

Original Article

DOI 10.1007/s12206-024-0632-9

Keywords:

· TCM
· Deep learning
· Wavelet threshold
· CNN

Correspondence to:

Yaonan Cheng
yaonancheng@163.com

Citation:

Guan, R., Cheng, Y., Jin, Y., Zhou, S., Gai, X., Lu, M. (2024). Research on tool wear prediction for milling high strength steel based on DenseNet-ResNet-GRU. *Journal of Mechanical Science and Technology* 38 (7) (2024) 3585~3596. <http://doi.org/10.1007/s12206-024-0632-9>

Received July 19th, 2023

Revised January 19th, 2024

Accepted March 20th, 2024

† Recommended by Editor
Hyung Wook Park

Research on tool wear prediction for milling high strength steel based on DenseNet-ResNet-GRU

Rui Guan^{1,2}, Yaonan Cheng¹, Yingbo Jin¹, Shilong Zhou¹, Xiaoyu Gai¹ and Mengda Lu¹

¹College of Mechanical and Power Engineering, Key Laboratory of Advanced Manufacturing and Intelligent Technology, Ministry of Education, Harbin University of Science and Technology, Harbin 150080, China, ²Harbin Vocational and Technical College, Harbin, China

Abstract Tool condition monitoring is an important basis to ensure workpiece quality and machining efficiency. It is also a key factor in improving machining efficiency, ensuring machining accuracy. Therefore, a new method for predicting tool wear based on DenseNet-ResNet-GRU is proposed. Firstly, statistical theory and an improved wavelet threshold denoising method are used to improve the signal quality. In addition, the asymptotic semi-soft threshold function is applied to reduce the noise of the cutting force signal. Secondly, DenseNet, ResNet, and GRU (gate recurrent unit) deep learning networks are integrated to create a new tool wear prediction model to realize the nonlinear mapping relationship between the tool wear amount and the cutting force characteristic. Finally, the tool wear prediction model is verified by high-strength steel experiment. The experimental results verify the accuracy and reliability of the method, which has a better training effect and higher prediction accuracy compared with the CNN-GRU model.

1. Introduction

With the continuous development of artificial intelligence, internet of things and other technologies, intelligent manufacturing has become an important development direction of modern manufacturing [1, 2]. With the comprehensive promotion of intelligent manufacturing informatization and the deep integration of industrialization, cutting tools have become one of the influencing factors for efficient manufacturing performance. The water chamber head of the AP1000 nuclear power plant is the key component of the steam generator, which has the characteristics of heavy weight, large volume, complex profile and special material [3]. The material of the water chamber head is 508III steel. Since 508III steel is a difficult-to-cut material with high strength and good plasticity, and the cutting data during the milling process is large, the milling tools are subjected to thermo-mechanical loads with low cycles and high shocks. In addition, the rake face and the chip, the flank face and the workpiece generate severe extrusion and friction, resulting in severe wear under the action of force and heat [4]. At present, at the water chamber head machining site, the replacement of tools mainly depends on the on-site judgment of experienced workers. The on-site processing experience of the workers is very demanding, and the labor cost is high. Sometimes the tool is replaced in advance when it is far from reaching the end of its service life. Although the quality of workpiece machining is guaranteed, the company's production costs are increased [5, 6]. Kennametal, a well-known American tool company, found that machine tools equipped with tool condition monitoring systems can save up to 30 % of machining costs [7, 8]. Therefore, the application of artificial intelligence in the tool processing industry and the continuous monitoring of tool condition in the processing can effectively improve the automation level of the workshop, reduce the tool cost, and ensure the processing quality of the product as much as possible.

If the cutting data is the same in the milling process, tool wear will lead directly to an increase in cutting force. The cutting force propagation path is short, so the cutting force signal is less

disturbed than other monitoring signals. Because of these advantages, the use of cutting force to monitor tool wear has become a hot research topic. Tool condition monitoring means that after receiving the output signal from the sensor, the original signal is processed using various algorithms to determine the condition of the cutting tool. The signal received from the sensor is usually time series data, showing a nonlinear change law, and the amount of data is huge. However, the density of the value data is low, and the original signal contains ambient noise. The effect is not good without data preprocessing. The original physical signal needs to be amplified, filtered, normalized, denoised, etc [9-11]. The main process of analyzing signals by tool wear prediction method are machine learning methods and deep learning methods. Deep learning omits the steps of feature extraction and selection of machine learning, and enters all the information of the signal into the deep learning model for learning. Deep learning models commonly include AutoEncoder (AE), convolutional neural networks (CNN), sparse coding (SC), and deep belief networks (DBN). Through multi-layer processing, simple models can be used for complex classification.

Most scholars have used different deep learning models to study tool wear prediction. Liu et al. [12] proposed an ensemble model combining CNN with bidirectional long short-term memory (LSTM). The cutting force, vibration and sound signals are denoised and sampled by the wavelet threshold and used as input to realize tool wear monitoring. This model is ideal in terms of accuracy and stability. Li et al. [13] proposed a tool monitoring model with residual-dense network (RDN) to effectively and accurately monitor the tool wear value in machining process. By acquiring a time-domain signal of the tool vibration, the signal is wavelet denoised, and then segmented to obtain more homogeneous samples. Finally, it is fed into the model to adaptively extract features and perform tool wear prediction. Compared with other traditional neural networks, it has high prediction accuracy. Xu et al. [14] proposed a multiscale convolutional gating loop unit to process raw sensor data and tool wear prediction. Wang et al. [15] established a model based on deep heterogeneous GRU and local feature extraction for monitor tool wear

monitoring. Zhao et al. [16] developed a convolutional bidirectional LSTM. Scholars have used various methods to extensively study tool wear signal processing and tool monitoring models to determine reasonable deep learning models to accurately predict tool wear. However, there are still some technical bottlenecks in practical applications [17, 18]. Due to the special properties of 508III steel material, the cutting process is more complicated. It is still an important issue that needs to be solved urgently. For example, the monitoring signals are pre-processed to reduce the influence of the external environment. And the reasonable deep learning method is determined to monitor the amount of tool wear during the milling process.

In summary, advances in data processing and artificial intelligence technology provides a good theoretical basis and algorithmic conditions for realizing high-precision and high-reliability intelligent tool monitoring. However, conventional denoising techniques are not effective in eliminating the original signal noise. Machine learning, on the other hand, is prone to problems such as vanishing or exploding gradients, where the model no longer converges and accuracy is low. Therefore, it established a multiscale DenseNet-ResNet-GRU of deep learning for tool wear prediction model in this paper. The cutting force signal matrix with the multiscale convolution kernel as the input is used for convolution. According to the respective characteristics of DenseNet and ResNet, the signal space features are extracted, and the time series features are extracted by using the GRU to test and verify the prediction accuracy of the model.

