Autonomous Intelligent Systems

ORIGINAL ARTICLE Open Access

An intelligent surface roughness prediction method based on automatic feature extraction and adaptive data fusion

Xun Zhang[1,2], Sibao Wang[1,2*], Fangrui Gao[1,2], Hao Wang[1,2], Haoyu Wu[1,2] and Ying Liu[3]

Abstract

Machining quality prediction based on cutting big data is the core focus of current developments in intelligent manufacturing. Presently, predictions of machining quality primarily rely on process and signal analyses. Process-based predictions are generally constrained to the development of rudimentary regression models. Signal-based predictions often require large amounts of data, multiple processing steps (such as noise reduction, principal component analysis, modulation, etc.), and have low prediction efficiency. In addition, the accuracy of the model depends on tedious manual parameter tuning. This paper proposes a convolutional neural network quality intelligent prediction model based on automatic feature extraction and adaptive data fusion (CNN-AFEADF). Firstly, by processing signals from multiple directions, time-frequency domain images with rich features can be obtained, which significantly benefit neural network learning. Secondly, the corresponding images in three directions are fused into one image by setting different fusion weight parameters. The optimal fusion weight parameters and window length are determined by the Particle Swarm Optimization algorithm (PSO). This data fusion method reduces training time by 16.74 times. Finally, the proposed method is verified by various experiments. This method can automatically identify sensitive data features through neural network fitting experiments and optimization, thereby eliminating the need for expert experience in determining the significance of data features. Based on this approach, the model achieves an average relative error of 2.95%, reducing the prediction error compared to traditional models. Furthermore, this method enhances the intelligent machining level.

Keywords: Surface roughness prediction, Short-time Fourier transform, Feature extraction, Data fusion, Convolutional neural network

1 Introduction

The manufacturing industry is demanding increasingly higher processing quality standards [1]. Traditional quality control methods, which primarily rely on experience and post-processing inspections, are insufficient for providing early warnings of quality issues. With the rapid development of intelligent manufacturing technology [2], the advent of big data has provided new ideas and methods for processing quality prediction [3]. Through the analysis and mining of massive data [4], real-time monitoring and quality prediction of the processing process can be achieved [5, 6]. This advancement enhances the automation [7] and intelligence [8] of the production process, improves product quality [9], and reduces production costs [10].

Surface roughness is a critical quality attribute of machine tool products [11]. Surface roughness prediction can be categorized into regression prediction [12, 13] and neural network prediction [14]. Conventional regression models have limited capability to process high-dimensional big data, and the typically small data volumes make them

*Correspondence: wangsibaocqu@cqu.edu.cn
[1]State Key Laboratory of Mechanical Transmission for Advanced Equipments, Chongqing University, Chongqing 400044, China
[2]College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China
Full list of author information is available at the end of the article

highly sensitive to noise during the manufacturing process. Neural networks have demonstrated significant advantages in handling nonlinear relationships and large-scale complex data [15, 16]. It can extract features and patterns in data analysis and is widely used to predict unknown data [17].

Studies [18–20] indicate that feature extraction and data fusion [21] of original data significantly affect the intelligence, train time, and prediction error of data-driven predictions. Regarding feature extraction, Wu et al. [22] performed feature extraction on data using autoencoders. Zeng et al. [23] applied convolution and pooling techniques to extract features from multi-source signals. Wang et al. [24] utilized convolutional neural networks and PCA feature recognition to extract features from multi-source signals. The aforementioned research methods employ autoencoders, convolution operations, and pooling operations in neural networks to extract features from large volumes of original data. These approaches, however, significantly increase the training time of the neural network model. Additionally, these feature extraction methods rely on human expertise and judgment, impeding the realization of rapid automatic feature extraction. Regarding data fusion, He et al. [25] selected sensitive features as input based on feature extraction from multi-sensor signals. Huang et al. [26] designed a feature matrix data fusion as the input of the neural network. Zhang et al. [27] performed feature extraction on multi-sensor signals by designing sliding windows. However, the aforementioned method does not account for the impact of different feature data on variations in neural network training and fitting. The fusion method essentially extracts features from the data and uses the entire dataset as feature input. Although this approach can improve prediction accuracy [28, 29], it also increases the model's training time. Existing data fusion methods fail to achieve the dual objectives of reducing data volume and enhancing data quality simultaneously. In addressing the aforementioned challenges, Maintaining a low prediction error while automatically extracting rich features and reducing data volume presents significant research value.

This paper uses five-axis milling of thin-walled parts as an example to study the aforementioned challenges. A feature extraction layer based on the time-frequency analysis method is established, employing the STFT for automatic feature extraction from signals collected in multiple directions. The time-frequency domain processing method converts the original signal into a time-frequency image, with the STFT window length parameter determined by Particle Swarm Optimization (PSO) [30]. Time-frequency domain images encapsulate rich feature and texture information, empowering convolutional neural networks (CNN) to comprehensively learn from the data. This method circumvents the necessity for collecting large volumes and various types of data, eliminates the need for signal denoising, and simplifies the feature extraction process. For the data fusion layer, the image data in three directions are weighted and fused by setting different fusion weights, which are also determined through PSO. This fusion method reduces the data volume while retaining time-frequency domain features from multiple directions.

