

A New Time-Space Attention Mechanism Driven Multi-Feature Fusion Method for Tool Wear Monitoring

Tingting Feng

Southwest Jiaotong University

Liang Guo (✉ guoliang@swjtu.edu.cn)

Southwest Jiaotong University

Hongli Gao

Southwest Jiaotong University

Tao Chen

Southwest Jiaotong University

Yaoxiang Yu

Southwest Jiaotong University

Changgen Li

Southwest Jiaotong University

Research Article

Keywords: Tool wear monitoring, Multi-feature fusion, Attention mechanism, Residual useful life prediction

Posted Date: September 8th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-864032/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Title page

Title :

A new time-space attention mechanism driven multi-feature fusion method for tool wear monitoring

Author information:

Name	Affiliation	E-mail
Tingting Feng	School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China	tinafeng@my.swjtu.edu.cn
Liang Guo (corresponding author)	1. School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China 2. Laboratory of Science and technology on integrated Logistics Support, National University of Defense Technology	guoliang@swjtu.edu.cn
Hongli Gao	School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China	
Tao Chen	School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China	
Yaoliang Yu	School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China	
Changgen Li	School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China	

Abstract :

In order to accurately monitor the tool wear process, it is usually necessary to collect a variety of sensor signals during the cutting process. Different sensor signals in the feature space can provide complementary information. In addition, the monitoring signal is time series data, which also contains a wealth of tool degradation information in the time dimension. However, how to fuse multi-sensor information in time and space dimensions is a key issue that needs to be solved. This paper proposes a new time-space attention mechanism driven multi-feature fusion method to realize the tool wear monitoring. Firstly, lots of features are established from different sensor signals

and selected preliminarily. Then, a new feature fusion model with time-space attention mechanism is constructed to fuse features in time and space dimensions. Finally, the tool degradation model is established according to the predicted wear, and the tool remaining useful life is predicted by particle filter. The effectiveness of this method is verified by a tool life cycle wear experiment. Through comparing with other feature fusion models, it is demonstrated that the proposed method realizes the tool wear monitoring more accurately and has better stability.

Keywords :

Tool wear monitoring, Multi-feature fusion, Attention mechanism, Residual useful life prediction

Declarations:

Funding:

This work was supported in part by the National Natural Science Foundation of China(51905452, 51775452), Local Development Foundation guided by the Central Government(2020ZYD012) and Planning Project of Science & Technology Department of Sichuan Province under Grant (2020YFN0062).

Conflicts of interest/Competing interests :

The contribution of this paper is our original work and all authors have agreed to the submission of this work to this journal. And there is no conflict of interest.

Availability of data and material:

Relevant data is not public.

Code availability:

The code is not public.

Ethics approval:

Not applicable.

Consent to participate :

Agree.

Consent for publication:

Agree.

A new time-space attention mechanism driven multi-feature fusion method for tool wear monitoring

Tingting Feng¹, Liang Guo^{1, 2, *}, Hongli Gao¹, Tao Chen¹, Yaoxiang Yu¹, Changgen Li¹

Abstract

In order to accurately monitor the tool wear process, it is usually necessary to collect a variety of sensor signals during the cutting process. Different sensor signals in the feature space can provide complementary information. In addition, the monitoring signal is time series data, which also contains a wealth of tool degradation information in the time dimension. However, how to fuse multi-sensor information in time and space dimensions is a key issue that needs to be solved. This paper proposes a new time-space attention mechanism driven multi-feature fusion method to realize the tool wear monitoring. Firstly, lots of features are established from different sensor signals and selected preliminarily. Then, a new feature fusion model with time-space attention mechanism is constructed to fuse features in time and space dimensions. Finally, the tool degradation model is established according to the predicted wear, and the tool remaining useful life is predicted by particle filter. The effectiveness of this method is verified by a tool life cycle wear experiment. Through comparing with other feature fusion models, it is demonstrated that the proposed method realizes the tool wear monitoring more accurately and has better stability.

Keywords : Tool wear monitoring, Multi-feature fusion, Attention mechanism, Residual

* Liang Guo
guoliang@swjtu.edu.cn

¹ School of Mechanical Engineering, Southwest Jiaotong University, Chengdu 610031, China

² Laboratory of Science and technology on integrated Logistics Support, National University of Defense Technology

useful life prediction

1. Introduction

Numerical control machine tool is the fundamental device used for execution in smart manufacturing systems, plays a significant role in the overall production system[1][2]. Cutting tool is the tooth of the numerical control machine tool, even if the small tool failure rate has a great negative impact on product quality[3][4]. More seriously, it can even cause unscheduled downtime. Therefore, appropriate planning of the tool according the accurate tool wear monitoring and residual useful life (RUL) prediction is meaningful for promoting processing efficiency and reducing production cost[5][6], which is the most effective way to achieve sustainable manufacturing paradigm.

Among many tool wear monitoring methods, the data-driven indirect method is used to monitor tool wear by establishing the correlation model between tool wear and sensor signals (such as cutting force, vibration, acoustic emission and current, etc.) [7]. It has the advantages of convenience and low cost. After years of research, numerous studies have shown that a single monitoring signal cannot exactly describe the state of machine work for monitoring in advanced manufacture[8–10]. With the rapid development of sensing and computing technologies, extracting different kinds features from different sensor signals to represent the tool wear status that makes RUL prediction more reliable[11][12]. In order to better fuse the information from different sensors, Sun et al.[13] proposed a non-linear Wiener-based prediction model. On the basis of Bayesian model, tool wear monitoring and RUL prediction are realized by force, vibration and acoustic emission signals. Base on statistical models, the method established the tool degradation model firstly, and then the parameters were updated by

the sensor signals during the cutting process. It is difficult to apply to the actual processing production, because it requires a large number of complex mathematical operations and has weak anti-noise ability. Thus, fusing features by machine learning is increasingly applied to tool wear monitoring. Wu et al.[14] introduced a random forests (RFs)-based method to fuse the features to predict tool wear accurately. Wu et al.[15] proposed a multi-sensor information fusion system for online RUL prediction of machining tools. The system used adaptive network based fuzzy inference system to fuse features. However, traditional machine learning is difficult to extract effective information directly from original features[16]. To achieve accurate wear monitoring, features need to be preprocessed by principal component analysis (PCA) and empirical mode decomposition (EMD). Deep neural network (DNN) is a good solution to this problem [17–19]. In reference [20], convolutional neural network (CNN) was used to extract the hidden information of features from spatial dimensions to realize tool wear monitoring and residual life prediction. In addition, the sensor signal is time series data, and the time scale also contains the tool wear degradation information. A special recurrent neural network(RNN), named Long Short-Term Memory (LSTM), can find the temporal dependency among data [21], to realize the feature fusion in the time dimension.

The methods mentioned above achieve feature fusion in spatial and temporal dimensions to some extent, but they do not take into account the different contributions of different features to tool degradation at different time steps. Attention mechanism can make the DNN pay more attention to the relevant information and ignore the irrelevant information, so as to extract the potential information in the data in a deeper

level. It is widely used in natural language processing(NLP)[22][23]. In this paper, we propose a new multi-feature fusion model for tool wear monitoring based on time-space attention mechanism. It makes up for the shortcomings of the aforementioned methods well. Signals during cutting were obtained from different sensors firstly. After preprocessing and preliminary selecting, the pre-selected features are fused using feature fusion model. The importance weights of multiple features are assigned from time and space dimensions by attention mechanism to predict tool wear accurately. Finally, particle filter based digital-analog linkage method is used to predict the RUL of the tool. The major contributions of this paper are summarized as follows.

(1) In order to avoid the problem that the dynamometer, which record cutting force signals, has high cost and negative impact on machining system rigidity. We abandon the cutting force signal with a good trendability. Only vibration, current, and sound signals are used to construct multi-feature in the time domain, frequency domain, and time-frequency domain, respectively.

(2) A time-space attention mechanism based multi-feature fusion model is established to capture the complex spatio-temporal relationship between tool wear and features. Channel attention mechanism pays attention to more meaningful channel information in the feature space by using inter-channel relationship of features. Hidden state attention mechanism assigns attention weight to the hidden state of different time steps, the most relevant information in the input sequence was stressed in the time dimension to predict the tool wear. The superior performance of the proposed method is verified through the experiment and compared with other feature fusion networks.