2. Milling experiments and data acquisition

2.1 Data type

The experiment is conducted in a room temperature environment. 508 III steel, the water chamber head material of the nuclear island nuclear power plant, is as the experiment material. The tool wear test was carried out on Dalian machine tool VDL-1000E, and the workpiece was a size of 170×105×80 mm. The experiment uses an indexable cutterhead with the brand name Zhuzhou Diamond, model FMR04-100-B32-RD16-06. The cut-

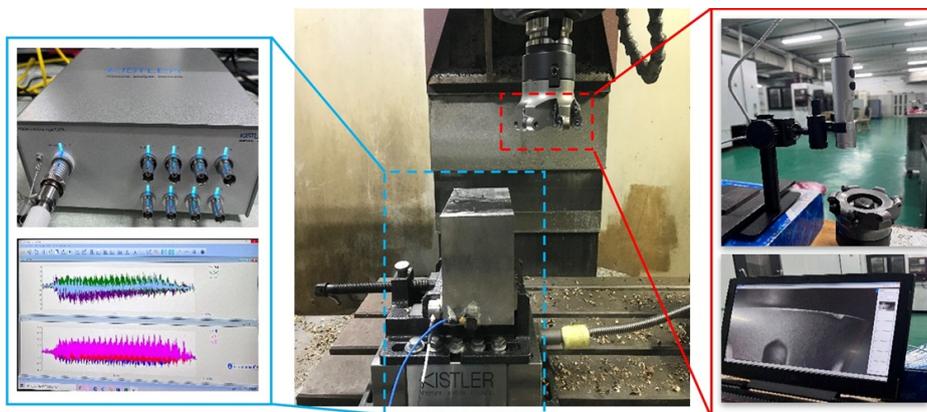


Fig. 1. Tool wear experiment equipment of milling high-strength steel.

terhead is installed symmetrically with 3 inserts. The insert material is cemented carbide coated with TiAlN. The model is WIDIA RDMT1605MOTXA. The dynamometer sensor is placed on the machine table and the fixture is mounted above the sensor. The workpiece is fixed using a fixture, the sensor is connected to the charge amplifier by a signal line, and finally stored and displayed by KISTLER software in the computer. Tool wear was measured using Supereyes industrial digital microscope, and the experimental setup is shown in Fig. 1.

2.2 Cutting force signal acquisition and tool wear monitoring

The parameters of this test are set to $v_c = 250$ m/min, $f_z = 0.3$ mm/z, $a_p = 1$ mm, and $a_e = 10$ mm. After each signal acquisition test at the same time interval, the amount of flank wear is monitored by an industrial camera to record changes over the entire life cycle of the tool. The topography and the amount of flank wear are measured after each pass. Whenever the pass length is 105 mm, the VB value is measured and recorded. Therefore, a total of 78 cutting force signals were monitored and 78 tool wear values were measured before tool failure, resulting in 78 samples and labels. Fig. 2 shows the tool wear curve measured by the test. It can be observed from the figure that as the cutting proceeds, the waveforms of the tool wear condition and X-direction cutting force vary differently with different amounts of tool wear. The tool wear shows different elevated trends. And it is divided into three stages of initial, normal, and severe wear. The wear value of 0-0.2 mm is the initial wear stage, the wear value of 0.2-0.3 mm is the normal wear stage, and the wear value of more than 0.3 mm is the severe wear stage.

3. Data preprocessing

Due to the complexity of the working environment of the tool in the milling process, there is inevitably noise interference in the monitoring signal. It will result in a large amount of useless information in the signal, affecting the authenticity of the cutting force signal. Therefore, it is necessary to preprocess the original

cutting force signal to eliminate invalid data and reduce noise.

3.1 Abnormal data processing based on statistics

The cutting force data itself has a certain regularity. Some external factors cause some data to deviate from the overall trend, thus affecting the accuracy of the test. Outliers are detected and rejected to eliminate the interference of erroneous data and ensure the validity of predicted data. Statistical-based anomaly detection is a relatively mature and widely used method. The sampling frequency set in this paper is 10000 Hz, and the data volume is huge. According to the literature, when the data volume is large, it is more reasonable to use the Laida

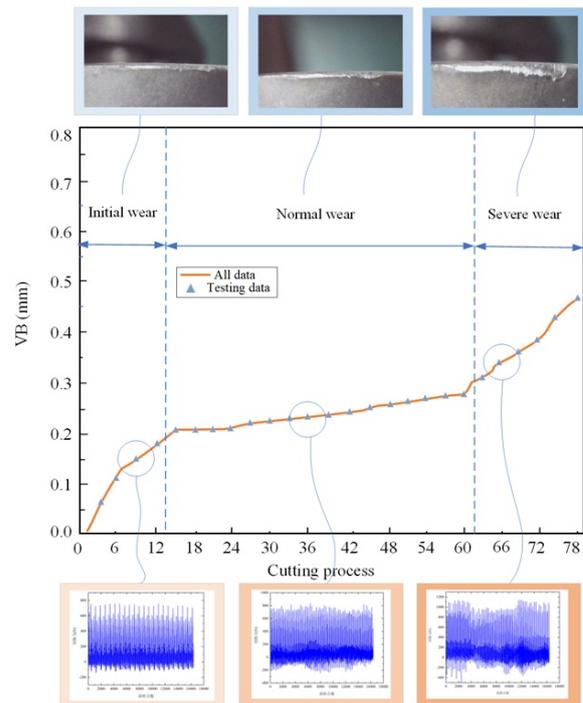
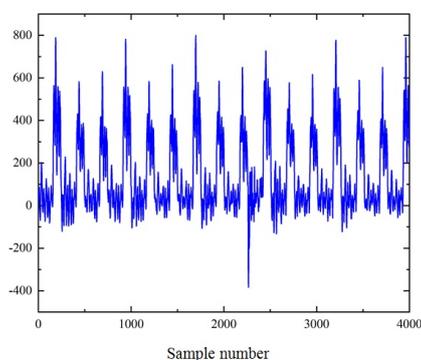
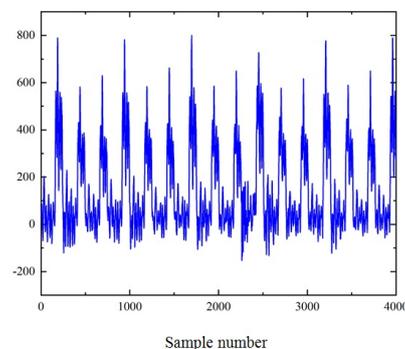


Fig. 2. Tool wear curve.



(a) Unprocessed data



(b) Processed data

Fig. 3. Diagram of before and after outlier processing.

Table 1. Denoising effect of different wavelet basis.

