



# Deep learning-based tool wear prediction and its application for machining process using multi-scale feature fusion and channel attention mechanism

Xingwei Xu, Jianwen Wang, Bingfu Zhong, Weiwei Ming\*, Ming Chen

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

## ARTICLE INFO

### Keywords:

Tool wear prediction  
Deep learning  
Multi-sensor feature fusion  
Channel attention mechanism

## ABSTRACT

Tool wear is a key factor in the cutting process, which directly affects the machining precision and part quality. Accurate tool wear prediction can make proper tool change at an early stage to reduce downtime and enhance product quality. However, traditional methods can not meet the high requirements of the intelligent manufacturing. Therefore, a novel method based on deep learning is proposed to improve the prediction accuracy of tool wear. The multi-scale feature fusion was implemented by the developed parallel convolutional neural networks. The channel attention mechanism combined with the residual connection was developed to consider the weight of the different feature map to enhance the performance of the model. The different tool wear prediction experiments were implemented to verify the superiority of the developed method, and the prediction results of tool wear are more robust and accurate than current methods. Finally, a tool wear monitoring system was developed and applied to the tapping process of the engine cylinder to ensure the quality of the engine cylinder and the stability of the machining process.

## 1. Introduction

Machining technology performs a key role in modern manufacturing industries as the growing demand for equipment and parts for the aerospace, automotive, and precision machinery, etc. A major problem during the machining process is the severe tool wear and the caused shortened tool life, which eventually leads to poor machining part quality and low productivity [1,2]. According to the statistics, the scrapped cutting tools were worth up to \$200 million only in the U.S during April 2018 [3]. Furthermore, about 20% downtime of a machine tool is caused by tool failures. Specifically, if the tool can not timely change when the tool wear exceeds the failure criterion, it will affect the surface quality of the machined workpieces [4,5]. However, if the tool is replaced too early, it will cause waste and decrease productivity [6]. Therefore, it is of considerable significance to predict tool wear accurately.

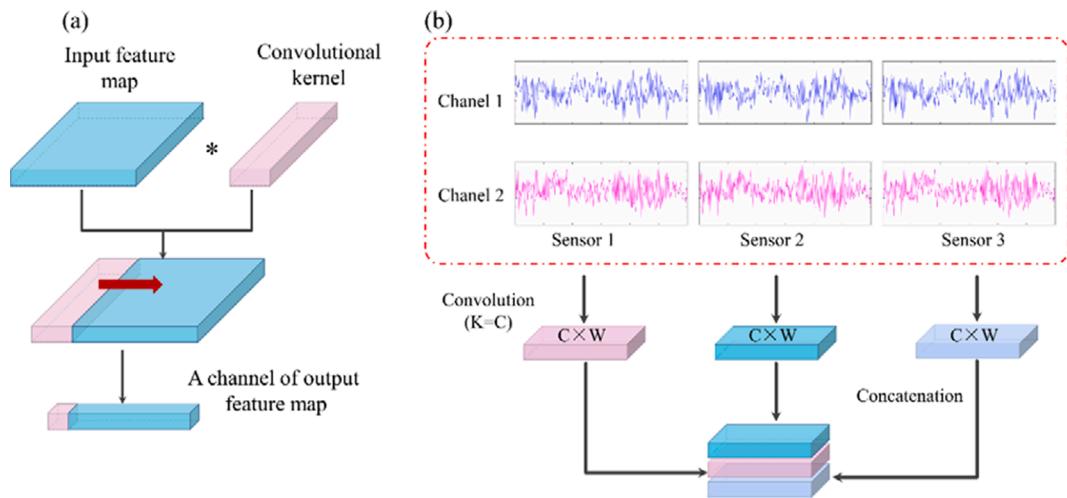
Current research on tool wear monitoring and prediction is often associated with data-driven techniques [7,8]. There are three stages for data-driven methods: signal acquisition, features extraction, and model construction and its application [3,9]. Signal acquisition is to collect the original signal from the CNC machine by sensors. Many researchers have

attempted to use multiple sensors such as the cutting force sensor, vibration sensor to collect data for tool wear prediction [10]. The cutting force sensor, vibration sensor, etc., are generally utilized to collect data. [11,12]. Wang and Azmi et al. [13,14] utilized force signals to illustrate the tool wear status. Vibration signal has also been widely applied on account of its easy installation, and the amplitude of the vibration signal changes as the tool wear change [10,15]. Feature extraction is about extracting the representative features from original signals for classification or regression [16]. In tool wear monitoring and prediction, the time domains (e.g., mean, root mean square, standard deviation, skew, etc.) [17–19] and frequency domain (e.g., harmonic ratio, spectral amplitudes, spectral density, etc.) are the main features for tool wear. After that, the extracted features are considered to be the input for the classifier, and the tool wear is then generated by the trained classifier.

At present, the model for tool wear prediction mainly focuses on machine learning methods, including artificial neural network (ANN) [20,21], support vector machine (SVM) [11,22–24], relevance vector machine (RVM) [25], hidden Markov model (HMM) [26–28] and so forth. In addition, the Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) [29,30] are the most popular deep learning model for tool wear prediction. Specifically, in [11], the SVM and

\* Corresponding author at: School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.

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



**Fig. 1.** The architecture of multi-sensor feature fusion: (a) convolution operation, (b) multi-sensor feature extraction and fusion.

multisensory data fusion methods were utilized to predict the tool wear for improving the reliability of the manufacturing system. In [17], CNN was employed to predict tool wear and verified its effectiveness on different datasets. Kong et al. [25] utilized an algorithm based on the KPCA\_IRBF for dimension reduction, and then RVM was utilized for the remaining useful life prediction of the cutting tool. Li et al. [28] proposed an improved hidden Markov model to describe the tool wear process. Cao et al. [18] first employed the translation-invariant wavelet and CNN for tool wear state estimation.

Although considerable success in tool wear monitoring and prediction has been achieved, there are some flaws in the traditional tool wear prediction. Firstly, machine learning methods above have to make feature engineering first and then utilize the machine learning methods to monitor and predict tool wear. So it highly relies on whether the extracted features are highly related to the tool wear conditions [18]. In addition, feature engineering is labor-intensive and time-consuming [31,32]. Secondly, the current deep learning methods for tool wear monitoring do not consider the weight of the different feature map, and the prediction accuracy is limited. Therefore, It is natural to think that the network can consider the importance of different feature map extracting from different signals to enhance the prediction accuracy of the model. Besides, due to the importance of feature engineering and multi-sensor feature fusion, it is necessary to focus on how to implement feature fusion from multi-sensors for tool wear prediction.

To address the above weaknesses, we proposed a deep learning-based method by using the convolutional network and channel attention mechanism. In the proposed method, a parallel CNN layer is firstly developed to extract and fuse features from multi-sensors information to realize multisensory feature fusion. Besides, inspired by the development of the attention mechanism of deep learning [33,34], the channel attention mechanism is developed for tool wear prediction. Specifically, we consider the weight of each channel of the feature map by developing the channel attention mechanism. Lastly, tapping is a general cutting operation in the advanced manufacturing industry. Tap wear has a significant effect on threaded hole quality, especially surface roughness and thread strength [35]. Barooah et al. [36] investigated the tap wear mechanisms and quantified wear by developing new measurement metrics. Monka et al. [37] studied a tap failure during the internal threads making. However, to the best of our knowledge, there is little research on tap wear prediction with deep learning methods [38]. Thus, the tap wear prediction is regarded as one example to validate the proposed method. In conclusion, The major highlights of our work are summarized below:

- (1) The developed approach presents an innovative method based on CNN of deep learning to achieve multi-sensor feature fusion. The proposed multi-scale CNN layer can extract and fuse different features from multi-sensors to implement feature fusion. The relationship between the extracted features from multi-sensors and tool wear level is found.
- (2) Channel attention mechanisms combining with the residual connections are developed to achieve tool wear prediction. The former considers the weight of the different feature map through learning such that the proposed method can promote the features that are useful and suppress the features that are not useful for tool wear prediction. The latter make the proposed model have easier gradient transmission and relatively fast training speed. Therefore, the proposed method can obtain higher prediction accuracy than the current mainstream methods.
- (3) The developed method is able to utilize for various tool wear prediction. To illustrate the generalization and stability of our proposed method, two typical tool wear prediction items are carried out, and the superiority of the developed model is verified. Finally, a general tool wear monitoring system was developed for implementation, which can be a good demonstration for tool wear prediction.

The rest of the paper is organized as follows: the basic structure of CNN and the proposed structure are introduced and described in Section 2. In Section 3, the performance of the developed model was verified by two experiments, and comparisons with some related works are illustrated. In addition, the application with the proposed method is demonstrated. The concluding remarks are presented in Section 4.

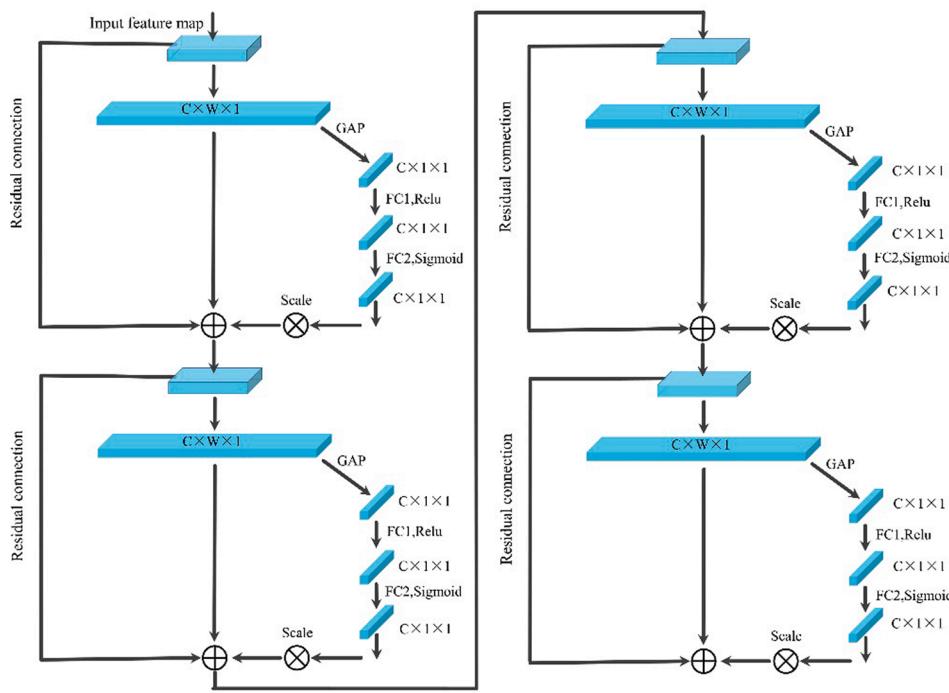
## 2. Theoretical method

### 2.1. Convolutional neural network

The developed method has some basic components of the traditional CNN, which consist of the convolutional layer, pooling layer. Specifically, the convolution operation can be defined as:

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

where  $X_j^l$  represents the  $j$ -th feature map of the layer  $l$ .  $\omega_{kj}^l$  represent the convolutional operation.  $x_k^{l-1}$  represent the  $k$ -th output of the feature map.  $M_j$  represents the size of the input.  $b_j^l$  represents the corresponding bias.