(3) The predicted tool wear is used to realize tool RUL accurately by digital-analog linkage method based on particle filtering. Meanwhile, aiming at the difficulty in determining the initial parameters of particle filter algorithm, the dynamic setting of the initial parameters is achieved by curve fitting.

This paper is organized as follows. The Section I introduces the literature reviews on tool wear monitoring and RUL prediction methods based on feature fusion. Section II describes the proposed wear prediction method in detail. We compare the proposed model with other methods to verify its superiority by experimental study in Section III, which is followed by concluding remarks in Section IV.

2. The proposed method

As shown in Fig. 1, the proposed tool wear monitoring method mainly includes three stages: preprocessing of the raw signals and the feature construction, construction of feature fusion model, RUL prediction by a digital-analog linkage method. The details of each stage are introduced as follows.

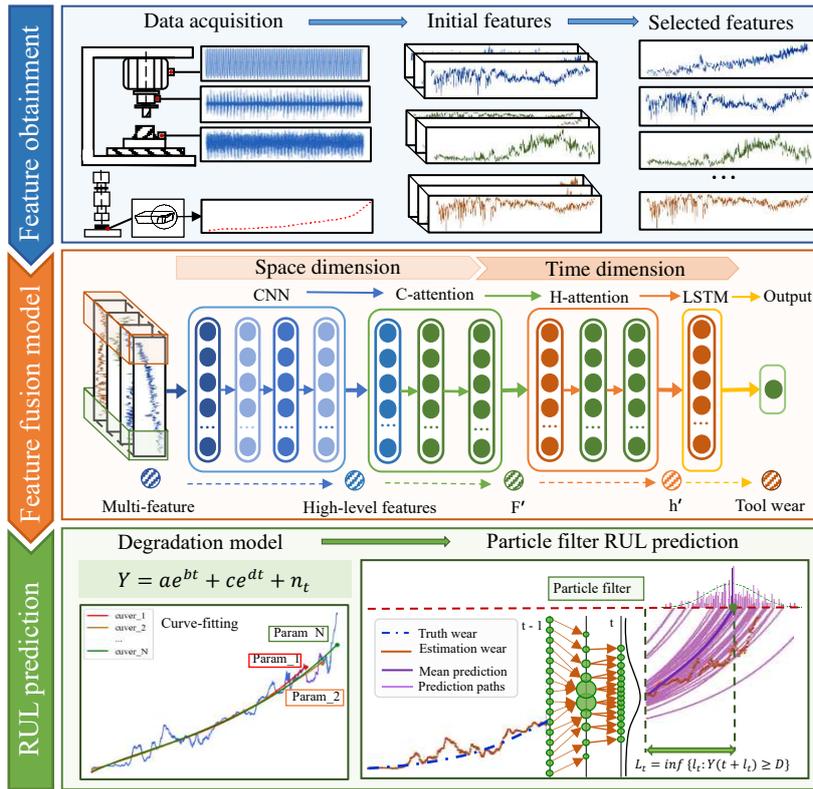


Fig 1. The proposed method.

2.1. Multi-feature construction

In order to characterize the degradation trend of tool more clearly, we construct features and select features preliminary after obtaining raw signals from different sensors.

2.1.1. Data obtaining and preprocessing

Usually, cutting force signals, spindle motor current signals, vibration signals and acoustic emission signals in the milling process are collected to characterize the cutter degradation process. However, the dynamometer, which record cutting force signals, is not suitable for practical application due to its high cost and negative impact on the rigidity of the machining system [23]. So, we use only vibration, current and acoustic emission signals to establish cutter degradation features.

The data come from different types of sensors and with some noise, directly feeding the raw sensor readings with high variance to the machine learning models may hinder the learning process and affect the model performance. To avoid this issue, we remove outliers and interpolate missing values.

2.1.2. Feature construction and preliminary selection

After preprocessing, large amounts of features from time domain, frequency domain and time-frequency domain are extracted to reflect cutter wear process comprehensively.

In time domain, we extracted 5 dimensional features and 6 dimensionless features. In the frequency domain, a data sequence is constructed by a full frequency spectrum and four sub-bands frequency spectra. We calculate the mean value, standard deviation, root mean square, skewness factor and kurtosis factor of every spectrum and its shape, and also calculate the spectrum entropy. In addition, time and frequency resolution of wavelet analysis is able to change adaptively depending on the frequency of the signal[24][25]. Hence, we generate features by performing haar wavelet package transform with a three-level decomposition. In the end, there are 11 features derived through time domain statistical analysis, 11 features in every frequency spectrum of frequency domain and 8 energy ratios of eight frequency sub-bands by performing haar wavelet package transform be acquired from every sensor signal.

However, some original features may be not informative for the degradation processes or even have a negative influence on the fusion result. Therefore, it is necessary to select original features preliminarily before fusion. In this paper, we calculate the Pearson correlation coefficient between the original feature and the tool

wear value to select the features. The equation of the Pearson correlation coefficient is defined as:

$$R_e(x_i, y) = \frac{cov(x_i, y)}{\sigma_{x_i} \sigma_y} \quad (1)$$

where x_i is the i th extracted statistics feature, y is the tool wear level, σ_{x_i} and σ_y are respectively standard deviation of x_i and y . In descending order, the features with correlation coefficient >0.6 or < -0.6 were selected as sensitive features. The final feature number extracted by each sensor is shown in Table 1.

Table 1 Feature number of each sensor.

2.2. Multi-feature fusion model construction

Deep learning is able to adaptive feature learning and nonlinear function approximation. In this paper, we construct a time-space attention mechanism driven multi-feature fusion model based on deep learning framework. It extracts the hidden

Channel	Vibration_x	Vibration_y	Vibration_z	Current	Sound	Total
Number	13	14	12	17	6	62

characteristics of tool wear state in space dimension, and improves prediction accuracy of time dimension in the same time.

2.2.1. Network framework

CNN is a depth-feedback neural network with convolution kernels of unit width. It connects each neuron to the feature map of the upper layer through weight matrix (convolution kernel) sparsely. Therefore, it is able to extract spatial information effectively and shares the weights[26][27]. In this model, a one-dimensional CNN layer is used to obtain the high-level features from the preliminary selected features. Then, the pooling layer is utilized after each 1D-CNN layer to compress the feature space and extract the significant local feature maps. The equation for this process is as follows,

$$\mathbf{c}_{i,k} = \text{ReLU}(\mathbf{W}_k * \mathbf{x}_i + \mathbf{b}_k), \quad (2)$$

$$\mathbf{p}_{n,k} = \text{pool}(\mathbf{c}_{i,k}), \quad (3)$$

where $*$ indicates the convolution operation, $\mathbf{c}_{i,k}$ represents the learned feature corresponding to the filter kernel \mathbf{W}_k on the sample sequence \mathbf{x}_i , $\text{ReLU}(\cdot)$ means the ReLU activation function and $\text{pool}(\cdot)$ is the pooling rule. In addition, in order to stress the significant features in space dimension, a channel attention mechanism is introduced. Finally, in order to accelerate convergence, we use the strategy of batch normalization (BN) after every layer.

To solve the problem of CNN ignoring the temporal dependency among data points in a given input sequence when dealing with the sequence data and the limitation of receptive field size. We build an RNN network after CNN further extracting the features to realize the tool wear monitoring in the time dimension. A special type of RNN named LSTM which models the dynamics of sequences by introducing the memory cells is used in this network. It can effectively ease the vanishing gradients problem of traditional RNN[28]. Its updating equations are given as follows:

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \mathbf{x}_t + \mathbf{V}_i \mathbf{h}_{t-1} + \mathbf{b}_i), \quad (4)$$

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \mathbf{x}_t + \mathbf{V}_f \mathbf{h}_{t-1} + \mathbf{b}_f), \quad (5)$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \mathbf{x}_t + \mathbf{V}_o \mathbf{h}_{t-1} + \mathbf{b}_o), \quad (6)$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} \mathbf{i}_t + \tanh \odot (\mathbf{W}_c \mathbf{x}_t + \mathbf{V}_c \mathbf{h}_{t-1} + \mathbf{b}_c), \quad (7)$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh(\mathbf{c}_t), \quad (8)$$

where at each time step t , hidden state \mathbf{h}_t is updated by current data at the same time step \mathbf{x}_t , the hidden state at the previous time step \mathbf{h}_{t-1} , the input gate \mathbf{i}_t , the forget gate \mathbf{f}_t , the output gate \mathbf{o}_t and a memory cell \mathbf{c}_t .

