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

Tingting Feng1 · Liang Guo1,2 · Hongli Gao1 · Tao Chen1 · Yaoxiang Yu1 · Changgen Li1

Received: 31 August 2021 / Accepted: 7 March 2022 / Published online: 21 March 2022 © The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

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 can provide complementary information in the feature space. In addition, monitoring signals are time series data, which also contains a wealth of time dimension tool degradation information. However, how to fuse multi-sensor information in time and space dimensions is a key issue that needs to be solved. In this paper, a new time–space attention mechanism driven multi-feature fusion method is proposed for tool wear monitoring and residual useful life (RUL) prediction. A time–space attention mechanism is innovatively introduced into the tool wear monitoring model, and features are weighted from two dimensions of space and time. It can more accurately capture the complex spatio-temporal relationship between tool wear values and features, so that the model can accurately predict wear values even if it gives up cutting force signals with good trends. The experimental results show that the correlation of the predicted wear and the actual wear is greater than 0.95, and the relative accuracy of the RUL predicted by the predicted wear combined with the particle filter can also be around 0.78. Compared with other feature fusion models, 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

1 Introduction
Numerical control machine tool is the fundamental device in smart manufacturing systems, which plays a significant role in the overall production system [1, 2]. Cutting tool as the tooth of the numerical control machine tool, even a small tool failure will have a great negative impact on product quality [3, 4]. More seriously, it may even cause unscheduled downtime. Therefore, the accurate tool RUL prediction and appropriate maintenance of the tool is meaningful for promoting processing efficiency and reducing production cost [5]. The prediction accuracy of tool RUL mainly depends on the reliability of tool wear condition monitoring method. Therefore, this is the research focus of tool prognostics health management (PHM).

Among numerous tool wear monitoring methods, the data-driven indirect method monitors tool wear by establishing the correlation model between tool wear and sensor signals (such as cutting force, vibration, acoustic emission, and current) [6], which has been widely investigated because of convenience and lower cost. The steps of this method are as follows: collecting appropriate sensor signals, selecting appropriate methods to process the signals, extracting features sensitive to the tool wear, and establishing the predictive model [5]. At the same time, selecting effective evaluation criteria is indispensable for evaluating the performance of the model [7].

With the rapid development of sensing and computing technologies, numerous studies have found that the amount of information represented by the fusion of multiple features from different sensor signals is much larger...
than that of single sensor features. Multi-feature fusion can overcome the problem that single sensor features are difficult to accurately describe the condition of the machine in advanced manufacture [8–11]. However, how to fuse multi-sensor information is a key issue that needs to be solved of current indirect tool wear monitoring methods. Lee et al. [12] proposed a generic fuzzy logic algorithm for validation and fusion of uncertain sensor data. Equeter et al. [13] used gamma process monitoring tool wear for providing tool RUL at several cutting speeds. Sun et al. [14] proposed a nonlinear 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. These methods are based on statistical models, which are difficult to apply to the actual processing production, because they require a large number of complex mathematical operations and has weak anti-noise ability.

As a very popular method in recent years, machine learning does not require complicated mathematical calculations and has good noise immunity. Thus, fusing features by machine learning is increasingly applied to tool wear monitoring. Hotait et al. [15] proposed a methodology of the real-time monitoring (IRT-OPTICS) for the detection of defect by fusing three domain features (time, frequency, and scale). Srinivasan et al. [16] integrated vibration and acoustic sensor to tool condition monitoring by linear support vector machines. Wu et al. [17] proposed a multi-sensor information fusion system for online RUL prediction of tools. The system used adaptive network-based fuzzy inference system to fuse features. However, it is difficult for traditional machine learning to directly extract effective information from a large number of original features [18]. To achieve accurate wear monitoring, some complex preprocessing of features, such as principal component analysis (PCA) and empirical mode decomposition (EMD), are needed to be carried out.

Deep neural network (DNN) is a good solution to the above problem [19–21]. In reference [22], 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 [23], so as to realize the feature fusion in the time dimension. These methods achieve feature fusion in spatial and temporal dimensions to some extent, but they do not consider the different contributions of different features to tool degradation at different time steps.

