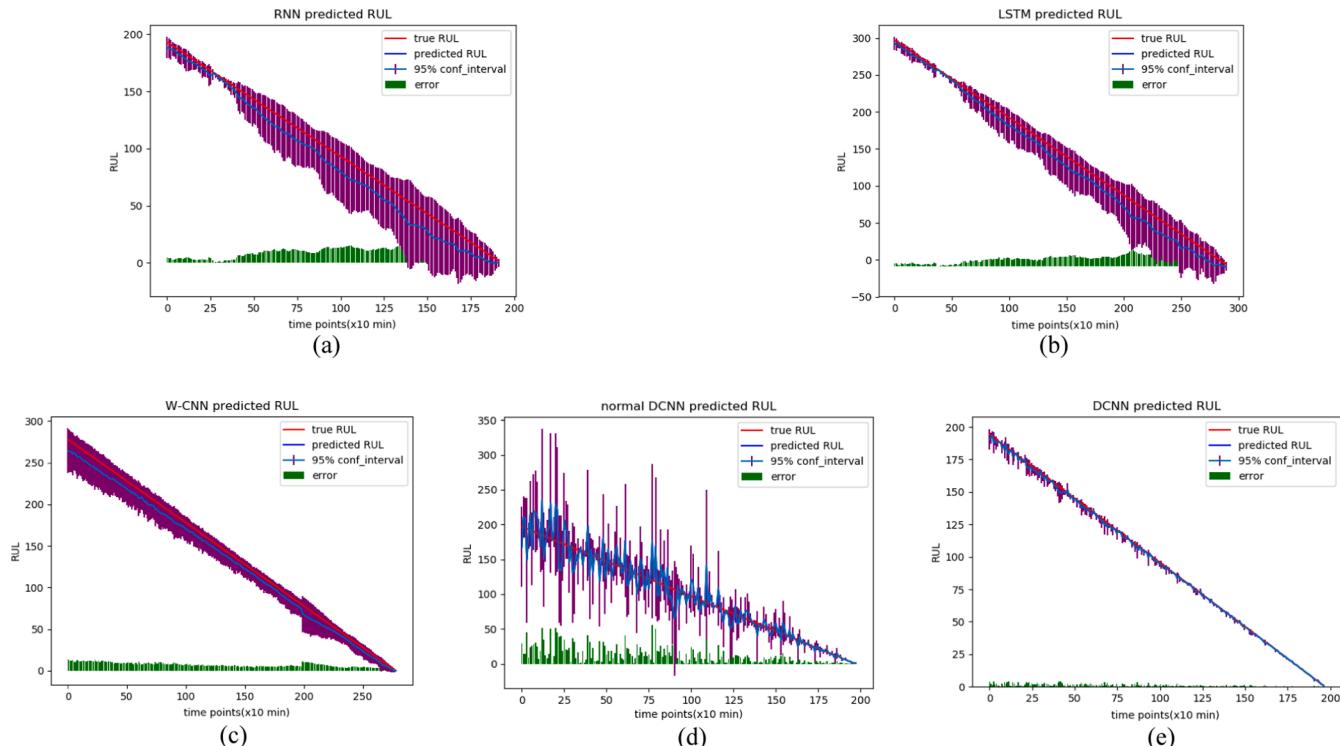
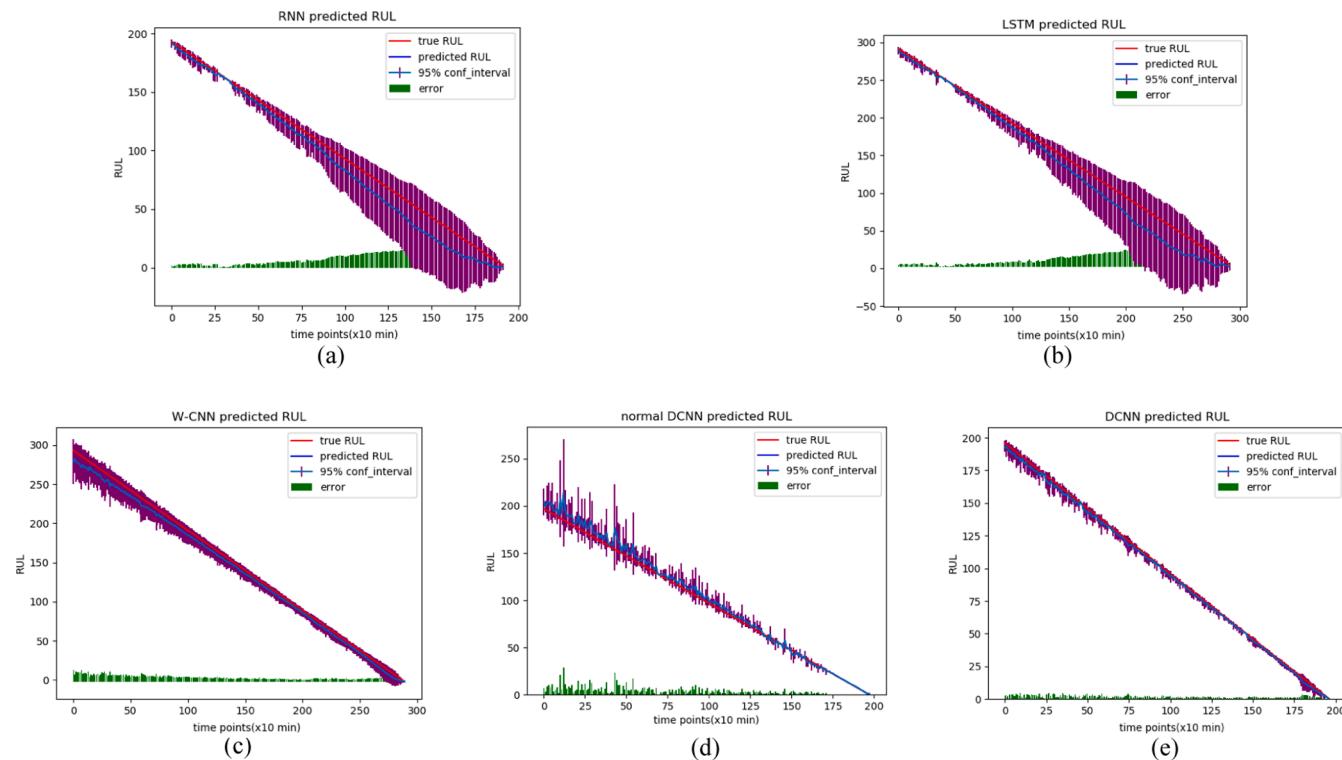


Fig. 9. Prediction results of five models without denoising. The blue line represents the prediction trend, and the red line represents the actual trend. (a) RNN prediction result, (b) LSTM prediction result, (c) W-CNN prediction result, (d) normal DCNN predicted result, and (e) DCNN prediction result. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3

Evaluation indexes for each prediction model.

Models	RMSE	R ²	Adjusted_R ²	Score	MAPE	RA
RNN	0.04649	0.86942	0.86706	0.84124	0.17613	0.75512
LSTM	0.04570	0.84356	0.84162	1.51304	0.18176	0.74806
W-CNN	0.02926	0.93444	0.93312	0.26876	0.09093	0.93115
N-DCNN	0.01445	0.93899	0.93151	0.25037	0.05724	0.95962
DCNN	0.01818	0.95846	0.95812	0.23231	0.04378	0.96622

**Fig. 10.** Prediction results of the five models with sliding median denoising. (a) RNN prediction, (b) LSTM prediction, (c) W-CNN prediction, (d) normal DCNN predicted, and (e) DCNN prediction.**Table 4**

Evaluation indexes for each prediction model.