High-level feature extracted from CNN is fed into a layer of LSTM. It generates a hidden state sequence and a sequence of cell states. And return the hidden states of all time step. In order to further enhance the prediction ability of the model, a hidden state attention mechanism is used to assign weight for hidden state in time dimension. The weighted hidden state is input to the next layer of LSTM. Finally, the wear prediction is completed through a fully-connected layer. The structure of model is detailed as shown in Fig. 2.

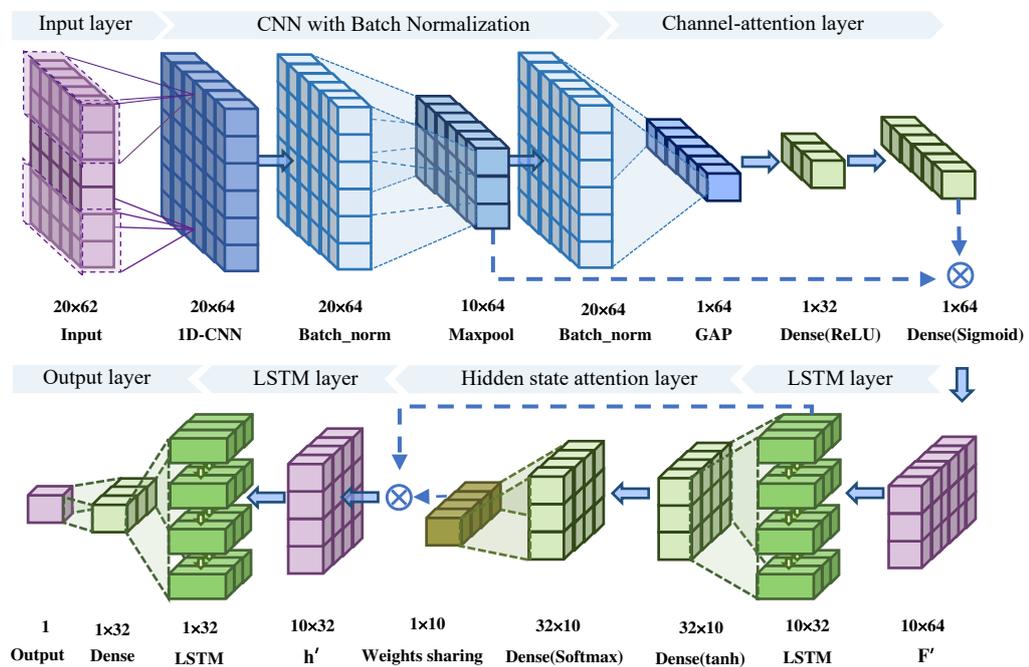


Fig 2. The structure of time-space attention model.

2.2.2. Time-space attention mechanism

A) Channel attention mechanism

The feature maps obtained from the above convolutional operation have different importance. Therefore, the importance degree of informative feature map is enhanced by a channel attention mechanism after pooling layer. As shown in Fig. 3, the channel attention mechanism assigns importance to each channel feature map by introducing an

attention function. Firstly, the state space information is aggregated through global average pooling (GAP), which greatly improves the expression ability of the network. Then, two fully-connected layers are used to generate the attention weights. It is computed as,

$$M = \sigma(W_1(W_0 * F_{avg} + b_0) + b_1), \quad (9)$$

where σ is the sigmoid function, W_0, W_1, b_0, b_1 are the weight and bias of the first and second fully-connected layers respectively, F_{avg} represents the squeezed state maps from GAP. Finally, the weighted attention weight and the output come from 1D-CNN is used as the last 1D-CNN layer.

$$F' = M \otimes F, \quad (10)$$

where F and F' represent the input and refined output of channel-attention mechanism respectively. \otimes denotes multiplication where the channel attention weights are broadcasted along the spatial dimension.

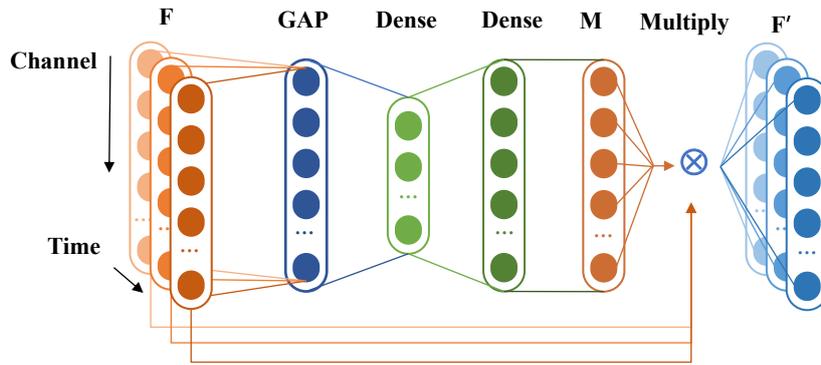


Fig 3. Channel-attention weight.

B) Hidden state attention mechanism

Not all of hidden states contribute equally to the task of tool wear prediction. Therefore, hidden state attention mechanism is introduced to assign attention weight to the hidden states output by the LSTM layer. The structure is shown in Fig. 4. We input

the hidden state obtained from the LSTM layer into a Dense layer with Softmax activation function to generate the attention weight vector \mathbf{a}_t . And then, the weight values of different channels are averaged in the same time step for weight sharing. After this, the hidden state vector \mathbf{h}_t is multiplied with the weight vector \mathbf{a}'_t to realize the attention assignment in the time dimension. The model is computed as

$$\mathbf{a}_t = f_a(\mathbf{h}_t) = \text{softmax}(\mathbf{W}_a \mathbf{h}_t + \mathbf{b}_a), \quad (11)$$

$$\mathbf{a}'_t = \text{avg}(\mathbf{a}_t), \quad (12)$$

$$\mathbf{h}'_t = \mathbf{a}'_t \otimes \mathbf{h}_t. \quad (13)$$

The output of the attention mechanism model is used as the input of the next LSTM layer to complete the tool wear prediction.

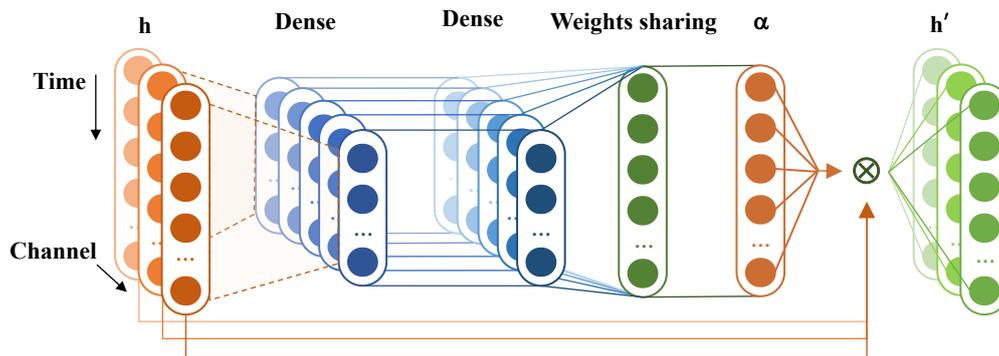


Fig 4. Hidden state attention weight.

2.3. Residual useful life prediction

After obtaining predicted tool wear, we verify its reliability in tool remaining useful life prediction by using the particle filter based digital-analog linkage method.

2.3.1. Tool degradation model

The predicted wear values are used as the observation values for tool health indicators. Combined with the degradation law of CNC machine tools, a double

exponential model was established as the observation equation, which can be expressed as follows,

$$Y = ae^{bk} + ce^{dk} + n_k, \quad (14)$$

where Y is the health states of the cutter at time k ; a , b , c and d are the model parameters; n_k is the observation noise.

2.3.2. Particle filter

Then, particle filter algorithm is used to update the model parameters and predict the tool failure curve. Particle filtering is based on Bayesian theory, and introduces Monte Carlo sampling to replace the complex integral operation in Bayesian filtering with sample mean[29][30]. The posterior probability density function of the system state is calculated as follow :

$$p(x_k | y_{1:k}) \approx \sum_{i=1}^M w_k^i \delta(x_k - x_k^i), \quad (15)$$

where x_k^i , w_k^i denote the state, normalized weight of i th particle at time k respectively, $\delta(\cdot)$ is the Dirac delta function.