Attention mechanism is a state-of-the-art technology widely used in natural language processing (NLP) [24, 25]. It can instruct the DNN to pay more attention to the relevant information and ignore the irrelevant information, so as to extract the potential information in the data at a deeper level. 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 first. 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 is used for cutting force monitoring, has high cost and negative impact on machining system rigidity. We abandon the cutting force signal with good trendability. Only vibration, current, and sound signals are used to construct multiple features 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. The hidden state attention mechanism assigns attention weights to the hidden state of different time steps, and the most relevant information in the input sequence is stressed in the time dimension to predict the tool wear. The superior performance of the proposed method is verified through an 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, curve fitting is used to dynamically set initial parameters.
This paper is organized as follows. Section 1 introduces literature reviews on tool wear monitoring and RUL prediction methods based on feature fusion. Section 2 describes the proposed wear prediction method in detail. In Sect. 3, we compare the proposed model with other methods to verify its superiority by experimental study, which is followed by concluding remarks in Sect. 4.

2 The proposed method

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

2.1 Multi-feature construction

In order to characterize the degradation trend of tools more clearly, we construct 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 is used for cutting force monitoring, is not suitable for practical application due to its high cost and negative impact on the rigidity of the machining system [25]. Therefore, we use only vibration, current, and sound signals to establish cutter degradation features.
The data that come from different sensors are with some noise, and 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, outliers and interpolate missing values are removed.

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 comprehensively reflect the cutter wear process.

In the 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-band 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 signals [26, 27]. Hence, in time–frequency domain, we generate features by performing haar wavelet packet transform with a three-level decomposition. In the end, the original feature set of each channel contains 11 time-domain features, 55 frequency-domain features of a full frequency spectrum and four sub-band frequency spectra, and 8 time–frequency domain features obtained through haar wavelet packet transform. The original feature set has a total of 74 × 5 = 370 features.

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 conducting 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:

\[
R_x(x_i, y) = \frac{\text{cov}(x_i, y)}{\sigma_x \sigma_y} \tag{1}
\]

where \(x_i\) is the \(i\)th extracted statistics feature, \(y\) is the tool wear level, and \(\sigma_x\) and \(\sigma_y\) indicate standard deviation of \(x_i\) and \(y\), respectively. The features with correlation coefficient > 0.6 or < −0.6 were selected as sensitive features.

2.2 Multi-feature fusion model construction

Deep learning is able to adapt feature learning and nonlinear function approximation. In this paper, we construct a time–space attention mechanism driven multi-feature fusion model based on the deep learning framework. It can extract the hidden characteristics of tool wear state in space dimension and improve prediction accuracy of time dimension at 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 maps of the upper layer through weight matrix (convolution kernel) sparsely. Therefore, it is able to extract spatial information effectively and shares weights [28]. In this model, a one-dimensional CNN layer is used to obtain 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:

\[
c_{i,k} = \text{ReLU}(W_k * x_i + b_k), \tag{2}
\]

\[
p_{n,k} = \text{pool}(c_{i,k}), \tag{3}
\]

where \(\ast\) indicates the convolution operation, \(c_{i,k}\) represents the learned feature corresponding to the filter kernel \(W_k\) on the sample sequence \(x_i\), \(\text{ReLU}(\cdot)\) 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 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 gradient problem of traditional RNN [29]. Its updating equations are given as follows:

\[
i_t = \sigma(W_x x_t + V_t h_{t-1} + b_t). \tag{4}
\]
where at each time step \( t \), hidden state \( h_t \) is updated by the current data at the same time step \( x_t \), the hidden state at the previous time step \( h_{t-1} \), the input gate \( i_t \), the forget gate \( f_t \), the output gate \( o_t \), and a memory cell \( c_t \).

\[
f_t = \sigma(W_f x_t + V_f h_{t-1} + b_f),
\]

\[
o_t = \sigma(W_o x_t + V_o h_{t-1} + b_o),
\]

\[
c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c x_t + V_c h_{t-1} + b_c),
\]

\[
h_t = o_t \odot \tanh(c_t),
\]

where at each time step \( t \), hidden state \( h_t \) is updated by the current data at the same time step \( x_t \), the hidden state at the previous time step \( h_{t-1} \), the input gate \( i_t \), the forget gate \( f_t \), the output gate \( o_t \), and a memory cell \( c_t \).

of all time steps are returned. In order to further enhance the prediction ability of the model, a hidden state attention mechanism is used to assign weights to hidden states in time dimension. The weighted hidden states are input to the next layer of LSTM. Finally, the wear prediction is completed through a fully connected layer. The structure of the proposed model is detailed as shown in Fig. 2.