Models	RMSE	R ²	Adjusted_R ²	Score	MAPE	RA
RNN	0.05402	0.81629	0.81067	0.84980	0.20015	0.80443
LSTM	0.06371	0.71503	0.70102	1.10866	0.25046	0.62761
W-CNN	0.01878	0.95463	0.95363	0.16019	0.08336	0.93641
N-DCNN	0.02053	0.93016	0.92887	0.37122	0.10948	0.89699
DCNN	0.01360	0.96455	0.96419	0.08439	0.03647	0.98353

The comprehensive qualitative and quantitative analyses reveal that the predicted values of DCNN are closest to the expected values when sliding median denoising is performed.

3. Hard threshold denoising

Hard threshold denoising is effective in improving the accuracy. As shown in Fig. 11, the predicted curves are fit better than the previous denoising results, which illustrates that the range of the 95% confidence interval decreases, and the error between the predicted and the actual values is also smaller, both of which indicate that the prediction accuracy of each model has been improved. All the predicted results from the qualitative analysis are close. The quantitative analysis in Table 5 shows that DCNN can achieve the optimal values of each evaluation index

except for the Score index, because the score function gives a large penalty for overestimation and a small penalty for underprediction (as shown in equation (15)). The optimal values are RMSE = 0.05753, $R^2=0.99382$, Adjusted_R² = 0.99338, Score = 0.11500, MAPE = 0.04597, and RA = 0.95403.

The comprehensive analysis shows that DCNN can obtain the ideal predicted result with hard threshold denoising and that continuous convolution operations are helpful for improving the prediction ability of the model.

4. Soft threshold denoising

The large ranges of the 95% confidence interval are derived from the RNN, W-CNN and normal DCNN prediction in Fig. 12, which indicates

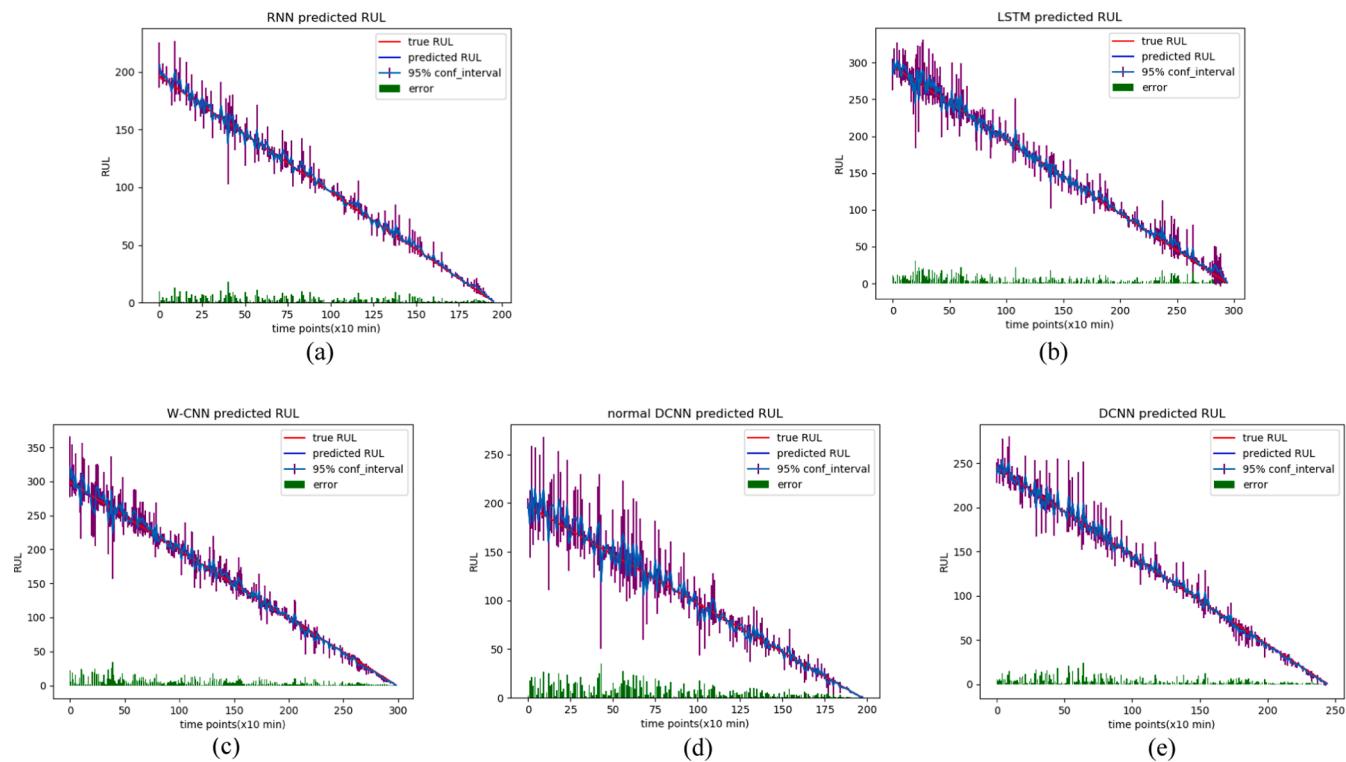


Fig. 11. Prediction results of the five models with hard threshold denoising. (a) RNN prediction, (b) LSTM prediction, (c) W-CNN prediction, (d) normal DCNN predicted, and (e) DCNN prediction.

Table 5

Evaluation indexes for each prediction model.

Models	RMSE	R ²	Adjusted_R ²	Score	MAPE	RA
RNN	0.06675	0.85055	0.84953	0.11734	0.07778	0.92919
LSTM	0.08419	0.85873	0.85776	0.11500	0.07589	0.91618
W-CNN	0.07915	0.95199	0.95035	0.29335	0.04766	0.95233
N-DCNN	0.08064	0.91848	0.90865	0.27112	0.05498	0.93636
DCNN	0.05753	0.99382	0.99338	0.14581	0.04597	0.95403

that the prediction curves of these models fit worse than the other models. The errors of LSTM and DCNN are smaller, which shows that both can obtain better prediction results. The prediction errors of RNN, LSTM and W-CNN increase with increasing bearing running instability. The normal DCNN and DCNN can obtain a more precise prediction value at a later stage of bearing operation.

The quantitative analysis from Table 6 shows that the values of the evaluation index, which are shown in bold font, are closer to the ideal results. By DCNN, RMSE = 0.05287, R^2 =0.99441, Adjusted_ R^2 = 0.99437, Score = 0.14785, MAPE = 0.03404, and RA = 0.96597.

The qualitative and quantitative analyses show that DCNN has better prediction ability with soft threshold denoising.

5. Singular value decomposition (SVD) denoising

Fig. 13 shows that the gentlest range of the 95% confidence interval is obtained from the W-CNN prediction, which illustrates that the fluctuation in the prediction curve is the gentlest. However, DCNN and W-CNN are almost equal in terms of the whole areas of the 95% confidence interval and predicted errors. The RNN, LSTM and normal DCNN models result in the larger areas of the 95% confidence interval and predicted errors; this phenomenon indicates that the prediction curves of RNN and LSTM substantially differ from the actual curves. The prediction errors of RNN, LSTM and normal DCNN are larger, while the time is near the end of the bearing life.