**Fig. 2.** A number of channel attention mechanism with residual connection.

The pooling layer is a sub-sampling process, which can reduce the size of the feature. The general pooling operations include max pooling and average pooling. The max-pooling, in this paper, is utilized to acquire the maximum local features, which is defined as:

$$P_j = \max_{x_j^i \in S} X_j^i \quad (2)$$

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

## 2.2. The proposed method

### 2.2.1. Multi-sensor information feature fusion

At present, multiple signals are often utilized for tool wear monitoring and prediction, such as cutting force signals, vibration signals, acoustic emission signals, and so forth. The signal from the multi-sensor has more meaningful information than a single signal, and multi-sensor feature fusion can enhance the performance of the model. Thus, we design a parallel structure of CNN to extract and fuse features from multi-sensor information, and the specific architecture is shown in Fig. 1. It is noted that all utilized CNN part are the one-dimensional CNN. Fig. 1(a) shows the convolution operation, which can be described that the convolutional kernel slides on the input so that the output feature map can be obtained. In each convolutional part, we often utilized several convolutional kernels to obtain different feature maps. Fig. 1(b) shows that there are different inputs from multiple sensors with multi-channel. The developed parallel convolution operation can extract features from different signals collecting by multiple sensors, respectively. And then, the multi-sensor feature fusion is implemented by concatenating the extracted feature from CNN.

### 2.2.2. Channel attention mechanism with residual connection

In the tool wear prediction, the collected data from multiple sensors are composed of different channels. Therefore, It is natural to think that the network can consider the weight of the different feature map extracting by different signals to enhance the prediction accuracy of the model. Our proposed method can automatically acquire the weight of the different feature maps through learning. And then, the proposed method promotes the features that are useful and suppress the features

that are not useful according to the obtained weight of each channel. The method can be achieved by the following steps:

Step1: After the convolution operation in Section 2.2.1, we assume that the output feature map is  $x = [x_1, x_2, x_3, \dots, x_C]$ , the size of the  $x$  is  $W \times H$ , and the number of channels is  $C$ . Each channel of the  $x$  was squeeze by global average pooling (GAP) so that the squeezed feature have global information, and the GAP operation can be expressed as:

$$z_C = F_{sq}(X_c) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_c(i,j) \quad (3)$$

where  $z_C$  represent the squeezed value of the  $c$ -th channel of the input feature map.  $x_c(j,k)$  represent the value of the  $c$ -th channel in the position  $j$  and  $k$ .  $F_{sq}$  represent the squeeze operation.  $W$  and  $H$  represent the size of the feature map.

Step2: After the squeeze operation, the shape of the output feature map becomes  $C \times 1 \times 1$ . Then, the following two fully connected layers (FCL) are adopted to learn the weight of each channel. The first layer (FCL layer) can reduce the dimension of the squeezed feature map, and can be mathematically expressed as:

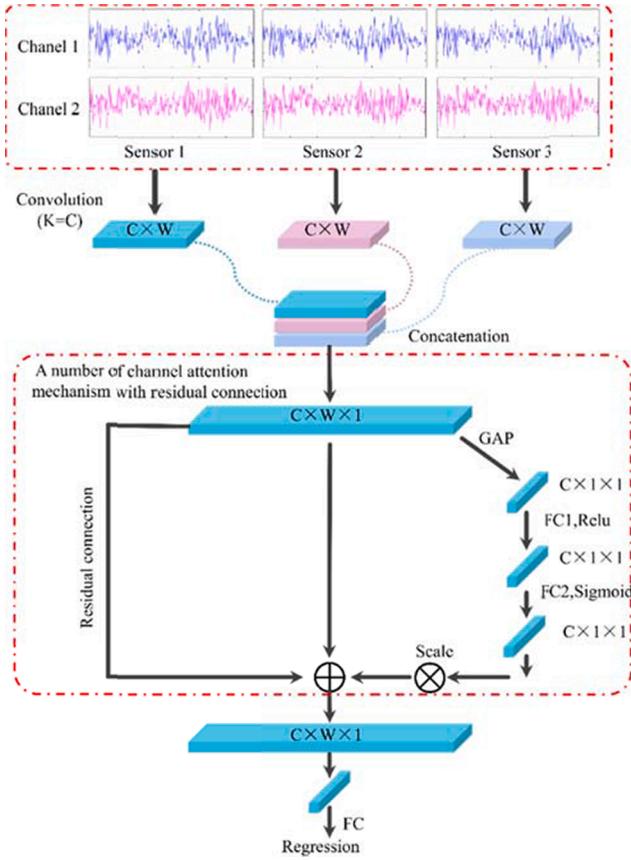
$$\begin{cases} s_{c1} = A_{FC1}(W_1 \times z + b_1) \\ A_{FC1}(x) = \max(x, 0) \end{cases} \quad (4)$$

where  $z$  represents the value of the feature map after squeeze operation,  $W_1$  and  $b_1$  represent the weight and bias of the layer  $FC1$ , respectively.  $A_{FC1}(\cdot)$  is the  $ReLU$  activation function. The second layer ( $FC2$  layer) can learn the weight of each channel of  $s_{c1}$  and map its value to 0–1 through the sigmoid function, which can be mathematically expressed as:

$$\begin{cases} s = A_{FC2}(W_2 \times s_{c1} + b_2) \\ A_{FC2}(x) = \frac{1}{1 + e^{-x}} \end{cases} \quad (5)$$

where  $W_2$  and  $b_2$  represent the weight and bias of the  $FC2$  layer, respectively.  $A_{FC2}(\cdot)$  is sigmoid activation function.

Step 3: Add the weight obtained by step 2 it to the original feature map  $x$ , which can be mathematically expressed as:



**Fig. 3.** The architecture of the proposed method.

$$x_c = F_{scale}(x, s) = [s_1 \times x_1, s_2 \times x_2, \dots, s_i \times x_i, \dots, s_c \times x_c] \quad (6)$$

where  $F_{scale}(x, s)$  refers to channel-wise multiplication between the scalar  $s$  and the feature map  $x$

The above steps can achieve data recalibration. Besides, the deep Residual Networks (DRN) has more advantages than the traditional CNN [39]. Thus, a simple residual connection of the DRN is adopted in our proposed architecture, which is written as:

$$H(x) = f(x) + x \quad (7)$$

where  $x$  is the input,  $f(x)$  is the mapping relation between input and output,  $H(x)$  is the output of the residual connection. So the output of the channel attention mechanism combined with the residual connection can be expressed as

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

The architecture of the channel attention mechanism combined with the residual connection is presented in Fig. 2.

#### 2.2.3. Fully connected layer and tricks of the proposed method

The fully connected layer (FCL) is designed to generate output  $y_i^{(l)}$ , and can be written as:

$$\begin{cases} z^l = W^l x^{l-1} + b^l \\ y^{(l)} = \sigma(z^l) \end{cases} \quad (9)$$

where  $W^l$  is the weight matrix,  $x^{l-1}$  refers to the output vector of the former layer,  $\sigma$  is the activation function.

Besides, the dropout layer [40] and batch normalization (BN) layer [41] were utilized to increase the training speed and alleviate the overfitting issues. The dropout is defined as:

$$\begin{cases} r_i^l \sim Bernoulli(p) \\ \bar{K}_i^l = \frac{r_i^l * K_i^l}{p} z_i^l(j) = \bar{K}_i^l * x_j^l \end{cases} \quad (10)$$

where  $p$  represents the dropout rate,  $r_i^l$  follows Bernoulli distribution, and  $z_i^l(j)$  represents the output of  $j$ th input after convolution. The BN is expressed as:

$$\begin{cases} X_{n \times m} = \{x_1, x_2, \dots, x_m\} \\ Y = \{y_i = BatchNorm_\gamma, \beta(x_i)\}, i = 1, 2, \dots, m \\ \mu_\beta = \frac{1}{M} \sum_{i=1}^M x_i \\ \sigma_\beta^2 = \frac{1}{M} \sum_{i=1}^M (x_i - \mu_\beta)^2 \\ \hat{x}_i = \frac{x_i - \mu_\beta}{\sqrt{\sigma_\beta^2 + \epsilon}} \\ y_i = \gamma \hat{x}_i + \beta \equiv BatchNorm(x_i) \end{cases} \quad (11)$$

where the  $m$  and  $n$  represent the size of the input matrix  $X$ .  $\mu_\beta$  and  $\sigma_\beta$  refer to the mean and variance.  $\beta$  and  $\gamma$  are the learnable parameters.  $\epsilon$  is a default value, which is a minimum value.

#### 2.3. Architecture of the proposed method

According to the above discussion, the proposed architecture is developed and presented in Fig. 3. First, the parallel structures of CNN are developed to extract features from different sensors, respectively. And then, the extracted features from various sensors are concatenated for multi-sensor feature fusion. Next, a number of channel attention mechanism with the residual connection are designed to consider the weight of the different feature map to improve the prediction accuracy. Furthermore, the residual connection is designed in each of the channel attention mechanism parts. Finally, the FCL is developed to generate the results of tool wear.