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 degrees of informative feature maps are enhanced by using a channel attention mechanism after the pooling layer. As shown

Fig. 2 The structure of time–space attention model

Fig. 3 Channel-attention weight
in Fig. 3, the channel attention mechanism assigns different 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:

\[ M = \sigma(W_1(W_0 \ast F_{avg} + b_0) + b_1), \quad (9) \]

where \( \sigma \) 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, and \( F_{avg} \) represents squeezed state maps from GAP. Finally, the weighted feature maps and the outputs come from 1D-CNN are multiplied and as the inputs of the next layer.

\[ F' = M \otimes F, \quad (10) \]

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

(B) Hidden state attention mechanism

Actually, not all hidden states contribute equally to the task of tool wear prediction. Therefore, hidden state attention mechanism is introduced to assign attention weights to the hidden states output by the LSTM layer. The structure is shown in Fig. 4. We input the hidden states obtained from the LSTM layer into a dense layer with Softmax activation function to generate the attention weight vector \( a_t \). Then, the weight values of different channels are averaged in the same time step for weight sharing. Subsequently, the hidden state vector \( h_t \) is multiplied with the weight vector \( a_t \) to realize the attention assignment in the time dimension. The model is computed as:

\[ a_t = f_a(h_t) = softmax(W_{a}h_t + b_{a}), \quad (11) \]
\[ a'_t = \text{avg}(a_t), \quad (12) \]
\[ h'_t = a'_t \otimes h_t. \quad (13) \]

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

### 2.3 Residual useful life prediction

After obtaining the predicted tool wear, its reliability is verified in tool RUL 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. According to the reference [30], 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 \( Y \) denotes the health state of the cutter at time \( k; \ a, b, c, \) and \( d \) are the model parameters; and \( n_k \) represents the observation noise.

#### 2.3.2 Particle filter

Then, particle filter algorithm is used to update 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 operations in Bayesian filtering with sample mean \([31, 32]\). The posterior probability density function of the system state is calculated as follows:

\[
p(x_k|y_{1:k}) \approx \sum_{i=1}^{M} w^i_k \delta(x_k - x^i_k),
\]

where \(x^i_k\) and \(w^i_k\) denote the state and normalized weight of \(i\) th particle at time \(k\), respectively, and \(\delta(\cdot)\) is the Dirac delta function.

In practical application, it is difficult to extract \(x^i_k\) directly from the posterior probability distribution. Therefore, the importance sampling density \(q(x_0:k|z_{1:k})\) is used to extract samples to overcome this problem. In general, we take the prior distribution of the system state as the importance probability density function, as shown in Eq. (16):

\[
q(x_k|x_{k-1}, z_k) = p(x_k|x_{k-1}).
\]

Based on the latest observation information \(y_k\), the updated particle weights will be calculated through the likelihood function,

\[
w^i_k = w^i_{k-1} p(x_k|x^i_k).
\]

In addition, the setting of the initial parameter set \(\{x^i_0\}_{i=1}^{M}\) 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 \(D\), the tool failure is judged. The time from the predicted starting point \(k\) to the point of failure is the remaining life of the tool \(L_k\). The residual life calculation process is shown in the formula:

\[
L_k = \inf \{ l_k : Y(k + l_k) \geq D, k > 0 \},
\]

where \(\inf \{ \cdot \} \) represents the lower limit of the variable and \(l_k\) is the time required from the current time \(t\) to the 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 of RUL 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 \(a_0, b_0, c_0, d_0\) by curve fitting of the obtained wear value \(y\)