2 Predictive model framework and parameter optimization method

Vibration is the key influencing factor on surface roughness during machine tool processing [28]. During the processing of thin-walled aviation blades by a five-axis CNC milling machine, vibration signals are collected by vibration sensors. The steps of CNN-AFEADF, including vibration signal feature extraction, data fusion, and parameter optimization, are shown in Fig. 1.

2.1 Overall framework of the model

The intelligent prediction method of surface roughness is composed of Input layer, Feature Extraction layer, Data Fusion layer, Convolution layer, Pooling layer, Fully Connected layer, and Output layer.

Input layer: the input layer receives the collected time-domain signals, which are then forwarded to the Feature Extraction Layer.

Feature Extraction layer: the signal generated by the five-axis milling of thin-walled aircraft blades is the unsteady. Hilbert-Huang Transform (HHT), Wavelet Transform (WT), and STFT are the methods used to process the unsteady signal to get the feature map. In Sect. 4.2 of this paper, a comparative analysis of the time-frequency image data sets obtained by the three methods will be conducted. The following is a brief introduction to the theory of the three methods.

HHT comprises Empirical Mode Decomposition (EMD) and Hilbert Transform (HT). Ultimately, all the acquired IMF (Intrinsic Mode Function) components undergo HT. The formula is given by Eq. (1).

$$H[h_i(t)] = \frac{1}{\pi}\int_{-\infty}^{+\infty}\frac{h_i(\tau)}{t-\tau}d\tau. \tag{1}$$

WT uses wavelet functions to decompose signals, which have localized characteristics. The mathematical representation of WT is given by Eq. (2).

$$CWT(a,b) = \frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}x(t)\varphi^*\left(\frac{t-b}{a}\right)dt. \tag{2}$$

STFT is a Fourier transform with a fixed window length. The process of STFT is to multiply the signal by an analysis window function $h(t)$ before the Fourier transform. The mathematical representation of STFT is given by Eq. (3).

$$STFT(t,f) = \int_{-\infty}^{+\infty}x(\tau)h(\tau-t)\mathrm{e}^{-\mathrm{j}2\pi f\tau}d\tau. \tag{3}$$

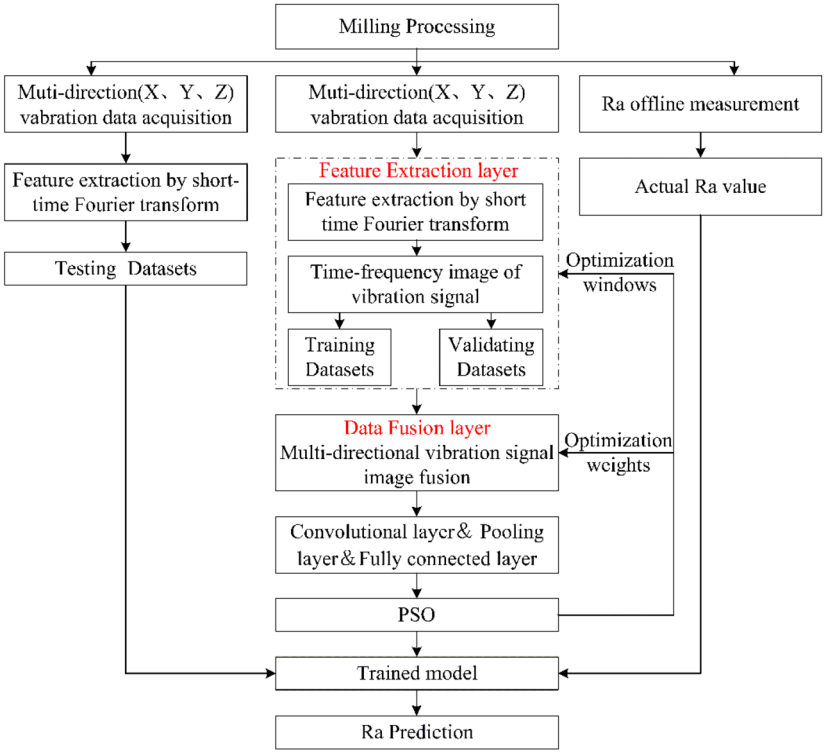


Figure 1 Modeling and experimental process

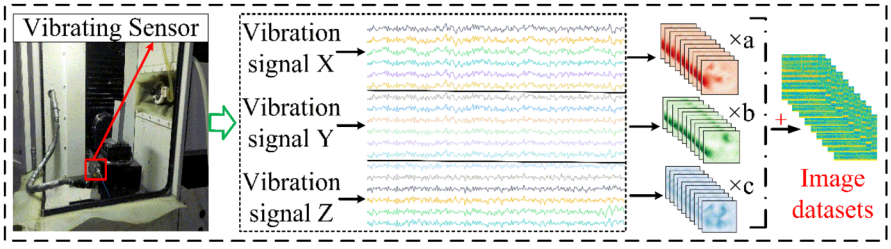


Figure 2 Image fusion process

Data Fusion layer: the fusion layer serves to decrease the data dimensionality and enhance the feature richness of individual image data. Firstly, the time-frequency images in the three directions of X, Y, and Z are obtained in the feature layer, and the corresponding time-frequency images in the three directions are fused. Then, the fused image is processed by the Lanczos interpolation algorithm in order to reduce the image size. To set the initial weight, such as Eq. (4), the sum of the three weights a, b, and c is equal to 1 to ensure that the pixel value of the weighted pixel matrix does not exceed 255. The fusion process is shown in Fig. 2, and the weight parameters are optimized through the PSO. The fusion layer employs RGB images as input, offering richer feature information than traditional grayscale pixel matrices[29]. The STFT converts vibration signals into RGB time-frequency domain images, which simultaneously capture time, frequency, and intensity information. This multi-dimensional data is essential for accurately predicting surface roughness, as roughness is directly correlated with vibration. Consequently, RGB time-frequency domain images provide a more comprehensive reflection of these characteristics.