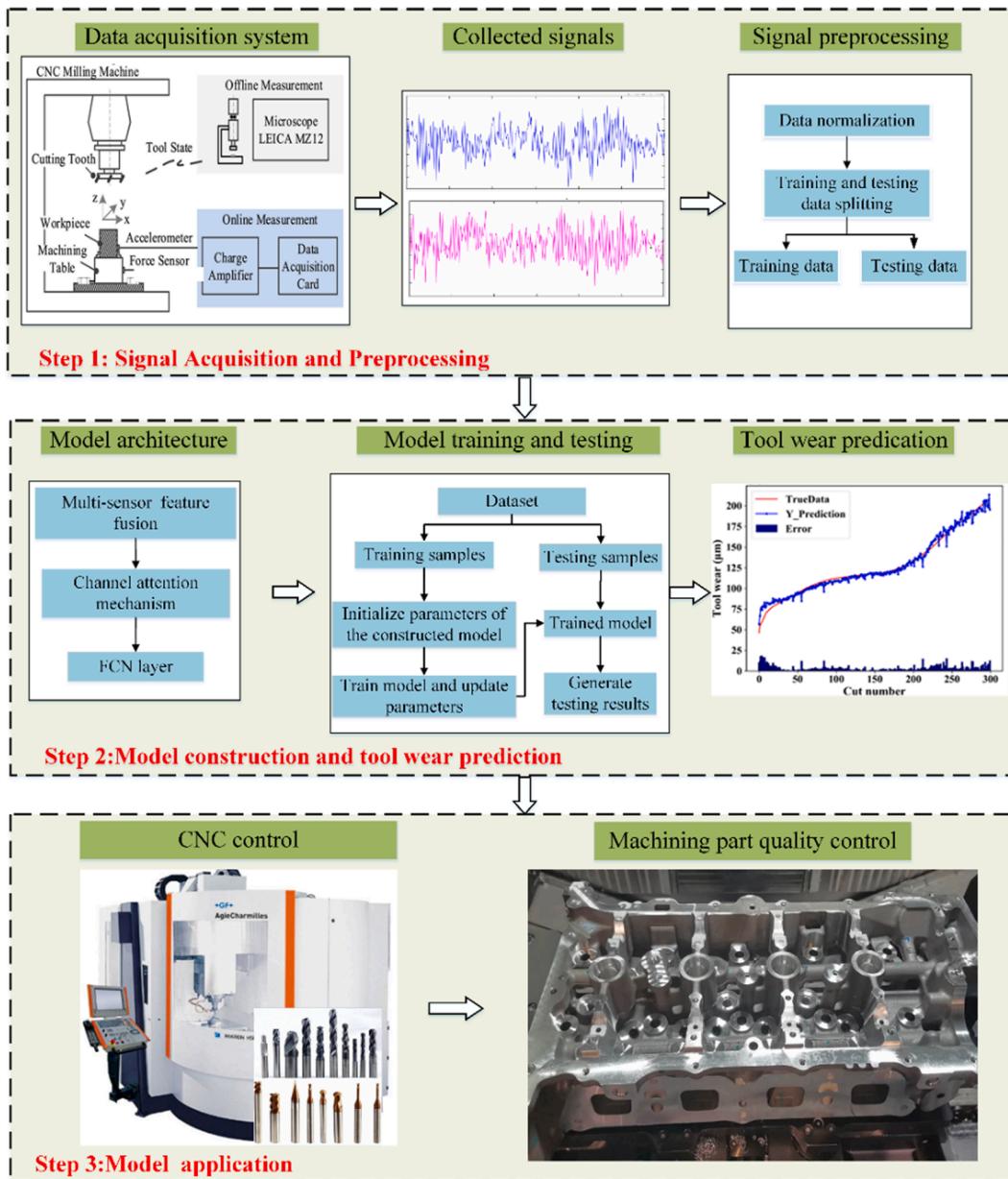
#### 2.4. Loss function and metrics

The mean squared error is selected as the loss function for this prediction problem, and it is defined as:

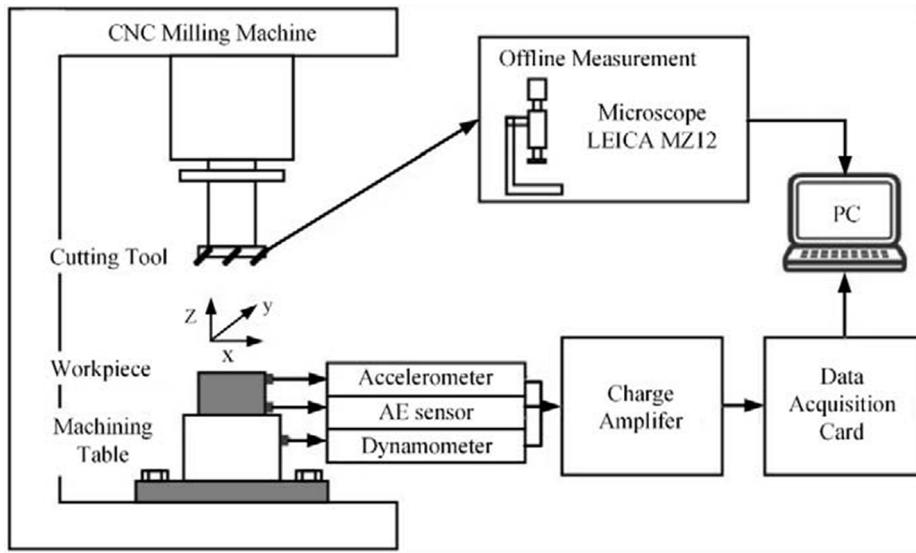
$$y = \frac{\sum_{i=1}^n (y_i - y'_i)^2}{n} \quad (12)$$

where  $y_i$  represent the actual data while  $y'_i$  represent the predicted data, and  $n$  represent the number of samples.

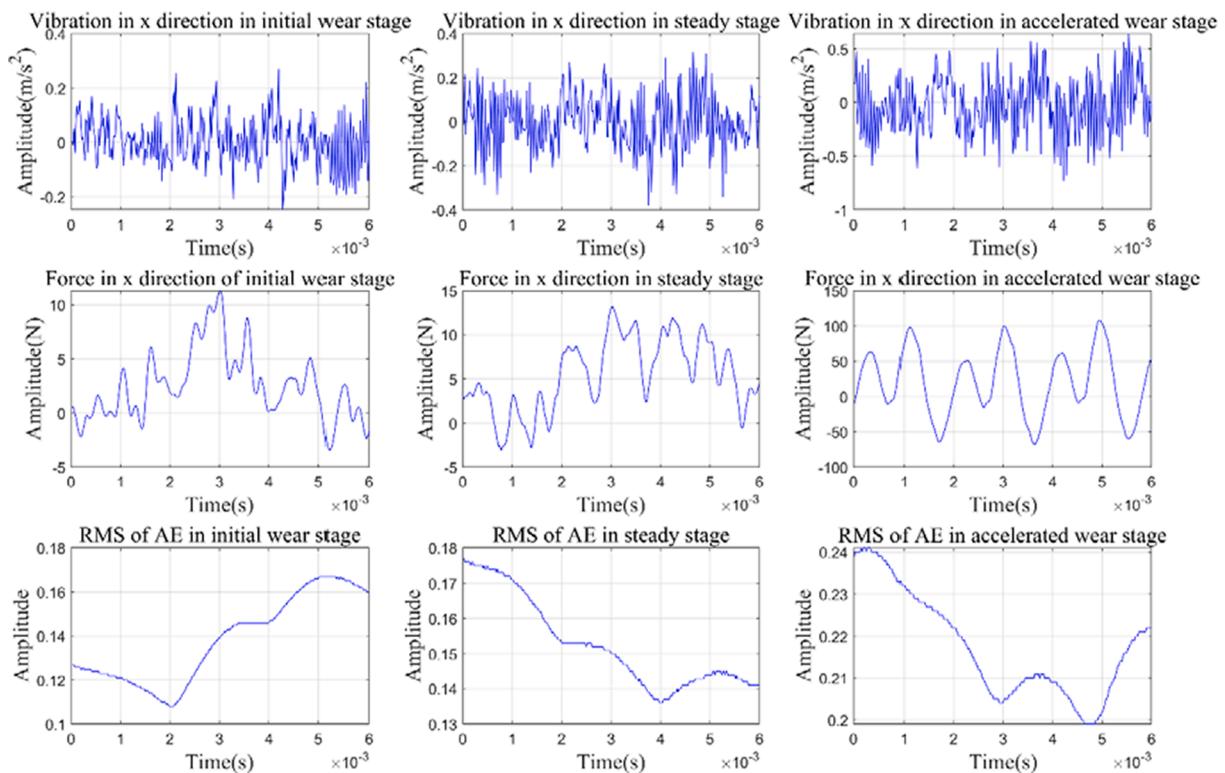
Besides, the adaptive moment estimation (Adam) [42] was utilized as the optimization algorithm of the developed model, and the learning rate is set to be 0.001 [43]. The root mean squared error (RMSE), mean



**Fig. 4.** The procedure for the proposed method for tool wear prediction.



**Fig. 5.** Schematic diagram of experimental setup for 2010 PHM Dataset.



**Fig. 6.** The examples of the different signals in different tool wear stage.

**Table 1**

The division of training and testing dataset.

Number	Training dataset	Testing dataset
1	C1, C4	C6
2	C4, C6	C1
3	C6, C1	C4

absolute error (MAE), and the coefficient of determination ( $R^2$ ) are the metrics for tool wear prediction, which are defined as:

$$\left\{ \begin{array}{l} MAE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{\bar{y}_i} \right| \\ RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n \left( \frac{y_i - \hat{y}_i}{\bar{y}_i} \right)^2} \\ R^2 = 1 - \frac{\sum_{i=1}^n \left( \frac{y_i - \hat{y}_i}{\bar{y}_i} \right)}{\sum_{i=1}^n \left( \frac{y_i - \hat{y}_i}{\bar{y}_i} \right)^2} \end{array} \right. \quad (13)$$

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

## 2.5. Tool wear prediction with the proposed method

The procedure for tool wear monitoring and forecasting with the proposed method is presented in Fig. 4. It contains data acquisition and preprocessing, model construction and training, and model application. First, different signals were collected from the CNC machine by multiple sensors and then normalized the collected data into the range  $[-1, 1]$ . Then the dataset is split into training and testing data. It should be noted that the normalized data are used directly as the model input. Next, the proposed model for tool wear monitoring and prediction starts to be built. And then, train and test the proposed model by the training and testing data. Once the proposed model is trained and satisfied the requirement of metrics, the predicted tool wear is obtained and put the results into practical applications.

## 3. Model verification and application

### 3.1. Verification of the proposed model with the benchmark data

#### 3.1.1. Dataset introduction.

To evaluate the performance of the developed model objectively, the public dataset from the “prognostic data challenge 2010” was selected [44]. Fig. 5 shows the schematic diagram of the experimental platform. Three ball-nose cutters of tungsten carbide named C1, C4, and C6 were used to cutting the workpiece. The cutting parameters were as follows: the depth of cut was set to be 0.125 mm; the spindle speed was set to be 10,400 r/min; the feed rate was set to be 1555 mm/min.

