



A novel multi-scale CNN and attention mechanism method with multi-sensor signal for remaining useful life prediction

Xingwei Xu ^{*}, Xiang Li, Weiwei Ming ^{*}, Ming Chen

State Key Laboratory of Mechanical Systems and Vibration, Shanghai Jiao Tong University, Shanghai 200240, China
School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China



ARTICLE INFO

Keywords:

Remaining useful life
Deep learning
Tool wear prediction
Turbofan RUL prediction

ABSTRACT

Remaining useful life prediction is crucial in smart manufacturing systems due to many advantages of early prognostics, i.e., downtime reduction, service time prolongation, ultimate work efficiency improvement, and cost-saving. However, the conventional methods highly depend on feature selection and extraction, the accuracy and generalization can not be guaranteed. Inspired by the development of deep learning, a new method is proposed. Firstly, a parallel one-dimensional convolutional neural network (CNN) and the pooling layer were developed to extract and fuse features from the multiple signals. The dilated convolution with the residual connection and attention mechanism were specially developed to deal with the features from the pooling layers. After that, the regression layer was designed to generate the remaining useful life (RUL). Moreover, to verify the prognostic performance of the proposed method, two experiments, including cutting tool wear prediction and turbofan engine RUL prediction, were conducted. The proposed model was compared with the related works, and the results showed that the new method was more robust and accurate than currently published methods.

1. Introduction

As an essential part of intelligent manufacturing, RUL prediction has been deeply developed and spread in mechanical industries. Accurate prognostic can make it possible to identify equipment problems at an early stage to reduce downtime, prolong service time, and ultimately enhance safety (Chen et al., 2019; Wang et al., 2019). Generally, the main methods for RUL prognostic are based on physical methods and data driving methods (Cai et al., 2021; Deng et al., 2020; Ren et al., 2017). Recently, with the development of artificial intelligence, parallel computing, and sensor technology, data-driven methods for RUL prognostic are more and more popular (Souza et al., 2021).

Therefore, the data-driven methods are widely applied to RUL prognostic with its procedure to build the data-driven model using the monitoring data from sensors (e.g. vibration, force, or power) through machine learning methods or deep learning (DL) methods. The typical procedure of machine learning for RUL prognostic is composed of two parts, i.e., feature extraction&selection and feature recognition (Javed et al., 2016; Mao et al., 2020), and many researchers have achieved extraordinary outcomes. For example, in (Kumar et al., 2019), the HMM and polynomial regression methods had utilized for estimating the

cutting tool wear state and RUL. In (Jun-Hong et al., 2009), time-domain and frequency-domain features with a total of 16 features have been extracted and selected to predict the tool wear state. Wang et al. (Wang et al., 2017) utilized the support vector regression model and multi-sensor data fusion to perform classification and regression of the tool wear. Kong et al. (Kong et al., 2019) developed an algorithm based on the KPCA_IRBF technique and used relevance vector machine model to predict the remaining useful life of the cutting tool.

Besides, DL has been widely used in RUL prognostic due to its powerful abilities in terms of feature learning and automatic feature extraction from raw data without any domain knowledge (Che et al., 2020; Jia et al., 2022; Shao et al., 2018; Zhang et al., 2019). At present, Recurrent Neural Network (RNN) (Yu et al., 2019) and Long Short-Term Memory (LSTM) are the dominant methods in RUL prognostic. Chen et al. (Chen et al., 2020) utilized the convolutional RNN to build health indicators before predicting the RUL of bearings. Zhang et al. (Zhang et al., 2018) used LSTM to estimate the RUL of the machine, which achieved a good result. In (An et al., 2019), RNN was used for bearing intelligent fault diagnosis with considerably good results. Zhao et al. (Zhao, Zhang, et al., 2019) had proposed a creative method for RUL prediction and first employed the complete ensemble empirical mode

* Corresponding authors.

E-mail addresses: xuxingwei@sjtu.edu.cn (X. Xu), mingseas@sjtu.edu.cn (W. Ming).

decomposition to extract features from raw data before adopting LSTM for RUL prediction. Wu et al. (Wu et al., 2018) utilized vanilla LSTM neural networks for RUL prediction in different working conditions and strong noising environments.

All aforementioned methods had demonstrated great performance in prognostic. However, it is difficult to extract the representative features for health prognostic without domain knowledge. In addition, researchers have proven that many features may probably be only suitable for a particular condition at a certain stage with machine learning methods (Deng et al., 2021; Tran et al., 2012). The accuracy and generalization of the method could not be guaranteed due to the fact that different researchers may extract different features. Besides, feature extraction and selection are labor-extensive, time-consuming, and complicated.

DL methods are a powerful tool for prognostic because it is available

for automatic feature learning without domain knowledge. However, the point is that the conventional RNN and LSTM of deep learning do not take spatial information into consideration. In addition, recurrent architectures of RNN and LSTM have been discovered to be very difficult to train and introspect (Pascanu et al., 2013). Moreover, the researchers who worked on RNN demonstrated that the architecture for better results could not be easily figured out (Jozefowicz et al., 2015). These issues somehow affected the performance of RNN and LSTM on prediction.

Inspired by the Temporal Convolutional Network structure (Zhao, Che, et al., 2019; Zhao, Gao, et al., 2019), a new deep learning model was proposed for RUL prognostic. Firstly, a parallel 1D CNN was developed for the automatic feature extraction from multiple signals and the pooling layer is utilized to reduce the dimension of the extracted feature. Secondly, the dilated convolution with the residual connection

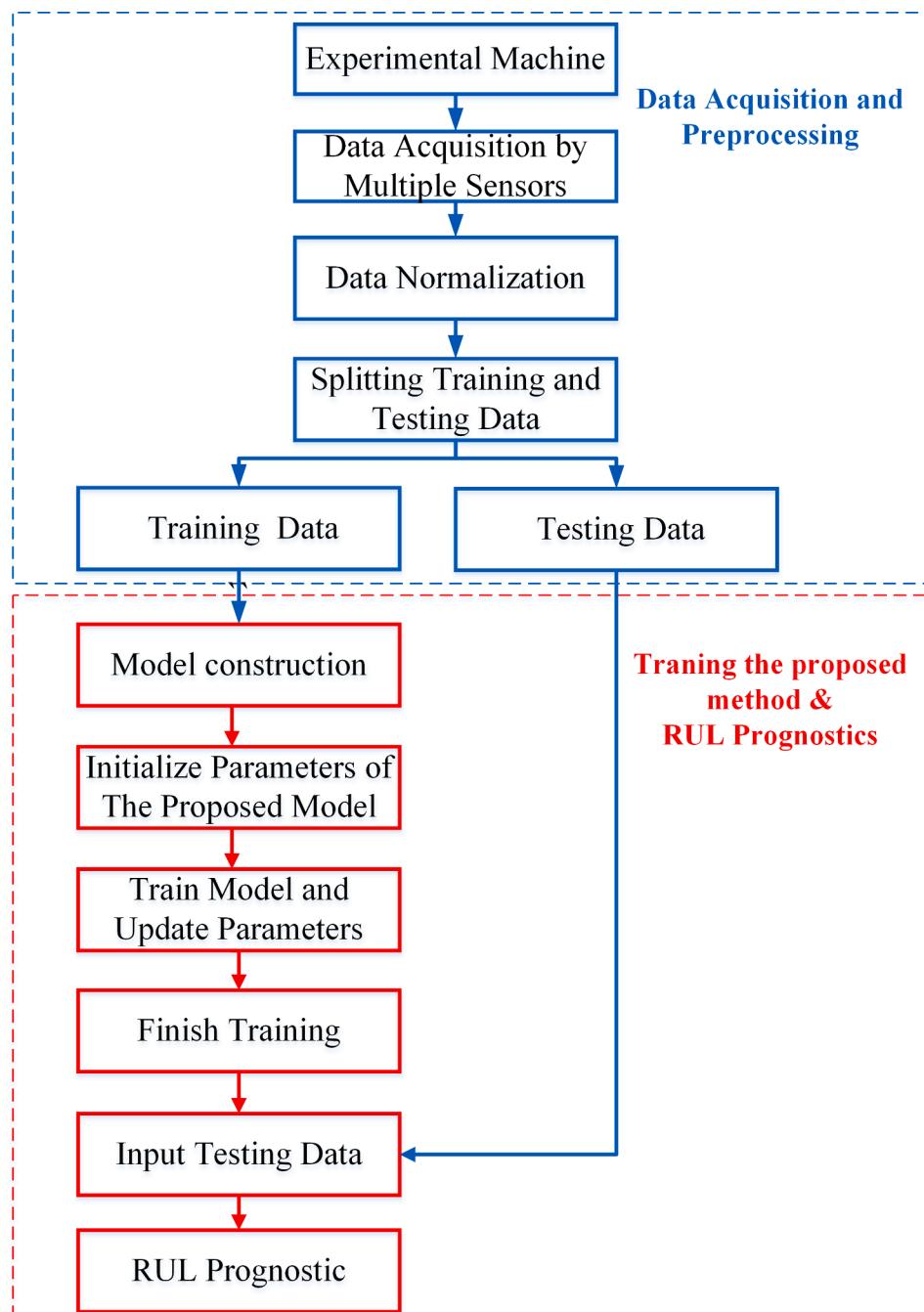


Fig. 1. Flow chart of the proposed method for RUL prediction.

is specially developed for RUL prognostic. Finally, two experiments were conducted to verify the proposed method. The main contributions of our work are listed as follows:

- (1) Instead of shallow machine learning, the proposed deep learning architecture can extract high-dimensional features adaptively and automatically without prior expertise. In addition, the developed signal pre-processing method combined with the 1D CNN has the function of automatic feature extraction and can achieve multi-sensor feature fusion.
- (2) The standard convolution will lead to an increasing number of network layers, and there is a problem of information loss during the polling operation. Thus, the developed new framework named deep residual dilated convolution can solve the problem of information loss and increase the receptive field.
- (3) Attention mechanisms are adopted in the proposed method for RUL prediction such that the proposed method can promote the features that are useful and suppress the features that are not useful.
- (4) Two experiments, including cutting tool wear prediction and turbofan engine RUL prediction, were conducted to validate the proposed method. The predicted results were compared with the related works, and the results showed that the proposed model was more robust and accurate than the current methods. In addition, the proposed method can widely be utilized in various machine health prognostics as long as retraining the proposed network.