In practical application, it is difficult to extract x_k^i directly from the posterior probability distribution. Therefore, the importance sampling density $q(x_{0:k} | z_{1:k})$ is used to extract samples to avoid the difficulty of directly extracting samples from the posterior probability density. In general, we take the prior distribution of system state as the importance probability density function, as shown in Equation 16,

$$q(x_{0:k} | z_{1:k}) = p(x_k | x_{k-1}). \quad (16)$$

Based on the latest observation information y_k , the updated particle weight will be calculated through the likelihood function,

$$w_k^i = w_{k-1}^i p(z_k | x_k^i). \quad (17)$$

In addition, the setting of initial parameter set $\{x_0^i\}_{i=1}^M$ of the model has a great influence on the final prediction result. We dynamically determine the initial parameter set of the model by curve fitting according to the observed values. It realizes the real-time accurate prediction of the remaining life.

2.3.3. RUL prediction

When the curve reaches the failure threshold D and the tool failure is judged. The time from the predicted starting point t to the point of failure is the remaining life of the tool L_k . The residual life calculation process is shown in the formula,

$$L_k = \inf \{l_k: Y(k + l_k) \geq D, k > 0\}, \quad (18)$$

where $\inf \{\cdot\}$ represents the lower limit of the variable, and l_k is the time required from the current time t to tool failure point.

In the end, the probability density function of N particles is used to express the time distribution of health indicator reaching the failure threshold.

The detailed procedure for residual useful life prediction is shown as follows,

Pseudo code: RUL prediction

Input: The wear values y predicted by multi-feature fusion model, and the degradation model.

Curve fitting: Obtained the initial parameter set of the tool degradation model $\mathbf{a}_0, \mathbf{b}_0, \mathbf{c}_0, \mathbf{d}_0$ by curve fitting of the obtained wear value y .

Particle filter:

Particle set initialization: when $k = 0$.

for $i = 1, 2, \dots, M$ **do**

Generating sampling particle: Generated sample particle $\{x_0^i\}_{i=1}^M$ from $p(x_0)$.

end

for $k = 1, 2, \dots$ **do**

Importance sampling:

for $i = 1, 2, \dots, M$ **do**

Sampling particle generation: Generated sampling particle $\{x_k^i\}_{i=1}^M$ from the importance probability density.

Particle weights calculation: Calculate and normalize w_k^i .

end

Resampling: Resampling the particle set $\{x_k^i, w_k^i\}$ results in $\{x_k^i, 1/M\}$.

Output: Calculate the state estimate $\hat{x}_k = \sum_{i=1}^M x_k^i w_k^i$ at time k .

end

RUL prediction: After obtaining the observed system state value y_k at time k and the estimated model parameter value $\mathbf{a}_k, \mathbf{b}_k, \mathbf{c}_k, \mathbf{d}_k, \mathbf{L}_k$ is obtained according to the life calculation formula $L_k = \inf\{l_k: Y(k + l_k) \geq D, k > 0\}$.

Output: tool residual useful life L_k at time k .

3. Experimental study and prediction result

In this section, the performance of the proposed method has been verified by a milling cutter experiment. The tool life cycle data were collected by multi-sensors. The prognostic wear results will be shown in Section 3.3. And the RUL prediction results show in Section 3.4.

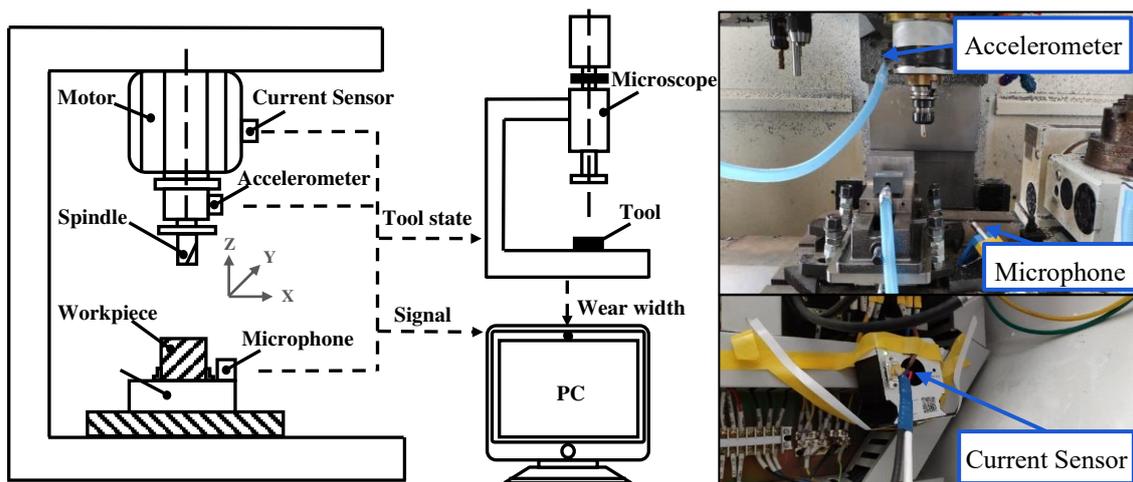


Fig 5. Sensor placement.

3.1. Experiment introduction

We use DAHENG VMC850 5-axis CNC to carry out the life cycle wear experiment of milling cutter. APMT1135 carbide milling cutter is used to mill 45# steel workpiece. During the experiment, we collected relevant signals through triaxial accelerometer PCB 356A15, current sensor Anyway CSA201-P030T01, acoustic emission sensor PAC WD and free-field microphone Bruel&Kjaer 4966. The sensor

placement is shown in Fig. 5. After a tool cutting is complete, we disassemble the tool and measure the wear width by a vision microscope. The obtained tool wear statuses are shown in Fig. 6. According to the milling cutter life criterion in GBT16460-1996, when the wear width (average wear width) of flank face $\geq 300 \mu\text{m}$, the tool is judged to have failed.

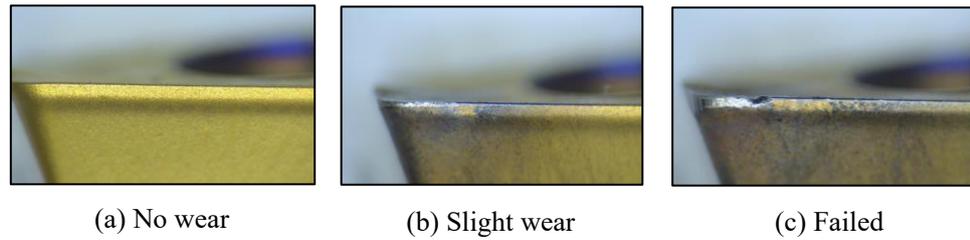


Fig. 6. Tool wear status.

3.2. Dataset establishment and model training

As shown in Table 2, the signal of the first four tools collected under condition 1 are used as the training dataset, and the cutter 5 and 6 are used as the verification dataset. Meanwhile, the robustness of the model was verified by the data of the cutting tools under condition 2. Only 10s of data under stable cutting condition is collected during each cutting process, and each piece of data is taken as a sample.

Table2 Tool working conditions.

Cutting condition	Spindle speed	Feed rate	Depth of cut	Cutters	
Condition1	4500rpm	400mm/min	1.75 mm	C1_1	C1_2
				C1_3	C1_4
				C1_5	C1_6
Condition2	3500rpm	300mm/min	1.5 mm	C2_1	