This dataset consists of seven-channel data, including the cutting force signal, the vibration signal, and AE-RMS, and the sampling rate is 50 kHz. The workpiece material and workpiece length are stainless steel (HRC52) and 108 mm, respectively. After finishing each cutting stroke,

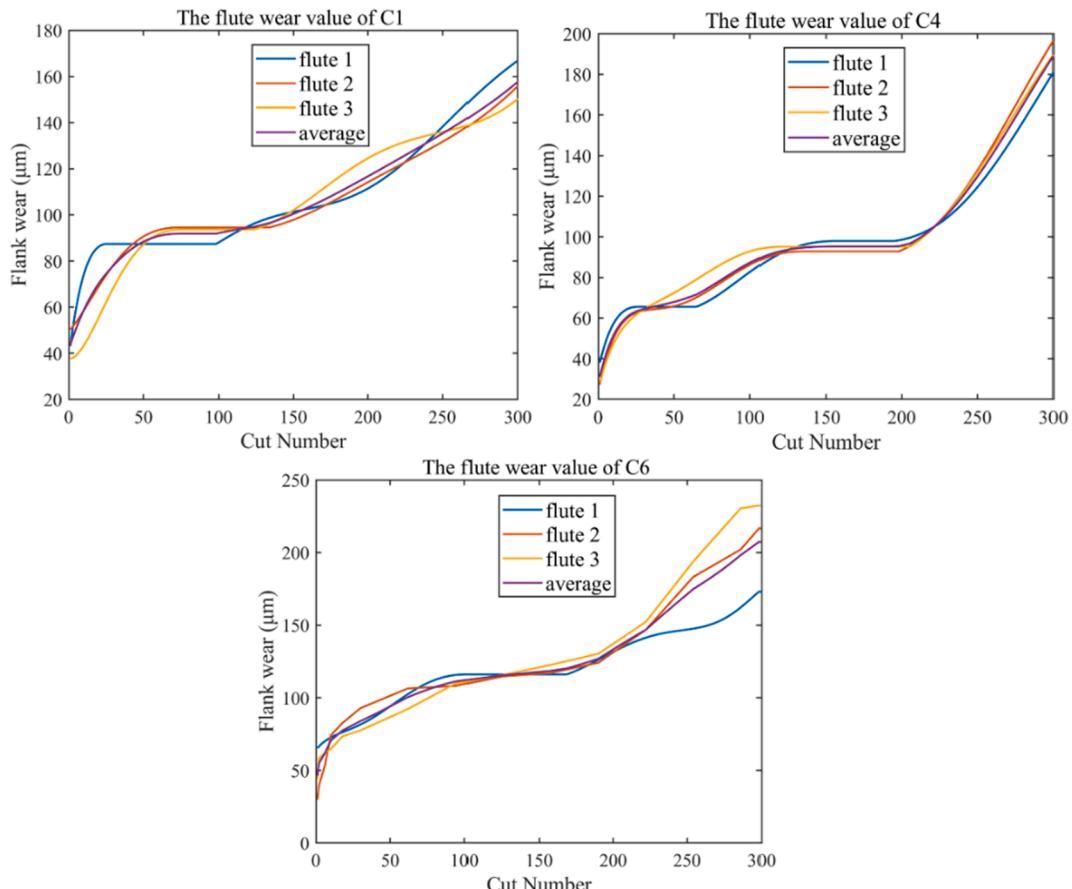
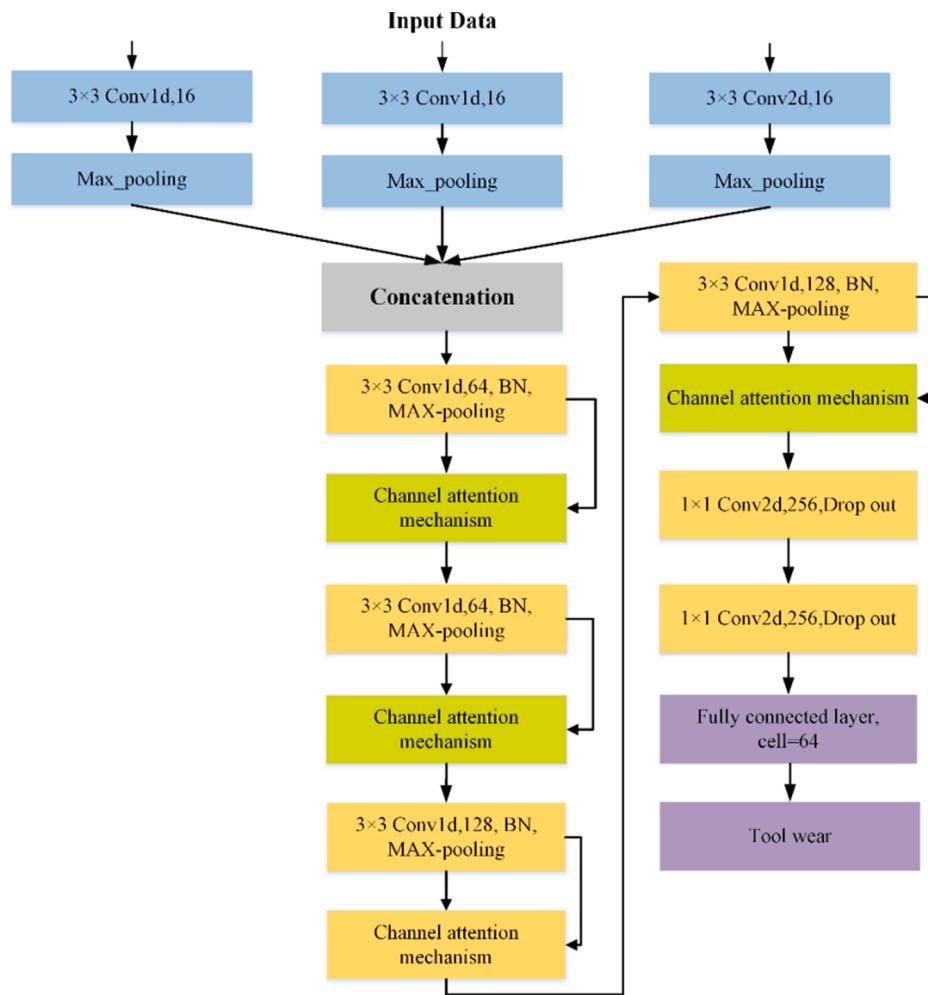


Fig. 7. . The measured tool wear of the three cutters.



**Fig. 8.** The specific parameters of the developed method.

the VB of the cutters were observed by utilizing the LEICA MZ12 microscope. Each cutter consists of 315 cutting stroke, and the collected signal from each cutting stroke are considered to be a sample. In this paper, 900 samples were randomly selected from three cutters for testing our proposed model. Fig. 6. shows the examples of seven-channel signals in the tool wear process.

### 3.1.2. Parameter configuration

The cross-validation method is adopted to verify the generalization of the proposed method. For example, if the training data is C1 and C4, the C6 will be regarded as the testing data. The details about the division of the training/testing dataset are illustrated in Table 1. The input of the model is the seven-channel signal data, and the corresponding labels are the average VB, which are presented in Fig. 7. The parameters of the developed model are presented in Fig. 8. The major operation in Fig. 8 contain 1D CNN, Max pooling, BN, dropout, channel mechanism with residual connection, and the FCL layer, which are explained in Section 2. More specifically, the parallel CNN is utilized to extract features from different signals, respectively, The kernel size of the CNN is  $3 \times 3$ . Then, the residual connection and channel attention mechanism are utilized after each CNN layer. Lastly, the fully connected layer is to generate the predicted tool wear results.

### 3.1.3. Results discussion

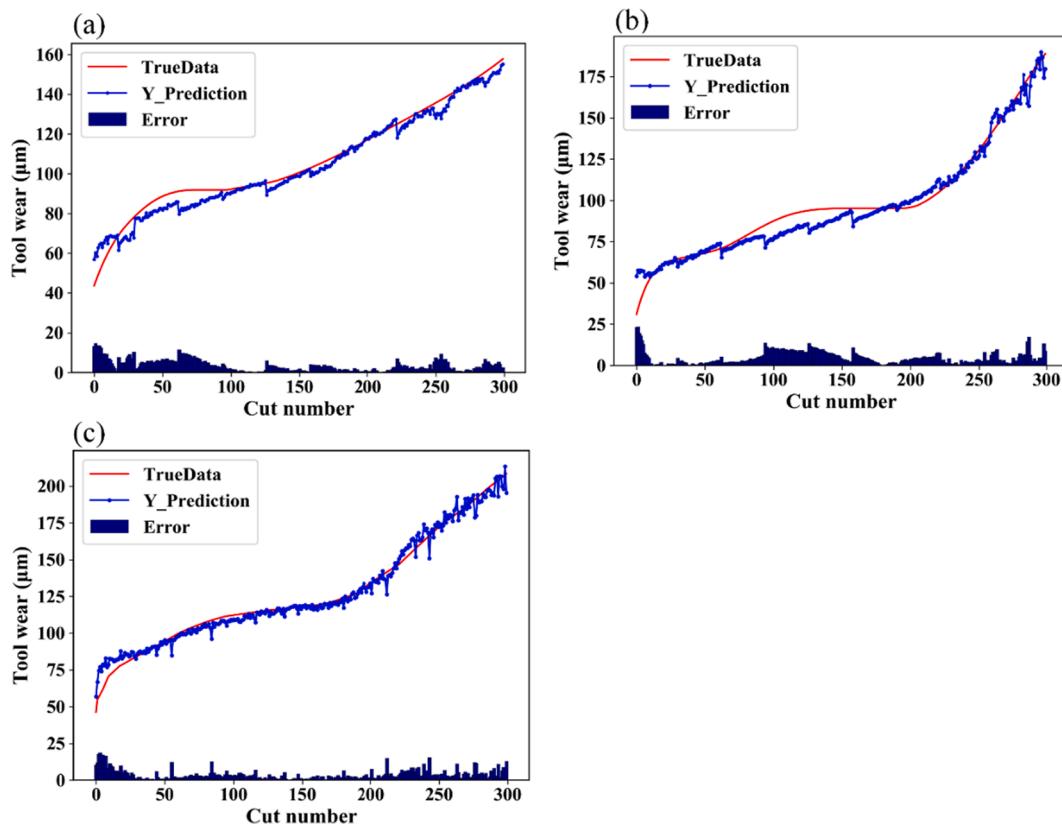
The predicted tool wear results of the developed model in the various testing dataset are illustrated in Fig. 9. The true data represent the actual tool wear measured offline by the microscope, and the predicted tool

wear are obtained by the proposed method. From the compared results, it is clearly observed that the predicted tool wear for three cutters are highly closed to the measured tool wear.