Wavelet type	Order	Evaluation indicators		Wavelet type	Order	Evaluation indicators	
		SNR	RMSE			SNR	RMSE
Db	2	22.397	0.492	Coif	3	24.514	0.386
	3	24.279	0.396		4	25.301	0.352
	4	25.035	0.363		5	25.760	0.334
	5	24.413	0.390	Sym	2	22.397	0.492
	6	25.003	0.365		3	24.279	0.396
	7	24.364	0.392		4	23.854	0.416
	8	25.156	0.358		5	23.477	0.435
	9	24.692	0.378		6	25.183	0.357
	10	24.405	0.391		7	24.800	0.373
					8	24.373	0.392
Doif	1	23.234	0.447				
	2	23.556	0.431				

criterion for processing [19]. Taking the time domain signal of the X-directional force part of the 6th milling process as an example, the data before and after processing are shown in Fig. 3.

3.2 Denoising based on the wavelet threshold method

In the milling wear monitoring test, the signal is easily disturbed by noise during transmission, hard to receive the signal correctly on the receiving end. However, the wavelet transform meets the requirements of signal noise reduction. Because it can adjust the resolution in the time and frequency domains according to the characteristics of the signal itself [20, 21].

The signal-to-noise ratio (SNR) is a common method for evaluating the denoising effect, and a high SNR represents a good denoising effect. The unit is decibel and SNR is defined as Eq. (1).

$$SNR = 10 \times \lg \left[\frac{\sum_{i=1}^N s_i^2}{\sum_{i=1}^N (s_i - f_i)^2} \right]. \quad (1)$$

A small root mean square error (RMSE) represents effective noise reduction and is expressed as.

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (s_i - f_i)^2}. \quad (2)$$

Where s_i is the original signal. f_i is the effective signal. N is the number of points.

1) The number of decomposition layers and wavelet bases selection.

In a large number of experiments, the number of decomposition layers is mostly between 3 and 6 layers, and in this paper, 3 layers are decomposed according to experience. Unlike Fourier analysis, there are many types of wavelets that satisfy the conditions, and dB wavelets, coif wavelets and sym wavelets

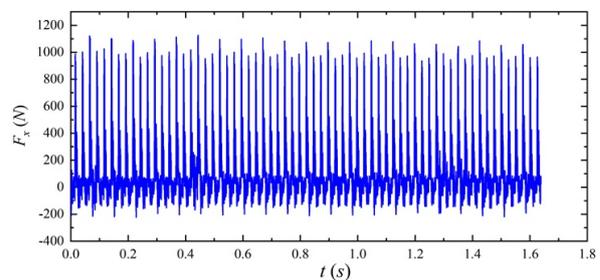


Fig. 4. Time domain signal of the x-directional force during the 58th milling.

are often used in engineering.

The time domain signal in the X-direction force of the 58th milling process is shown in Fig. 4. The same data were processed by decomposition layer 3, soft threshold function, rigrsure threshold, and different kinds and orde wavelet bases. The noise reduction effect is shown in Table 1. The coordinate plots in Fig. 5 were drawn to compare and analyze the SNR and RMSE.

From the SNR calculation results, the coif wavelet family has the largest value of coif5. Both db and sym wavelet families fluctuate significantly. The maximum value of the db wavelet family is at db8, and the max sym wavelet family is at sym6. From the RMSE it can be seen that the coif wavelet family RMSE decreases with the increase of the wavelet order. The result of the db wavelet family is that there is a certain fluctuation after the first decline, and the smaller values are obtained at db4, db6 and db8, respectively. The sym wavelet family is generally on a downward trend. After considering different factors, the coif5 wavelet base with the highest SNR and the lowest RMSE was selected as the wavelet base for the subsequent noise reduction work.

2) Selection of wavelet thresholds

There are four common threshold guidelines for wavelet analysis: sqtwolog threshold, rigrsure threshold, heursure threshold and minimaxi threshold [22].

Set the wavelet base to coif5. Using the soft threshold func-

Table 2. Denoising effect of several wavelet thresholds.

Threshold	SNR	RMSE
Sqtwolog	17.224	0.893
Rigrsure	25.760	0.334
Heursure	17.782	0.837
Minimaxi	19.242	0.708

Table 3. Denoising effect of different threshold function.

Threshold function	SNR	RMSE
Hard threshold function	30.836	0.186
Soft threshold function	25.760	0.334
Progressive semi-soft threshold function	67.021	0.003

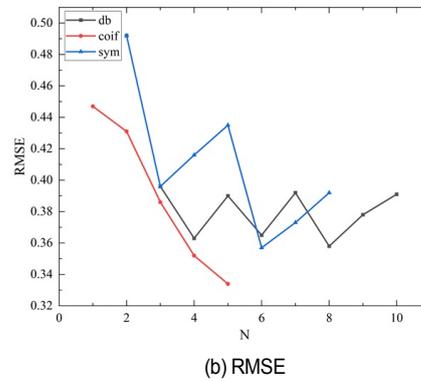
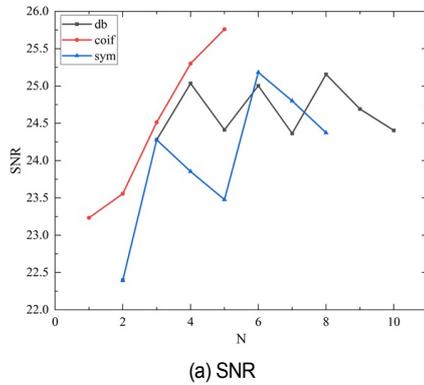


Fig. 5. A contrast of the denoising effects of several wavelet basis.

tion, the results of denoising with different thresholds are shown in Table 2.

In Table 2, the SNR and RMSE of the heursure threshold and the sqtwolog threshold are similar. Both denoising effects are slightly inferior to the minimaxi threshold. The rigrsure threshold has the best noise reduction. Therefore, the rigrsure threshold is used for 508III steel cutting force signal noise reduction processing.

3) Selection of threshold functions

Common threshold functions are divided into soft threshold functions and hard threshold functions [23].

However, both the hard threshold function and the soft threshold function have their limitations. Therefore, a progressive semi-soft threshold function $f(\omega)$ is used as follows.

$$\tilde{\omega}_{j,k} = f(\omega) = \begin{cases} \text{sign}(\omega_{j,k}) \left(|\omega_{j,k}| - \frac{2\lambda^2}{|\omega_{j,k}| + \lambda e^{|\omega_{j,k}| - \lambda}} \right) & |\omega_{j,k}| > \lambda \\ 0 & |\omega_{j,k}| \leq \lambda \end{cases} \quad (3)$$

Where, $\omega_{j,k}$ is the wavelet coefficient, λ is the threshold, and $\tilde{\omega}_{j,k}$ is the modified wavelet coefficient.

The rigrsure threshold is selected, and the signal is denoised by the *coif5* wavelet basis function and different threshold functions.