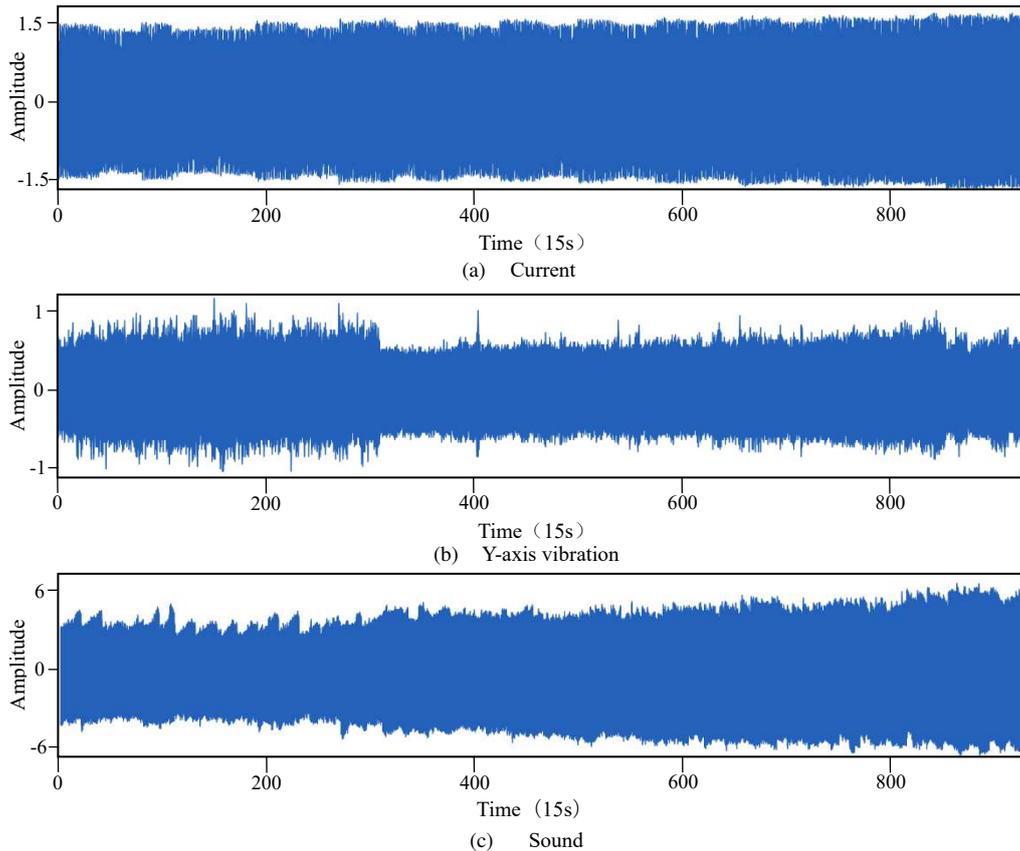


Fig 7. Monitoring signals of C1_1 during the whole operating life.

Fig. 7 shows the current signal, Y-axis vibration signal and sound signal of C1_1 in its whole life cycle. Seen from the figure, there is almost no trendability, especially the current signal. Some features of different sensing signals in time domain, frequency domain and time frequency domain are shown in Figure 8. Through the preliminary selection of correlation analysis, some features that have no trendability at all, such as (b) and (c), can be removed. However, not all selected features have a good correlation

with the wear value as (f). In order to accurately characterize the tool degradation trend and predict the wear value, features need to be fused to further extract the hidden information of time and space dimensions.

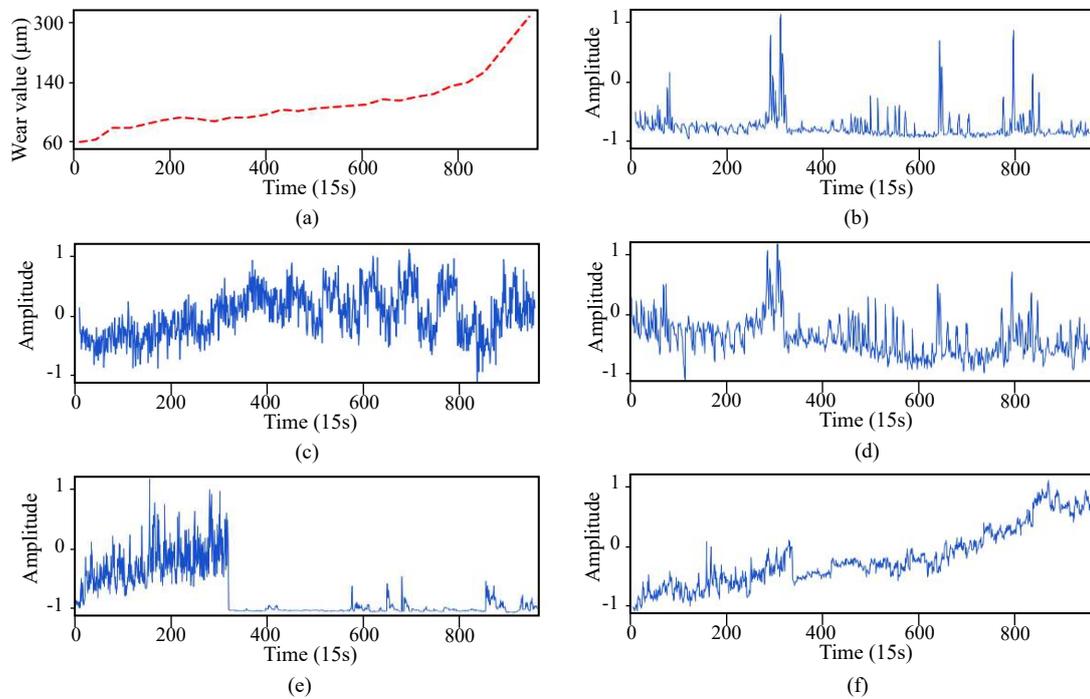


Fig 8. Tool wear and extracted features: (a) Wear curve, (b) and (c) Features without trendability at all, (d) and (e) Features have some trendability, (f) Feature have good trendability.

In order to obtain more useful temporal information to improve the prediction performance[31], a fixed-size time window is used to segment the features. As shown in Fig. 9, W represents the width of the time window, s is the shifting size, and n is the number of features. In this paper, W and s are 20 and 1 respectively. In the end, divide the whole data into the training dataset, validation dataset and test dataset.

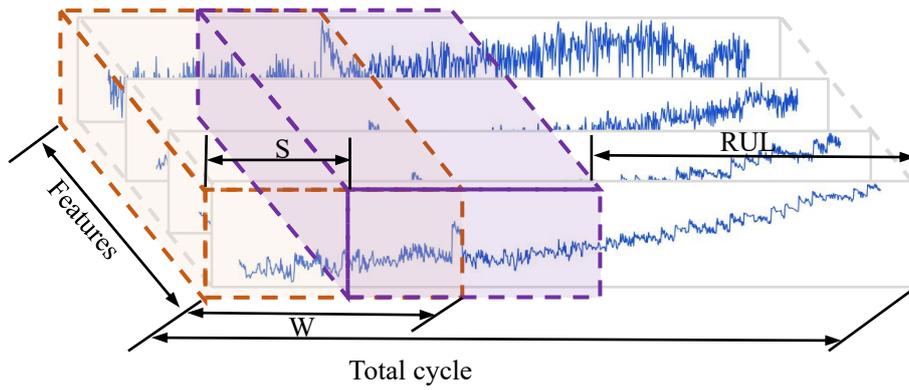


Fig 9. Time window segment the features.

The train dataset and validation dataset are fed into the monitoring model proposed above to train model. Multi-feature is the input to the model and wear value as the output. The mean square error (MSE) between the outputs and the correct results was chosen as the training loss, and chooses the Adam as the optimizer. The initial learning rate is set to 0.01 and the mini-batch size was 64. Each training epoch was monitored by the loss on the validation set. If the loss stops after 4 epochs, the learning rate will be reduced; if the loss does not drop after 10 epochs, the training will be finished early.

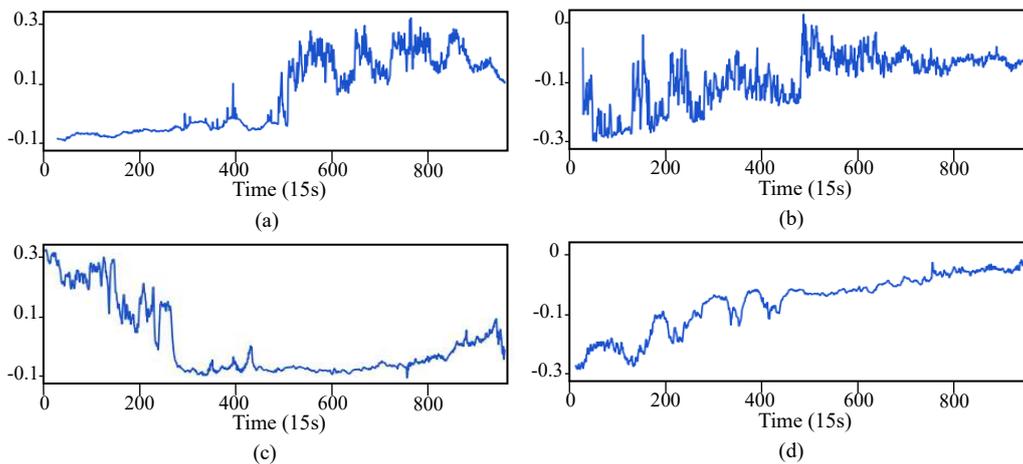


Fig 10. The features fused by time-space attention mechanism fusion model:(a) and (b)

Features after spatial dimension information extraction, (c) and (d) Features after time dimension information extraction.