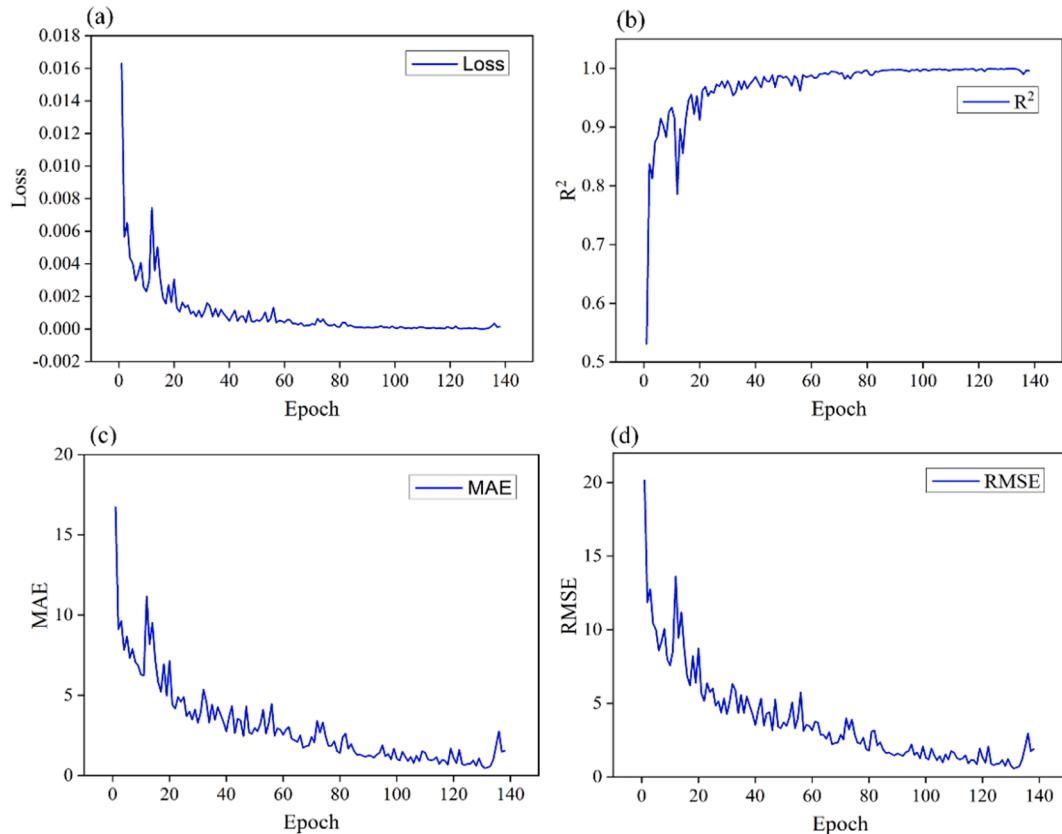
To better demonstrate the effectiveness of the proposed method, evaluation criteria including MAE, RMSE,  $R^2$ , and loss in the training process were computed and plotted in Fig. 10. In Fig. 10(a), the training loss converged at zero at very tiny epochs. Fig. 10 (b) shows the accuracy was close to 1. Fig. 10(c) and (d) show that the MAE and the RMSE converged. Thus, all the evaluation criteria can converge easily and early. Therefore, the developed model is very effective for tool wear prediction.

### 3.1.4. Related works comparison and analysis

The dataset utilized in this part is very famous in tool wear monitoring and prediction, and many state-of-the-art results have been reported in recent years. To demonstrate the performance of the proposed method more objectively, the latest related works are summarized and compared. Table 2 and Fig. 11 show the MAE and RSME results of the state-of-the-art models under three different testing cutters. From Fig. 11, it is seen that RMSE and MAE for three different testing cases with the proposed method are the smallest one among all the compared methods. Besides, the TDConvLSTM developed in [45] is the second-best because the authors consider both the temporal dependency and spatial correlation of the input signals. In addition, CNN [45] and RNN [46] are also competitive. To quantify, the RMSE of the proposed method for C1, C4, and C6 are 5.54, 5.27, and 4.90, respectively. The MAE values for C1, C4, and C6 are 3.89, 4.53, and 3.52, respectively.



**Fig. 9.** The predicted tool wear results of the various testing cutters: (a) predicted tool wear of C1, (b) predicted tool wear of C4, (c) predicted tool wear of C6.



**Fig. 10.** The iteration curves during the training process of C6: (a) Loss, (b)  $R^2$ , (c) MAE, (d) RMSE.

**Table 2**  
Performance of different algorithms for three cutters.

Model Feature fusion		MAE			RMSE		
		C1	C4	C6	C1	C4	C6
MLP [46]	No	24.5	18.0	24.8	31.2	20.0	31.4
SVR [46]	No	15.6	17.0	24.9	18.5	19.6	31.5
CNN [45]	No	15.32	14.34	17.36	18.59	18.80	21.85
RNN [46]	No	13.1	16.7	25.5	15.6	19.7	32.9
CNN + LSTM [45]	No	11.18	9.39	11.34	13.77	11.85	14.33
TDCovLSTM [45]	No	6.99	6.96	7.50	8.33	8.39	10.22
Developed model	Yes	<b>3.89</b>	<b>4.53</b>	<b>3.52</b>	<b>5.54</b>	<b>5.27</b>	<b>4.90</b>

In the proposed method, the channel attention mechanism with the residual connection is utilized to learn the importance of the extracted feature from different signal sequences and further enhance the prediction accuracy of tool wear. Fig. 12 illustrates the comparisons of tool wear prediction results of different methods. It can be seen that the tool wear prediction results using the proposed method with the channel attention mechanism are more closed and stable to the actual tool wear. In addition, the metrics of different methods, both with and without channel attention mechanism are illustrated in Fig. 13. It is seen that the current mainstream methods with channel mechanism achieved a better prediction accuracy. In conclusion, the comprehensive comparison results presented above illustrate that the proposed architecture is more promising than others in tool wear prediction and able to predict reliable and accurate tool wear in different cases.

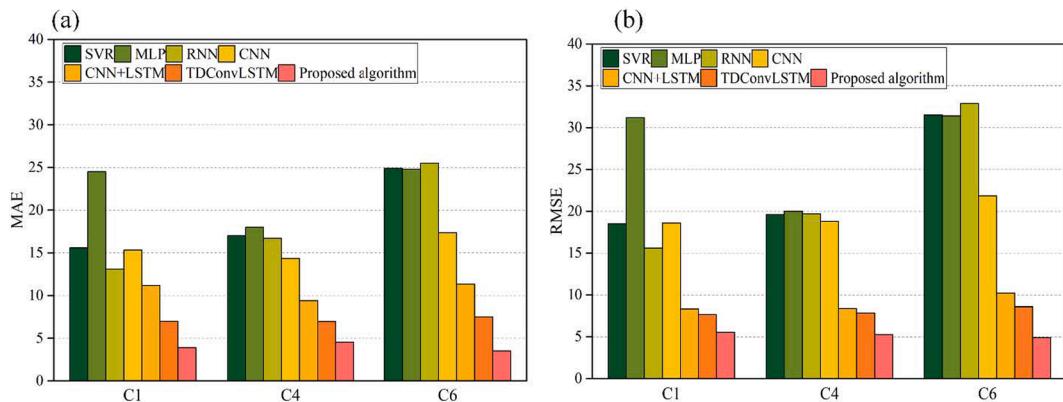


Fig. 11. The performance of the proposed model for three cutters: (a) MAE, (b) RMSE.

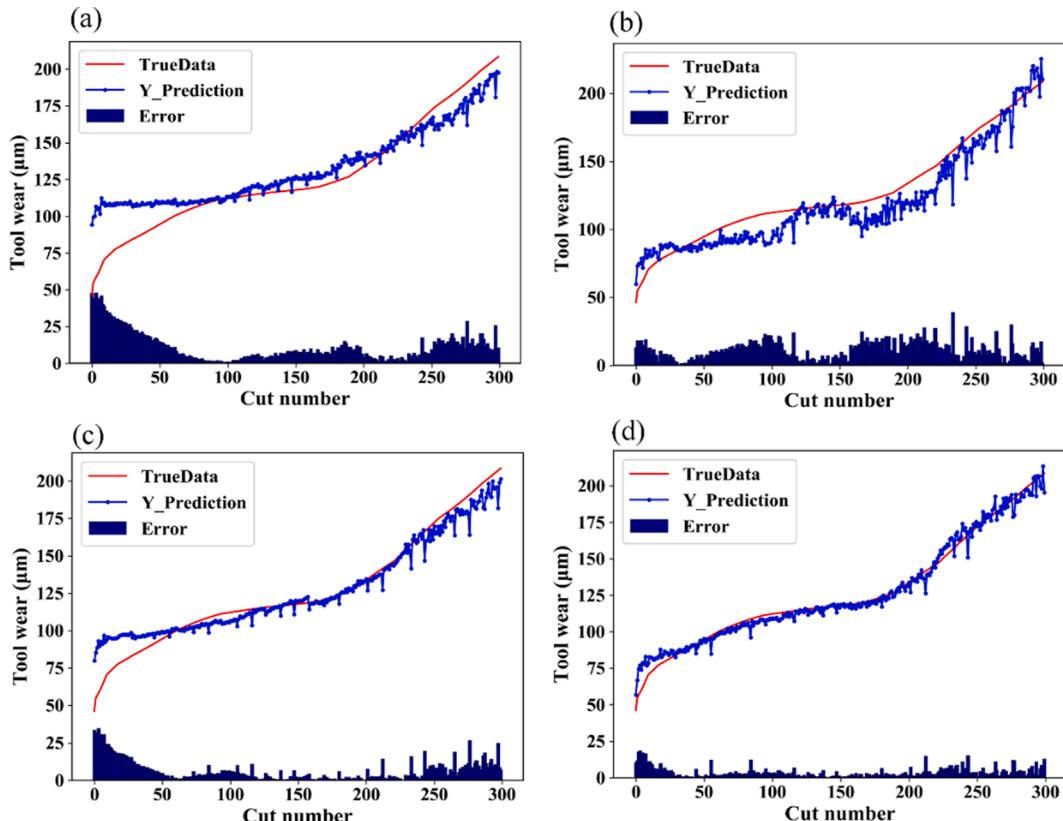


Fig. 12. The prediction results with different models of C6: (a) CNN, (b) LSTM, (c) proposed model without channel attention mechanism, (d) proposed model with channel attention mechanism.

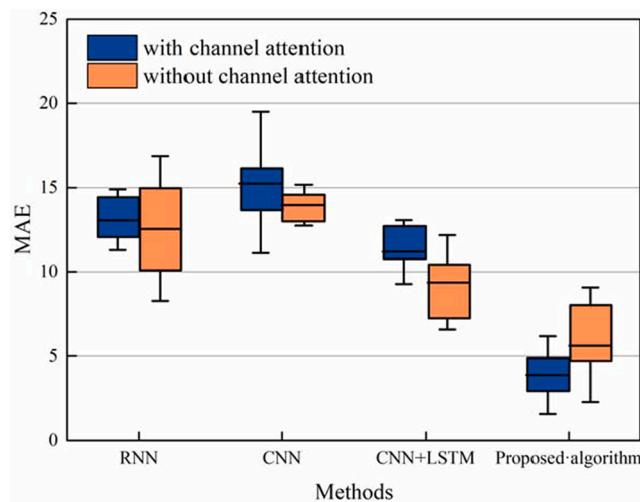


Fig. 13. Effect analysis of the channel attention mechanism.

### 3.2. Verification of the proposed model with experimental data

#### 3.2.1. Experimental platform and data acquisition

The machining experiment was carried out on a HURCO VMX42 five-axis vertical machining center, and the experimental platform is presented in Fig. 14. In this experiment, three taping cutters named T1, T2, T3 are used for threading. The basic parameters of the tap are illustrated in Fig. 15. The workpiece material of tapping was AlSi7Mg. The tapping

parameters were as follows: The depth of taping was set to 18 mm; The spindle speed was set 1800 (rpm); The feed rate was set to 1 mm/r.

An accelerometer is mounted on the machine spindle to collect the vibration signal with three directions (x, y, z) during the tapping process at a sampling rate of 10 kHz. After five holes were tapped, the tap tool was removed from the CNC, and the VB was measured by a high-precision Keyence VHX-500 3D microscope system. The above-mentioned steps were then repeated until each tap failed. Finally, the three tools T1, T2, and T3 consist of 184, 200, 199 samples, respectively.