2. Methodology

2.1. The procedure for RUL prediction

The procedure for RUL prediction with the proposed method is presented in Fig. 1. It mainly consists of two parts. The first part includes data acquisition and data pre-processing, while the second part consists of model construction and RUL prediction. In the first part, the input data are acquired through different sensors. And the input data are pre-processed before feeding into the proposed model. In the second part, the proposed model is constructed and trained by the input data for RUL prediction, both online and offline.

2.2. Convolutional neural networks

CNN has received much attention on various applications such as computer vision and target recognition (Krizhevsky et al., 2012). Among different network types of CNN, 2-dimensional CNN (Conv2D) is the most commonly used in computer vision due to its powerful capability to capture and extract 2D data features. However, one-dimensional CNN is more suitable for time series modelling X. Zhao et al., 2019 (Abdeljaber et al., 2017; Zhao, Jiang, et al., 2019). So the 1D CNN is utilized in this paper, and the basic structure of 1D CNN is composed of convolutional layers and pooling layers. Mathematically, the convolution operation can be written as:

$$X_j^l = \sum_{k \in M_j} x_k^{l-1} \cdot w_{kj}^l + b_j^l \quad (1)$$

$$u_j^l = \sigma(X_j^l) \quad (2)$$

where X_j^l represents the j th feature map of layer l , x^{l-1} represents the k output feature map of the former layer, w_{kj}^l represents the convolution operation between the k th feature map of the former layer, M_j represents the size of the input feature, and b_j^l represents the corresponding bias. The pooling operation can be expressed as:

$$P_j = \max_{x_j^l \in S} X_j^l \quad (3)$$

where S represents the pooling window size, and X represents the input feature map.

2.3. Dilated convolution

In sequence modeling, the input data from sensors are very long due to a rather high sampling frequency. When the input data is very long, the standard convolution will lead to an increasing number of network layers, which makes the model more complicated. In addition, the standard CNN gains a larger receptive field by adding a pooling layer. However, there is a problem of information loss after the pooling operation. To solve the problem of information loss and increase the receptive field, the dilated convolution is adapted. Suppose that the convolutional filters are $f = (f(1), f(2), \dots, f(k))$, the one-dimensional sequence input is $x \in \mathbb{R}^n$. So, the dilated convolution F on the element s of the sequence can be expressed as:

$$F(s) = (x *_d f)(s) = \sum_{i=0}^{k-1} f(i) \cdot x_{s-d \cdot i} \quad (4)$$

where d represents the dilation rate (DR), k represents the filter size, and $(s-d \cdot i)$ stands for the data of the previous layer and before the previous layer (Bai et al., 2018). Fig. 2 shows the diagram of dilated convolution and its dilation rate. As a particular structure of CNN, the dilated convolution injects holes into the standard convolution to enhance the receptive field. This is an extraordinary performance in that the convolution operations can be performed on different scales according to the dilation rate (DR). when the DR = 1, the dilated convolution is the ordinary convolution operation.

3. The proposed method

3.1. Multi-sensor signal pre-processing and feature fusion

At present, multi-sensor raw signal data are often the input for RUL prediction. Therefore, various data pre-processed techniques are proposed to extract useful features from raw signals. In the proposed method, the multi-sensor data are regarded as the metrics and normalized the data into [0,1]. Under this pattern, the raw signal x is decomposed into a sequence of T segments evenly. The features from the multi-sensor are extracted from each local window, and the procedure is depicted as follows:

- (1) In RUL prediction, we suppose there are n different sensors to acquire data, and the collected data from n sensors at time step t can be mathematically expressed as:

$$x_i(t) = [x_i^1(t), x_i^2(t), \dots, x_i^n(t)] \quad (5)$$

- (2) All the collected data are normalized into [0,1] by the following equation:

$$x_j = \frac{x_j - \text{Min}_x}{\text{Max}_x - \text{Min}_x} \quad (6)$$

where x_j represents the signal data from the j th sensor, and the Min_x and Max_x represent the minimum and maximum values of the input x , respectively.

- (3) Convert all the signal data into matrixes, and the equation can be written as:

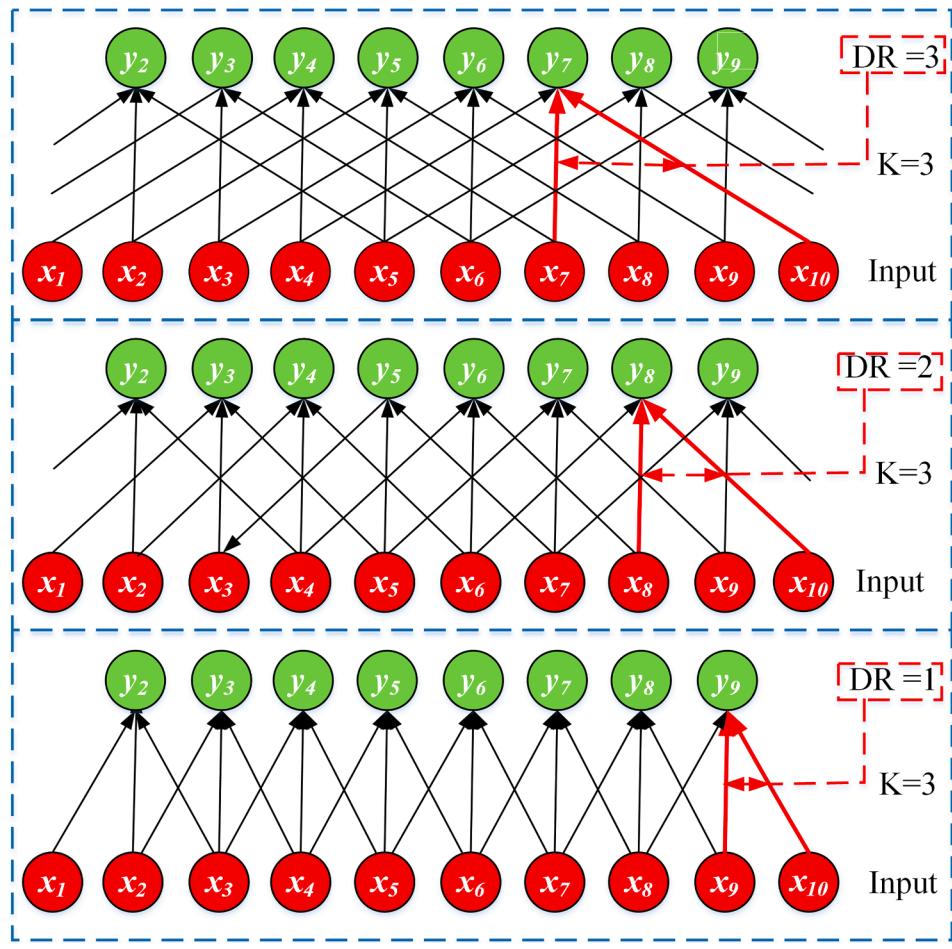


Fig. 2. Dilated convolution, (a) DR = 1, (b) DR = 2, (c) DR = 3.

$$x_i(t) = x_i(t) = \begin{bmatrix} x_{i1}^1 & x_{i1}^2 & \dots & x_{i1}^n \\ x_{i2}^1 & x_{i2}^2 & \dots & x_{i2}^n \\ \vdots & \vdots & \ddots & \vdots \\ x_{im}^1 & x_{im}^2 & \dots & x_{im}^n \end{bmatrix} \quad (7)$$

where m represents the length of each signal data, and n represents the number of the sensors.

(4) Finally, the convolution operation and pooling operation are performed to extract the representative feature from the converted matrixes by equation (1) and equation (4), and then, the extracted features are concatenated.

3.2. Deep residual dilated convolution and attention mechanism

Deep residual networks (DRN) demonstrated its excellent performance in classification for image recognition (He et al., 2016). Compared to conventional CNN, the DRN has a higher training speed and easier gradient transmission (Chen et al., 2019). Thus, the residual connection of DRN is adopted in our proposed architecture, which is written as:

$$H(x) = f(x) + x \quad (8)$$

where x represents the input, $f(x)$ represents a residual connection.

Then, we consider the advantage of DRN and dilated convolution, simultaneously. The multi-scale deep residual dilated convolution is proposed, and the architecture is presented in Fig. 3. In the proposed multi-scale structure, each residual connection consists of the dilated

convolution layer, Swish layer, and dropout layer.

After that, the attention mechanism is developed to consider the importance of learned features from the above steps so that the proposed method can automatically promote the features that are useful and suppress the features that are not useful. Mathematically, the importance of the extracted feature can be obtained as follows:

$$u_t = \tanh(W_u h_t + b_u) \quad (9)$$

where \tanh is the activation function, h_t is the extracted feature from the previous layer, and W_u and b_u are the dynamic weight matrix and bias. Then, all the obtained attention weights are normalized by using the following formula:

$$u_t = \frac{\exp(u_t^T)}{\sum_i^t \exp(u_i^T)} \quad (10)$$

By merging the weights of internal attention, the new representative feature of each time series data are obtained:

$$x_{new} = \sum_1^t x_t u_t = [u_1 \times x_1, u_2 \times x_2, \dots, u_t \times x_t] \quad (11)$$

3.3. Generating the prediction results

The representative feature pass to the regression layer to generate the prognostic results, which can be written as:

Deep residual dilated convolution

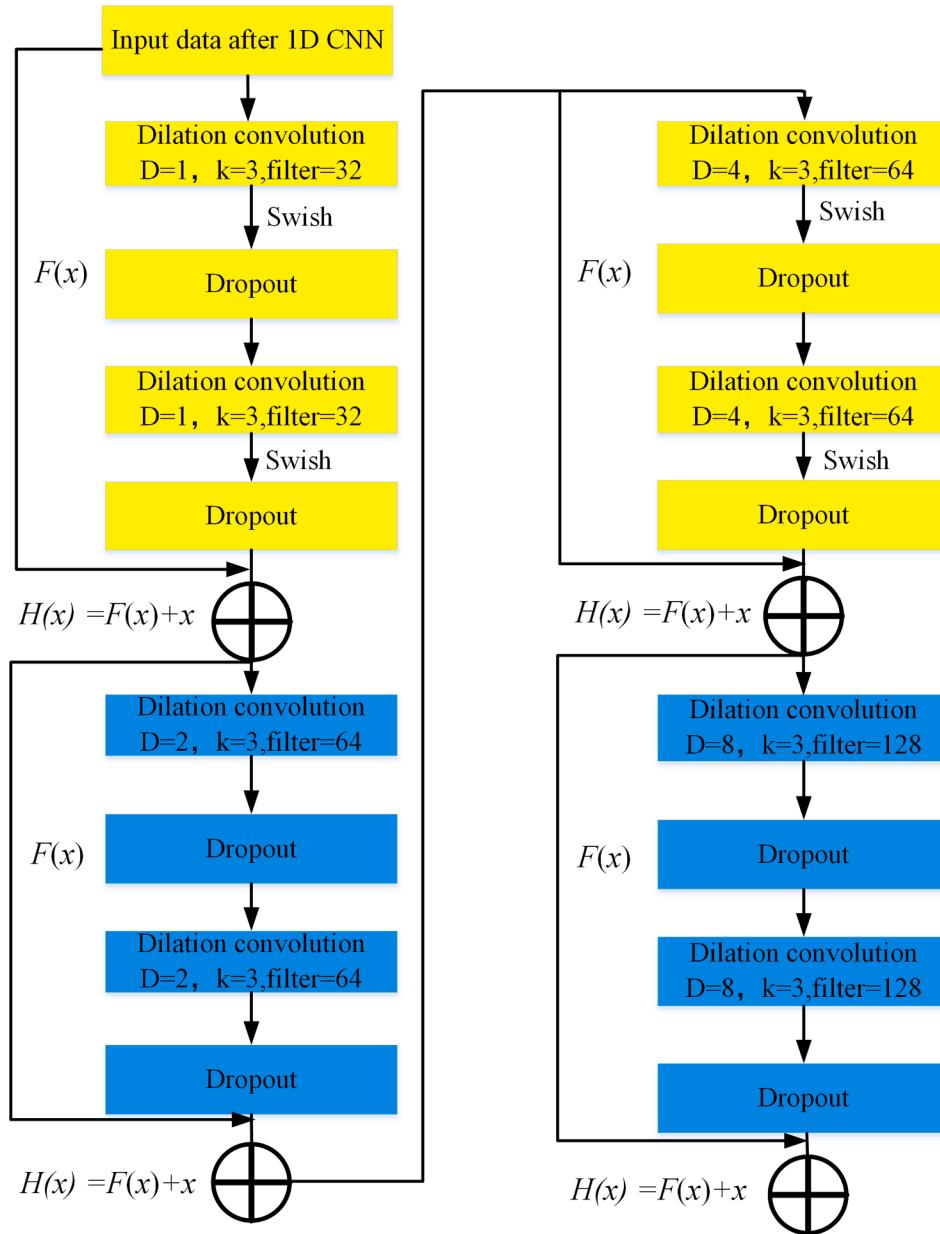


Fig. 3. The proposed multi-scale architecture of deep residual dilated convolution.

$$\begin{cases} z_j^l = W_{ji}^{l-1} x_i^{l-1} + b_j^{l-1} \\ y_i^{(l)} = \sigma(z_j^{(l)}) = \frac{1}{1 + e^{-x}} (z_j^{(l)}) \end{cases} \quad (12)$$

where x^{l-1} is the output vector of layer $l-1$, W^l is the weight matrix, and σ is the sigmoid function.

3.4. Tricks of the proposed method

To increase the nonlinearity of the proposed model, the nonlinear activation function swish layer is employed in our proposed method, as defined in Eq. (8):

$$\text{Swish}(x) = x \times \left(\frac{\beta}{1 + e^{-x}} \right) \quad (13)$$

where x is the input of the activation function, and β is the learning

parameter. To avoid the overfitting issues, the dropout layer is utilized, which is defined as (Srivastava et al., 2014):

$$\begin{cases} r_i^l \text{Bernoulli}(p) \\ \bar{K}_i^l = \frac{r_i^l * K_i^l}{p} \\ z_i^l(j) = \bar{K}_i^l * x_j^l \end{cases} \quad (14)$$

where p represents the Dropout rate, r_i^l follows Bernoulli distribution, and $z_i^l(j)$ represents the output of j th input x_j after convolution.

3.5. The proposed architecture

In this section, the proposed method is constructed for RUL prediction. The input data is the one-dimensional data from multi-sensor, and are normalized and reshaped into matrixes according to section3.1. The

main architecture and specific parameters of the proposed method are presented in Fig. 4. After the signal pre-processing, the convolutional layer and the pooling layer are used to extract features and reduce the dimension of features, respectively. Then, the dilated convolution with residual connections and attention mechanism are designed to process the extracted feature and acquire the importance of the extracted feature, respectively. Finally, the fully connected is built to generate the predicted results of RUL. Specifically. In the proposed architecture, the four layers of Conv1D and Max-Pooling are designed to extract features and reduce the feature dimension, respectively. The kernel size of all Conv1D is 3×3 , and the filters for four layers of Conv1D are 32, 64, 128, and 128, respectively. Then, the deep dilated convolution is designed with four residual connections. It should be noted that each residual connection has its dilation convolution, and the dilation factors $d = 1, 2, 4$, and 8, respectively. To make our model more suitable for solving more complicated problems and overfitting problems, the dropout and the swish layer are adopted in our model. Then the attention mechanism is utilized after all the convolutional layer and pooling layer. Finally, the fully connected neural networks are to generate prognostic results.

3.6. Loss function and evaluation metrics of the proposed model

After completing the design of the proposed structure, the proposed model needs to be trained by using the labeled dataset to obtain the inner parameters. The training of the developed model is to minimize the loss function by optimization algorithms, and the loss function can be expressed as:

$$J(w, b) = J(\theta) = \frac{\sum_{i=1}^n (y_i - y'_i)^2}{n} \quad (15)$$

where n represents the number of testing samples, y_i represents the true data, and y'_i represents the predicted data. In order to minimize the loss function to get the optimized w and b , the optimization algorithms need to be selected. The adaptive moment estimation (Adam) optimization algorithm has higher computational efficiency than the traditional

optimization algorithm. It can adjust the learning rate adaptively and have a higher training speed (Cai et al., 2020; Chen et al., 2019; Xu et al., 2020). Therefore, the Adam is selected as the optimization algorithm, and the updating process is expressed as:

$$\left\{ \begin{array}{l} v_{i+1} = \beta_1 v_i + (1 - \beta_1) \frac{\partial}{\partial W_{ji}^l} (\theta_i) \\ s_{i+1} = \beta_2 s_i + (1 - \beta_2) \nabla \frac{\partial}{\partial W_{ji}^l} (\theta_i^2) \\ v_{i+1}^c = \frac{v_{i+1}}{1 - \beta_1^i} \\ s_{i+1}^c = \frac{s_{i+1}}{1 - \beta_2^i} \\ \theta_{i+1} = \theta_i - \alpha \frac{v_{i+1}^c}{\sqrt{s_{i+1}^c + \epsilon}} \end{array} \right. \quad (16)$$

where θ represents the parameter w and b , α represents the learning rate, v_{i+1} represents a partial first-order moment estimate, s_{i+1} represents a partial second-order moment estimation, and the β_1 and β_2 represent exponential decay rates of moment estimation. In addition, β_1 and β_2 are often set to be 0.9 and 0.999.

The evaluation metrics: mean absolute error (MAE), root mean squared error (RMSE), and coefficient of determination (R^2) were adopted to evaluate the performance of the developed algorithm, as is defined as:

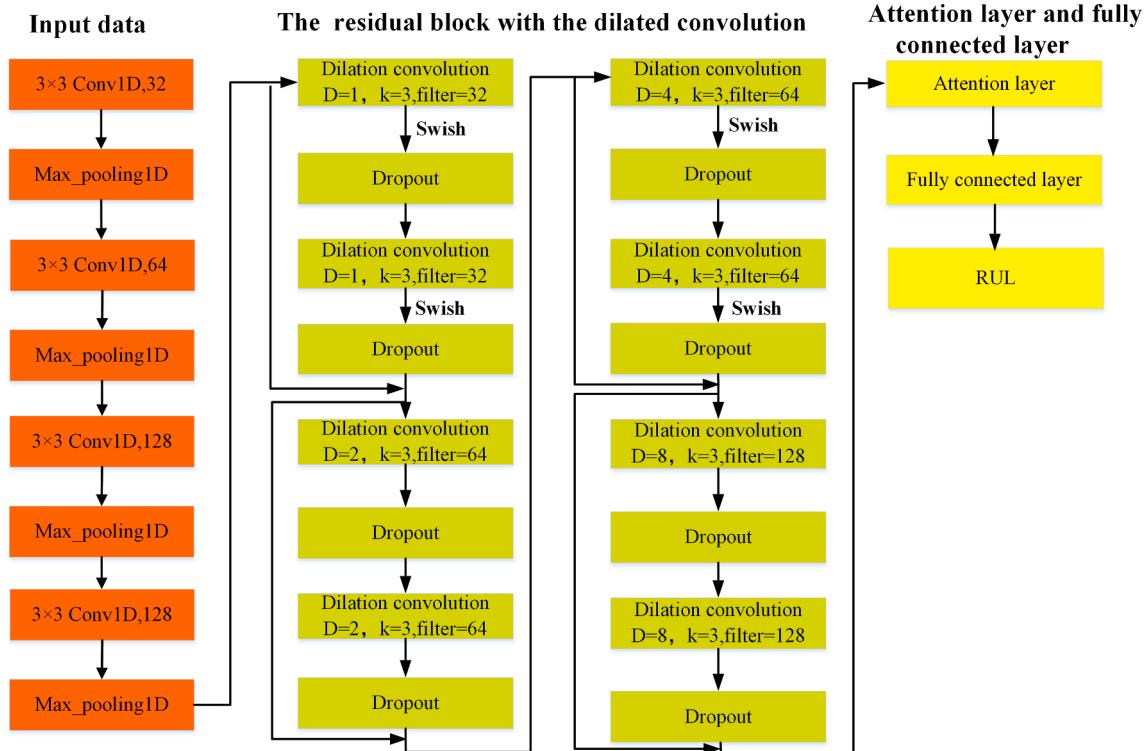


Fig. 4. Detailed architecture and parameters.

$$\left\{ \begin{array}{l} MAE = \frac{1}{n} \sum_{i=1}^n \left| \tilde{y}_i - y_i \right| \\ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\tilde{y}_i - y_i \right)^2} R^2 \end{array} \right.$$

$$= 1 - \frac{\sum_{i=1}^n \left(\tilde{y}_i - y_i \right)}{\sum_{i=1}^n \left(\tilde{y}_i - \bar{y}_i \right)} \quad (17)$$

where \tilde{y}_i is the actual data while y_i is the predicted data, \bar{y}_i is the average predicted data, and n is the number of testing samples.