Table 7 shows the values of the evaluation index for each prediction model. The best evaluation indexes are RMSE = 0.03814, R^2 =0.99285, Adjusted_ R^2 = 0.99281, Score = 0.14089, MAPE = 0.02894, and RA = 0.98318 achieved by the DCNN, except for RMSE and MAPE, which means that the gentler fluctuation of the predicted curve produces better values of RMSE and MAPE.

The comprehensive analyses in Fig. 13 and Table 7 show that DCNN has better prediction ability with SVD denoising.

6. 3 sigma criterion denoising

Fig. 14 shows that the prediction curves of the RNN and LSTM models substantially differ from the actual curves, which indicates low model prediction accuracy. Conversely, the prediction curves of W-CNN, normal DCNN and DCNN fit the actual curve, and the smaller areas of the 95% confidence interval and predicted error indicate higher prediction accuracy. At the later stage of bearing operation, the fault occurs frequently and is overlapped, while the instability of the bearing operation increases, which produces the larger prediction error from RNN and LSTM.

The quantitative analysis in Table 8 shows that the optimal values for each evaluation index are RMSE = 0.00525, R^2 = 0.99762, Adjusted_ R^2 = 0.99760, Score = 0.11116, MAPE = 0.00558, and RA = 0.99442, which are all achieved by the DCNN model.

According to the qualitative and quantitative analyses for the three

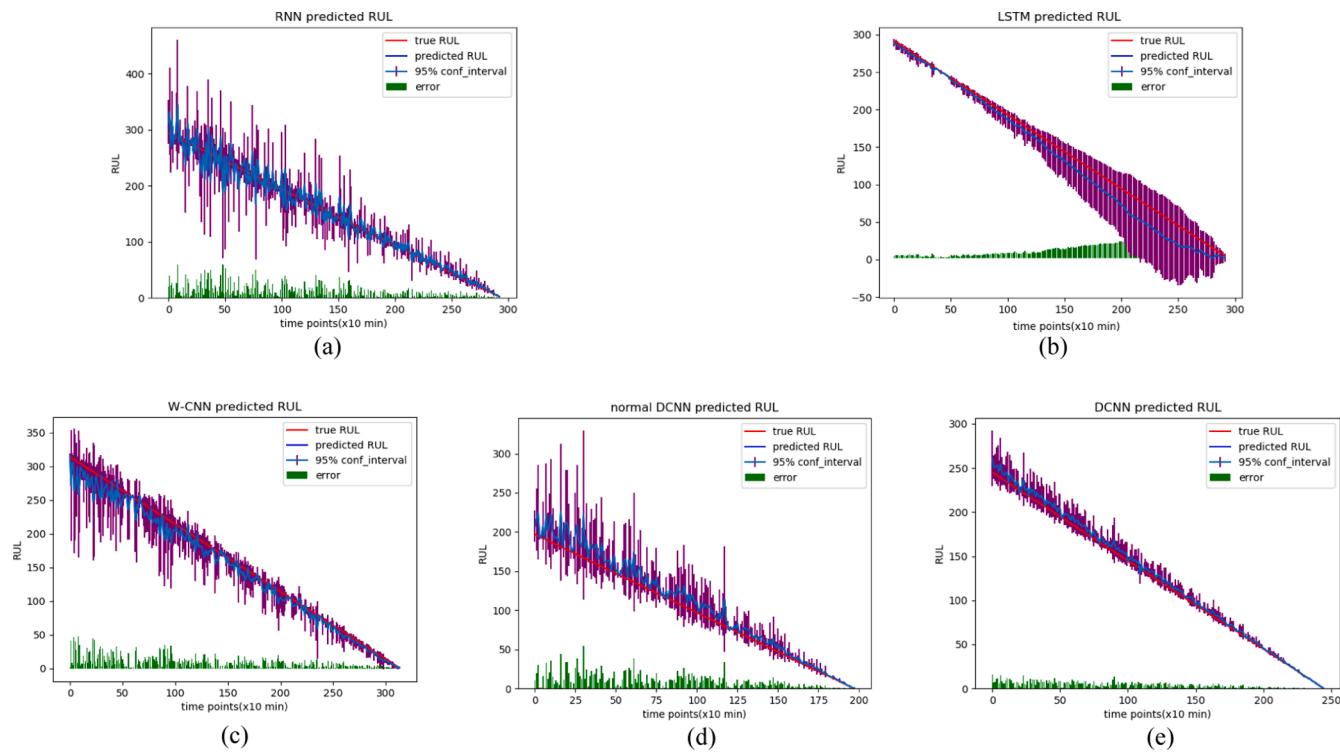


Fig. 12. Prediction results of the five models with soft threshold denoising. (a) RNN prediction, (b) LSTM prediction, (c) W-CNN prediction, (d) normal DCNN prediction, and (e) DCNN prediction.

Table 6
Evaluation indexes for each prediction model.

Models	RMSE	R ²	Adjusted_R ²	Score	MAPE	RA
RNN	0.16936	0.89590	0.88959	0.19090	0.09087	0.90915
LSTM	0.05725	0.98545	0.98128	0.16948	0.05087	0.94100
W-CNN	0.13034	0.95932	0.95190	0.26009	0.07486	0.92514
N-DCNN	0.13936	0.92168	0.91939	0.20498	0.07245	0.92892
DCNN	0.05287	0.99441	0.99437	0.14785	0.03404	0.96597

methods of data preprocessing, the predicted values of DCNN are closest to the actual values with 3 sigma criterion denoising.

According to Fig. 9 to Fig. 14 and Tables 3–8, compared with the prediction results of the RNN, LSTM, W-CNN and normal DCNN models, DCNN can achieve the best prediction results with different data denoising methods. To further compare the influence of the denoising methods on the prediction results of the model, the evaluation index values of the model with the best prediction effect (DCNN model) and various denoising methods are selected for comparison. The results are shown in Table 9.

It can be determined from Table 9 that the denoising method is helpful for improving the prediction accuracy of the models. Compared with the corresponding evaluation index values of DCNN with various denoising methods, it is determined that the best evaluation index value of DCNN can be achieved by using 3 sigma criterion denoising. The best value of each evaluation index is shown in bold font. The Score of the DCNN is the best when sliding median denoising is employed. By comparing Fig. 9(d) to Fig. 14(d), it can be determined that the prediction curve of DCNN is lower than the actual curve with sliding median denoising, as shown in Fig. 10(d). The prediction value is lower than the actual value, the model shows under prediction, the Score function gives a small penalty for underprediction, and the Score value is the best at this time.

According to the comprehensive comparison, the best prediction result can be obtained by using 3 sigma criterion denoising combined

with DCNN.

b. Experiment 2

The degradation trend of each bearing is different as shown in Fig. 7. Thus, the operation data of the second, third and fourth bearings are selected as the experimental testing data to verify the generalization ability of the five models and evaluate the performance of the model dependence on each dataset. Experiment 1 showed that 3 sigma criterion denoising maximizes the prediction accuracy of DCNN. To verify the generalization ability of the model with the best prediction performance (DCNN with 3 sigma denoising), 3 sigma criterion denoising is still adopted in Experiment 2, and the parameters and structures of the five models are consistent with those in Experiment 1 when the 3 sigma criterion is applied for denoising. The generalization ability of the five models is verified by the three different datasets.

As shown in Fig. 15, the differences between the prediction curves and the actual curves of the RNN and LSTM models are obvious, and the ranges of the 95% confidence interval are larger; that is, the models' prediction accuracies are low. At the end of the bearing life, the prediction errors are large, especially for bearing 3 and bearing 4, because the degradation trends of these four bearings are different, as shown by the RMS trends in Fig. 7. For bearings with different degradation trends, RNN and LSTM can achieve poor prediction values, which illustrates that the generalization abilities of these two models are worse.