In order to reflect the real machining process precisely, the raw data collected from the vibration sensor are utilized to test our approach. The vibration data in the time domain and frequency domain in different wear stages of T3 are shown in Fig. 16. It is seen that the amplitude changes with the tool wear changes. Fig. 17 shows the tool wear evolution of T3 during the tapping process.

#### 3.2.2. Results discussion and model comparison

The cross-validation method is also utilized in this experiment to verify the generalization of the developed model. The details about training and testing data are shown in Table 3. The input of the developed model for this experiment is the vibration signal. The detailed parameters of the model are also as same as the first experiment presented in Fig. 8.

The developed method is retrained and retested by the collected data, and the prediction results of three cutters were obtained. Fig. 18 shows the predicted results and true tool wear of three cutters. The actual tool wear in Fig. 18 were obtained by a microscope, and the predicted tool wear was obtained by the proposed method. It is found that the predicted wear among the three cutters are all able to follow the

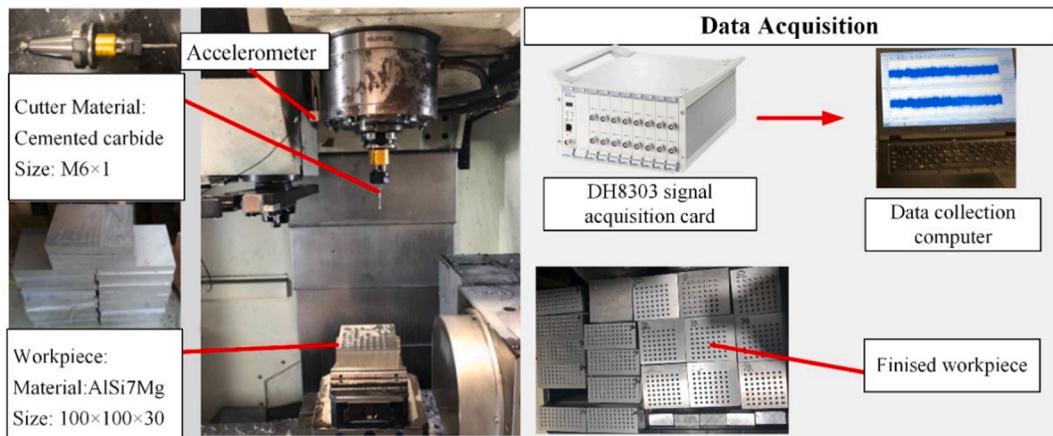


Fig. 14. The experimental setup and data acquisition system.

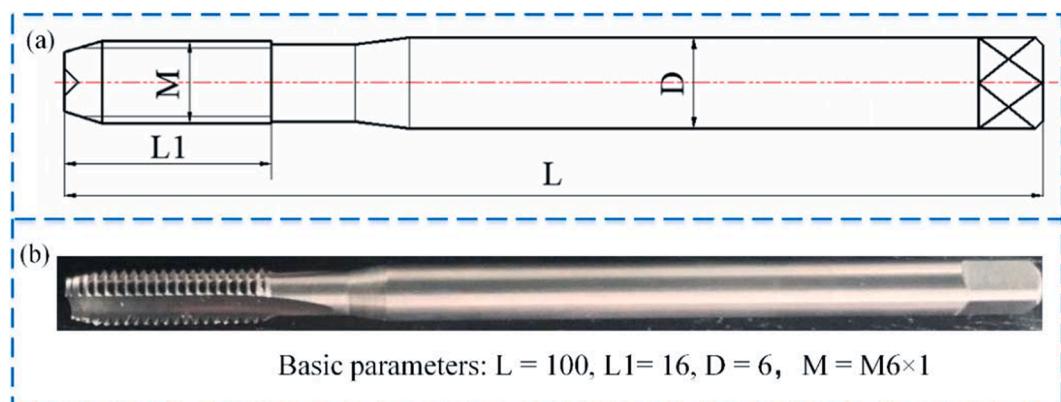
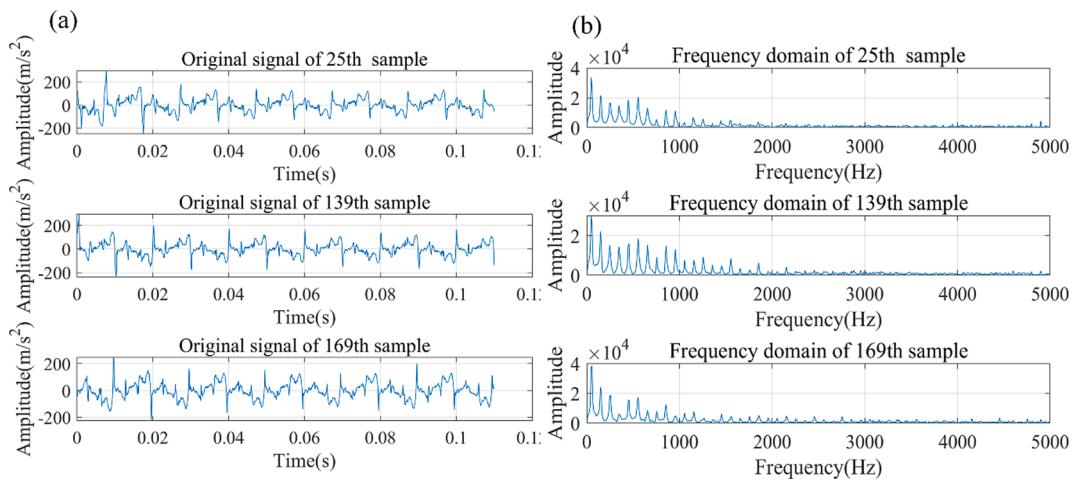
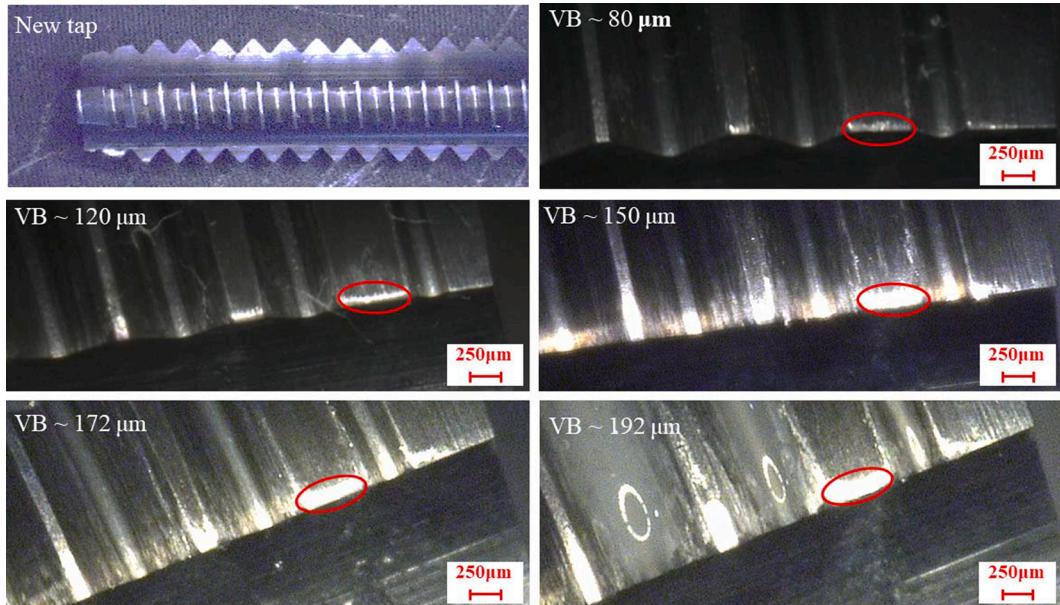


Fig. 15. The basic parameters of the tap tool: (a) two-dimensional diagram of tap, (b) real image and the basic parameters of tap.



**Fig. 16.** The collected vibration data of tap tool: (a) time domain, (b) frequency domain.



**Fig. 17.** The evolution process of the tap tool wear of T3.

**Table 3**

The division of training and testing dataset.

Number	Training dataset	Testing dataset
1	T2, T3	T1
2	T1, T3	T2
3	T1, T2	T3

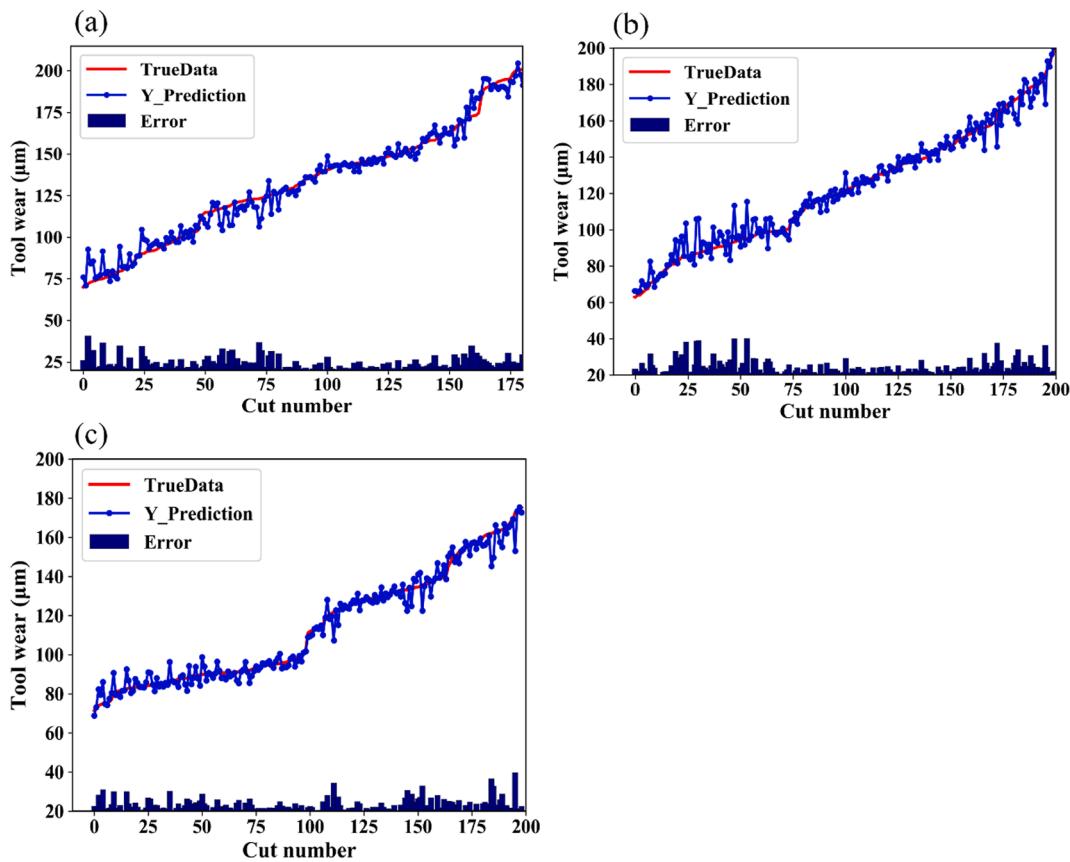
trend of actual tool wear values well, and the error between actual values and the predicted values are tiny. So the developed model can predict the tool wear accurately.