**Particle filter:**

**Particle set initialization:** when \(k = 0\)

for \(i = 1, 2, \ldots, M\) do

**Generating sampling particle:** Generated sample particle \(\{x^i_0\}_{i=1}^{M}\) from \(p(x_0)\)

end

for \(k = 1, 2, \ldots, D\) do

**Importance sampling:**

for \(i = 1, 2, \ldots, M\) do

**Sampling particle generation:** Generated sampling particle \(\{x^i_k\}_{i=1}^{M}\)

from the importance probability density

**Particle weights calculation:** Calculate and normalize \(w^i_k\)

end

**Resampling:** Resampling the particle set \(\{x^i_k, w^i_k\}\) results in \(\{x^i_k, 1/M\}\)

**Output:** Calculate the state estimate \(\hat{x}_k = \sum_{i=1}^{M} x^i_k w^i_k\) at time \(k\)

end

**RUL prediction:** After obtaining the observed system state value \(y_k\) at time \(k\) and the estimated model parameter value \(a_k, b_k, c_k, d_k\), \(L_k\) 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 is collected by multi-sensors. The prognostic wear results will be shown in Sect. 3.3. And the RUL prediction results are in Sect. 3.4.

#### 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 cutters are used to mill 45# steel workpieces. During the experiment, we collect relevant signals through the triaxial accelerometer PCB 356A15, the current sensor CSA201-P030T01, and the free-field microphone Bruel & Kjaer 4966. The detailed information of the sensors is shown in Table 1. The sensor placement is shown in Fig. 5. The triaxial accelerometer is installed on the spindle of the machine tool and measures the vibration.

| Sensor type   | Sensor model     | Sampling frequency | Sampling duration |
|---------------|------------------|--------------------|-------------------|
| Accelerometer | PCB 356A15      | 10 kHz             | > 10 s            |
| Current sensor| CSA201-P030T01   |                    |                   |
| Microphone    | Bruel & Kjaer 4966|                    |                   |

Table 1: Sensor information
of the spindle in the X, Y, and Z directions due to milling. The current sensor is installed on the U phase of the spindle motor. And the free-field microphone is installed on the worktable on the right side of the workpiece to avoid affecting the milling. After a tool cutting is completed, we disassemble the tool and measure the wear width with 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 $\geq 300 \, \mu m$, the tool is judged to have failed.

### 3.2 Dataset establishment and model design

As shown in Table 2, the signals of the first 4 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 also verified by the data of the cutting tools under condition 2. Only 10 s of data under stable cutting condition is collected during each cutting process, and each piece of data is taken as a sample.

Figure 7 shows the current signal, Y-axis vibration signal, and sound signal of C1_1 in its whole life cycle. It can be observed in this figure that almost all the signals have no trendability, especially the current signal.

After features extracting and preliminary selecting, the feature number extracted by each sensor is shown in Table 3.

Some features of different sensing signals in time domain, frequency domain, and time frequency domain are shown in Fig. 8. Through the preliminary selection of correlation analysis, some features that have no trendability at all, such as Fig. 8b, 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.

In order to obtain more useful temporal information to improve the prediction performance [33], 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. Following the references [34, 35], S is set to be 1. By comparing the prediction results of different W = 1, 10, 20, and 30, we find that the larger the W, the higher the prediction accuracy. However, an excessively large W will generate high-dimensional input, requiring more memory and computing time. Considering comprehensively, we set W to 20. In the end, the whole dataset is divided into the training dataset, validation dataset, and test dataset.
As introduced in Sect. 2.2, the proposed time–space attention mechanism model is composed of CNN-based spatial attention mechanism and LSTM-based temporal attention mechanism. The CNN-based spatial attention mechanism consists of a convolutional layer, a maximum pooling layer with a pooling size of 2, a global average pooling layer, and two fully connected layers, Dense1 and Dense2; otherwise, two batch normalization layers follow after the convolutional layer and maximum pooling layer. The stride of the convolutional layer is 5. The LSTM-based temporal attention mechanism includes two LSTM layers, LSTM1 and LSTM2. Besides, the hidden state attention mechanism consists of two fully connected layers (Dense3 and Dense4) and a weight sharing module. Finally, the model gets the final prediction result through a fully connected layer Dense5. The complete structure of the model is shown in Table 4, in which F means the number of filters, K represents the kernel size, and P denotes the parameters in each layer.