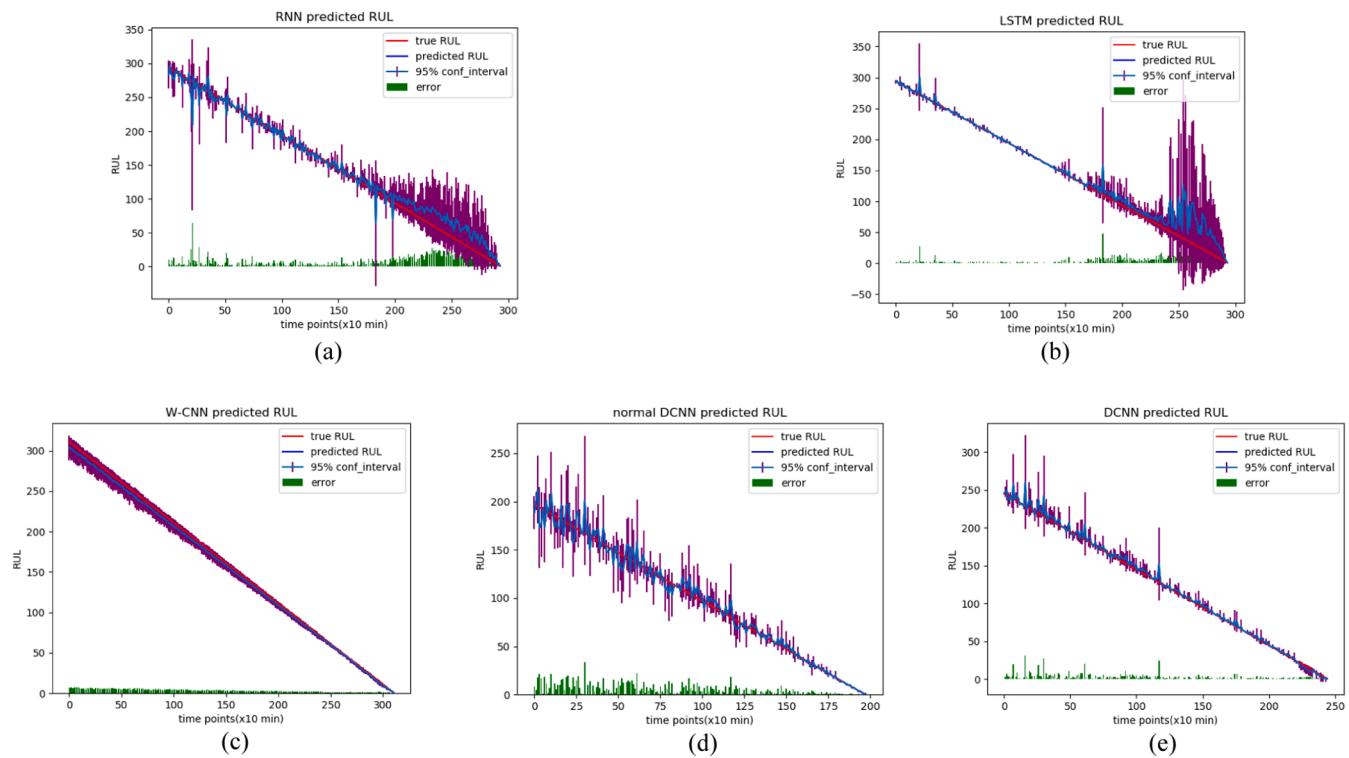


Fig. 13. Prediction results of the five models with SVD denoising. (a) RNN prediction, (b) LSTM prediction, (c) W-CNN prediction, (d) normal DCNN prediction, and (e) DCNN prediction.

Table 7
Evaluation indexes for each prediction model.

Models	RMSE	R ²	Adjusted_R ²	Score	MAPE	RA
RNN	0.12788	0.97729	0.97714	0.89219	0.19964	0.80035
LSTM	0.17556	0.95721	0.95692	1.40562	0.29364	0.70636
W-CNN	0.03814	0.98196	0.97984	0.19593	0.02894	0.97106
N-DCNN	0.09415	0.96523	0.96011	0.53838	0.16699	0.89267
DCNN	0.05072	0.99285	0.99281	0.14089	0.04681	0.98318

The prediction curves of W-CNN and N-DCNN show agreement with the actual curves for bearing 2 and bearing 3, and the areas of the 95% confidence interval and prediction error are smaller. However, for bearing 4, the prediction curve trend of W-CNN is different from the actual curve, and the 95% confidence interval and prediction error are obvious. Thus, the prediction accuracy and generalization abilities of W-CNN and N-DCNN are not satisfactory.

For DCNN, the predicted curves are almost the same as the actual curves, and the 95% confidence interval and prediction error are small, although in the later stage, the degradation trends of bearings (shown in Fig. 7) are obviously different, and the prediction error of DCNN is still small, which indicates high prediction accuracy.

The evaluation indexes of each model for different bearings are shown in Table 10. The optimal evaluation index values for different bearings are shown in bold font; all are achieved by DCNN.

A large gap exists between the evaluation index values of RNN and LSTM and that of the other models; that is, the prediction results are the worst. The analysis results of Table 10 are consistent with those of Fig. 15(a) and (b). Thus, the prediction abilities of RNN and LSTM are poor, which reflects the weak generalization ability.

The evaluation index values of W-CNN and N-DCNN are better than those of RNN and LSTM; the prediction ability of W-CNN and N-DCNN is also better; and the results from Table 10 are the same as those in Fig. 15. The DCNN model has the best evaluation index values for different

bearings compared with other models chosen for the experiment.

The comprehensive analysis results from Fig. 15 and Table 10 show that the prediction ability of RNN and LSTM is poor, and DCNN can obtain better prediction results for bearings with different degradation trends; that is, the generalization ability of DCNN is the best.

c. Experiment 3

From the previous two experiments, when the 3 sigma criterion is employed for data denoising, stratified sampling is utilized for dataset partitioning, DCNN is the prediction model, and the best accuracy and generalization ability of the prediction results are achieved. To verify the influence of stratified sampling on the prediction results of the model, the dataset is divided according to the traditional time series data partitioning method, which is based on the ratio 8:2, while keeping the denoising method and the prediction model unchanged. The dataset of four bearings is applied.