### 3.2.3. Model comparison and analysis

To be more objective, the published deep learning methods, including MLP, CNN, and LSTM, were selected to compare the prediction results of tool wear. The MLP model was composed of three hidden layers, and the parameters were the same as those of the proposed method. The structure and the parameters of the 1D CNN were the same as those of the proposed method. The only difference was that the 1D CNN did not have the channel attention mechanism and residual

connection. The structure and the parameters of the LSTM networks were the same as those in literature [47].

MAE and RSME results of three testing cutters with all the above-mentioned methods are listed in Table 4. It can be observed the results of RMSE and MAE of the three cutters with the proposed method are the smallest among all the methods. In addition, the RMSE of the proposed method of T1, T2, and T3 were 4.03, 4.62, and 4.41, respectively. The MAE values of T1, T2, and T3 were 2.64, 3.92, and 3.95, respectively. In the compared method, it can be seen that the MLP performed worst because of the limitation of its feature ability. The 1D CNN performed well because CNN has a remarkable feature extraction ability, and it is quite suitable for one-dimensional data. Based on the evaluation criteria, the proposed method has a good predictive ability for tool wear, and it is robust and effective to put the proposed method into practical application.



**Fig. 18.** The predicted tool wear results of different cutters: (a) T3 is the testing data, (b) T2 is the testing data, (c) T1 is the testing data.

**Table 4**  
Performance of the proposed model for three cutters.

Algorithms	MAE			RMSE		
	T1	T2	T3	T1	T2	T3
MLP	24.5	22.34	21.18	28.2	24.6	24.85
LSTM	7.25	8.55,	8.07	9.73	10.85	10.08
1D CNN	5.12	6.53	7.94	7.43	7.89	9.25
Proposed algorithm	2.64	3.92	3.95	4.03	4.62	4.41

### 3.3. Tool wear monitoring system development and engineering application

#### 3.3.1. Tool wear monitoring system development

During the machining process on the automatic production line, it is time-consuming and impractical to stop the machine tool for measuring the tool wear. Thus, a tool wear monitoring system with the proposed method is developed for monitoring and predicting the tool wear in real-time. Fig. 19 shows the procedure and function of developing a tool wear monitoring system for implementation. The vibration sensors are required to collect the vibration signal for the input of the proposed method. The internal data of the machine tool were obtained via Object Linking and Embedding for Process Control (OPC UA). The vibration signal and the internal data of the machine tool are stored in a database and displayed in the interface.

#### 3.3.2. Engineering application

The developed tool wear monitoring system was put into the tapping process to monitor tool wear in an engine cylinder head production line of SAIC General Motors Corporation Limited. The object of this study was the cylinder head C10T-CVG of SAIC General Motors, as is shown in Fig. 20(c). The tapping process of the engine cylinder was carried out on

a five-axis horizontal machining center, and the in-situ scene of cutting operation is presented in Fig. 20. The material of the cylinder head, cutting tool, and the cutting parameters are as same as in Section 3.2.1. The accelerometer is mounted on the machine spindle to measure the vibration signal on-line, and the collected signal is regarded as the input of the developing model and displayed in the developed software.

By using the developed tool monitoring system, the internal data of the machine tool can obtain by OPC UA and display in the interface online. So the cutting process is monitored, and the whole monitoring results during the tapping process are shown in Fig. 21. It can be vividly observed that the tool wear can be predicted every 1 s, and the predicted tool wear result are displayed in the interface. Besides, the cutting parameters, coordinate, and the spindle information of the machine tool, etc, are obtained by OPC UA and displayed in real-time in the developed software during the cutting process. When the too wear achieve the threshold of the tool, the tool can be changed properly to ensure the quality of the engine cylinder and the stability of the machining process.

## 4. Conclusions

To achieve a higher accuracy of tool wear prediction, a deep learning-based method by using the convolutional network and channel attention mechanism is proposed. The proposed method can realize multi-sensor feature fusion and consider the weight of the different feature map. The main conclusions are drawn as follows:

- (1). The parallel structure of CNN was developed to implement multi-sensor feature fusion such that the various feature of the different signals are fully used for tool wear prediction, and the designed structure of the network can be a new way for achieving multi-scale feature fusion.

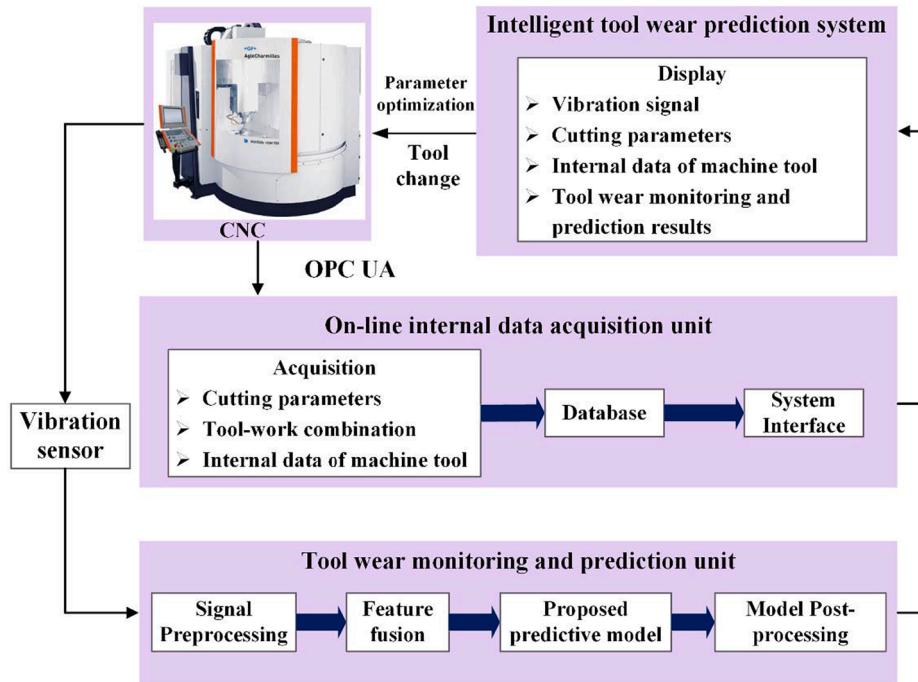


Fig. 19. Diagram of the intelligent tool wear monitoring system for implementation.

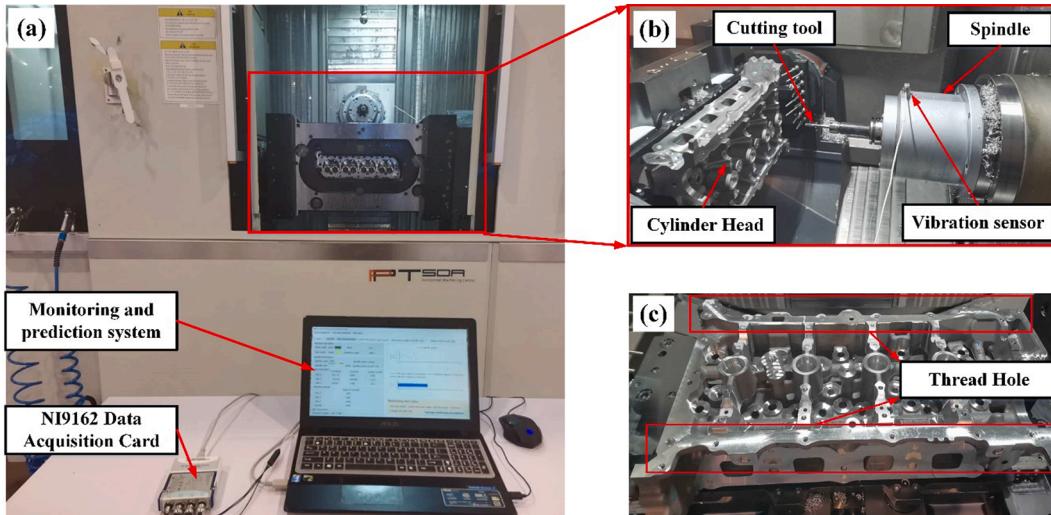


Fig. 20. . The in-situ scene of cutting operation of the engine cylinder head production line.

(2). The channel attention mechanism with the residual connection is utilized to select the useful features to improve the tool wear prediction results. Different tool wear prediction experiments with the proposed method are conducted. As a result, the proposed method has shown its superiority in prediction accuracy by comparing with the state-of-the-art methods, and the effectiveness and generalization are validated. So this research provides a new method for tool wear prediction, and the proposed method is promising for the modern manufacturing industry.

(3). The developed tool monitoring system is implemented in the engine cylinder head production line, which combined CNC, OPC UA, sensors with the proposed method. As an engineering application of the proposed method, the developed tool monitoring system can be a reference for the whole tool monitoring community.

For future work, it is necessary to further improve the tool wear monitoring system and validate the proposed method in various

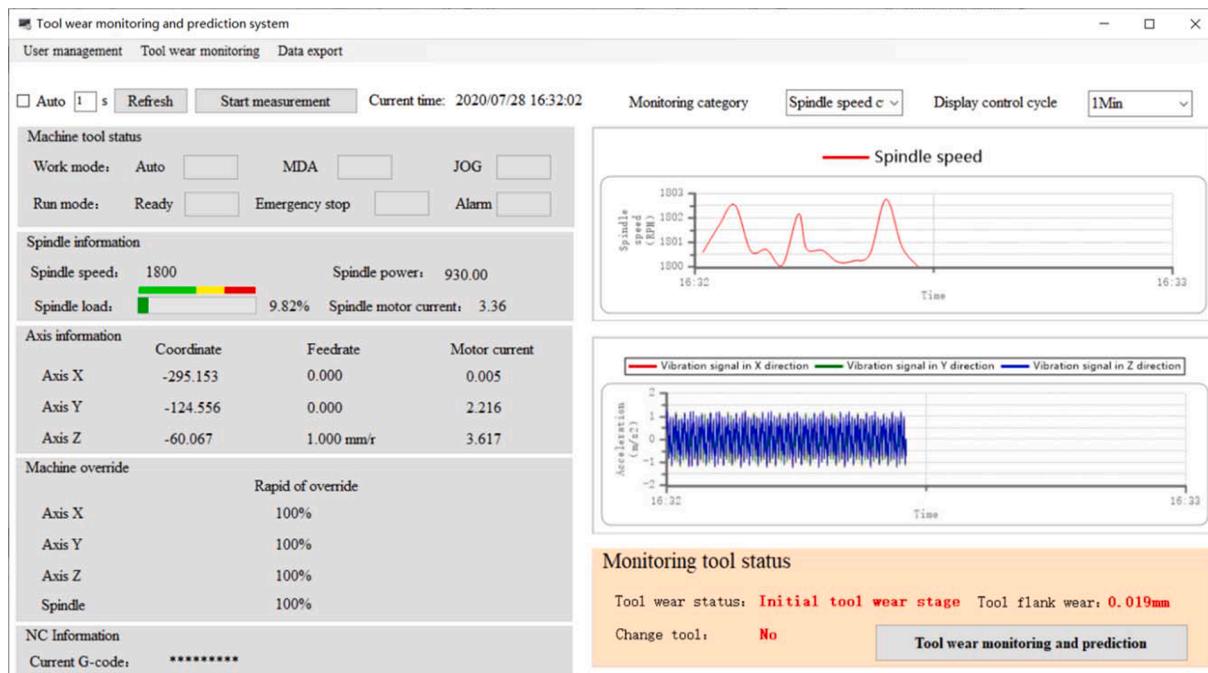


Fig. 21. The developed tool wear monitoring system and its testing results.

scenarios.