$$\begin{cases} IF = aX + bY + cZ, \\ a + b + c = 1. \end{cases} \tag{4}$$

Convolution layer: convolutional layers are used to extract information from the input image. The size of the im-

age is assumed to be $W \times H$, and the size of the convolution kernel is assumed to be $F \times F$, stride = S, and Padding = P. The size of the output matrix after passing through the convolution layer is shown in Eq. (5) and Eq. (6).

$$\mathrm{conv}W_{output} = \frac{convW_{input} - F + 2P}{S} + 1, \tag{5}$$

$$\mathrm{convH}_{output} = \frac{convH_{input} - F + 2P}{S} + 1. \tag{6}$$

In the process of forward propagation, the convolutional layer performs a convolution operation between the feature filter and the input local region, followed by an activation function. The output of the convolutional layer can be regarded as the feature map obtained by extracting features from the feature matrix. The forward propagation process of the convolution layer is shown as Eq. (7), and the backpropagation equation is shown as Eq. (8). The specific details of the equation can be found in the literature[31].

$$X_j^l = f(\sum_{i \in M_j} (X_j^{l-1} \times K_{ij}^l) + b_j^l), \tag{7}$$

$$\frac{\partial L}{\partial K_{ij}^l} = \sum_{u,v} ((\delta_j^l)_{u,v} (\mathrm{P}_j^{l-1})_{u,v}), \quad \frac{\partial L}{\partial \beta_j^l} = \sum_{u,v} (\delta_j^l)_{u,v}. \tag{8}$$

Pooling layer: after the convolution calculation, the dimensionality of the parameters in the pooling layer remains high. The function of the pooling layer is to reduce the dimensionality of the features obtained by the convolution layer. The pooling layer is calculated by sliding a matrix window of $n \times n$ size. Through the pooling operation, image features that are not affected by position are selected, the feature dimension is reduced, the amount of calculation is reduced and the receptive field of subsequent features is increased. The forward propagation calculation of the pooling layer is as shown in Eq. (9), and the backpropagation calculation of the pooling layer is as shown in Eq. (10).

$$X_j^l = f(\beta_j^l \bullet \mathrm{down}(X_j^{l-1}) + b_j^l), \tag{9}$$

$$\frac{\partial L}{\partial \beta_j^l} = \sum_{u,v} (\delta_j^l \mathrm{down}(X_j^{l-1}))_{u,v}, \quad \frac{\partial L}{\partial \beta_j^l} = \sum_{u,v} (\delta_j^l)_{u,v}. \tag{10}$$

Fully connected layer: the fully connected layer converts the feature matrix of the pooling layer into a one-dimensional feature vector. High-level features extracted from convolutional and pooling layers are integrated and mapped. The neurons in the fully connected layer are connected to the neurons in the previous layer, and the output calculation process is as Eq. (11). The parameters of the fully connected layer are also updated through the backpropagation algorithm, and the calculation process is shown in Eq. (12). To reduce the problem of over-fitting and improve the generalization ability of the model, regularization (Dropout) is introduced in the fully connected layer.

$$X^l = \sigma(W^l X^{l-1} + b^l), \tag{11}$$

$$\frac{\partial L}{\partial W_j^l} = \delta_j^l (X^{l-1})^{\mathrm{T}}, \quad \frac{\partial L}{\partial b^l} = \delta^l. \tag{12}$$

Output layer: the function of the output layer is to convert the feature mapping learned by the network into specific numerical output. The goal of the regression prediction task is to predict continuous numerical values.

2.2 Model parameter optimization

In the field of mechanical intelligence diagnosis and prediction, many methods use the sigmoid function as the activation function of CNN. However, the sigmoid function has the defects of gradient disappearance and slow convergence in the training process. The direct linear unit (ReLU) function can effectively overcome these shortcomings. Therefore, the ReLU function is used as an activation function for the convolutional layer, the pooled layer, and the fully connected layer.

Intelligent prediction models need to adjust model parameters during the training process so that model training can converge faster and more effectively to obtain the optimal solution. The Adam [32] optimizer combines the idea of the momentum method of Gradient Descent and the idea of the adaptive learning rate of RMSprop.

However, the Adam optimizer can only optimize the internal parameters of the intelligent prediction model. The external parameters that are set cannot be optimized. Therefore, the PSO algorithm is proposed as an external parameter optimization method for intelligent prediction models. The Adam optimizer is combined with the PSO algorithm to form internal and external optimization of the model.