4. Experimental verification

4.1. Case 1: Tool wear and RUL prediction

To evaluate the prognostic performance of the proposed method, the dataset from a milling machine was utilized to predict the tool wear and its RUL. The data was provided by SIMTECH Institute in Singapore (Li et al., 2009; PHMSociety, 2010). The experimental platform is presented in Fig. 5. The workpiece was stainless steel. A Ball-nose cutting tool of tungsten carbide with 3 flutes was used to cut the workpiece, and a total of three cutters were tested in a down milling, namely, C1, C4, C6. The Milling parameters were as follows: the spindle speed was 10,400 r/min; the feed rate was 1,555 mm/min; the depth of cut was 0.125 mm.

During the cutting process, the cutting force was measured by a

Kistler quartz 3-component platform dynamometer; the vibration signal in three directions was measured by a Kistler Piezo accelerometer; the Kistler acoustic emission (AE) sensor was used to measure the root mean square (RMS) of AE signal. All the signals were collected by using a DAQ NI PCI1200 at a sampling rate of 50 kHz. Eventually, the collected signals were composed of seven-channel information (cutting force in three directions, vibration signal in three directions, and AE-RMS). The workpiece length in the feed direction L was 108 mm, and the cutter's flank wear was measured using a LEICA MZ12 microscopy system after finishing each cut. Each cutter consists of 315 cut cycles. The signal of each cut is regarded as a sample, and a total of 900 samples were randomly selected from three cutters. Fig. 6. shows the examples of seven-channel signals in the tool wear process. It can be seen from the figure that as the tool wear increases, the amplitude of the signal increases.

4.1.1. Data pre-processing and RUL target function

The input data is the seven-channel signal data, and the output is the average flank wear and RUL. The training and testing data are normalized into the range [0, 1]. The subsequence length of a sample in this experiment is 20,000 points, and the corresponding training labels are the average flank wear and RUL. Fig. 8. shows the average flank wear obtained by the microscope. The training label RUL are able to obtain by the following formulas:

$$RUL(t) = t_f - t \quad (18)$$

where t_f represents the time for which the tool reaches the threshold of the tool life, and t represents the tool wear of the current time during the cutting process (Fig. 7.). The t_f has corresponded to the threshold of tool wear. In the engineering application, the wear threshold can obtain by the tool manufacturer or obtain experimentally. In this experiment, the threshold is related to the amount of wear obtained at the cycle of 300.

During the experiment, the method of cross-validation was adopted. For example, C1 and C4 are considered to be the training data, and the remainder is the testing data, as illustrated in Table 1. The proposed method is implemented in Python with TensorFlow deep learning framework. During the training process, the initial learning rate is set to be 0.001 (Chen et al., 2019; Xu et al., 2021), and the mini-batch is adopted, and the size of a mini-batch is set to be 32 (Zhao et al., 2020).

4.1.2. Results discussion

After training the proposed method, the prediction results of tool wear and RUL are achieved and presented in Fig. 9 and Fig. 10, respectively. It is observed that the predicted tool wear and RUL for each

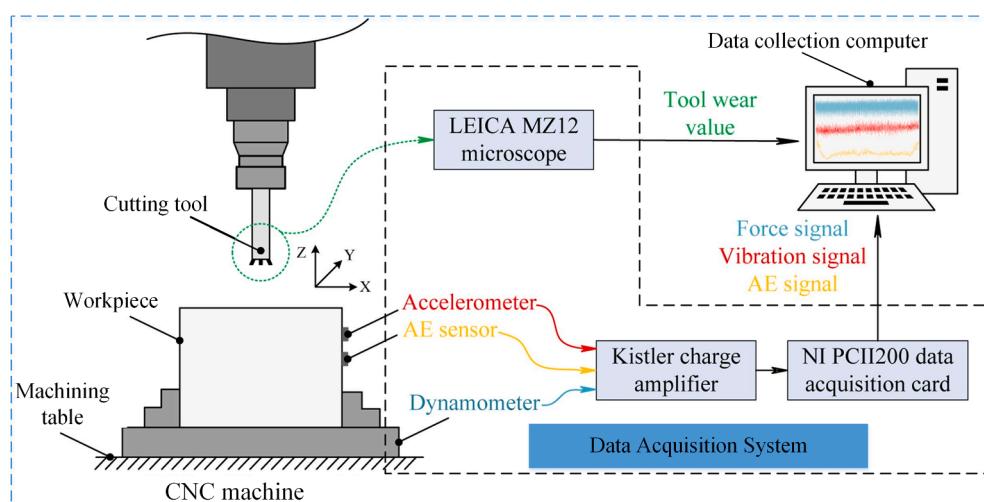


Fig. 5. The schematic diagram of the experimental setup.

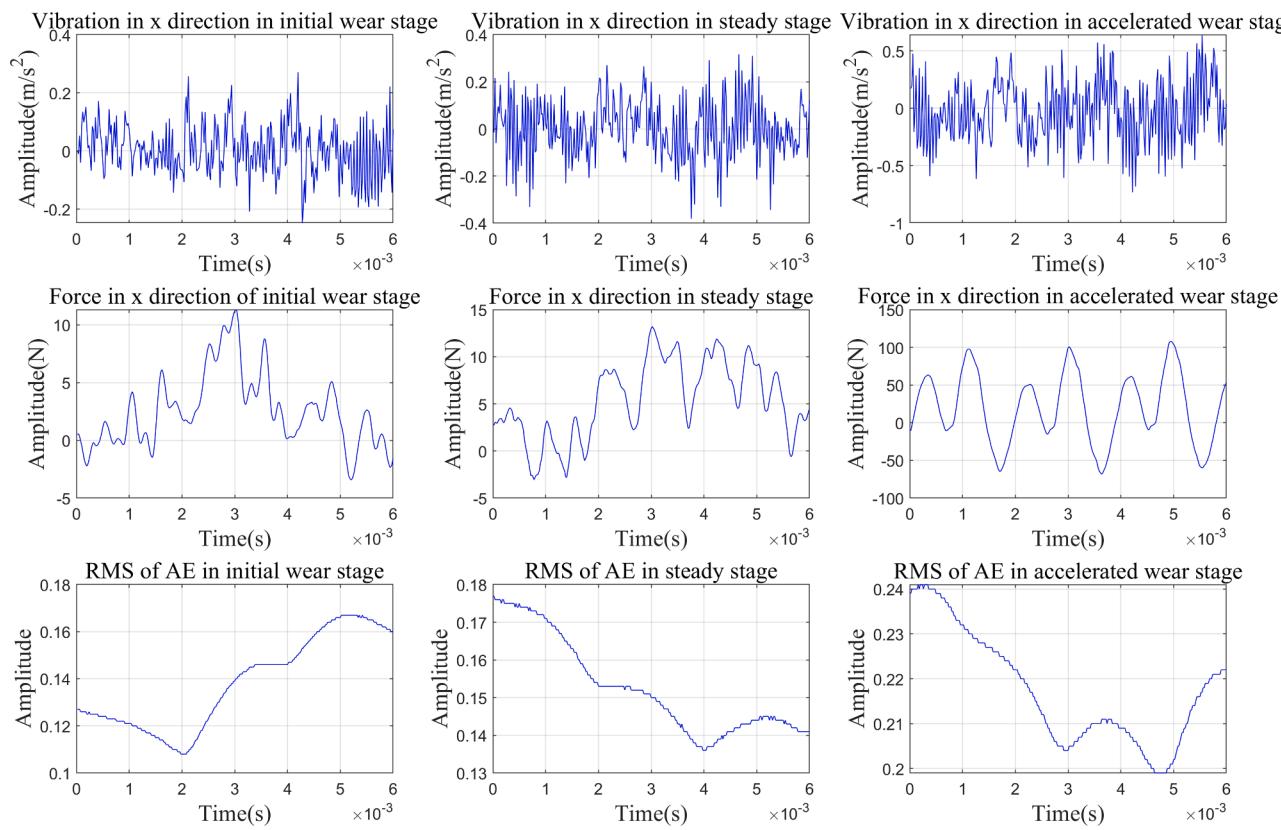
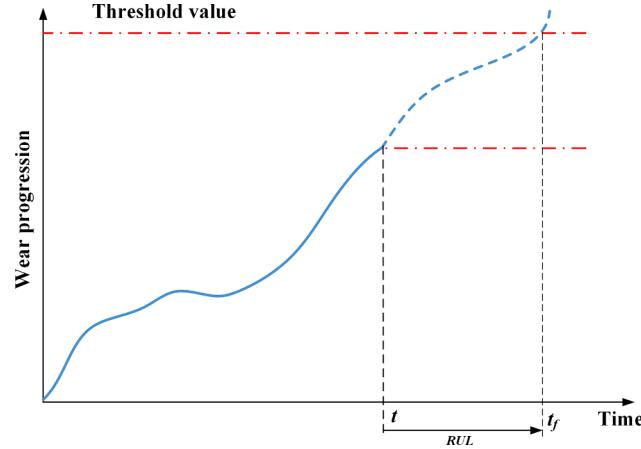


Fig. 6. Examples of the different signals of C6.

Fig. 7. RUL calculation at time t .

testing case are close to the trend of true tool wear and RUL well, and the errors between true values and the predicted ones of tool wear are very small. In that case, the effectiveness of the proposed method is proven by the predicted results.

To illustrate the effectiveness of the proposed method, evaluation criteria during the training and testing process were computed and plotted in Fig. 11. From Fig. 11(a), we can vividly see that the training loss became very small and converges after a few training epochs. The testing loss decreased continuously and ultimately tended to be stable. It indicated that the proposed model is very effective and easy to train. In Fig. 11(b), we can see that the accuracy is enhanced with the iteration process increased, and it is closed to 1. Besides, the testing MAE and RMSE were small at the end of the iteration process. It illustrates that the error of the proposed method on the testing part is very small. In

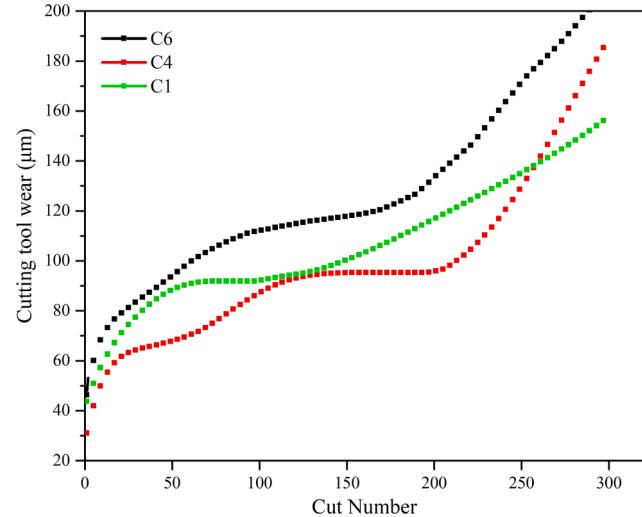


Fig. 8. The cutting tool flank wear of three cutters.