As shown in Table 3, SNR obtained by using the threshold functions of hard threshold, soft threshold, and the semi-soft threshold are 30.8355 dB, 25.7596 dB and 67.0212 dB, respectively. The threshold function used in this paper can reach twice the hard threshold and has a lower RMSE, so it has a better denoising effect than the other methods.

4. Prediction models

Deep learning can adaptively extract features to avoid the tedious manual extraction of features [24]. This section attempts to build a tool wear prediction model based on deep learning.

4.1 Improved convolutional neural networks (CNNs)

Theoretically, the deeper the layers number of a CNN is, the more neurons it has, and the better the fitting effect is. However, with the increase in the number of layers, problems such as gradient explosion and gradient disappearance occur. This negatively affects backpropagation training. In response to these problems, CNNs have a lot of improved structures. This paper adopts ResNet and DenseNet in the improved structure. These two structures show good results in practical applications, so this paper includes them as part of the predictive functional model.

1) ResNet

ResNet is to add a cross-layer identity map between CNNs to solve the gradient degradation problem. Due to the addition of a path to identity mapping, the model can choose the best path to improve the model training speed when updating parameters through backpropagation. The residual module is as follows.

$$y = \begin{cases} F(x, W_i) + x \\ F(x, W_i) + W_s x \end{cases} \quad (4)$$

Where, x is the mapping input. W_i and W_s are convolutional.

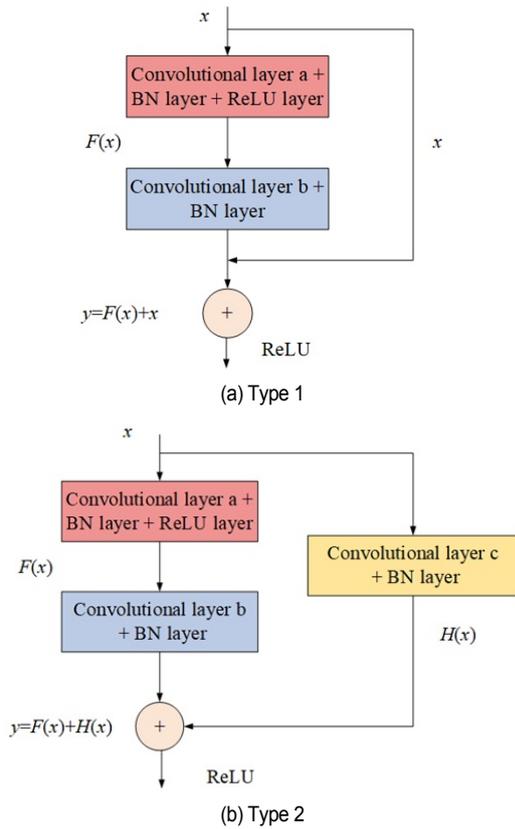


Fig. 6. ResNet module structure.

$F(x, W_i)$ is the residual mapping. The same location elements of the two maps are added together to get the output y . The structure of this residual module is shown in Fig. 6.

The main branch consists of two convolutional layers and two BN layers, which are superimposed alternately. ReLU is the activation function, and the branch is the input of this ResNet module. As shown in Fig. 6(a), when the input and output dimensions of the main branch are inconsistent, the branch contains one convolutional layer and one BN layer. This aligns the output size of the branch with output size of the main branch, as shown in Fig. 6(b). After the tensors of the input residual block have passed through the main and branch respectively, the two new tensors are added. Finally, ReLU is the activation function forms the output of the residual block.

2) DenseNet

DenseNet is a way to alleviate the problem of vanishing gradients by connecting any two layers in a hopping manner, thereby training deeper networks. The DenseNet is given by Eq. (5).

$$x_i = H_i([x_0, x_1, \dots, x_{i-1}]) \tag{5}$$

Where: X_i is the output feature of the DenseNet module. $[x_0, x_1, \dots, x_{i-1}]$ is the output layer 0 to $i-1$ to be connected, $H_i(\cdot)$ represents BN, ReLU, pooling and convolution. There-

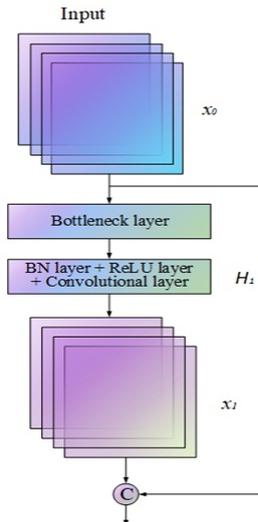


Fig. 7. DenseNet module structure.

fore, the features extracted by DenseNet are concatenated in series with the features of the previous layer to achieve feature reuse to reduce redundancy. The dense block designed in this article contains one layer, the structure of which is shown in Fig. 7.

As shown in Fig. 7, the DenseNet module first establishes a bottleneck layer, which can act as a dimensionality reduction. It includes a BN layer, a ReLU activation function, and a convolutional layer with a convolution kernel of 1×1 . Then set the BN layer again, and after activation by the ReLU activation function, connect a convolutional layer with a convolution kernel of 3×3 . Connect the output features with the input features in the channel dimension, which forms the output of a dense block.

The connection between the two DenseNet modules is through the transition layer. The Transition layer reduces the feature map size. The transition layer consists of a convolutional layer with a convolution kernel of 1×1 . and an average pooling layer of 2×2 .

4.2 GRU model structure

GRU is an improved recurrent neural network (RNN). It has a simple internal structure and requires less computation to upgrade the internal state, making it easier to train. The internal structure of the individual GRUs is shown in Fig. 8.

The reset gate equation is as follows:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \tag{6}$$

Where: σ is the sigmoid function. W_r is the reset gate weight matrix. x_t is the input of the t -th time step. h_{t-1} is the hidden information stored from the previous time step. The equation for the update gate is as follows:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \tag{7}$$

Table 4. Parameter settings of multiscale DenseNet-ResNet-GRU model.

Title	Structure	Model information	Value
Convolutional layer 1,3	Convolutional layer	Convolution parameters	The number of convolution kernels = 64; convolution kernel size = 3×1; step = 1×1;
Convolutional layer 2, 4	Convolutional layer	Convolution parameters	The number of convolution kernels = 64; convolution kernel size = 5×1; step = 1×1;
Max pooling layer 1-4	Max pooling layers	Max pooling	Convolution kernel size = 3×3; step = 2×2;
ResNet module 1, 2	Master branch convolutional layer a	Convolution parameters	The number of convolution kernels = 64; convolution kernel size = 3×3; step = 1×1 ; ;
	Master branch convolutional layer b	Convolution parameters	The number of convolution kernels = 64; convolution kernel size = 3×3; step = 1×1;
	Branch branches	--	--
ResNet module 3, 4	Master branch convolutional layer a	Convolution parameters	The number of convolution kernels = 128; convolution kernel size = 3×3; sep = 2×2;
	Master branch convolutional layer b	Convolution parameters	The number of convolution kernels = 128; convolution kernel size = 3×3; step = 1×1;
	Branch convolutional layer c	Convolution parameters	The number of convolution kernels = 128; convolution kernel size = 1×1; step = 2×2;
GRU layer 1-3	GRU	Number of neurons	64
Fully connected layer 1	Fully connected layer	Number of neurons	64
Fully connected layer 2	Fully connected layer	Number of neurons	32
Fully connected layer 3	Fully connected layer	Number of neurons	1

Where, W_2 is the update gate weight matrix. The other variables are the same as the reset gate. The candidate hidden layer state equation is:

$$\tilde{h}_t = \tanh(W \cdot [r_t * h_{t-1}, x_t]). \quad (8)$$

Where: W is the candidate state weight matrix. r_t is the reset gate and h_{t-1} is the previous time step hidden information. Corresponding elements of these two matrices are multiplied to get $r_t * h_{t-1}$.