The training dataset and validation dataset are fed into the monitoring model proposed above to train model. Multiple features are input to the model, while wear values are used as the output. The mean square error (MSE) between the outputs and the correct results was chosen as the training loss and sets 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 does not decrease after 4 epochs, the learning rate will be reduced; if the loss does not drop after 10 epochs, the training will be finished early.

Figure 10a, b are the features after the spatial dimension information extraction. Compared with the pre-selected

| Cutting condition | Spindle speed | Feed rate (radial)/(radial) | Depth of cut (axial)/(axial) | Cutters |
|-------------------|--------------|-----------------------------|----------------------------|---------|
| Condition 1       | 4600 rpm     | 400 mm/min                  | 1.5 mm                     | C1_1    |
|                   |              |                             |                            | C1_2    |
|                   |              |                             |                            | C1_3    |
|                   |              |                             |                            | C1_4    |
|                   |              |                             |                            | C1_5    |
|                   |              |                             |                            | C1_6    |
| Condition 2       | 3500 rpm     | 300 mm/min                  | 1.75 mm                    | C2_1    |

Fig. 7 Monitoring signals of C1_1 during the whole operating life. (a) Current. (b) Y-axis vibration. (c) Sound
features in Fig. 8, the trendability is enhanced but has severe fluctuation, which still cannot characterize the tool degradation trend. However, it is found in Fig. 10c, d that the features, which are obtained after the time dimension extraction, have more obvious trendability, lower fluctuation, and strong correlation with the tool wear value. And there is a 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

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. We denote our proposed method as time–space attention model, and the comparison methods are denoted as hidden state attention model [36], channel attention model, and without attention model [37], respectively.

The tool wear prediction results of the cutter C1_5 and C1_6 by using four different models are shown in Fig. 11a, b, respectively. 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. Furthermore, 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.

Table 3 Feature number of each sensor

| Channel   | Vibration_x | Vibration_y | Vibration_z | Current | Sound |
|-----------|-------------|-------------|-------------|---------|-------|
| Total number | 13 | 14 | 12 | 17 | 6 |
| Time domain | 1 | 0 | 1 | 3 | 1 |
| Frequency domain | 9 | 11 | 9 | 13 | 4 |
| Time–frequency domain | 3 | 3 | 2 | 1 | 1 |

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, and (f) feature have good trendability.
The channel attention mechanism just makes up for this defect, and it enhances the ability of model to extract spatial dimension information. Therefore, in the late 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 although under the different working condition, the model can still achieve the monitoring of tool wear value well. This proves that the model has good robustness.

For the sake of quantitatively and visually assessing the prediction performance of the proposed model, MAE, RMSE, and correlation factor are introduced to evaluate the prediction accuracy of model. The peak-to-peak value (p-p value) is used to evaluate the stability of the model. The four criteria are as follows:

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |l_i - p_i|
\]  

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (l_i - p_i)^2}
\]  

\[
\text{Correlation factor} = \frac{\sum_{i=1}^{n} (p_i - \bar{p})(l_i - \bar{l})}{\sqrt{\sum_{i=1}^{n} (p_i - \bar{p})^2 \sum_{i=1}^{n} (l_i - \bar{l})^2}}
\]

where \(l\) is the actual wear value, \(p\) is the prediction result, \(e\) is the relative error, and \(\text{norm}_2\) is the 2-norm.

The criteria of the four networks are tabulated in Fig. 13. It can be clearly seen that the tool wear monitoring network based on 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 useless information, enhances the key information, and improves the feature fusion ability of model.

### 3.4 RUL prediction result

Take cutter C1_5 as an example. As shown in Fig. 14a, when the curves of 1000 particles reach the failure threshold D, the tool failure is judged. It can be found that in Fig. 14b, the distribution of the predicted RULs of 1000 particles is approximately normally distributed. Hence, we calculate the mean of the particle RUL distribution at each moment as the predicted value at this moment.