Comparisons from Fig. 14(e), Fig. 16(a), Fig. 15 Bearing 2(e), Fig. 16(b), Fig. 15 Bearing 3(e), Fig. 16(c), Fig. 15 Bearing 4(e) and Fig. 16(d) show that the prediction ability of the model drops sharply when the traditional time series data partitioning method is applied. The model can learn the data characteristics from the training dataset. Therefore, in the early stage of prediction, the predicted value of the model fits well with the actual value. However, in the later stage of the bearing

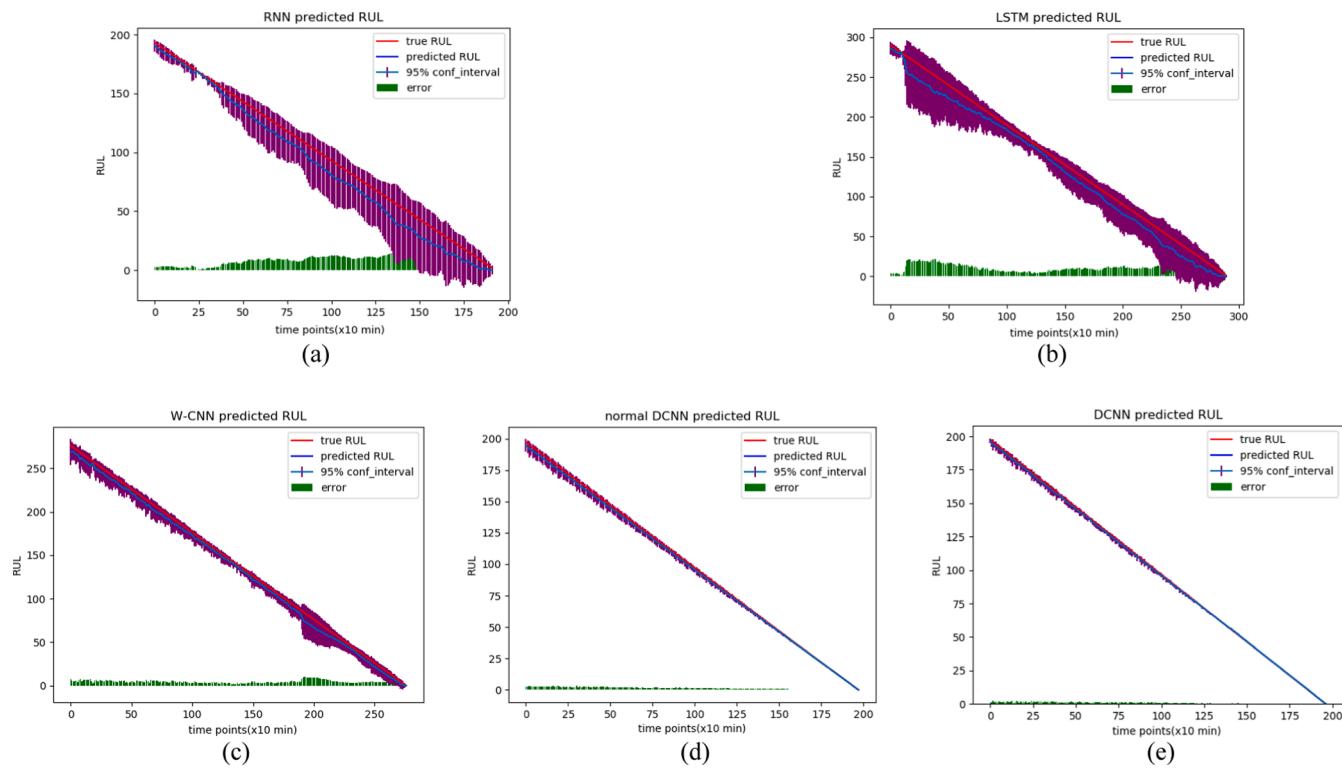


Fig. 14. Prediction results of the five models with 3 sigma criterion denoising. (a) RNN prediction, (b) LSTM prediction, (c) W-CNN prediction, (d) normal DCNN prediction, and (e) DCNN prediction.

Table 8
Evaluation indexes for each prediction model.

Models	RMSE	R ²	Adjusted_R ²	Score	MAPE	RA
RNN	0.04523	0.89045	0.88977	0.84072	0.16927	0.82133
LSTM	0.05402	0.87159	0.86953	2.87601	0.21764	0.79189
W-CNN	0.01315	0.98753	0.98738	0.18968	0.03368	0.98327
N-DCNN	0.01169	0.97462	0.96988	0.21210	0.09444	0.98589
DCNN	0.00525	0.99762	0.99760	0.11116	0.00558	0.99442

Table 9
Evaluation indexes for DCNN WITH DIFFERENT DENOISING METHODS.

Denoising method	RMSE	R ²	Adjusted_R ²	Score	MAPE	RA
Without denoising	0.01818	0.95846	0.95812	0.23231	0.04378	0.96622
Sliding median	0.01360	0.96455	0.96419	0.08439	0.03647	0.98353
Hard threshold	0.05753	0.99382	0.99338	0.14581	0.04597	0.95403
Soft threshold	0.05287	0.99441	0.99437	0.14785	0.03404	0.96597
SVD	0.05072	0.99285	0.99281	0.14089	0.04681	0.98318
3sigma denoising	0.00525	0.99762	0.99760	0.11116	0.00558	0.99442

operation, the fluctuation of the data increases, and these data do not exist in the training set (shown in Fig. 7). Thus, the model cannot carry out feature learning, which causes poor prediction accuracy in the later stage of the model. The evaluation index of the model prediction performance with each dataset is shown in Table 11.

These results illustrate that stratified sampling can improve the prediction accuracy and enhance the generalization ability of the model.

4) RESULTS ANALYSIS

The analysis of the results of Experiment 1, as shown in Tables 3–8, indicate that the predictive evaluation index values of the DCNN model are closest to the ideal values; therefore, the prediction accuracy of the

DCNN model is the highest. The prediction evaluation index values of the five models are improved by 3 sigma criterion denoising preprocessing, which shows that the 3 sigma criterion has the best denoising effect. Therefore, the 3 sigma criterion denoising preprocessing method combined with the DCNN model optimizes the four evaluation indexes, and the optimal values are RMSE = 0.00525, R² = 0.99762, Adjusted_R² = 0.99760, and Score = 0.11116.

The analysis of the results of Experiment 2, as shown in Tables 9 and 10, compares the corresponding evaluation index values of each model. The corresponding evaluation scales of the RNN, LSTM, W-CNN and N-DCNN models change substantially, which indicates that RNN, LSTM, W-CNN and N-DCNN depend heavily on the dataset and are vulnerable to the influence of the data structure and have weak generalization