Fig. 10 (a) and (b) are the features after spatial dimension information extraction. Compared with the pre-selected features in Fig. 8, the trendability is enhanced but has great fluctuation, still cannot characterize the tool degradation trend. After time dimension extraction (as shown in the (c), (d)), the feature more obvious trendability and lower fluctuation. And there is strong correlation with the tool wear value. It indicates that extracting the hidden information of features in time and space dimensions simultaneously can effectively represent the tool degradation trend.

3.3. Tool wear prediction results and comparison analysis

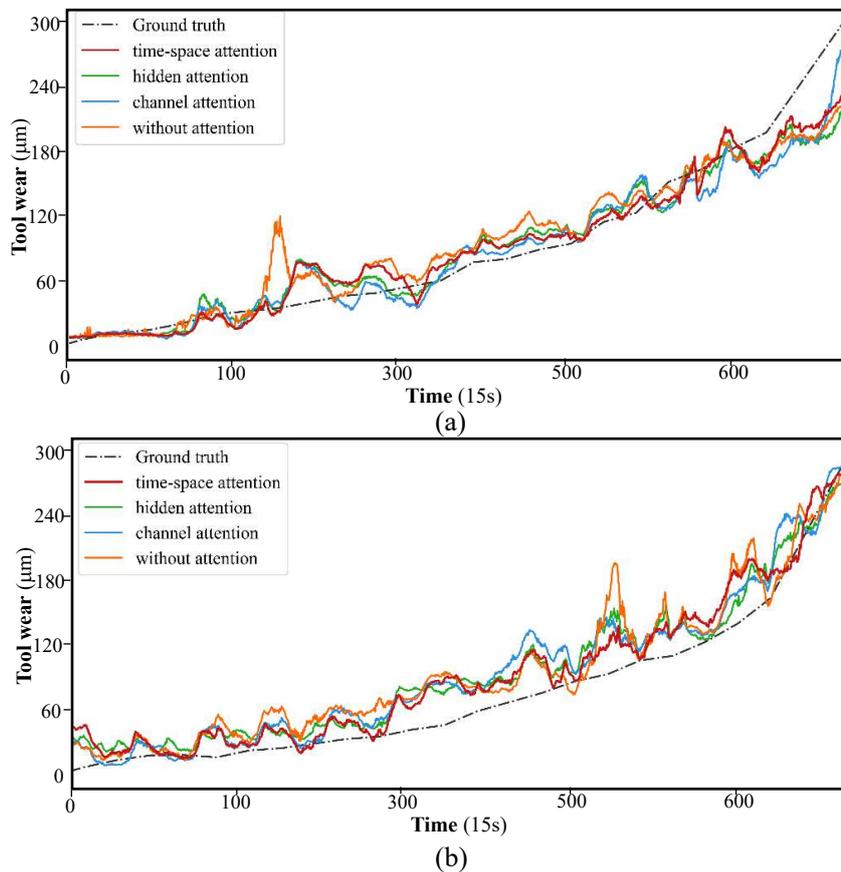


Fig 11. Tool wear prediction results of (a) C1_5 and (b) C1_6.

In order to illustrate the effectiveness of the proposed time-space attention mechanism in the tool wear monitoring model. We compared it with the other three

models named channel attention model, hidden state attention model and without attention model.

As shown in Fig. 11, (a) and (b) present the different model tool wear prediction result of cutter C1_5 and C1_6 using different models. It can be seen that although the predicted results fluctuate locally, all the models can monitor the overall trend of tool wear, which proves the feasibility of the multi-feature fusion model integrated with CNN and RNN proposed in this paper. The network with the time-space attention mechanism has obvious advantages in the accuracy and stability of the prediction. By comparison, it can be seen that the model with the hidden state attention mechanism can better fit the tool wear trend in the early stage, but when the tool wear suddenly changes in the late stage, its prediction ability decreases obviously. Channel attention mechanism just makes up for this defect, it enhances the ability of model to extract spatial dimension information. Therefore, in the later stage of degradation, the time-space attention model can also accurately monitor the tool wear.

At the same time, the data of C2_1 are fed into monitoring model. In Fig. 12, we can see that under different working conditions, the model can still achieve the monitoring of tool wear value well. This proves that the model has good robustness.

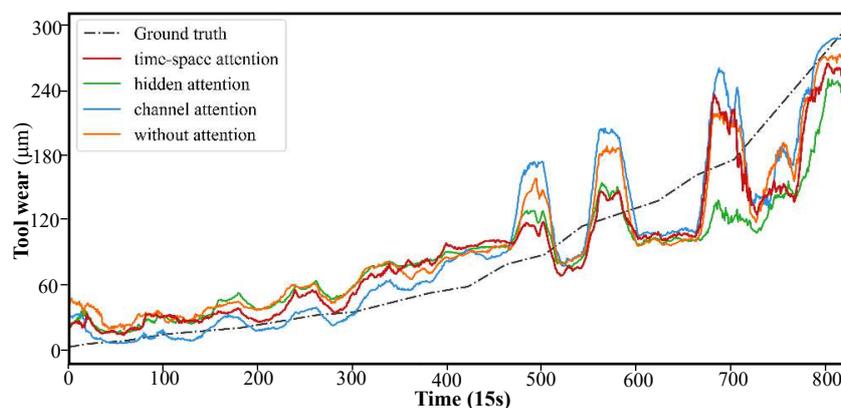


Fig 12. Tool wear prediction results of C2_1.

For the sake of quantitatively and visually assessing the prediction performance of the proposed model, relative error and correlation factor are introduced to evaluate the prediction accuracy of model, and Root Mean Square Error (RMSE) and error standard deviation (σ) are used to evaluate the stability of the model. The four criteria are as follows,

$$\text{Relative error} = \frac{\text{norm}_2(l-p)}{\text{norm}_2(p)} \quad (19)$$

$$\text{Correlation factor} = \frac{\sum_{i=1}^n (fp_i - \bar{p})(l_i - \bar{l})}{\sqrt{\sum_{i=1}^n (p_i - \bar{p})^2 \sum_{i=1}^n (l_i - \bar{l})^2}} \quad (20)$$

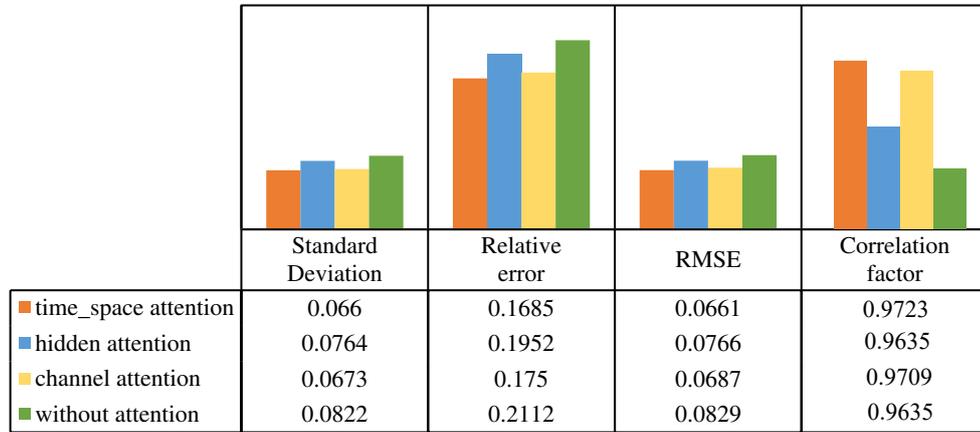
$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (l_i - p_i)^2} \quad (21)$$

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (e_i - \bar{e})^2} \quad (22)$$

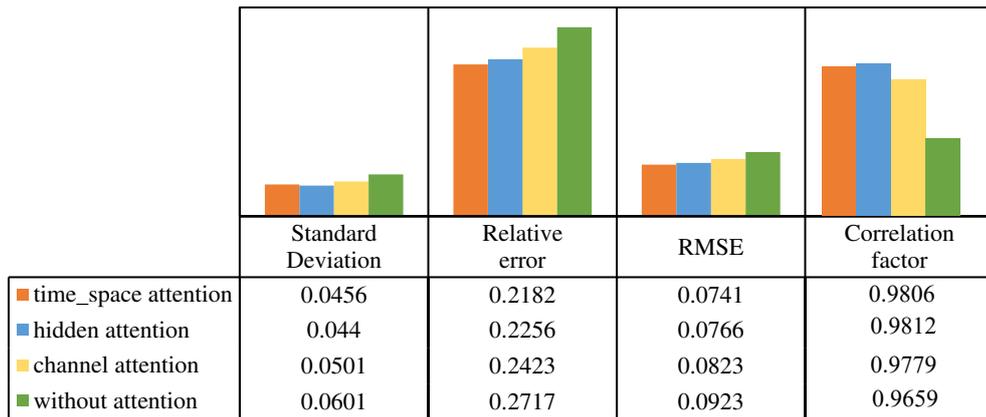
where l is the actual wear value, p is the prediction result, e is the relative error and norm_2 is the 2-norm.