The tool RUL prediction results are shown in Fig. 15. As can be seen from the above results, at the beginning of the prediction

| Table 4 | Time–space attention mechanism model architecture |
|---------|--------------------------------------------------|
| Layer   | Spatial attention | P | Layer | Temporal attention | P |
| Conv    | 64 | 5 | 19,904 | LSTM1 | 32 | 12,416 |
| Pooling1| Maximum pooling   | - | 2080 | Dense3 | 32 | 110 |
| Pooling2| GAP                | - | 2112 | Dense4 | 32 | 110 |
| Dense1  | 32 | - | Weight sharing | LSTM2 | 32 | 8320 |
stage, the prediction has a large deviation from the true RUL. 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 increase, and the predicted results tend to be stable and gradually approach 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. 10 The features fused by time-space attention mechanism fusion model: (a and b) features after spatial dimension information extraction and (c and d) features after time dimension information extraction

Fig. 11 Tool wear prediction results of (a) C1_5 and (b) C1_6
We also evaluated the performance of the four networks. After comparing the significance of different criteria [7], we choose relative accuracy and standard deviation ($\sigma$) of error as the criteria, which are given by:

$$\text{Relative accuracy} = 1 - \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\text{RUL}_i - \text{RUL}'_i}{\text{RUL}_i} \right|,$$

(23)

where $\text{RUL}_i$ and $\text{RUL}'_i$ denote the predictive RUL and ground truth RUL at time $i$, respectively.

As shown in Table 5, the network based on time–space attention mechanism is superior to other networks in both accuracy and stability.

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

(24)

Fig. 12 Tool wear prediction results of C2_1

Fig. 13 Tool wear performance estimation results of four different prognostics networks. (a) C1_5. (b) C1_6
Fig. 14 RUL prediction result. (a) Particle distribution. (b) Particle RUL distribution

Fig. 15 RUL prediction result
Table 5 RUL performance estimation results of four different prognostics networks

|                        | Relative accuracy | Standard deviation |
|------------------------|-------------------|--------------------|
| Time–space attention   | 0.7890            | 0.4432             |
| Hidden attention       | 0.7495            | 0.4917             |
| Channel attention      | 0.6873            | 0.4925             |
| Without attention      | 0.7458            | 0.4964             |

The [bold] means to highlight the best performing method (the method proposed in the article) among the four methods.

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 values. Finally, the RUL 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. The comparative analysis with channel attention model, hidden state attention model [36], and without attention model [37] were respectively conducted. The comparison results show that the proposed time–space attention model has greater advantages both in the accuracy and stability of tool wear monitoring and RUL prediction. At the same time, the model is applied to the tool data under different working conditions, which verifies that the model has good robustness.

Although the proposed method can accurately predict the wear value and RUL of the milling cutter under different working conditions in the experiment, how to enhance the robustness of this method under actual processing conditions such as lack of data and variable working conditions remains to be further studied. In our future study, we would like to further explore more effective multi-feature fusion methods, introduce reliability analysis of tool RUL, and establish a complete tool health monitoring system to tackle a variety of emerging real-world problems of tool RUL prediction.

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).

Data availability Relevant data is not public.

Code availability The code is not public.

Declarations

Ethics approval Not applicable.

Consent to participate Agree.

Consent to publication Agree.

Conflict of interest 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.

References

1. Liu Y, Guo L, Gao H, You Z, Ye Y, Zhang B (2022) Machine vision based condition monitoring and fault diagnosis of machine tools using information from machined surface texture: a review. 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, Ly 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. You Z, Gao H, Guo L, Liu Y, Li J (2020) On-line milling cutter wear monitoring in a wide field-of-view camera. Wear 460–461:203479. https://doi.org/10.1016/j.wear.2020.203479
5. Yang Y, Guo Y, Huang Z, Chen N, Li L, Jiang Y, He N (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
6. 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
7. Saxena A, Celaya J, Saha B, Saha S, Goebel K (2010) Metrics for offline evaluation of prognostic performance. Int J Progn Health Manag 1. https://doi.org/10.36001/ijphm.2010.v1i1.1336
8. Lei Y, Li N, Gontarz S, Lin J, Radkowski S, Dybala J (2016) A model-based method for remaining useful life prediction of machinery. IEEE Trans Reliab 65:1314–1326. https://doi.org/10.1109.jrre.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. https://doi.org/10.1155/2020/8871981