The final hidden state equation is:

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t. \quad (9)$$

Where, h_t is the output information.

4.3 Multiscale DenseNet-ResNet-GRU model

In this paper, a hybrid CNN-RNN tool wear prediction model is proposed. The convolutional portion of the model is DenseNet and ResNet. Each has two parallel channels. Therefore, there are four parallel channels. Single-scale convolutional kernels cannot extract richer features. Thus, every two parallel channels use convolution kernels of 3×1 and 5×1 to extract local information features of different scales from the input matrix. DenseNet enables feature reuse, and ResNet slows down network degradation. Therefore, the different advantages of these two networks can be used to extract different spatial features of samples. The concatenation layer is integrated and input to the deep GRU to extract sample sequence features.

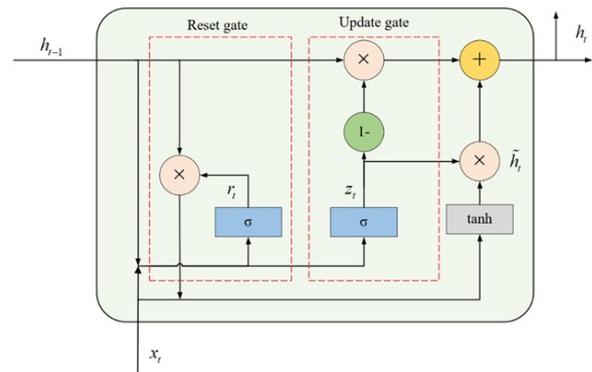


Fig. 8. GRU module structure.

After the deep GRU model, a fully connected layer and a regression layer are constructed to predict the amount of tool wear. The specific network structure is shown in Fig. 9.

Parameter settings of multiscale DenseNet-ResNet-GRU model. This is shown in Table 4.

4.4 Results analysis

The cutting force data is normalized and constructed as an input matrix to improve the model training effect. The normalization equation is as follows:

$$y = (x - x_{\min}) / (x_{\max} - x_{\min}). \quad (10)$$

Where, x is the data to be normalized. According to each set of experimental data, 78 milling records are included. Every

two milling records are selected as the testing set, and the rest serves as a training set. If 78 samples are numbered from 1-78, this paper selects 3, 6, 9, ... 72, 75, 78 as the testing set, as shown in Fig. 2. The model hyperparameter is set to learning rate $\eta = 0.001$, minibatch = 32, number of iterations = 100, and gradient threshold = 1. Finally, the Adam optimization algorithm is used for optimization. To verify the good performance of DenseNet-ResNet-GRU model, the CNN-GRU model is compared with it. The initialization parameters are the same as the multiscale DenseNet-ResNet-GRU model. The model training phase performs validation on a fixed period, useful for determining whether the model is overfitting. Therefore, the training loss and the corresponding validation results of RMSE need to be compared. During the training process, the network is validated every 20 iterations of both models, and the training results are shown in Fig. 10. After the iteration of the two models, the verification sets RMSE and Loss are shown in Table 5 below.

As shown in Fig. 10 and Table 5, when the number of iterations is 100, the model convergence trend is basically stable. At the beginning and the end of the iteration, both RMSE and Loss during the training process of the multiscale DenseNet-ResNet-GRU model are small. To observe the effect more obviously, the two models are analyzed in the same graph during training in Fig. 11.

Table 5. Validation effect of the validation set for both models.

Model	RMSE	Loss
CNN-GRU	0.0096	4.64×10^{-5}
Multiscale DenseNet-ResNet-GRU	0.0067	2.25×10^{-5}

As shown in Fig. 11, the multiscale DenseNet-ResNet-GRU model has a better convergence effect than the CNN-GRU model.

To examine the effectiveness of the proposed method further, the same dataset is fed into four other models: BPNN, ResNet, DenseNet and GRU. Fig. 12 shows the error prediction results of the actual and predicted values of the tool wear for the six models, and the calculation results of the four indicators are shown in Table 6.

As shown in Fig. 12 and Table 6, compared with the other five models, it is obvious that the output of the multi-scale DenseNet-ResNet-GRU model has the smallest fluctuation from the true value and better fitting effect. Machine learning, such as BPNN, may have excellent performance in some specific cases, but the newly added samples will affect the successfully learned network, and the generalization ability of the model is limited by parameter selection. Therefore, the error is larger. The MAE and RMSE of GRU are 5.5734 and

Table 6. Comparison of tool wear prediction results.

Model	Results			
	MAE	RMSE	MAPE	R ²
Multiscale DenseNet-ResNet-GRU	2.1635	2.3752	0.0265	0.9994
CNN-GRU	2.8646	3.2150	0.0328	0.9989
ResNet	3.8925	4.3522	0.0346	0.9785
DenseNet	4.1542	5.5216	0.0452	0.9617
BPNN	5.3856	5.8613	0.0562	0.9565
GRU	5.5734	6.2430	0.0698	0.9554

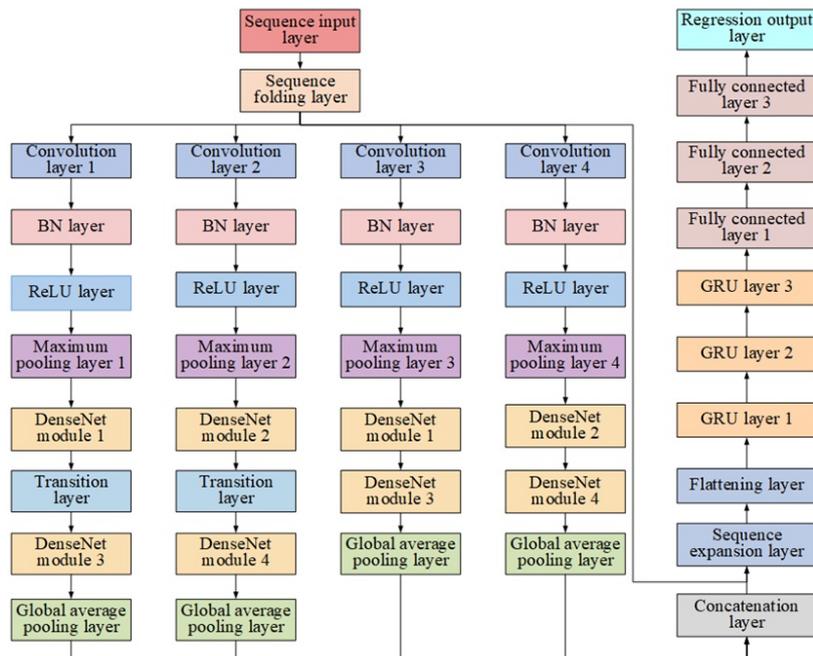


Fig. 9. Multiscale DenseNet-ResNet-GRU model network structure.