Table 1
The configuration of training data and testing data.

No	Training data	Testing data
1	C4, C6	C1
2	C1, C6	C4
3	C1, C4	C6

conclusion, the proposed method can converge easily and early, the prediction results are very reliable. Therefore, the developed model is very effective for RUL prediction.

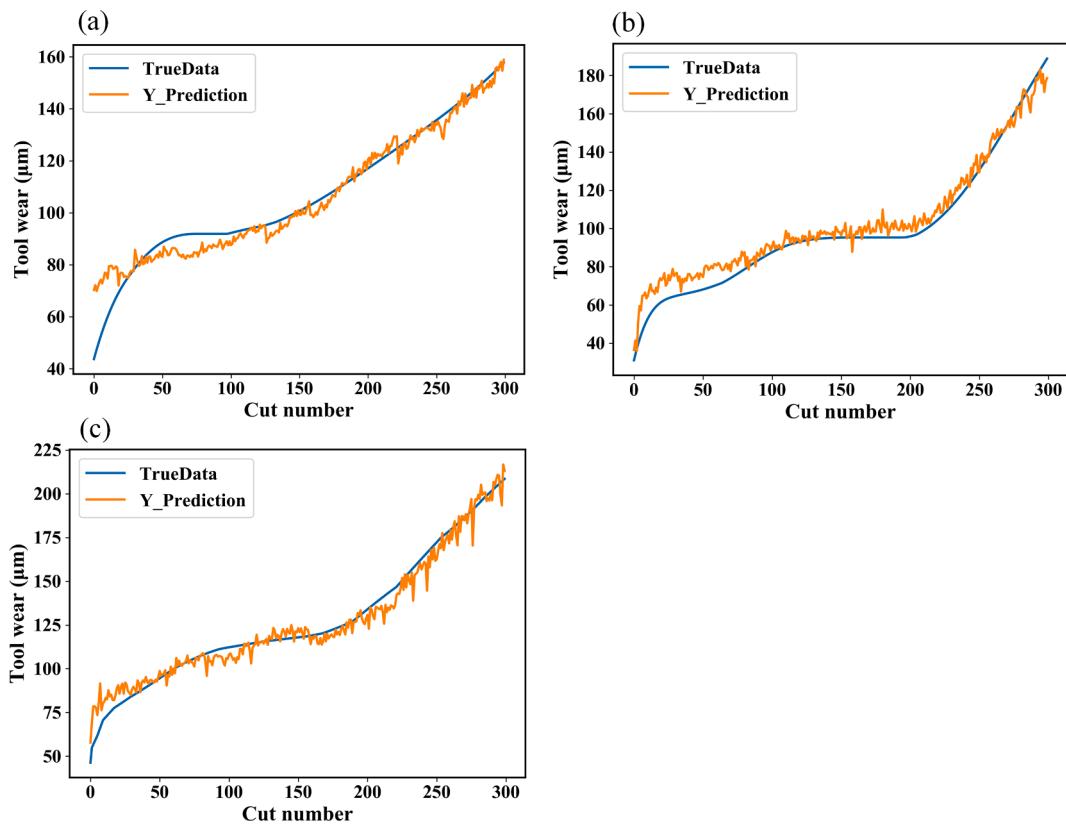


Fig. 9. The predicted tool wear of three testing cutters: (a) C1 is the testing data, (b) C4 is the testing data, (c) C6 is the testing data.

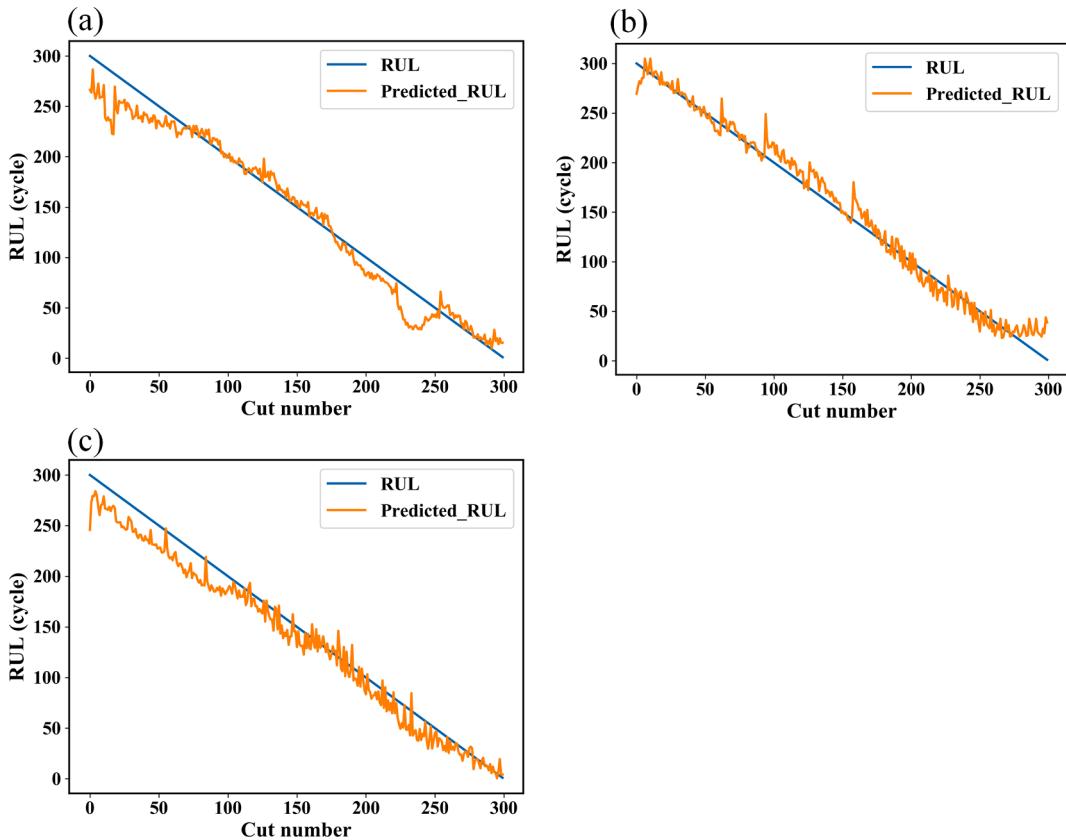


Fig. 10. The predicted RUL of three testing cutters: (a) C1 is the testing data, (b) C4 is the testing data, (c) C6 is the testing data.

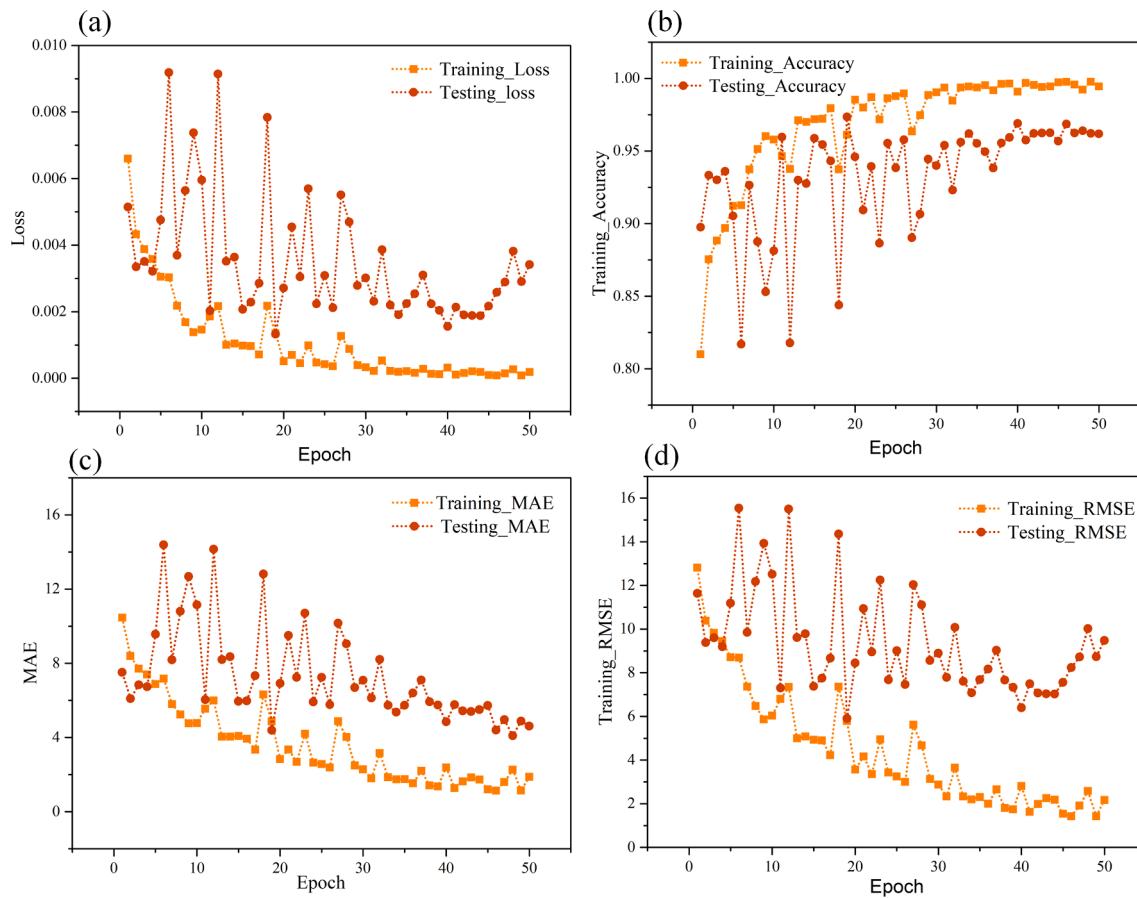


Fig. 11. The iteration curves during the training and testing process of C6: (a) Loss, (b) R^2 , (c) MAE, (d) RMSE.