PSO [30] is a swarm-based optimization algorithm inspired by nature. PSO initializes a group of particles with random positions and velocities that fall within the search space. During the iteration, each particle updates its velocity and position according to the following Eqs. (13) and (14):

$$v_i = \omega v_i + c_1 r_1 (p_i - x_i) + c_2 r_2 (g_i - x_i), \tag{13}$$

$$x_i = x_i + v_i. \tag{14}$$

The number of particles is set to 20, the dimension is set to 4, and the number of iterations is set to 500. A "4"-dimensional array of random initial positions is created, which represents three weight values and window length values respectively. The weight elements in the initial position are uniformly distributed random numbers between

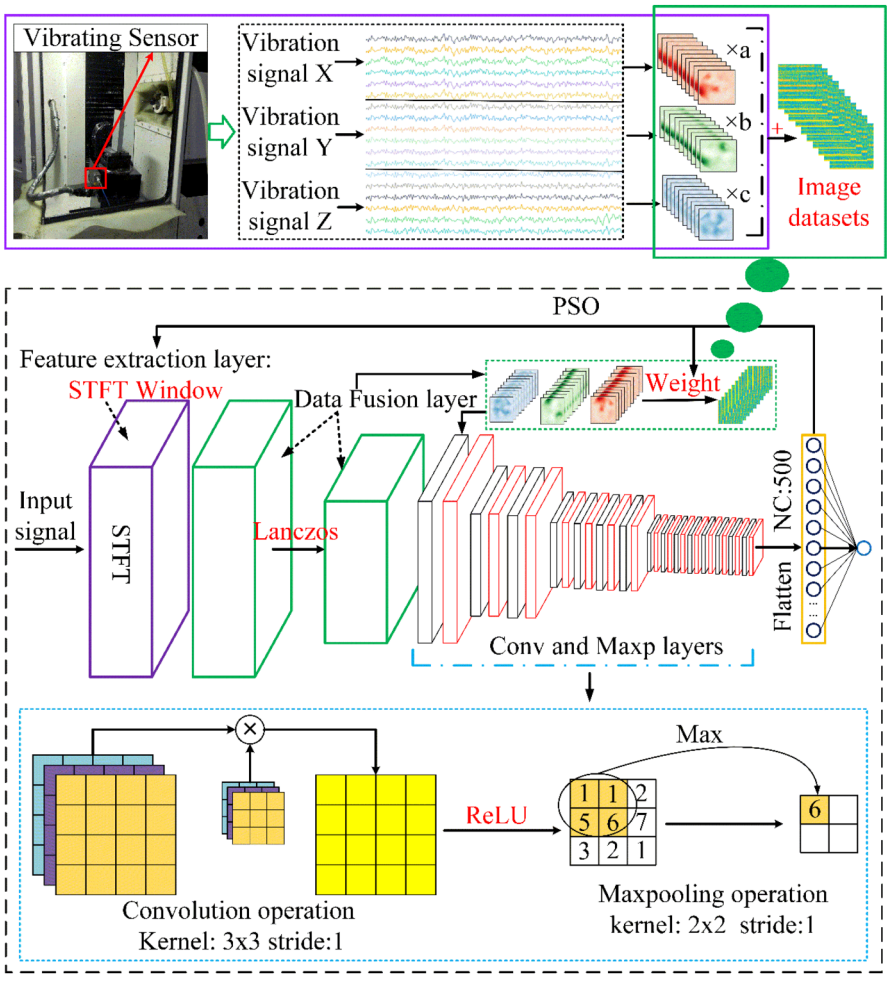


Figure 3 CNN-AFEADF model

the intervals [0, 1). The window length value parameter is also defined.

2.3 Model configuration

The input data for the input layer comprises the collected time-domain signal. The original time domain signal is sent to the feature layer through the input layer. In the feature extraction layer, a non-stationary signal processing method is used to extract the time domain signal into a time-frequency feature image. After completing the feature extraction, all the obtained images in the X, Y, and Z directions are sent to the fusion layer, and the X_i, Y_i, Z_i images are fused correspondingly. The input image size is the original image size of 540 × 570 × 3.

After completing the image fusion, the Lanczos interpolation algorithm is used in the fusion layer. The image is resized to 227 × 227 × 3. This operation can further reduce the computational complexity of the convolution neural network and reduce the model fitting time. After completing feature extraction and data fusion, four convolutional layers and four pooling layers are alternately connected. The convolution kernel and pooling kernel sizes are set to 3 × 3 and 2 × 2 respectively, and the stride is set to 1. The fully connected layer has 500 neurons and Dropout is set to 0.3. The network framework details and parameter settings are shown in Fig. 3.

3 Experimental configuration

Blades are used as milling workpieces, and the blade material is aluminum alloy. The blank size is 40 × 15 × 90 mm^3. When finishing, choose a half-price 1 mm TiSiN-coated ball cutter, and the finishing depth of the cut is 0.2 mm. As shown in Fig. 4, the vibration sensor is installed on the workbench of the five-axis milling machine. Vibration