#### CRediT authorship contribution statement

**Xingwei Xu:** Methodology, Software, Data curation, Writing - original draft. **Jianwen Wang:** Conceptualization, Validation. **Bingfu Zhong:** Writing - review & editing. **Weiwei Ming:** Visualization, Investigation. **Ming Chen:** Supervision.

#### Declaration of Competing Interest

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

#### Acknowledgment

The funding sources are the National Science and Technology Major Project of China (No. 2019ZX04007001).

#### References

- [1] T. Mohanraj, J. Yerchuru, H. Krishnan, R.S. Nithin Aravind, R. Yameni, Development of tool condition monitoring system in end milling process using wavelet features and Hölder's exponent with machine learning algorithms, Measurement (2020).
- [2] K. Zhu, Y. Zhang, A generic tool wear model and its application to force modeling and wear monitoring in high speed milling, Mech. Syst. Sig. Process. 115 (2019) 147–161.
- [3] Z. Li, R. Liu, D.J.J.o.M.P. Wu, Data-driven smart manufacturing: Tool wear monitoring with audio signals and machine learning, 48 (2019) 66–76.
- [4] M. Kuntoglu, H. Sağlam, Investigation of progressive tool wear for determining of optimized machining parameters in turning, Measurement 140 (2019) 427–436.
- [5] C. Wang, Z. Bao, P. Zhang, W. Ming, M. Chen, Tool wear evaluation under minimum quantity lubrication by clustering energy of acoustic emission burst signals, Measurement 138 (2019) 256–265.
- [6] Y. Yang, Y. Guo, Z. Huang, N. Chen, L. Li, Y. Jiang, N. He, Research on the milling tool wear and life prediction by establishing an integrated predictive model, Measurement 145 (2019) 178–189.
- [7] K. Javed, R. Gouriveau, X. Li, N. Zerhouni, Tool wear monitoring and prognostics challenges: a comparison of connectionist methods toward an adaptive ensemble model, J. Intell. Manuf. 29 (2016) 1873–1890.
- [8] J. Dou, C. Xu, S. Jiao, B. Li, J. Zhang, X. Xu, An unsupervised online monitoring method for tool wear using a sparse auto-encoder, Int. J. Adv. Manuf. Technol. (2019).
- [9] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R.X. Gao, Deep learning and its applications to machine health monitoring, Mech. Syst. Sig. Process. 115 (2019) 213–237.
- [10] J. Ratava, M. Lohtander, J. Varis, Tool condition monitoring in interrupted cutting with acceleration sensors, Rob. Comput. Integrat. Manuf. 47 (2017) 70–75.
- [11] J. Wang, J. Xie, R. Zhao, L. Zhang, L. Duan, Multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing, Rob. Comput. Integrat. Manuf. 45 (2017) 47–58.
- [12] X. Xu, Z. Tao, W. Ming, Q. An, M. Chen, Intelligent monitoring and diagnostics using a novel integrated model based on deep learning and multi-sensor feature fusion, Measurement 165 (2020).
- [13] G. Wang, Y. Yang, Q. Xie, Y. Zhang, Force based tool wear monitoring system for milling process based on relevance vector machine, Adv. Eng. Softw. 71 (2014) 46–51.
- [14] A.I. Azmi, Monitoring of tool wear using measured machining forces and neuro-fuzzy modelling approaches during machining of GFRP composites, Adv. Eng. Softw. 82 (2015) 53–64.
- [15] Y. Hui, X. Mei, G. Jiang, T. Tao, C. Pei, Z. Ma, Milling tool wear state recognition by vibration signal using a stacked generalization ensemble model, Shock Vib. 2019 (2019) 1–16.
- [16] Y. Lin, X. Li, Y. Hu, Deep diagnostics and prognostics: an integrated hierarchical learning framework in PHM applications, Appl. Soft Comput. 72 (2018) 555–564.
- [17] F. Aghazadeh, A. Tahani, M. Thomas, Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process, Int. J. Adv. Manuf. Technol. 98 (2018) 3217–3227.
- [18] X.-C. Cao, B.-Q. Chen, B. Yao, W.-P. He, Combining translation-invariant wavelet frames and convolutional neural network for intelligent tool wear state identification, Comput. Ind. 106 (2019) 71–84.
- [19] Y. Chen, Y. Jin, G. Jiri, Predicting tool wear with multi-sensor data using deep belief networks, Int. J. Adv. Manuf. Technol. 99 (2018) 1917–1926.
- [20] A. Bustillo, L.N. López de Lacalle, A. Fernández-Valdvielso, P.J.J.o.C.D. Santos, Engineering, Data-mining modeling for the prediction of wear on forming-taps in the threading of steel components, 3 (2016) 337–348.
- [21] Á. Arnaiz-González, A. Fernández-Valdvielso, A. Bustillo, L.N. López de Lacalle, Using artificial neural networks for the prediction of dimensional error on inclined surfaces manufactured by ball-end milling, Int. J. Adv. Manuf. Technol. 83 (2015) 847–859.
- [22] M. Elangovan, V. Sugumaran, K.I. Ramachandran, S. Ravikumar, Effect of SVM kernel functions on classification of vibration signals of a single point cutting tool, Expert Syst. Appl. 38 (2011) 15202–15207.
- [23] D. Kong, Y. Chen, N. Li, S. Tan, Tool wear monitoring based on kernel principal component analysis and v-support vector regression, Int. J. Adv. Manuf. Technol. 89 (2016) 175–190.
- [24] T. Benkedjouh, K. Medjaher, N. Zerhouni, S.J.J.o.M. Rechak, Health assessment and life prediction of cutting tools based on support vector regression, 26 (2015) 213–223.

- [25] D. Kong, Y. Chen, N. Li, C. Duan, L. Lu, D. Chen, Relevance vector machine for tool wear prediction, *Mech. Syst. Sig. Process.* 127 (2019) 573–594.
- [26] D. Kong, Y. Chen, N. Li, Force-based tool wear estimation for milling process using Gaussian mixture hidden Markov models, *Int. J. Adv. Manuf. Technol.* 92 (2017) 2853–2865.
- [27] K. Zhu, Y.S. Wong, G.S. Hong, Multi-category micro-milling tool wear monitoring with continuous hidden Markov models, *Mech. Syst. Sig. Process.* 23 (2009) 547–560.
- [28] W. Li, T. Liu, Time varying and condition adaptive hidden Markov model for tool wear state estimation and remaining useful life prediction in micro-milling, *Mech. Syst. Sig. Process.* 131 (2019) 689–702.
- [29] W. Cai, W. Zhang, X. Hu, Y. Liu, A hybrid information model based on long short-term memory network for tool condition monitoring, *J. Intell. Manuf.* (2020).
- [30] X. Wu, J. Li, Y. Jin, S. Zheng, Modeling and analysis of tool wear prediction based on SVD and BiLSTM, *Int. J. Adv. Manuf. Technol.* 106 (2020) 4391–4399.
- [31] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, J. Wang, Machine health monitoring using local feature-based gated recurrent unit networks, *IEEE Trans. Ind. Electron.* 65 (2018) 1539–1548.
- [32] G. Qiu, Y. Gu, Q. Cai, A deep convolutional neural networks model for intelligent fault diagnosis of a gearbox under different operational conditions, *Measurement* 145 (2019) 94–107.
- [33] Z. Yang, D. Yang, C. Dyer, X. He, E. Hovy, Hierarchical attention networks for document classification, in: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2016.
- [34] Y. Chen, G. Peng, Z. Zhu, S. Li, A novel deep learning method based on attention mechanism for bearing remaining useful life prediction, *Appl. Soft Comput.* 86 (2020).
- [35] A.G.D. Val, J. Fernández, M. Arizmendi, F. Veiga, J.L. Urfízar, A. Berriozábal, A. Axpe, P.M. Diéguez, On line diagnosis strategy of thread quality in tapping, *Procedia Eng.* 63 (2013) 208–217.
- [36] R.K. Barooah, A.F.M. Arif, J.M. Paiva, S. Oomen-Hurst, S.C. Veldhuis, Wear of form taps in threading of Al-Si alloy parts: mechanisms and measurements, *Wear* 442–443 (2020).
- [37] P. Monka, K. Monkova, V. Modrak, S. Hric, P. Pastucha, Study of a tap failure at the internal threads machining, *Eng. Fail. Anal.* 100 (2019) 25–36.
- [38] A. Gil del Val, J. Fernández, P.M. Diéguez, M. Arizmendi, F. Veiga, Real time diagnosis charts of thread quality in tapping operations, *Mater. Sci. Forum* 797 (2014) 71–77.
- [39] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [40] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R.J.T.j.o.m.l.r. Salakhutdinov, Dropout: a simple way to prevent neural networks from overfitting, 15 (2014) 1929–1958.
- [41] S. Ioffe, C.J.a.p.a. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, (2015).
- [42] D.P. Kingma, J.J.a.p.a. Ba, Adam: a method for stochastic optimization, (2014).
- [43] Z. Chen, Y. Chen, L. Wu, S. Cheng, P. Lin, Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions, *Energy Convers. Manage.* 198 (2019).
- [44] X. Li, B. Lim, J. Zhou, S. Huang, S. Phua, K. Shaw, M. Er, Fuzzy neural network modelling for tool wear estimation in dry milling operation, in: Annual conference of the prognostics and health management society, 2009, pp. 1–11.
- [45] H. Qiao, T. Wang, P. Wang, S. Qiao, L. Zhang, A time-distributed spatiotemporal feature learning method for machine health monitoring with multi-sensor time series, *Sensors (Basel)* 18 (2018).
- [46] R. Zhao, J. Wang, R. Yan, K. Mao, Machine health monitoring with LSTM networks, in: 2016 10th International Conference on Sensing Technology (ICST), (IEEE2016), pp. 1–6.
- [47] Z. Tao, Q. An, G. Liu, M. Chen, A novel method for tool condition monitoring based on long short-term memory and hidden Markov model hybrid framework in high-speed milling Ti-6Al-4V, *Int. J. Adv. Manuf. Technol.* 105 (2019) 3165–3182.