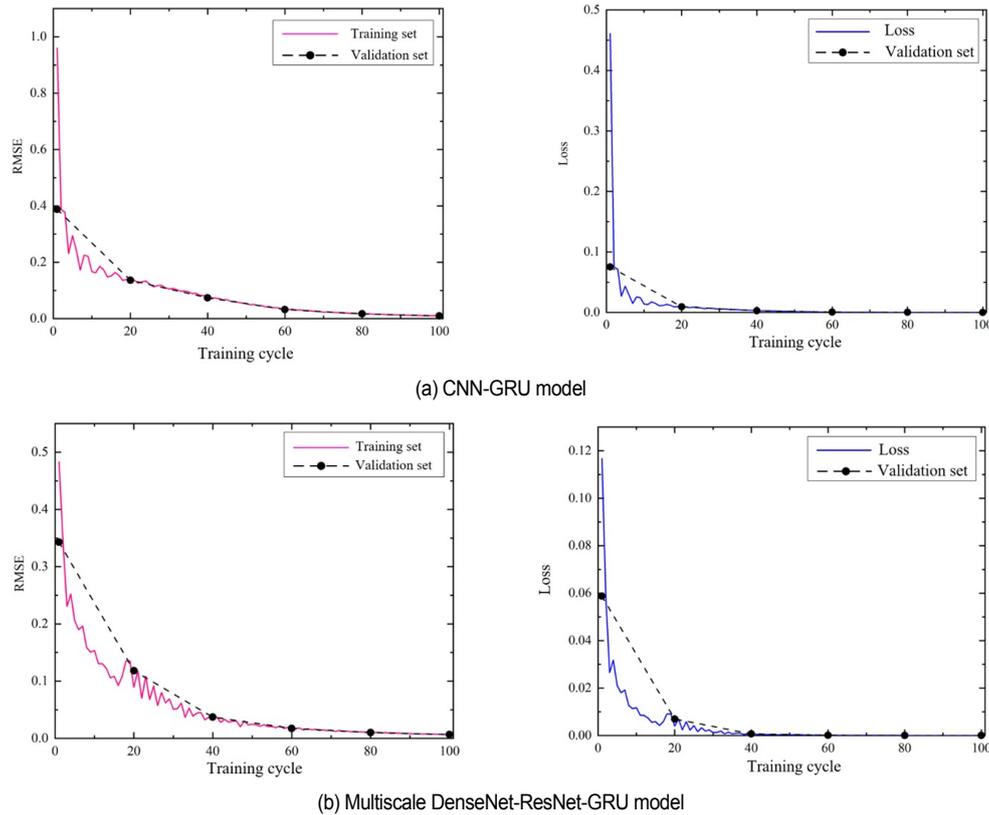


Fig. 10. Convergence curves of RMSE and loss function.

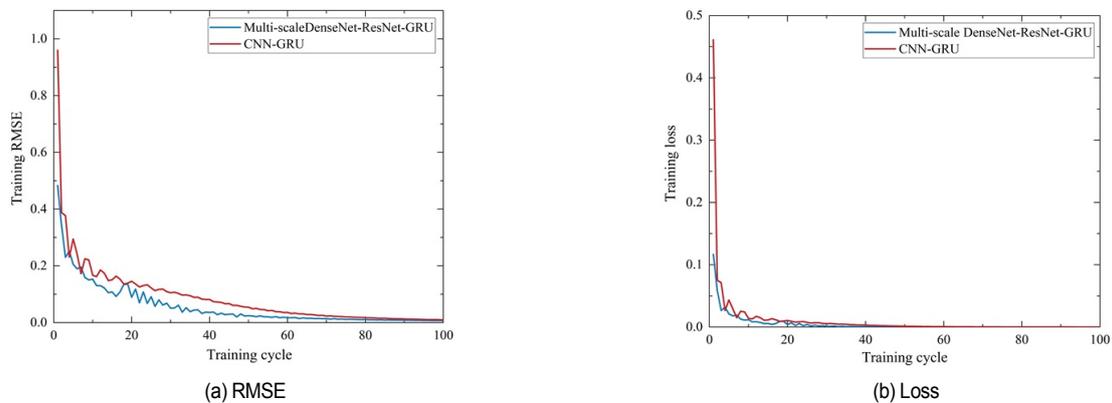


Fig. 11. Comparison of the impact of training.

6.2430, respectively. GRU does not work well without feature extraction. Because it is difficult to capture time-domain information from a large amount of data without considering multi-sensor fusion of spatial information. There is not much difference between DenseNet and ResNet in terms of MAE, RMSE, MAPE and R^2 in tool wear prediction. CNN-GRU can mine individual or fused signal features to reflect changes in tool wear. The mined features can be fed into the GRU to capture more timing features, thereby effectively reducing the error of tool wear. As a result, the performance of the CNN-GRU model has been improved. The fusion of DenseNet and

ResNet further solves the gradient degradation problem of CNN and alleviates the gradient disappearance. Therefore, the MAE, RMSE and MAPE of the multi-scale DenseNet-ResNet-GRU model are smaller than those of the other five models, and the R^2 is slightly larger than that of CNN-GRU. The results show that the method predicts a good fit. Therefore, it can be further verified that the multi-scale DenseNet-ResNet-GRU model is more superior in tool wear prediction.