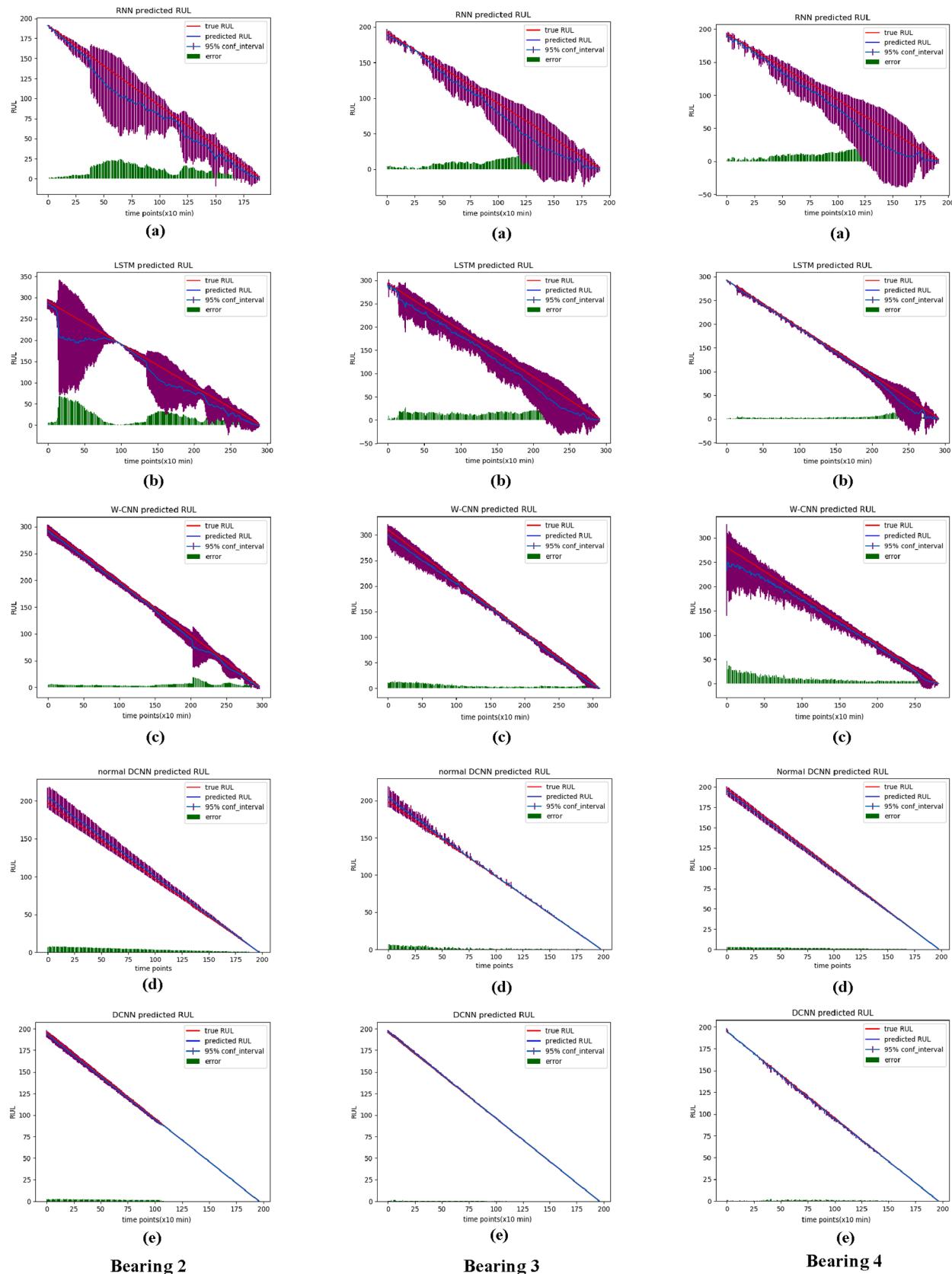
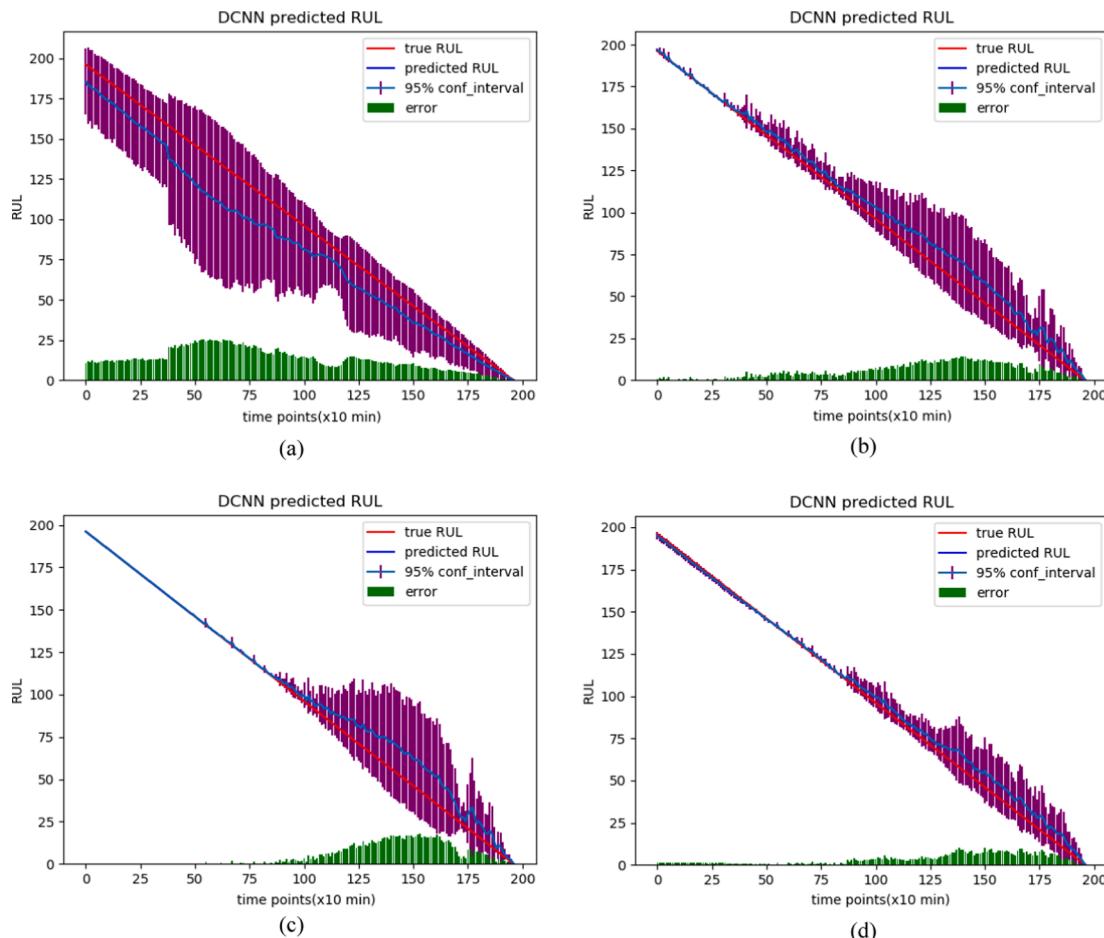


Fig. 15. Prediction results of the five models with 3 sigma criterion denoising of the second bearing operation data. (a) RNN model, which is the same as the RNN model with 3 sigma criterion denoising in Experiment 1, (b) LSTM model, which is the same as the counterpart with 3 sigma criterion denoising in Experiment 1, (c) W-CNN model, which is the same as the counterpart with 3 sigma criterion denoising in Experiment 1, (d) normal DCNN model, which is the same as the counterpart with 3 sigma criterion denoising in Experiment 1, and (e) DCNN model, which is the same as the counterpart with 3 sigma criterion denoising in Experiment 1.

Table 10

Evaluation indexes of each prediction model for different bearings.

Models	Bearing	RMSE	R ²	Adjusted_R ²	Score	MAPE	RA
RNN	Bearing- 2	0.09422	0.78507	0.77975	1.46681	2.92509	0.28350
LSTM		0.09298	0.69056	0.68759	2.25156	3.09458	0.21462
W-CNN		0.02053	0.97152	0.97122	0.22664	0.26471	0.75399
N-DCNN		0.01624	0.97649	0.97623	0.28353	0.27047	0.78165
DCNN		0.00732	0.99502	0.99498	0.12082	0.06575	0.98872
RNN	Bearing- 3	0.12654	0.73580	0.73011	1.71596	2.82767	0.31924
LSTM		0.41063	0.75473	0.74509	2.60768	1.23618	0.55281
W-CNN		0.02764	0.96119	0.96082	0.41557	0.26107	0.84510
N-DCNN		0.01444	0.97251	0.96924	0.32095	0.18233	0.94401
DCNN		0.01084	0.99345	0.99339	0.21061	0.05975	0.99016
RNN	Bearing- 4	0.05573	0.91098	0.90971	0.89462	2.51832	0.21104
LSTM		0.03680	0.94875	0.94553	0.95170	1.56312	0.83582
W-CNN		0.02707	0.96111	0.96066	0.49921	1.86387	0.55251
N-DCNN		0.09774	0.98263	0.97875	0.18002	0.10161	0.90122
DCNN		0.00772	0.99548	0.99544	0.13116	0.08924	0.99418

**Fig. 16.** Prediction results with traditional time series data partitioning method. (a) Bearing 1, (b) Bearing 2, (c) Bearing 3, and (d) Bearing 4.**Table 11**

DCNN evaluation indexes with different bearing data.