The proposed model can extract features automatically. To show how the proposed method affects, the extracted features of the fully connected (FC) layer are shown. Fig. 12 displayed the feature of 2th and 295th samples for C1. It can be seen that the features in the FC1 layer were linear separable before putting into the regressional layer. In addition, the amplitude of the extracted feature of FC1 and FC2 changes when the input signals are in various stages of the tool wear, which can interpret the proposed method can follow the trend of the tool wear mechanism.

4.1.3. Related work comparison

Furthermore, to make the results more objective, the latest related works are summarized and compared. Table 2 and Fig. 13 show the MAE and RMSE results under three different testing cutters for all methods. The compared algorithms, including SVR, CNN, RNN, MLP, and CNN-LSTM. From Fig. 13, it is seen that the values of RMSE and MAE for three different testing cases with the proposed method are the smallest ones among all compared methods. The TDConvLSTM proposed in (Qiao et al., 2018) is the second-best because the authors consider both the temporal dependency and spatial correlation of the input signals. In addition, CNN (Qiao et al., 2018) and RNN (Zhao et al., 2016) are also competitive. To quantify, the RMSE of the proposed method for C1, C4, and C6 are 5.62, 6.64, and 6.20, respectively. The MAE values for C1, C4, and C6 are 4.50, 5.43, and 4.63, respectively. In conclusion, the comprehensive comparison results presented above illustrate the proposed architecture is more promising than others in tool wear prediction and able to predict reliable and accurate tool wear in different cases.

4.2. Case 2: Turbofan engine RUL prognostic

To further evaluate the generalization and robustness of the

proposed model, another experiment on the RUL prognostic of the turbofan engine has been done. The dataset is provided by the US Army Research laboratory, which is called C-MAPSS dataset (Ramasso & Saxena, 2014), and it is available through NASA. The C-MAPSS data is widely used by many scholars and experts for the validation of their proposed method. The dataset contains four sub-datasets labeled from FD001 to FD004, and each row of the data is collected during a single operating time cycle by 21 sensors.

4.2.1. Data pre-processing and RUL target function

In this paper, dataset FD001 is selected to validate the proposed method. In dataset FD001, there are 100 engines, and the data of 1–80 engines of turbofan are selected as the training data, while the data of 81–100 engines are selected as the testing data as what the literature (Chen et al., 2019) did. The multi-sensor signal was used as the input of the proposed method when the above data with the same minimum and maximum values were removed in advance. Finally, the data of 17 sensors are selected to validate our proposed method, and then the selected data are normalized by the z-score method as the first experiment did. As for the label, the conventional function to evaluate the RUL values is the target, which degrades linearly along with time. In practice, the degradation of equipment can be ignored at the beginning of running. In addition, the RUL decreases when the equipment approaches the end-of-Life (Babu et al., 2016; Heimes, 2008). In this paper, we set the maximum limit as 130-time cycles (Babu et al., 2016; Heimes, 2008). Fig. 14 (a) and (b) present the sensor data of the first engine and the corresponding labeled RUL, respectively.

4.2.2. Results discussion and related work comparison

From the above, we can get that the input data is the signal data for 17 sensors, and the output is the predicted RUL. The subsequence length

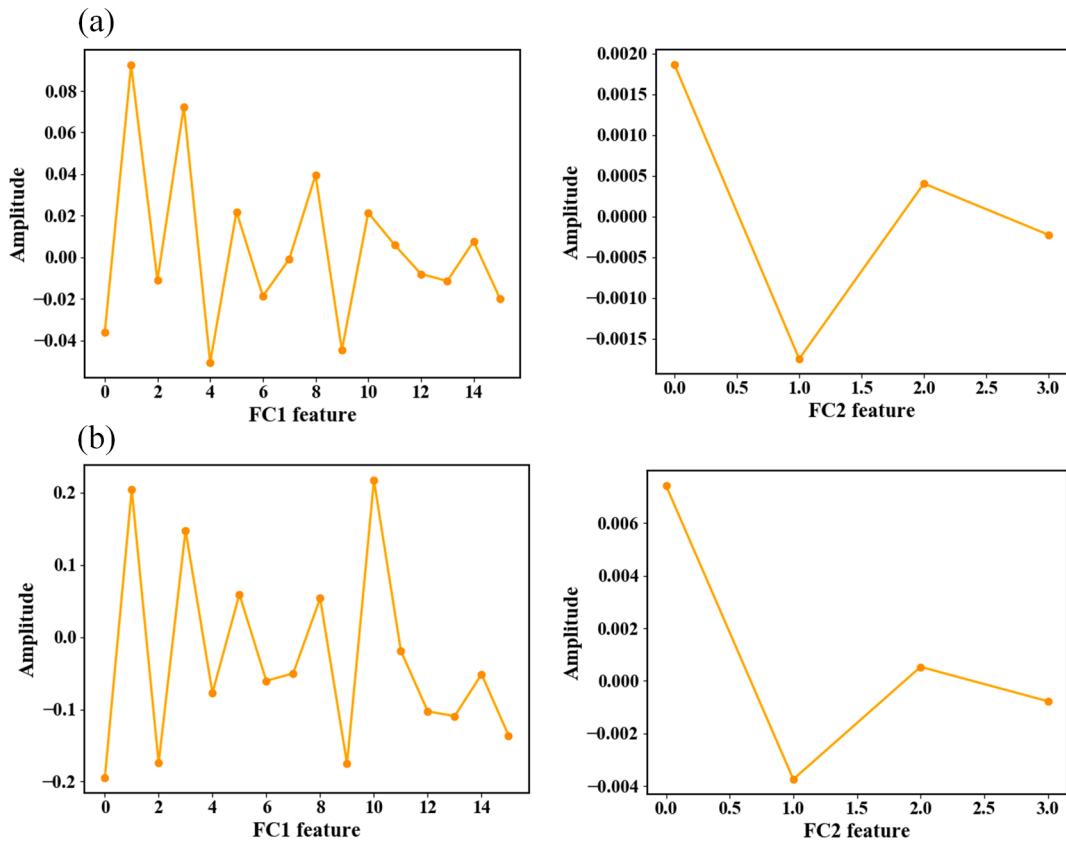


Fig. 12. The extracted features in the full-connected layer of the proposed method:(a) the feature of the second sample, (b) the feature of the 295th sample.

Table 2
Algorithm comparison results.

Algorithms	Feature extraction	MAE			RMSE		
		C1	C4	C6	C1	C4	C6
SVR (Zhao et al., 2016)	Yes	15.6	17.0	24.9	18.5	19.6	31.5
MLP (Zhao et al., 2016)	Yes	24.5	18.0	24.8	31.2	20.0	31.4
RNN (Zhao et al., 2016)	No	13.1	16.7	25.5	15.6	19.7	32.9
CNN (Qiao et al., 2018)	No	15.32	14.34	17.36	18.59	18.80	21.85
CNN + LSTM (Qiao et al., 2018)	No	11.18	9.39	11.34	13.77	11.85	14.33
TDConvLSTM (Qiao et al., 2018)	NO	6.99	6.96	7.50	8.33	8.39	10.22
Proposed algorithm	No	4.50	5.43	4.63	5.62	6.64	6.20

of the sample in this experiment is set to be 30 points per sample. Therefore the size of the input data is 30×17 . The output is a sequence that represents the predicted RUL. We retrained the proposed architecture, and the main structure and parameters for this experiment are also the same as those presented in Fig. 4. In addition, the difference is that only two layers of 1D CNN are used to extract features. During the training process, the loss function and the learning rate are also the same as those described in Section 3.

To analyze the results, the RMSE of the developed algorithm is listed in Table 3, and the prediction results are presented in Fig. 15. Fig. 15 (a) presents the predicted RUL and real RUL of 85, 87, 89, 99 engines, respectively. It can be seen from Fig. 15 that the predicted RUL can follow the trend of the actual RUL of engines with a small deviation.

The C-MAPSS dataset utilized in this part is very famous in prognostic research, and many state-of-the-art results have been reported in recent years. To make the prediction results more objective, the latest research results on the datasets of C-MAPSS were summarized in Table 3. It can be observed that many latest proposed approaches have shown great performance on this prognostic problem, including LSTM, GRU, etc. For better visualization, the RMSE of each engine with different models is plotted in Fig. 16. Obviously, the RMSE of the proposed method is smaller than those of other published methods. In addition, the root-mean-square (RMS) (Chen et al., 2019) of the RMSE of all engines are listed in table 0.3 and Fig. 17. To be more quantitative, the RMS of LSTM is 34.6, and that of GRU is 32, which is rather considerable. However, the RMS of the proposed method is close to 15.3, which is the best one among all the compared models. It should be noticed that the training splitting of training data and the testing data are what the literature (Chen et al., 2019) did. And the compared methods such as LSTM, GRU, and the proposed method are implemented with identical training and testing datasets. Therefore, the compared results are very objective.

Based on the compared results, it is concluded that the proposed method is promising for prognostic problems and able to predict accurate RUL, and it can be easily concluded that our proposed method has superiority over the existing methods.

5. Conclusions

In this study, a novel method has been proposed for RUL prognostic in the mechanical industry. The method does not require any domain knowledge but is available for automatic extraction of the features by 1D CNN. Besides, the multi-scale deep residual dilated convolution and attention mechanism have been designed for RUL prognostic, which has demonstrated its great performance in two experiments. The main conclusions can be drawn as follows:

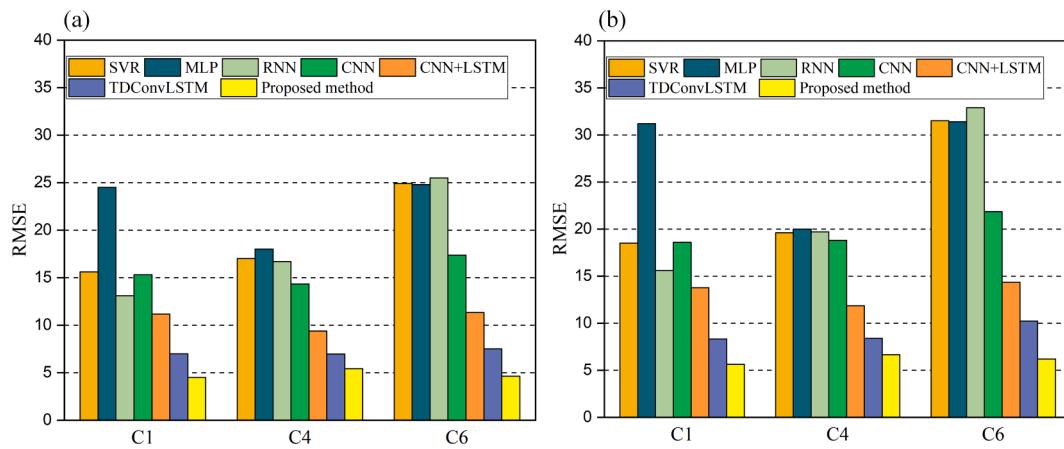


Fig. 13. Model performance for tool wear prediction of three testing cutters with different models: (a) MAE, (b) RMSE.