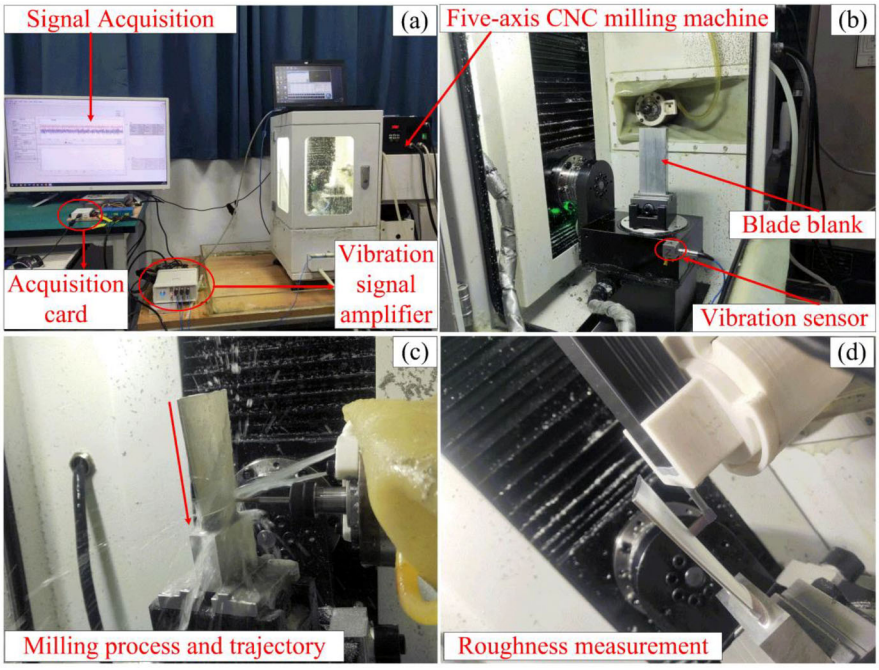


Figure 4 Experimental process and vibration singal collection

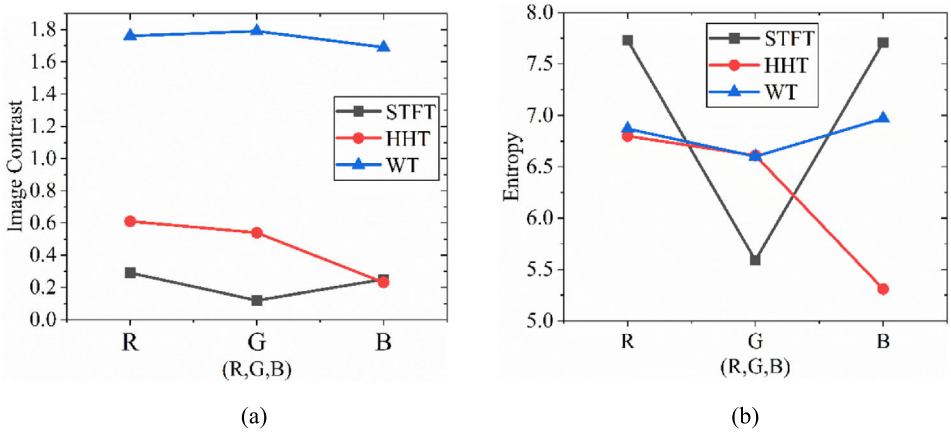


Figure 5 (a) Contrast results. (b) The result of entropy

sensors convert vibration signals from acceleration signals into electrical signals. The vibration signal is amplified 10 times by the amplifier. Then the capture card is connected and the sampling frequency is set to 5 kHz. The sampled data is transferred to the computer host. The blade machining process and tool path are shown in Fig. 4(c). The blade is segmented into 48 regions from top to bottom. In each region, three roughness measurements are taken, and the average value is considered as the roughness for that region. The roughness measurement procedure is depicted in Fig. 4(d).

4 Results and discussion

4.1 Feature extraction and data fusion

Various methods are employed to process non-stationary signals, including HHT, WT, and STFT. Each of these methods can generate time-frequency images. The image is processed through the Gabor filter. The contrast and entropy of the image are analyzed. The results are shown in Fig. 5. Contrast describes the degree of difference between different areas in an image. High contrast values represent significant differences in different areas of the image. In the field of target detection, higher contrast is often required.