Data	RMSE	R ²	Adjusted_R ²	Score	MAPE	RA
Bearing 1	12.8243	0.35684	0.34015	1.97053	6.05754	0.19403
Bearing 2	23.9602	0.13573	0.12788	3.06619	8.44318	0.16466
Bearing 3	9.24223	0.15221	0.15003	3.19507	5.28914	0.13902
Bearing 4	4.19556	0.24178	0.23052	2.41967	3.23409	0.27007

ability. The change in the value of each evaluation index, which corresponds to the DCNN model, is the smallest. This finding shows that the DCNN model has strong adaptability to the dataset and the best generalization ability. The ranges of the evaluation index values for the DCNN model predictions are listed as follows: RMSE (0.00525, 0.01084), R^2 (0.99345, 0.99762), Adjusted_ R^2 (0.99339, 0.99760), Score (0.11116, 0.21061), MAPE (0.00558, 0.08924), and RA (0.98872, 0.99442). The DCNN model has the smallest variation range and the best generalization ability.

The analysis of the results of Experiment 3, as shown in Fig. 16 and Table 11, illustrate that stratified sampling can enable the model to completely learn the data features, improve the prediction accuracy and enhance the generalization ability of the model.

Based on this analysis, the method of 3 sigma criterion preprocessing and stratified sampling combined with the DCNN model improves the prediction accuracy and enhances the generalization ability.

5. Conclusions

Based on the strong learning and feature recognition ability of CNN, this paper proposes a method for RUL prediction for bearings based on CNN theory, which has high predictive ability and strong generalization ability. The 3 sigma criterion method for preprocessing, which effectively removes gross monitored errors from the data and improves the model prediction accuracy, achieves values of evaluation indexes that are closer to the ideal results than those produced by other denoising methods. Stratified sampling is employed to construct the training set and testing set, and ensures that the prediction model completely learns the data features, which is different from the traditional time series data partitioning method. Therefore, the prediction accuracy of DCNN with this method is better than traditional time series partitioning. To avoid feature loss, the DCNN model without the pooling layer, which consists of three convolutional layers and two fully connected layers, is constructed. The model verification experiment shows that the denoising method based on the 3 sigma criterion and stratified sampling combined with DCNN has the highest prediction accuracy and strongest generalization ability.

In real applications, multisensor and multifeature monitoring are usually adopted to improve RUL prediction accuracy. To adapt the model for multifeature datasets, we plan to apply CNN feature extraction to build predictive models and achieve multifeature input prediction while ensuring satisfactory prediction accuracy and model generalization ability.

CRediT authorship contribution statement

Hua Ding: Funding acquisition, Investigation, Project administration, Conceptualization. **Liangliang Yang:** Conceptualization, Data curation, Methodology, Software, Validation, Writing - original draft, Writing - review & editing. **Zeyin Cheng:** Writing - original draft, Writing - review & editing. **Zhaojian Yang:** Funding acquisition, Project administration, Formal analysis, Resources.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work is supported by Shanxi Science and Technology Foundation Condition Platform Project under grant 201805D141002 and Shanxi Key Research & Development Projects under grant 201903D121064.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.measurement.2020.108878>.

References

- [1] F.T. Wang, X.F. Liu, G. Deng, et al., Remaining life prediction method for rolling bearing based on the long short-term memory network, *Neural Process Lett.* 50 (3) (2019) 2437–2454.
- [2] S. Khan, T. Yairi, A review on the application of deep learning in system health management, *Mech. Syst. Signal Process.* 107 (2018) 241–265.
- [3] P. Kundu, A.K. Darpe, M.S. Kulkarni, Weibull accelerated failure time regression model for remaining useful life prediction of bearing working under multiple operating conditions, *Mech. Syst. Signal Process.* 134 (2019), 106302.
- [4] J. Sun, H. Zuo, W. Wang, et al., Application of a state space modeling technique to system prognostics based on a health index for condition-based maintenance, *Mech. Syst. Signal Process.* 28 (2012) 585–596.
- [5] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, D. Siegel, Prognostics and health management design for rotary machinery systems-reviews, methodology and applications, *Mech. Syst. Signal Process.* 42 (2014) 314–334.
- [6] Y. Le, N. Li, S. Gontarz, et al., A model-based method for remaining useful life prediction of machinery, *IEEE Trans. Reliab.* 65 (2016) 1314–1326.
- [7] N. Li, Y. Lei, J. Lin, et al., An improved exponential model for predicting remaining useful life of rolling element bearings, *IEEE Trans. Ind. Electron.* 62 (2015) 7762–7773.
- [8] X. Zhang, H. Yan, Image denoising and enhancement algorithm based on median filtering and fractional order filtering, *J. Northeastern Univ. (Nat. Sci.)* 41 (4) (2020) 482–487.
- [9] S. Zhang, Y. Zhang, H. Liu, An improved EMD-based hard thresholding denoising algorithm, *Computer Measure. Control.* 22 (11) (2014) 3659–3661.
- [10] D.L. Donoho, De-noising by soft-thresholding, *IEEE Trans. Inf. Theory* 41 (3) (2002) 613–627.
- [11] X. Fan, T. Wu, Y. Zhi, et al., Denoising method for raman imaging data based on singular value decomposition and median absolute deviation, *Spectrosc. Spectral Anal.* 40 (2) (2020) 436–440.
- [12] H. Yan, X. Deng, Application of median filter in digital image denoising, *Computer Era.* 2 (2020) 47–49.
- [13] G. Yan, N. Zhou, S. Zhang, De-noising method of EEG signal based on MMTD and wavelet hard-threshold, *J. Syst. Simul.* 30 (4) (2008) 1490–1495.
- [14] J. Tian, S. Song, Processing method for mine wire rope damage signal based on improved particle swarm optimization wavelet threshold, *Coal Eng.* 52 (4) (2020) 103–107.
- [15] Z. Qiu, S. Yue, J. Yue, et al., Noise reducing for bridge monitoring data based on singular value decomposition method, *Geotechn. Invest. Surv.* 45 (012) (2017) 36–39.
- [16] M. Pecht, J. Gu, Physics-of-failure-based prognostics for electronic products, *Trans. Inst. Meas. Control.* 31 (2009) 309–322.
- [17] F.O. Heimes, Recurrent Neural networks for remaining useful life estimation, in: Proceedings of International Conference on Prognostics and Health Management, Denver, 1–6 October 2008.
- [18] O.F. Eker, F. Camci, I.K. Jennions, Major challenges in prognostics: study on benchmarking prognostics datasets, in: European Conference of the Prognostics and Health Management, Dresden, Germany, pp. 148–155, July 2012.
- [19] R. Zhao, R. Yan, J. Wang, Learning to monitor machine health with convolutional bi-directional LSTM networks, *Sensors* 17 (2) (2017) 273.
- [20] M. Jouin, R. Gouriveau, D. Hissel, M.C. Pera, N. Zerhouni, Degradation analysis and aging modeling for health assessment and prognostics of PEMFC, *Reliab. Eng. Syst. Saf.* 148 (2016) 78–95.
- [21] M. Jouin, R. Gouriveau, D. Hissel, M.C. Pera, N. Zerhouni, Particle filter-based prognostics: review, discussion and perspectives, *Mech. Syst. Signal Process.* 72–73 (2016) 2–31.
- [22] J.B. Ali, B.C. Morello, L. Saidi, S. Malinowski, Accurate bearing remaining useful life prediction based on Weibull distribution and artificial neural network, *Mech. Syst. Signal Process.* 56–57 (2015) 150–172.
- [23] X. Li, Q. Ding, J.Q. Sun, Remaining useful life estimation in prognostics using deep convolution neural networks, *Reliab. Eng. Syst. Saf.* 172 (2018) 1–11.
- [24] T. Benkedjouh, K. Medjaher, N. Zerhouni, S. Rechak, Remaining useful life estimation based on nonlinear feature reduction and support vector regression, *Eng. Appl. Artif. Intell.* 26 (2013) 1751–1760.
- [25] Z. Lei, Fault prognostics algorithm based on multivariate relevance vector machine and time series iterative prediction, *Procedia Eng.* 29 (2012) 678–686.
- [26] S.H. Shin, S.R. Kim, Y.H. Seo, Development of a fault monitoring technique for wind turbines using a hidden Markov model, *Sensors*. 18 (2018) 1790.
- [27] D. Pan, J.B. Liu, J. Cao, Remaining useful life estimation using an inverse Gaussian degradation model, *Neurocomputing*. 185 (2016) 64–72.
- [28] J. Chen, M.B. Salim, M. Matsumoto, A Gaussian mixture model-based continuous boundary detection for 3D sensor networks, *Sensors*. 10 (2010) 7632–7650.
- [29] Z. Tian, An artificial neural network method for remaining useful life prediction of equipment subject to condition monitoring, *J. Intell. Manuf.* 23 (2012) 227–237.
- [30] X. Zhang, L. Xiao, J. Kang, Degradation prediction model based on neural network with dynamic windows, *Sensors*. 15 (2015) 6996–7015.