The prediction results for the criteria of the four networks are tabulated in Fig. 13. It can be clearly seen that the tool wear monitoring network based the time-space attention mechanism is better than the attention mechanism using time or space dimension alone and the network without attention mechanism in terms of accuracy and stability of prediction. Moreover, after using the attention mechanism, the correlation between the predicted results and the real wear value is significantly improved, which further demonstrated that the attention mechanism proposed in this paper inhibits the useless

information, enhances the key information, and improves the feature fusion ability of model.



(a) C1_5

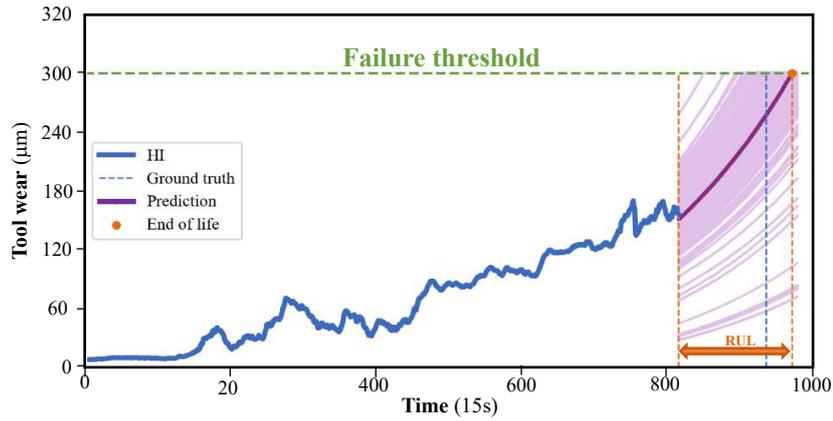


(b) C1_6

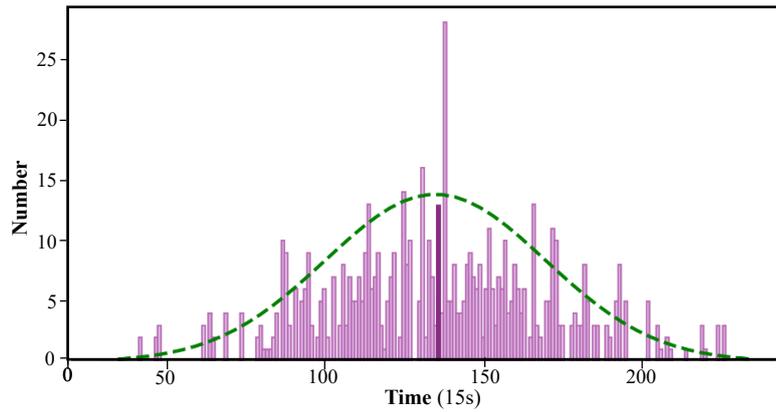
Fig 13. Tool wear performance estimation results of four different prognostics networks.

3.4. RUL prediction result

Taking cutter C1_5 as an example. As shown in the Fig. 14 (a), when the curves of 1000 particles reach the failure threshold D and the tool failure is judged. In Fig. 11 (b), the distribution of the predicted RULs of 1000 particles is approximately normally distributed. So, we take the mean of the particle RUL distribution at each moment to get the predicted value at this moment.



(a) Particle distribution



(b) Particle RUL distribution

Fig 14. RUL prediction result.

The tool RUL results of every time as shown in Fig. 15. As can be seen from the above results, at the beginning of the prediction stage, the prediction has a large deviation from the true life. This is because there is less data accumulation in the beginning and the model cannot be adequately trained. As time went on, the observed data increased, and the predicted results tended to be stable and gradually approached the truth. In general, it is feasible to use the tool wear curve established by the method proposed in this paper to predict the tool RUL, and it generally has a good downward trend. Especially in the later period of tool degradation, the prediction accuracy is greatly improved, which is of great significance to prevent serious tool wear and reduce economic loss.

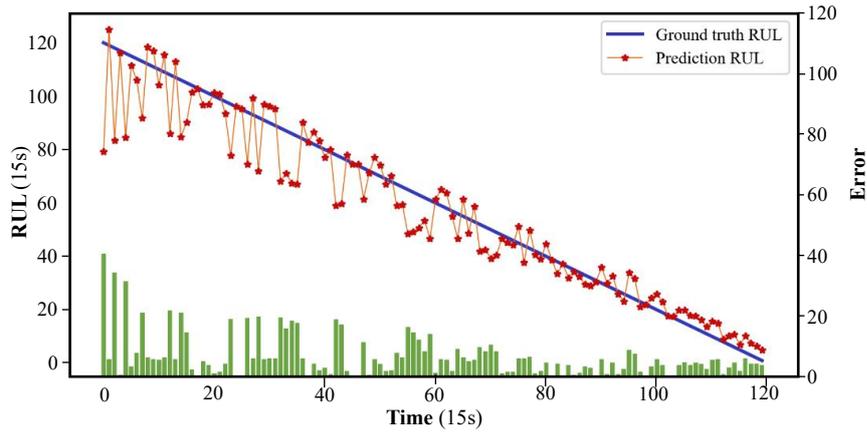


Fig 15. RUL prediction result.

We also evaluated the performance of the four networks. Relative accuracy and standard deviation (σ) of error are used as criteria, which are given by,

$$Relative\ accuracy = \frac{1}{n} \sum_{i=1}^n \frac{|RUL_i - RUL'_i|}{RUL_i}, \quad (23)$$

where RUL_i and RUL'_i denote the predictive RUL and ground truth RUL at time i .

Table.3 RUL performance estimation results of four different prognostics networks.

	Relative accuracy	Standard deviation
Time-space attention	0.2110	0.4432
Hidden attention	0.2595	0.4917
Channel attention	0.3127	0.4925
Without attention	0.2542	0.4964

As shown in Table 3, the network based on time-space attention mechanism is superior to other networks in both accuracy and stability.

4. Conclusion

In this paper, a multi-feature fusion method for tool condition monitoring is proposed. This method collects signals from different sensors and establishes multiple features. Then, based on the self-attention mechanism, the features are extracted from the space and time dimensions to obtain the tool wear prediction value. Finally, the remaining service life is predicted by the digital-analog linkage method based on

particle filtering. The performance of the proposed method is verified through the tool life cycle experiment, and shows the following conclusions.

- (1) The channel attention mechanism and hidden state attention mechanism used in the method assign importance to features from space and time dimensions. Even if the cutting force signal with a good trendability is abandoned. It can also effectively extract the hidden information in signals to achieve accurate prediction of tool wear.
- (2) Using the predicted wear value, the RUL can be accurately predicted by the digital-analog linkage method based on the particle filter. Especially in the later stage of tool degradation, with the increase of observation data, the accuracy is greatly improved.
- (3) Through comparative analysis with other models, the multi-feature fusion model based on time-space attention mechanism has obvious advantages in accuracy and stability. The accuracy of using the obtained wear value for the remaining life prediction is also higher than that of other models. At the same time, the model is applied to the tool data under different working conditions, which verifies that the model has good robustness.

The above conclusions demonstrate that the proposed method has the ability of multi-feature fusion, can effectively predict the tool wear of the milling cutter and realize the remaining life prediction. In the future, we will in-depth study the construction and optimization of health indicators for tool wear. It is expected to improve the performance of the feature fusion model for tool wear monitoring under different working conditions.