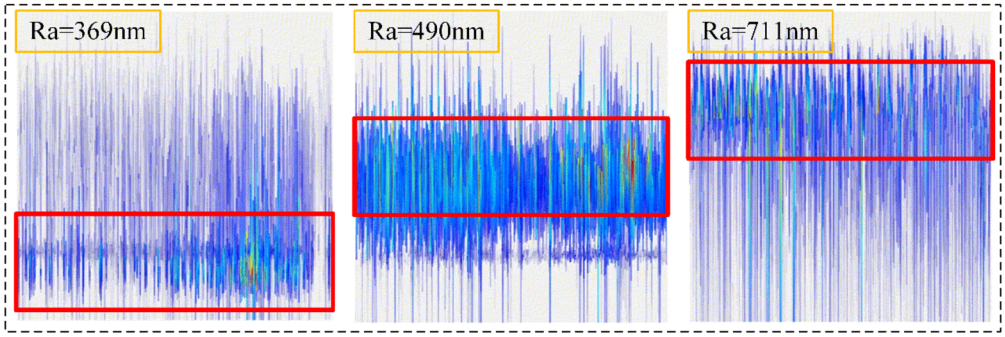


Figure 6 HHT time-frequency image

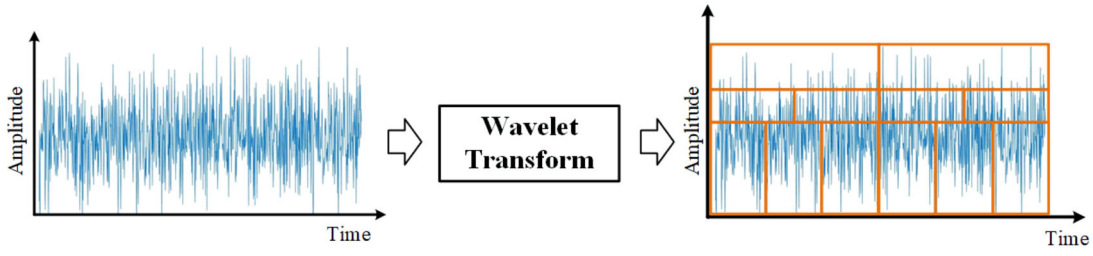


Figure 7 WT principle diagram

In image regression prediction, higher contrast appears as feature clutter. In contrast, lower contrast has stable characteristics. Entropy represents the amount of information in an image. The higher the entropy, the richer the texture information of the image. For the three types of images, it is necessary to obtain images with low contrast and high entropy. That is, images with regular features and rich feature information are selected as data sets. As shown in Fig. 5 (a), STFT exhibits superior contrast compared to both the WT and HHT.

As shown in Fig. 5(b), the entropy values of the R channel and B channel of the STFT are markedly higher than those of the WT and HHT. The value of HHT R channel entropy is 6.8, and the value of G channel entropy is 6.61. The value of WT R channel entropy is 6.87, and the value of G channel entropy is 6.6. The values of the R channel and G channel are almost equal. The value of the B channel entropy of the WTis 6.97. The value of HHT B channel entropy is 5.31. The B-channel entropy value of the WT is larger than the HHT. Therefore, the WTis better than the HHT. Based on comprehensive contrast analysis, the characteristic image obtained through the STFT proves to be the most optimal among the three methods.

The HHT is to perform EMD on the original signal and then perform HT to obtain the frequency domain diagram. The EMD process entails identifying the upper and lower envelopes of the signal and subsequently calculating their mean. These envelopes are determined by the local maxima and minima of the signal. Consequently, to achieve a coherent frequency domain representation, there are stringent requirements for the regularity and noise-free characteristics of the original signal throughout the processing. However, the processing process is affected by the machine tool status, processing environment, etc., and combined with image change analysis, the image obtained by the HHT changes irregularly, as shown in Fig. 6.

WT processing of time domain signal is shown in Fig. 7. The WT is capable of generating time-frequency images, which exhibit relatively regular changes overall, as shown in Fig. 8. This transform decomposes the signal using wavelet functions, resulting in a continuously varying window length that imparts localized characteristics, as shown in Fig. 9. However, the presence of excessive detail in the overall image can hinder the model's learning process described in this paper.

The STFT is a Fourier transform with a fixed window length, as shown in Fig. 10. As the roughness increases, the high-frequency features of the image increase and change regularly, as shown in Fig. 11. At the same time, the frequency features are continuous, and there is no continuous frequency change in the HHT and the WT. Images with the same set of surface roughness have stable features, as shown in Fig. 12.