10. Kuntoğlu M, Sağlam H (2021) Investigation of signal behaviors for sensor fusion with tool condition monitoring system in turning. Meas J Int Meas Confed 173. https://doi.org/10.1016/j.measurement.2020.108582

11. Guo L, Yu Y, Gao H, Feng T, Liu Y (2021) Online remaining useful life prediction of milling cutters based on multi-source data and feature learning. IEEE Trans Ind Informatics 3203. https://doi.org/10.1109/TII.2021.3118994

12. Lee MFR, Stanley K, Wu QMJ (2001) Implementation of sensor selection and fusion using fuzzy logic. Annu Conf North Am Fuzzy Inf Process Soc - NAFIPS 1:328–333. https://doi.org/10.1109/nafips.2001.944273

13. Equefer L, Letot C, Serra R, Dehombreux P (2016) Estimate of cutting tool lifespan through cox proportional hazards model. IFAC-PapersOnLine 49:238–243. https://doi.org/10.1016/j.ifacol.2016.11.041

14. 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 328–333. https://doi.org/10.1007/s00170-020-05264-3

15. Hotait H, Chiementin X, Mouchaweh MS, Rasolofondraibe L (2021) Monitoring of ball bearing based on improved real-time OPTICS clustering. J Signal Process Syst 93:221–237. https://doi.org/10.1007/s11265-020-01571-w

16. Srinivasan A, Dornfeld D, Bhinge R (2016) Integrated vibration and acoustic data fusion for chatter and tool condition classification in milling. In: 2016 International Symposium on Flexible Automation (ISFA). IEEE 263–266

17. Wu J, Su Y, Cheng Y, Shao X, Deng C, Liu C (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

18. Mao W, Feng W, Liu Y, Zhang D, Liang X (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

19. Guo L, Gao H, Huang H, He X, Li S (2016) Multifeatures fusion and nonlinear dimension reduction for intelligent bearing condition monitoring. Shock Vib. https://doi.org/10.1155/2016/4632562

20. Yu Y, Guo L, Gao H, Liu Y, Feng T (2021) Pareto-optimal adaptive loss residual shrinkage network for imbalanced fault diagnostics of machines. https://doi.org/10.1109/TII.2021.3094186

21. 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

22. 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

23. 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

24. Zheng Z, Huang S, Weng R, Dai X, Chen J (2020) Improving self-attention networks with sequential relations 28:1707–1716

25. Pal S, Heyns PS, Freyer BH, Theron NJ, Pal SK (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

26. 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

27. Yan GG, Member S, Lin K (2000) Wavelet packet feature extraction for vibration monitoring 47:650–667

28. 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

29. 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

30. Jin X, Sun Y, Que Z, Wang Y, Chow T (2016) Anomaly detection and fault prognosis for bearings. IEEE Trans Instrum Meas 65:2046–2054. https://doi.org/10.1109/TIM.2016.2570398

31. 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

32. 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

33. 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

34. Sateesh Babu G, Zhao P, Li X-L (2016) Deep convolutional neural network based regression approach for estimation of remaining useful life. Springer International Publishing, Cham, pp 214–228

35. Ragab M, Chen Z, Wu M, Kwoh C, Yan R, Li X (2021) Attention-based sequence to sequence model for machine remaining useful life prediction. Neurocomputing 466:58–68. https://doi.org/10.1016/j.neucom.2021.09.022

36. Liu C, Zhang L, Niu J, Yao R, Wu C (2020) Intelligent prognostics of machining tools based on adaptive variational mode decomposition and deep learning method with attention mechanism. Neurocomputing 417:239–254. https://doi.org/10.1016/j.neucom.2020.06.116

37. Li D, Yang L (2021) Remaining useful life prediction of lithium battery based on sequential CNN–LSTM method. J Electrochem Energy Convers Storage 18. https://doi.org/10.1115/1.4050886

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.