Declarations:

a. Funding :

This work was supported in part by the National Natural Science Foundation of China(51905452, 51775452), Local Development Foundation guided by the Central Government(2020ZYD012) and Planning Project of Science & Technology Department of Sichuan Province under Grant (2020YFN0062).

b. Conflicts of interest/Competing interests :

The contribution of this paper is our original work and all authors have agreed to the submission of this work to this journal. And there is no conflict of interest.

c. Availability of data and material :

Relevant data is not public.

d. Code availability:

The code is not public.

e. Ethics approval:

Not applicable.

f. Consent to participate :

Agree.

g. Consent for publication:

Agree.

Reference

1. Liu Y, Guo L, Gao H, et al (2022) Machine vision based condition monitoring and fault diagnosis of machine tools using information from machined surface texture : A review
Machine vision based condition monitoring and fault diagnosis of machine tools using information from machined surfac. Mech Syst Signal Process 164:108068.
<https://doi.org/10.1016/j.ymssp.2021.108068>

2. Wang Y, Zheng L, Wang Y (2021) Event-driven tool condition monitoring methodology considering tool life prediction based on industrial internet. *J Manuf Syst* 58:205–222.
<https://doi.org/10.1016/j.jmsy.2020.11.019>
3. Liao Z, Gao D, Lu Y, Lv Z (2016) Multi-scale hybrid HMM for tool wear condition monitoring. *Int J Adv Manuf Technol* 84:2437–2448. <https://doi.org/10.1007/s00170-015-7895-3>
4. Qiao H, Wang T, Wang P (2020) A tool wear monitoring and prediction system based on multiscale deep learning models and fog computing. *Int J Adv Manuf Technol* 108:2367–2384.
<https://doi.org/10.1007/s00170-020-05548-8>
5. An H, Wang G, Dong Y, et al (2019) Tool life prediction based on Gauss importance resampling particle filter. *Int J Adv Manuf Technol* 103:4627–4634.
<https://doi.org/10.1007/s00170-019-03934-5>
6. Yang Y, Guo Y, Huang Z, et al (2019) Research on the milling tool wear and life prediction by establishing an integrated predictive model. *Meas J Int Meas Confed* 145:178–189.
<https://doi.org/10.1016/j.measurement.2019.05.009>
7. Zhu D, Zhang X, Ding H (2013) Tool wear characteristics in machining of nickel-based superalloys. *Int J Mach Tools Manuf* 64:60–77.
<https://doi.org/10.1016/j.ijmachtools.2012.08.001>
8. Yaguo Lei, Naipeng Li, Szymon Gontarz, Jing Lin, Stanislaw Radkowski and JD (2016) A Model-Based Method for Remaining Useful Life Prediction of Machinery. *IEEE Trans Reliab* 65:1314–1326. <https://doi.org/10.1016/j.measurement.2020.108277>

9. Wang H, Tang G, Zhou Y, Huang Y (2020) A Novel Multiscale Deep Health Indicator with Bidirectional LSTM Network for Bearing Performance Degradation Trend Prognosis. *Shock Vib* 2020:.. <https://doi.org/10.1155/2020/8871981>
10. Guo L, Li N, Jia F, et al (2017) A recurrent neural network based health indicator for remaining useful life prediction of bearings. *Neurocomputing* 240:98–109.
<https://doi.org/10.1016/j.neucom.2017.02.045>
11. Yan T, Lei Y, Li N, et al (2021) Degradation modeling and remaining useful life prediction for dependent competing failure processes. *Reliab Eng Syst Saf* 212:107638.
<https://doi.org/10.1016/j.ress.2021.107638>
12. Li C, Guo L, Gao H, Li Y (2021) Similarity-Measured Isolation Forest: Anomaly Detection Method for Machine Monitoring Data. *IEEE Trans Instrum Meas* 70:..
<https://doi.org/10.1109/TIM.2021.3062684>
13. Sun H, Pan J, Zhang J, Cao D (2020) Non-linear Wiener process–based cutting tool remaining useful life prediction considering measurement variability. *Int J Adv Manuf Technol* 107:4493–4502. <https://doi.org/10.1007/s00170-020-05264-3>
14. Wu D, Jennings C, Terpenney J, et al (2017) A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests. *J Manuf Sci Eng Trans ASME* 139:.. <https://doi.org/10.1115/1.4036350>
15. Wu J, Su Y, Cheng Y, et al (2018) Multi-sensor information fusion for remaining useful life prediction of machining tools by adaptive network based fuzzy inference system. *Appl Soft Comput J* 68:13–23. <https://doi.org/10.1016/j.asoc.2018.03.043>

16. Mao W, Feng W, Liu Y, et al (2021) A new deep auto-encoder method with fusing discriminant information for bearing fault diagnosis. *Mech Syst Signal Process* 150:107233. <https://doi.org/10.1016/j.ymssp.2020.107233>
17. Guo L, Gao H, Huang H, et al (2016) Multifeatures Fusion and Nonlinear Dimension Reduction for Intelligent Bearing Condition Monitoring. *Shock Vib* 2016:.. <https://doi.org/10.1155/2016/4632562>
18. Yu Y, Guo L, Gao H, et al (2021) Pareto-optimal Adaptive Loss Residual Shrinkage Network for Imbalanced Fault Diagnostics of Machines. <https://doi.org/10.1109/TII.2021.3094186>
19. Zhao M, Kang M, Tang B, Pecht M (2019) Multiple Wavelet Coefficients Fusion in Deep Residual Networks for Fault Diagnosis. *IEEE Trans Ind Electron* 66:4696–4706. <https://doi.org/10.1109/TIE.2018.2866050>
20. Yang B, Liu R, Zio E (2019) Remaining useful life prediction based on a double-convolutional neural network architecture. *IEEE Trans Ind Electron* 66:9521–9530. <https://doi.org/10.1109/TIE.2019.2924605>
21. Zhou JT, Zhao X, Gao J (2019) Tool remaining useful life prediction method based on LSTM under variable working conditions. *Int J Adv Manuf Technol* 104:4715–4726. <https://doi.org/10.1007/s00170-019-04349-y>
22. Zheng Z, Huang S, Weng R, et al (2020) Improving Self-Attention Networks With Sequential Relations. 28:1707–1716
23. Pal S, Heyns PS, Freyer BH, et al (2011) Tool wear monitoring and selection of optimum cutting conditions with progressive tool wear effect and input uncertainties. *J Intell Manuf* 22:491–504. <https://doi.org/10.1007/s10845-009-0310-x>

24. Ocak H, Loparo KA, Discenzo FM (2007) Online tracking of bearing wear using wavelet packet decomposition and probabilistic modeling: A method for bearing prognostics. *J Sound Vib* 302:951–961. <https://doi.org/10.1016/j.jsv.2007.01.001>
25. Yen GG, Member S, Lin K (2000) Wavelet packet feature extraction for Vibration Monitoring. *J Sound Vib* 47:650–667
26. Aghazadeh F, Tahan A, Thomas M (2018) Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process. *Int J Adv Manuf Technol* 98:3217–3227. <https://doi.org/10.1007/s00170-018-2420-0>
27. Jia F, Lei Y, Lu N, Xing S (2018) Deep normalized convolutional neural network for imbalanced fault classification of machinery and its understanding via visualization. *Mech Syst Signal Process* 110:349–367. <https://doi.org/10.1016/j.ymssp.2018.03.025>
28. Sepp Hochreiter J, Schmidhuber U (1997) Long Short-Term Memory. *Neural Comput* 9:1735–1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
29. Sun H, Cao D, Zhao Z, Kang X (2018) A Hybrid Approach to Cutting Tool Remaining Useful Life Prediction Based on the Wiener Process. *IEEE Trans Reliab* 67:1294–1303. <https://doi.org/10.1109/TR.2018.2831256>
30. Qian Y, Yan R (2015) Remaining Useful Life Prediction of Rolling Bearings Using an Enhanced Particle Filter. *IEEE Trans Instrum Meas* 64:2696–2707. <https://doi.org/10.1109/TIM.2015.2427891>
31. Zhang C, Lim P, Qin AK, Tan KC (2017) Multiobjective Deep Belief Networks Ensemble for Remaining Useful Life Estimation in Prognostics. *IEEE Trans Neural Networks Learn Syst* 28:2306–2318. <https://doi.org/10.1109/TNNLS.2016.2582798>