In addition, compared with the other five models, the multi-scale DenseNet-ResNet-GRU model has a better prediction effect in the whole process of tool wear, but the error is slightly

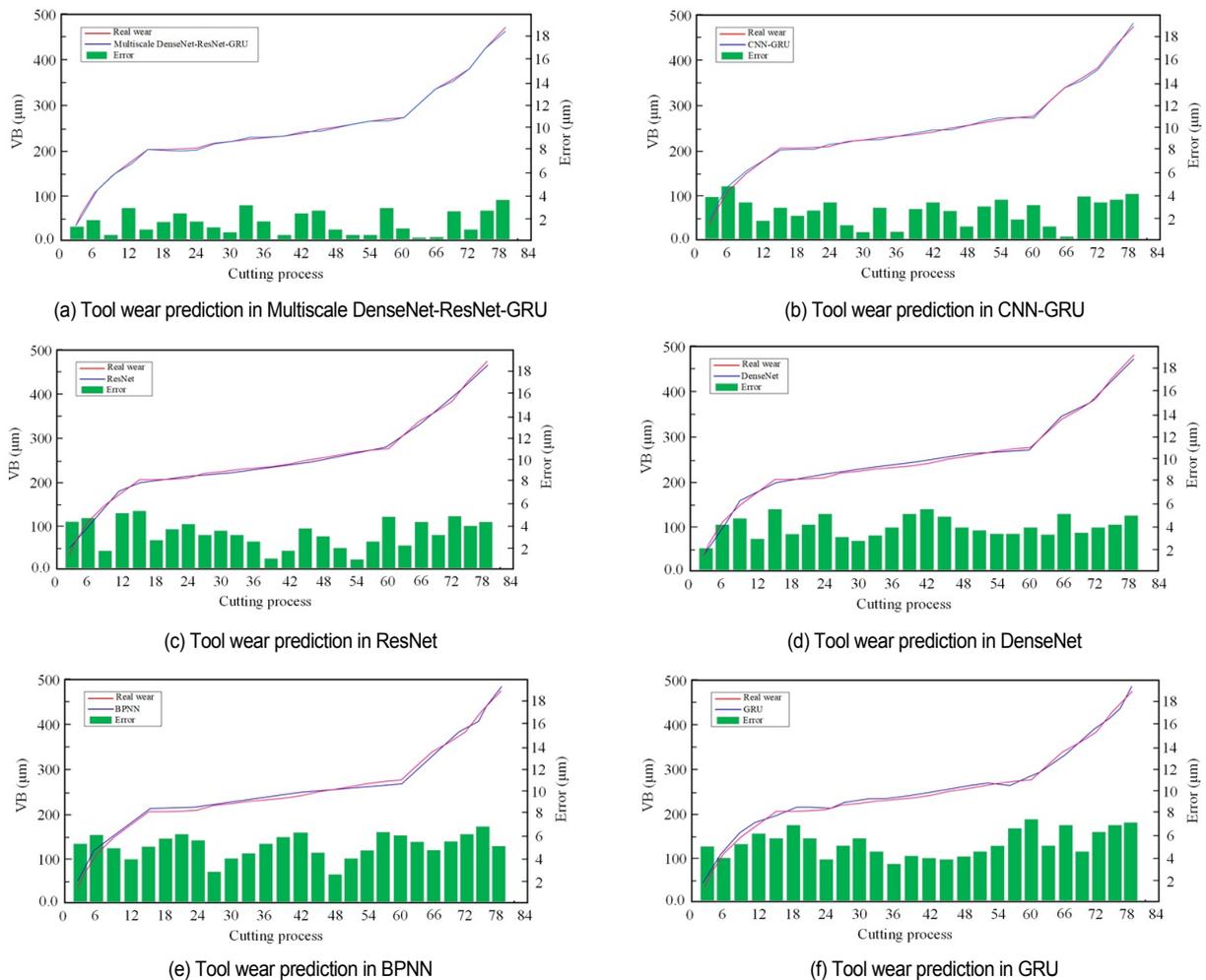


Fig. 12. Result of tool wear prediction.

larger in the later stage of tool severe wear. This is due to the severe wear stage of the tool, the amount of tool wear increases sharply, and the mechanical-thermal load between the tool and the workpiece produces violent vibration, which affects the prediction of the tool wear amount by using the cutting force signal to a certain extent. Therefore, how to reduce the error of the wear amount in the severe wear stage of the tool and predict the tool wear more accurately will be the direction of future research.

5. Conclusions

Based on the experiment of milling 508III steel material, this paper selected the cutting force signal to study milling tool wear prediction technology. To explore the mapping relationship between tool cutting force signal and tool wear, a multiscale DenseNet-ResNet-GRU wear prediction model is proposed, which achieves good results in predicting the wear of milling cutters. The conclusions are as follows.

1) The Laida criterion and wavelet threshold denoising is used to deal with outliers and noise in the cutting force signal. It

was determined that wavelet denoising was carried out by the *coif5* wavelet basis function, rigsure threshold, and progressive semi-soft threshold function, which achieved good results. It is conducive to improving tool wear prediction accuracy.

2) A multiscale DenseNet-ResNet-GRU tool wear prediction model was established. The 3×1 and 5×1 convolution kernels were used to convolve the input cutting force matrix at different scales. The sample space features were extracted based on the different advantages of DenseNet and ResNet. Sample sequence features are extracted by integration through the concatenation layer and input into the deep GRU. After the deep GRU model, a fully connected layer and a regression layer are constructed to predict the amount of tool wear.

3) The multiscale DenseNet-ResNet-GRU tool wear prediction model was used to predict the amount of tool wear. Compared with the CNN-GRU, BPNN, ResNet, DenseNet and GRU model, it is concluded that the RMSE, MAE, and MAPE values are the smallest. It is further verified that the Multiscale DenseNet-ResNet-GRU model is superior in tool wear prediction.

The tool wear prediction method in the milling process based

on multiscale DenseNet-ResNet-GRU has good effect on tool wear prediction, which can provide more accurate and effective information for the prediction of tool wear in the actual machining process and can also replace the tool in time. It provides reference and technical support for improving the accuracy of tool wear prediction, and is expected to be effectively applied and promoted in actual machining. Dynamometer sensor is placed on the machine table and the fixture is mounted above the sensor in this paper. If it is considered to be applied to the actual production, especially in the 5-axis machining center, the cable of the dynamometer can be damaged when the workpiece is moved and rotated. It will cause adverse consequences and safety problems. Therefore, it is necessary to use the wireless rotating cutting force dynamometers (RCD) to be directly mounted on the spindle of the machine tool through the spindle adapter according to the actual situation. It is suitable for five-axis machining, large-size, complex parts processing, and high-dynamic precision cutting applications.

Acknowledgments

The authors would like to acknowledge National Natural Science Foundation of China (No. 52175394) and the Joint Guidance Project of Heilongjiang Provincial Natural Science Foundation (No. LH2021E083) in the production of this work. The authors are grateful to the anonymous reviewers for valuable comments and suggestions, which helped to improve this study.