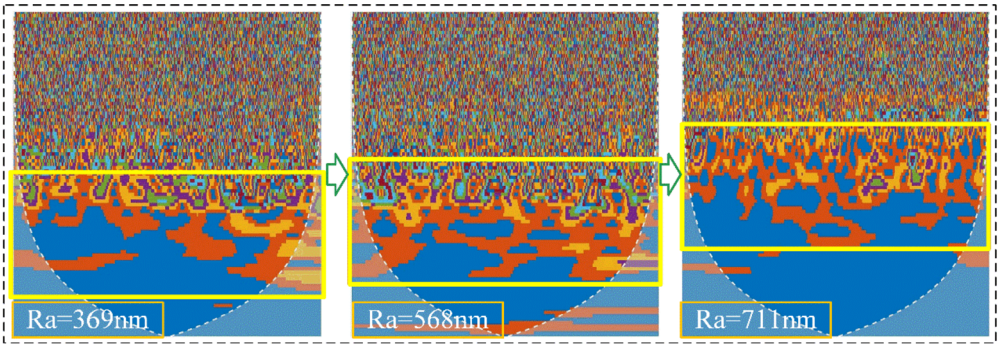


Figure 8 WT images with different surface roughness

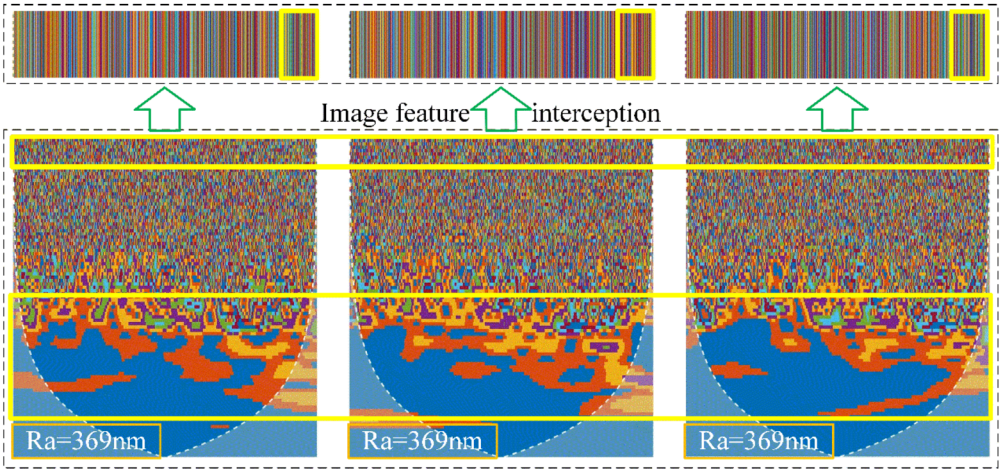


Figure 9 WT image with the same surface roughness

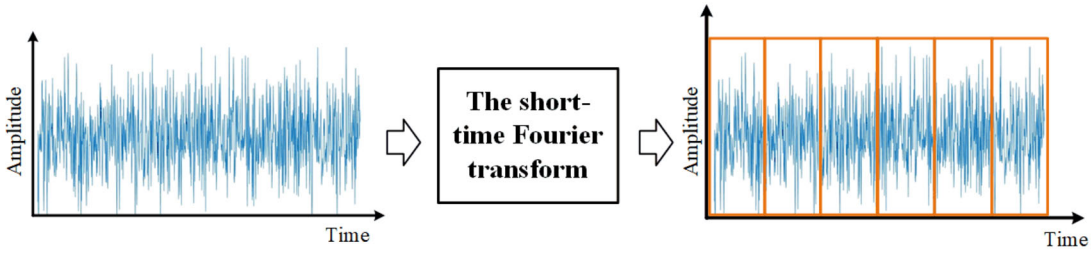


Figure 10 STFT principle diagram

To analyze the advantages and disadvantages of different time-frequency domain feature extraction methods, the training set is utilized as the model test input. The prediction curve is shown in Fig. 13. As shown in Figs. 11(a), (c), and (e), the predicted values of the model trained with the STFT dataset are closer to the true values. The prediction stability of the data set obtained by the STFT is better than that of the data set obtained by the HHT and the data set obtained by the WT. Additionally, the lowest accuracy in the three directions using HHT is 74%, for WT it is 75%, and for STFT it is 89%. Therefore, the model trained with the STFT dataset consistently demonstrates higher accuracy. Based on the above analysis, the time-frequency domain signal processing method can achieve automatic fea-

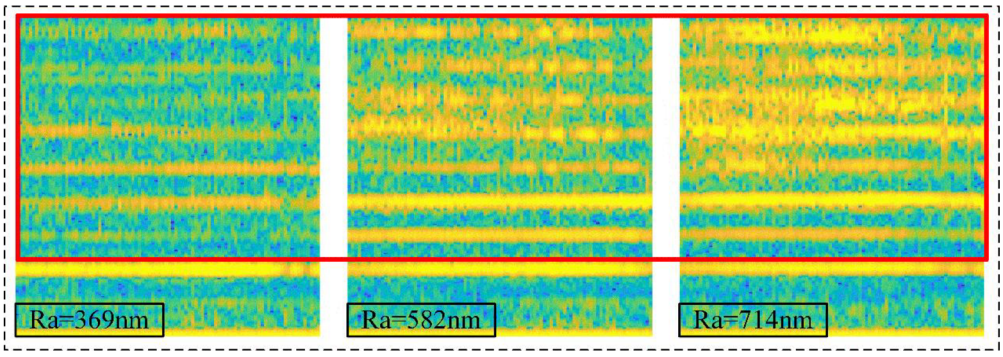


Figure 11 STFT images with different surface roughness

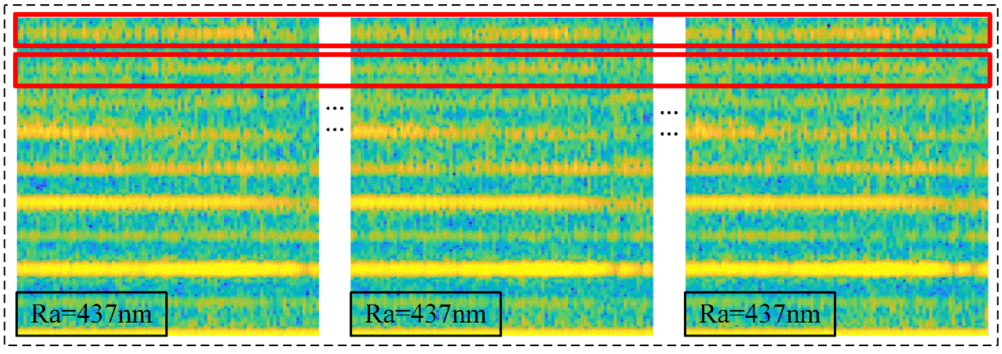


Figure 12 STFT image with the same surface roughness

ture extraction while maintaining lower prediction error. At the same time, among the three methods, automatic feature extraction through the STFT has the best stability of prediction accuracy.

As shown in Fig. 14, the vibration signal is processed by STFT to obtain image data in three directions. The image features in the three directions (X, Y, and Z) are distinct. To reduce the convergence time of the model, images in three directions are fused. The fusion of images in the X, Y, and Z directions not only reduces the data volume but also integrates time-frequency image features from all three directions into a single image. 1080 540 × 570 × 3 image data were obtained in each direction. Taking all data in three directions as input is traditional data fusion, such as Fig. 14. This paper fuses time-frequency images of size 540 × 570 × 3 in three directions into one image of size 540 × 570 × 3. At the same time, the fused image data size 540 × 570 × 3 is changed to 227 × 227 × 3 through the Lanczos interpolation algorithm. The training and fitting time for data of different sizes is shown in Fig. 15. The training time of 1420 × 570 × 3 is 12,018 s (Due to the limitations of computing server computing power and storage space, the 1620 × 570 × 3 training fitting time cannot be obtained), and the training time of 227 × 227 × 3 is 718 s. The training time of the model is effectively shortened by the data fusion operation.