- [31] R. Huang, L. Xi, X. Li, C.R. Liu, H. Qiu, J. Lee, Residual life predictions for ball bearings based on self-organizing map and back propagation neural network methods, *Mech. Syst. Signal Process.* 21 (2007) 193–207.
- [32] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, pp. 580–587, June 2014.
- [33] G. Yang, J. Yang, W. Sheng, F.E.F. Junior, S. Li, Convolutional neural network-based embarrassing situation detection under camera for social robot in smart homes, *Sensors.* 18 (2018) 530.
- [34] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, pp. 770–778, June 2016.
- [35] S.M.S. Islam, S. Rehman, M.M. Rehman, E.K. Dey, M. Shoyaib, Application of deep learning to computer vision: a comprehensive study, in: Proceedings of International Conference on Informatics, Electronics and Vision, Dhaka, Bangladesh, pp. 592–597, May 2016.
- [36] N. Kruger, P. Janssen, S. Kalkan, M. Laptev, Deep hierarchies in the primate visual cortex: what can we learn for computer vision, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (2013) 1847–1871.
- [37] N.G. Polson, V.O. Sokolov, Deep learning for short-term traffic flow prediction, *Transport. Res. Part C.* 79 (2017) 1–17.
- [38] C.J. Huang, P.H. Kuo, A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities, *Sensors* 18 (2018) 2220.
- [39] M. Baur, P. Albertelli, M. Monno, A review of prognostics and health management of machine tools, *Int. J. Adv. Manuf. Technol.* 107 (2020) 2843–2863.
- [40] L. Zhang, J. Lin, B. Liu, et al., A review on deep learning applications in prognostics and health management, *IEEE Access* 7 (2019) 162415–162438.
- [41] Y. LeCun, Y. Bengio, G. Hinton, Deep learning, *Nature* 521 (7553) (2015) 436–444.
- [42] J. Deutsch, D. He, Using deep learning based approach to predict remaining useful life of rotating components, *IEEE Trans. Syst. Man Cybern. Syst.* 48 (2018) 11–20.
- [43] P. Malhotra, V. V, A.R.G. Anand, L. Vig, P. Agarwal, G. Shroff, Multi-sensor prognostics using an unsupervised health index based on LSTM encoder-decoder, arXiv 2016, arXiv:1608.06154.
- [44] H. Qiao, T. Wang, P. Wang, S. Qiao, L. Zhang, A time-distributed spatiotemporal feature learning method for machine health monitoring with multi-sensor time series, *Sensors* 18 (2018) 2932.
- [45] J. Ma, H. Su, W. Zhao, B. Liu, Predicting the remaining useful life of an aircraft engine using a stacked sparse autoencoder multilayer self-learning, *Complexity* 2018 (2018) 1–13.
- [46] Z. Zhao, B. Liang, X. Wang, W. Lu, Remaining useful life prediction of aircraft engine based on degradation pattern learning, *Reliab. Eng. Syst. Saf.* 164 (2017) 74–83.
- [47] Y. Qian, R. Yan, R.X. Gao, A multi-time scale approach to remaining useful life prediction in rolling bearing, *Mech. Syst. Signal Process.* 83 (2017) 549–567.
- [48] J. Deutsch, M. He, D. He, Remaining useful life prediction of hybrid ceramic bearings using an integrated deep learning and particle filter approach, *Appl. Sci.* 7 (2017) 649.
- [49] L. Guo, N. Li, F. Jia, Y. Lei, J. Lin, A recurrent neural network based health indicator for remaining useful life prediction of bearings, *Neurocomputing.* 240 (2017) 98–109.
- [50] H. Ding, L. Yang, Z. Yang, A predictive maintenance method for shearer key parts based on qualitative and quantitative analysis of monitoring data, *IEEE Access* 7 (2019) 108684–108702.
- [51] J. Yoo, J.G. Beak, A novel image feature for the remaining useful lifetime prediction of bearings based on continuous wavelet transform and convolutional neural network, *Appl. Sci.* 8 (2018) 1102.
- [52] L. Ren, Y. Sun, H. Wang, L. Zhang, Prediction of bearing remaining useful life with deep convolution neural network, *IEEE Access* 6 (2018) 13041–13049.
- [53] Z. Chen, Y. Ni, Enhanced generalization of nonparametric model for magnetorheological dampers, *J. Vibrat. Shock.* 36 (06) (2017).
- [54] S. Mao, X. Pu, Y. Cheng, *Brief Course of Probability Theory and Mathematical Statistics*, Higher Education Press, 2012.
- [55] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, J. Lin, Machinery health prognostics: a systematic review from data acquisition to RUL prediction, *Mech. Syst. Signal Process.* 104 (2018) 799–834.
- [56] A. Saxena, J. Celaya, B. Saha, et al., Metrics for offline evaluation of prognostics performance, *Int. J. Prognost. Health Manage.* 1 (2010) 4–23.
- [57] AFNOR, Condition monitoring and diagnostics of machines-prognostics-Part1: General guidelines, NF ISO 13381-1: 2015, 5–8.
- [58] J. Lee, H. Qiu, G. Yu, J. Lin, Rexnord technical services: bearing dataset, Moffett Field, CA: IMS, Univ. Cincinnati. NASA Ames Prognostics Data Repository, NASA Ames, 2007.
- [59] W. Gousseau, J. Antoni, F. Girardin, J. Griffaton, Analysis of the rolling element bearing dataset of the center for intelligent maintenance systems of the University of Cincinnati, in: The Thirteenth International Conference on Condition Monitoring and Machinery Failure Prevention Technologies, Paris, France, 1–13, October2016.