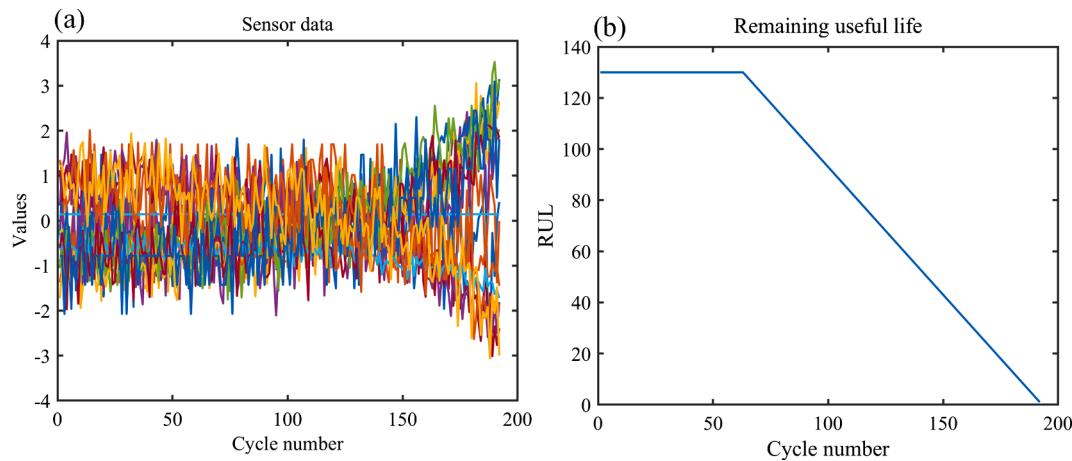


Fig. 14. The example of training data and labels: (a) signals of the first engine of FD001, (b) the corresponding label (RUL).

Table 3

The RMSE of 81–100 engines with different models.

No	LMD (Chen et al., 2019)	F-NMD (Chen et al., 2017; Chen et al., 2019)	T-NMD (Chen et al., 2017; Chen et al., 2019)	GRU (Chen et al., 2019)	GRU.s (Chen et al., 2019)	LSTM (Chen et al., 2019)	LSTM.s (Chen et al., 2019)	Proposed method
81	44.8	39.5	24.2	29.4	22	32.7	31.2	15.01
82	51.3	28.7	10.5	20	18.7	25.6	23.9	14.84
83	38.3	64.8	47.7	32.1	30	16.2	11	12.77
84	92.7	57	64.3	35.4	33.3	38.5	37.8	15.97
85	17.5	21.5	6.4	21.6	19.4	19.5	16.8	8.90
86	68.6	64.1	66.1	21.1	15	52.7	51.9	13.59
87	15.4	19.9	18.8	31.7	27.8	14.7	12.1	11.65
88	27.3	28.6	19.1	21.2	17.4	27.8	27	11.54
89	25.9	26	12.6	16.2	13.1	29.8	22.8	13.38
90	67.2	30.4	40	28.5	27.1	33.9	32.3	16.20
91	27	35.5	23.9	8.3	7.8	7.7	5.2	17.13
92	86.1	98.1	83.6	74	72.7	77.7	77	21.50
93	61	26.3	42.9	26.2	22.3	23.2	20.9	20.98
94	83	52	56	22.9	22.1	27.5	27.2	12.54
95	35.6	61.7	42.2	34.7	34.5	19	17.6	18.09
96	70.3	86.8	79	79.3	76.9	74.6	72.8	20.57
97	45.2	27.2	9.1	12.4	9.6	15.7	13.1	23.12
98	58.5	34.9	38.3	34	31.8	40.6	39.8	17.50
99	28.6	18.5	17.1	32.9	27.7	30.6	26	9.81
100	22.7	23.1	8.21	28.6	22.6	18.3	16.8	18.00
RMS	53.7	47.8	42.7	35	32.6	36.2	34.6	16.13

(1) Due to the fact that the signal data are one-dimensional and from multiple sensors, a novel signal pre-processing and feature extraction and fusion method is proposed, and the designed

signal pre-processing method and feature fusion method can be completed in a very simple way. Thus, the proposed signal pre-processing and feature extraction and fusion method may

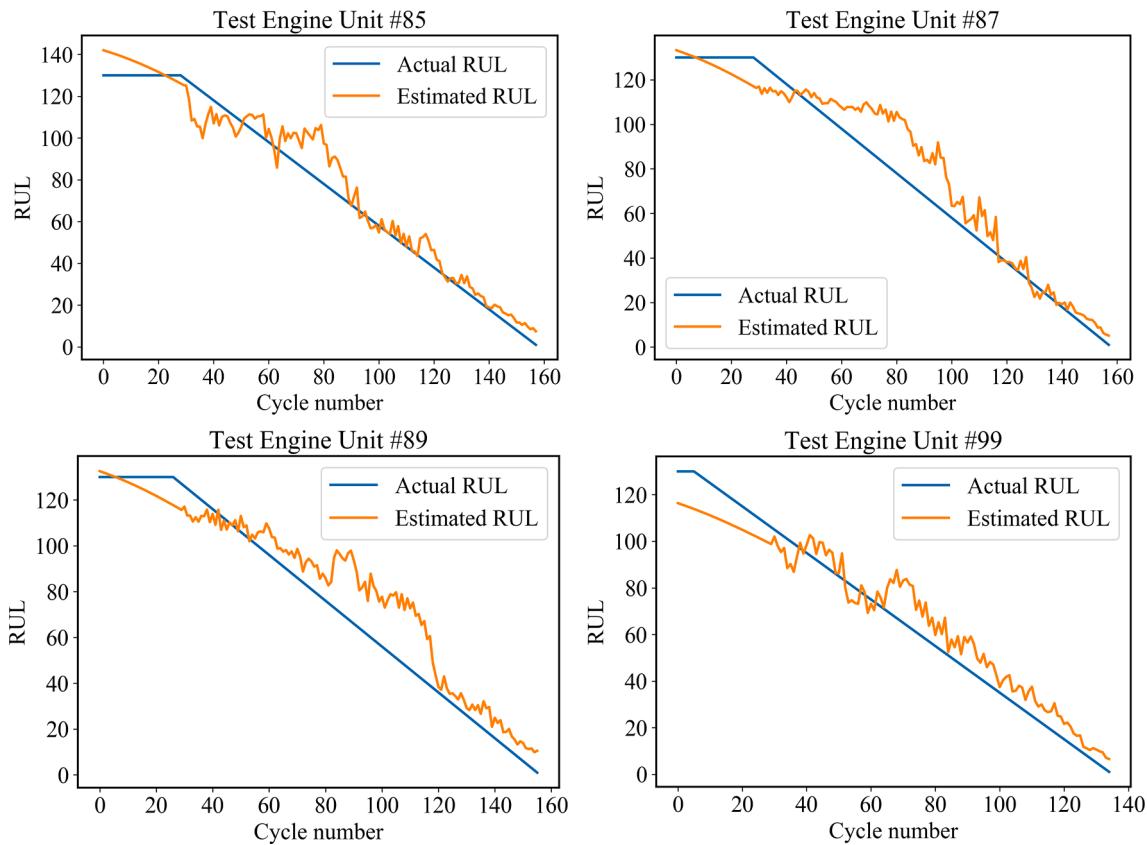


Fig. 15. The predicted RUL of different engines.

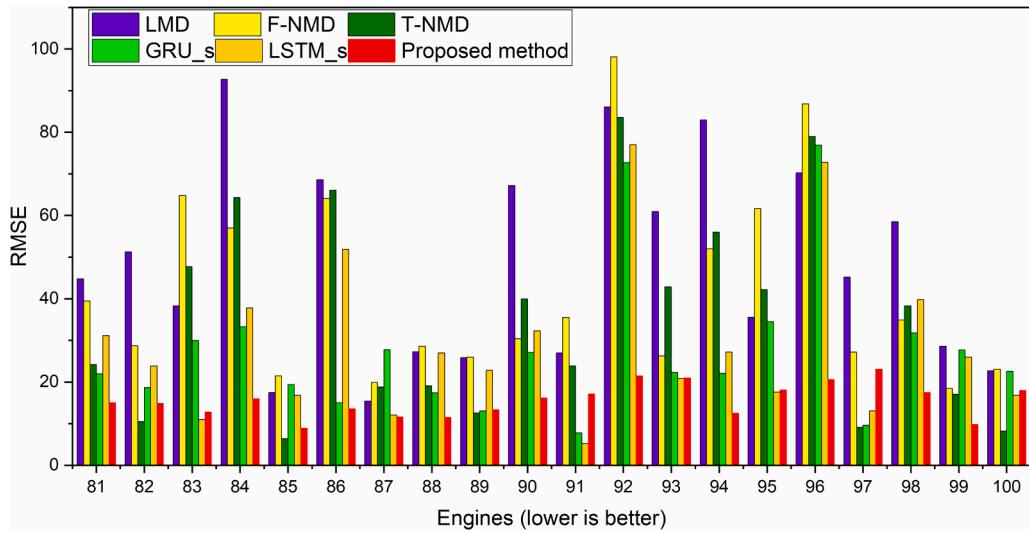


Fig. 16. The predicted error of 81–100 engines with different models.

become a new way to process the multi-sensor signal data and achieve multi-scale feature fusion.

- (2) Second, a novel method including residual connection and dilated convolution with attention mechanism was adopted. Different cases, i.e., tool wear and RUL prediction, and turbofan engine RUL prognostic are conducted to validate its effectiveness. As a result, both of them achieved an extraordinary outcome. Thus, the proposed method is very suited and promising for RUL prognostic.