References

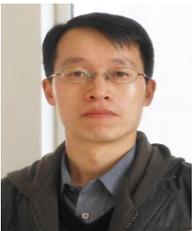
- [1] X. L. Liu, X. B. Li, M. N. Ding, C. X. Yue, L. H. Wang, Y. S. Liang and B. W. Zhang, Intelligent management and control technology of cutting tool life-cycle for intelligent manufacturing, *Journal of Mechanical Engineering*, 57 (10) (2021) 196-219.
- [2] G. Serin, B. Sener, A. M. Ozbayoglu and H. O. Unver, Review of tool condition monitoring in machining and opportunities for deep learning, *The International Journal of Advanced Manufacturing Technology*, 109 (3-4) (2020) 953-974.
- [3] Y. N. Cheng, Q. Y. Lv, C. Wang, C. Y. Li and Q. H. Yuan, Experimental study and simulation analysis of crack propagation of heavy-duty milling cemented carbide tool material, *Ferroelectrics*, 565 (1) (2020) 148-163.
- [4] Y. N. Cheng, R. Guan, Z. Z. Lu, M. Xu and Y. Z. Liu, A study on the milling temperature and tool wear of difficult-to-machine 508III steel, *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 232 (14) (2018) 2478-2487.
- [5] M. Ma, C. Sun, X. F. M. W. Zhang and R. Q. Yan, A deep coupled network for health state assessment of cutting tools based on fusion of multisensory signals, *IEEE Transactions on Industrial Informatics*, 15 (12) (2019) 6415-6424.
- [6] J. Duan, Research on tool wear condition recognition and prediction based on deep learning models, *Ph.D. Thesis*, Huazhong University of Science and Technology, China (2021) 45-60.
- [7] X. H. Mao, N. He and L. Li, Studies on tool wear monitoring based on cutting force, *Materials Science Forum*, 697-698 (2012) 268-272.
- [8] M. Aramesh, M. H. Attia, H. A. Kishawy and M. Balazinski, Estimating the remaining useful tool life of worn tools under different cutting parameters: a survival life analysis during turning of titanium metal matrix composites (Ti-MMCs), *CIRP Journal of Manufacturing Science and Technology*, 12 (2016) 35-43.
- [9] A. Siddhpura and R. Paurobally, A review of flank wear prediction methods for tool condition monitoring in a turning process, *The International Journal of Advanced Manufacturing Technology*, 65 (1-4) (2013) 371-393.
- [10] N. Dhobale, S. Mulik, R. Jegadeeshwaran and A. Patange, Supervision of milling tool inserts using conventional and artificial intelligence approach: a review, *Sound and Vibration*, 55 (2) (2021) 87-116.
- [11] M. H. Cheng, J. Li, P. Yan, H. S. Jiang, R. B. Wang, T. Y. Qiu and X. B. Wang, Intelligent tool wear monitoring and multi-step prediction based on deep learning model, *Journal of Manufacturing Systems*, 62 (2022) 286-300.
- [12] H. Y. Liu, S. Zhang, J. F. Li and X. N. Luan, Tool wear detection based on CNN-BiLSTM model, *China Mechanical Engineering*, 33 (16) (2022) 1940-1947.
- [13] Y. T. Li, Q. S. Xie, H. S. Huang and Q. P. Chen, Research on a tool wear monitoring algorithm based on residual dense network, *Symmetry*, 11 (6) (2019) 1-19.
- [14] W. X. Xu, H. H. Miao, Z. B. Zhao, J. X. Liu, C. Sun and R. Q. Yan, Multiscale convolutional gated recurrent unit networks for tool wear prediction in smart manufacturing, *Chinese Journal of Mechanical Engineering*, 34 (3) (2021) 143-158.
- [15] J. J. Wang, J. X. Yan, C. Li, R. X. Gao and R. Zhao, Deep heterogeneous GRU model for predictive analytics in smart manufacturing: application to tool wear prediction, *Computers in Industry*, 111 (2019) 1-14.
- [16] R. Zhao, R. Q. Yan, J. J. Wang and K. Z. Mao, Learning to monitor machine health with convolutional bi-directional LSTM networks, *Sensors*, 17 (2) (2017) 1-18.
- [17] J. H. Zhou, C. K. Pang, Z. Zhong and L. L. Frank, Tool wear monitoring using acoustic emissions by dominant-feature identification, *IEEE Trans. Instrumentation and Measurement*, 60 (2) (2011) 547-559.
- [18] C. M. Shi, G. Panoutsos, B. Luo, H. Q. Liu, B. Li and X. Lin, Using multiple-feature-spaces-based deep learning for tool condition monitoring in ultraprecision manufacturing, *IEEE Trans. Industrial Electronics*, 66 (5) (2019) 3794-3803.
- [19] D. C. Luo, Research on tool wear mechanism and prediction model of milling TC18 Titanium alloy based on deep learning, *Master's Thesis*, Guangxi University, China (2021) 12-18.
- [20] B. J. Hao, Research on tool condition monitoring technology in drilling, *Master's Thesis*, Nanjing University of Aeronautics and Astronautics, China (2019) 20-23.
- [21] P. He, R. B. He, J. Chen, X. Yang, J. Tang and S. Y. Zhou, Review on partial discharge denoising method of white noise based on wavelet thresholding, *Guangdong Electric Power*, 33 (11) (2020) 83-90.
- [22] F. M. Bayer, A. J. Kozakevicius and R. J. Cintra, An iterative

wavelet threshold for signal denoising, *Signal Processing*, 162 (2019) 10-20.

- [23] P. Lu, Z. Zhuo, W. H. Zhang, J. Tang, H. L. Tang and J. Q. Lu, Accuracy improvement of quantitative LIBS analysis of coal properties using a hybrid model based on a wavelet threshold de-noising and feature selection method, *Applied Optics*, 59 (22) (2020) 6443-6451.
- [24] Z. W. Huang, J. M. Zhu, J. T. Lei, X. R. Li and F. Q. Tian, Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations, *Journal of Intelligent Manufacturing*, 31 (4) (2020) 1-14.



Rui Guan is a doctoral student of Harbin University of Science and Technology. She is also a teacher in Harbin Vocational and Technical College. Her main research focuses on intelligent monitoring technology for cutting tool wear or breakage in cutting difficult-to-machine materials, metal cutting principles and tools.



Yaonan Cheng received his Ph.D. from Harbin University of Science and Technology, China. He is currently a Professor at College of Mechanical and Power Engineering, Harbin University of Science and Technology. His current research focuses on metal cutting theory and tool technology, intelligent manufacturing technology and efficient machining technology for difficult-to-machine materials.



Yingbo Jin is a master student of Harbin University of Science and Technology. His main research focuses on intelligent monitoring technology for difficult-to-machine materials, metal cutting principles and tools.



Shilong Zhou is a master student of Harbin University of Science and Technology, Harbin, China. His current research focuses on intelligent monitoring technology for difficult-to-machine materials, metal cutting principles and tools.



Xiaoyu Gai is a doctoral student of Harbin University of Science and Technology, Harbin, China. His current research focuses on intelligent monitoring technology for difficult-to-machine materials, metal cutting principles and tools.



Mengda Lu is a master student of Harbin University of Science and Technology. His main research focuses on intelligent monitoring technology for difficult-to-machine materials, metal cutting principles and tools.