CRediT authorship contribution statement

Xingwei Xu: Methodology, Software, Data curation, Writing – original draft. **Xiang Li:** Conceptualization, Validation. **Weiwei Ming:** Visualization, Investigation. **Ming Chen:** Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence

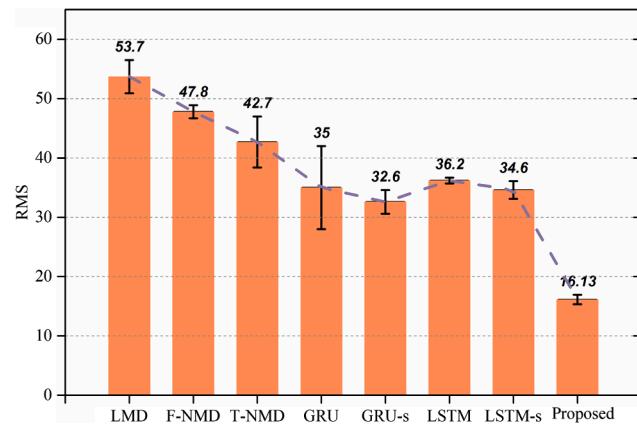


Fig. 17. The performance for RUL prediction with different models.

the work reported in this paper.

Acknowledgment

The work is supported by National Natural Science Foundation of China (No. 51875356). Besides, the author would like to thank the editor and the anonymous reviewers for their helpful comments.

References

- Abdeljaber, O., Avci, O., Kiranyaz, S., Gabbouj, M., & Inman, D. J. (2017). Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration*, 388, 154–170.
- An, Z., Li, S., Wang, J., & Jiang, X. (2019). A novel bearing intelligent fault diagnosis framework under time-varying working conditions using recurrent neural network. *ISA Transactions*.
- Babu, G. S., Zhao, P., & Li, X.-L. (2016). Deep convolutional neural network based regression approach for estimation of remaining useful life. In *International conference on database systems for advanced applications* (pp. 214–228). Springer.
- Cai, B., Fan, H., Shao, X., Liu, Y., Liu, G., Liu, Z., & Ji, R. (2021). Remaining useful life prediction methodology based on Wiener process: Subsea Christmas tree system as a case study. *Computers & Industrial Engineering*, 151.
- Cai, W., Zhang, W., Hu, X., & Liu, Y. (2020). A hybrid information model based on long short-term memory network for tool condition monitoring. *Journal of Intelligent Manufacturing*.
- Che, C., Wang, H., Ni, X., & Fu, Q. (2020). Domain adaptive deep belief network for rolling bearing fault diagnosis. *Computers & Industrial Engineering*, 143.
- Chen, Z., Chen, Y., Wu, L., Cheng, S., & Lin, P. (2019). Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions. *Energy Conversion and Management*, 198.
- Chen, G., Chen, J., Zi, Y., & Miao, H. (2017). Hyper-parameter optimization based nonlinear multistate deterioration modeling for deterioration level assessment and remaining useful life prognostics. *Reliability Engineering & System Safety*, 167, 517–526.
- Chen, L., Xu, G., Zhang, S., Yan, W., & Wu, Q. (2020). Health indicator construction of machinery based on end-to-end trainable convolution recurrent neural networks. *Journal of Manufacturing Systems*, 54, 1–11.
- Deng, Y., Huang, D., Du, S., Li, G., Zhao, C., & Lv, J. (2021). A double-layer attention based adversarial network for partial transfer learning in machinery fault diagnosis. *Computers in Industry*, 127.
- Deng, Y., Shichang, D., Shiyao, J., Chen, Z., & Zhiyuan, X. (2020). Prognostic study of ball screws by ensemble data-driven particle filters. *Journal of Manufacturing Systems*, 56, 359–372.
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).
- Heimes, F. O. (2008). Recurrent neural networks for remaining useful life estimation. In *2008 international conference on prognostics and health management* (pp. 1–6). IEEE.
- Javed, K., Gouriveau, R., Li, X., & Zerhouni, N. (2016). Tool wear monitoring and prognostics challenges: A comparison of connectionist methods toward an adaptive ensemble model. *Journal of Intelligent Manufacturing*, 29(8), 1873–1890.
- Jia, S., Deng, Y., Lv, J., Du, S., & Xie, Z. (2022). Joint distribution adaptation with diverse feature aggregation: A new transfer learning framework for bearing diagnosis across different machines. *Measurement*, 187.
- Jozefowicz, R., Zaremba, W., & Sutskever, I. (2015). An empirical exploration of recurrent network architectures. In *International Conference on Machine Learning* (pp. 2342–2350).
- Jun-Hong, Z., Chee Khiang, P., Lewis, F. L., & Zhao-Wei, Z. (2009). Intelligent Diagnosis and Prognosis of Tool Wear Using Dominant Feature Identification. *IEEE Transactions on Industrial Informatics*, 5(4), 454–464.
- Kong, D., Chen, Y., Li, N., Duan, C., Lu, L., & Chen, D. (2019). Relevance vector machine for tool wear prediction. *Mechanical Systems and Signal Processing*, 127, 573–594.
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105).
- Kumar, A., Chinnam, R. B., & Tseng, F. (2019). An HMM and polynomial regression based approach for remaining useful life and health state estimation of cutting tools. *Computers & Industrial Engineering*, 128, 1008–1014.
- Li, X., Lim, B., Zhou, J., Huang, S., Phua, S., Shaw, K., & Er, M. (2009). Fuzzy neural network modelling for tool wear estimation in dry milling operation. In *Annual conference of the prognostics and health management society* (pp. 1–11).
- Mao, W., Tian, S., Fan, J., Liang, X., & Sifan, A. (2020). Online detection of bearing incipient fault with semi-supervised architecture and deep feature representation. *Journal of Manufacturing Systems*, 55, 179–198.
- Pascanu, R., Mikolov, T., & Bengio, Y. (2013). On the difficulty of training recurrent neural networks. In *International conference on machine learning* (pp. 1310–1318).
- PHMSociety (2010). PHM data challenge 2010. <https://www.phmsociety.org/competition/phm10>.
- Qiao, H., Wang, T., Wang, P., Qiao, S., & Zhang, L. (2018). A Time-Distributed Spatiotemporal Feature Learning Method for Machine Health Monitoring with Multi-Sensor Time Series. *Sensors (Basel)*, 18(9).
- Ramasso, E., & Saxena, A. (2014). Performance Benchmarking and Analysis of Prognostic Methods for CMAPSS Datasets.
- Ren, L., Cui, J., Sun, Y., & Cheng, X. (2017). Multi-bearing remaining useful life collaborative prediction: A deep learning approach. *Journal of Manufacturing Systems*, 43, 248–256.
- Shao, H., Jiang, H., Zhang, H., & Liang, T. (2018). Electric Locomotive Bearing Fault Diagnosis Using a Novel Convolutional Deep Belief Network. *IEEE Transactions on Industrial Electronics*, 65(3), 2727–2736.
- Souza, R. M., Nascimento, E. G. S., Miranda, U. A., Silva, W. J. D., & Lepikson, H. A. (2021). Deep learning for diagnosis and classification of faults in industrial rotating machinery. *Computers & Industrial Engineering*, 153.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1), 1929–1958.
- Tran, V. T., Thom Pham, H., Yang, B.-S., & Tien Nguyen, T. (2012). Machine performance degradation assessment and remaining useful life prediction using proportional hazard model and support vector machine. *Mechanical Systems and Signal Processing*, 32, 320–330.
- Wang, X., Jiang, B., & Lu, N. (2019). Adaptive relevant vector machine based RUL prediction under uncertain conditions. *ISA Transactions*, 87, 217–224.
- Wang, J., Xie, J., Zhao, R., Zhang, L., & Duan, L. (2017). Multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing. *Robotics and Computer-Integrated Manufacturing*, 45, 47–58.
- Wu, Y., Yuan, M., Dong, S., Lin, L., & Liu, Y. (2018). Remaining useful life estimation of engineered systems using vanilla LSTM neural networks. *Neurocomputing*, 275, 167–179.
- Xu, X., Tao, Z., Ming, W., An, Q., & Chen, M. (2020). Intelligent monitoring and diagnostics using a novel integrated model based on deep learning and multi-sensor feature fusion. *Measurement*, 165.
- Xu, X., Wang, J., Zhong, B., Ming, W., & Chen, M. (2021). Deep learning-based tool wear prediction and its application for machining process using multi-scale feature fusion and channel attention mechanism. *Measurement*, 177.
- Yu, W., Kim, I. I. Y., & Mechebske, C. (2019). Remaining useful life estimation using a bidirectional recurrent neural network based autoencoder scheme. *Mechanical Systems and Signal Processing*, 129, 764–780.
- Zhang, W., Li, X., & Ding, Q. (2019). Deep residual learning-based fault diagnosis method for rotating machinery. *ISA Transactions*, 95, 295–305.
- Zhang, J., Wang, P., Yan, R., & Gao, R. X. (2018). Long short-term memory for machine remaining life prediction. *Journal of Manufacturing Systems*, 48, 78–86.
- Zhao, H. Y., Che, C., Jin, B., & Wei, X. P. (2019). A viral protein identifying framework based on temporal convolutional network. *Math Biosci Eng*, 16(3), 1709–1717.
- Zhao, W., Gao, Y., Ji, T., Wan, X., Ye, F., & Bai, G. (2019). Deep Temporal Convolutional Networks for Short-Term Traffic Flow Forecasting. *IEEE Access*, 7, 114496–114507.
- Zhao, X., Jiang, N., Liu, J., Yu, D., & Chang, J. (2019). Short-term average wind speed and turbulent standard deviation forecasts based on one-dimensional convolutional neural network and the integrate method for probabilistic framework. *Energy Conversion and Management*.
- Zhao, X., Jiang, N., Liu, J., Yu, D., & Chang, J. (2020). Short-term average wind speed and turbulent standard deviation forecasts based on one-dimensional convolutional neural network and the integrate method for probabilistic framework. *Energy Conversion and Management*, 203.
- Zhao, R., Wang, J., Yan, R., & Mao, K. (2016). Machine health monitoring with LSTM networks. In *2016 10th International Conference on Sensing Technology (ICST)* (pp. 1–6). IEEE.
- Zhao, S., Zhang, Y., Wang, S., Zhou, B., & Cheng, C. (2019). A recurrent neural network approach for remaining useful life prediction utilizing a novel trend features construction method. *Measurement*, 146, 279–288.