Additionally, to investigate the impact of the Lanczos interpolation algorithm on prediction accuracy, this study employs the algorithm to process four datasets of varying resolutions: 400 × 400, 350 × 350, 300 × 300, and 250 × 250. The prediction errors across these datasets consistently approximate 2.5%, as shown in *Fig. 16*. The results indicate that the Lanczos interpolation algorithm does not significantly influence the prediction error. Furthermore, in an effort to optimize computational efficiency and accelerate the model's training and fitting processes, an input size of 227 × 227—commonly utilized in image processing—is selected.

4.2 Determination of fusion weight parameters and window length parameters

In Sect. 4.3, the fusion operation is performed on the image data. However, the fusion weight and window length parameters are a randomly defined set. As shown in Fig. 17, the value of the window length affects the frequency component of the image. Different frequency components

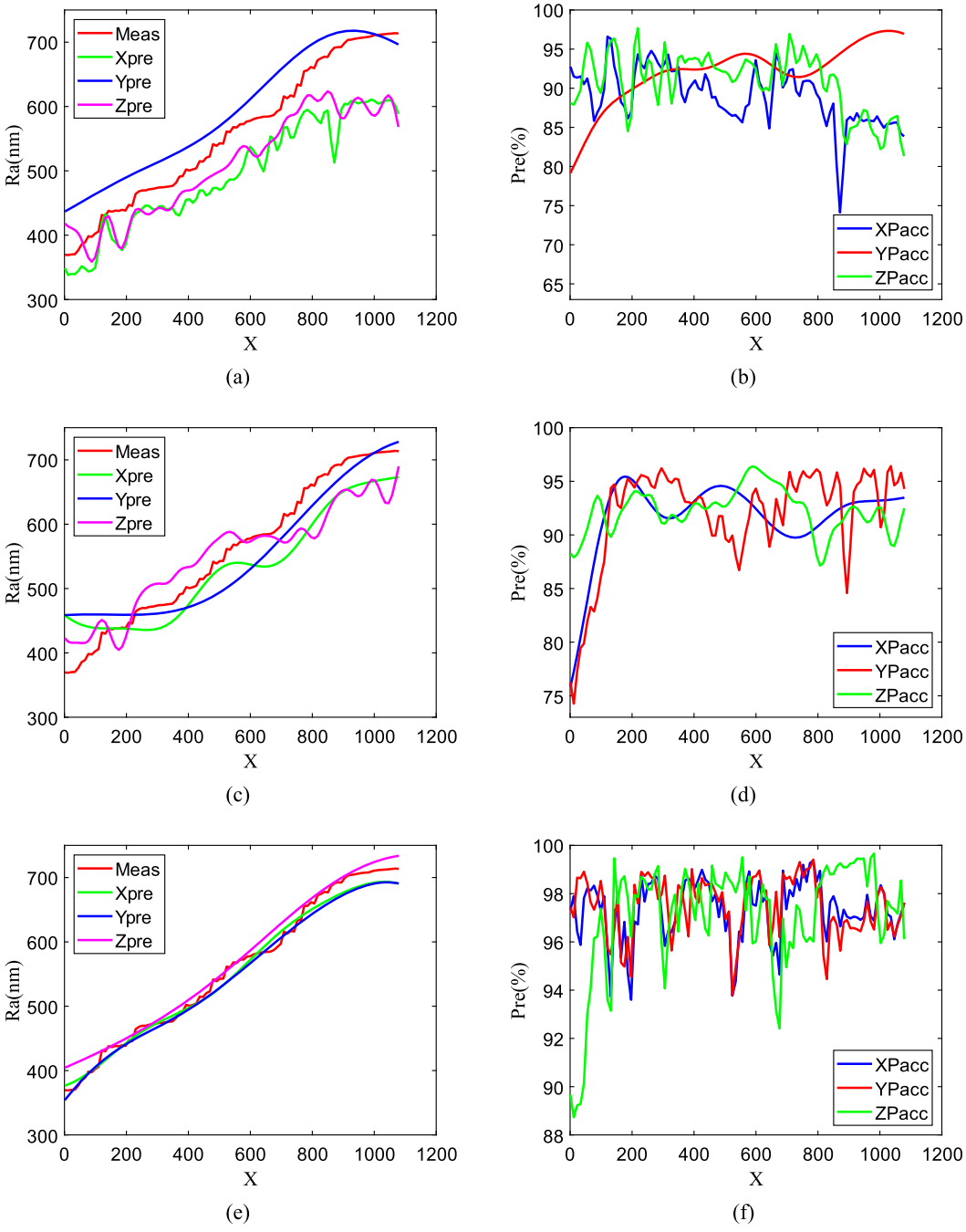


Figure 13 Prediction results. (a) The HHT prediction results. (b) HHT prediction accuracy. (c) The WT prediction results. (d) The WT prediction accuracy result. (e) The HHT prediction accuracy results. (f) The HHT prediction accuracy result

prove that the texture and detail characteristics of the picture are different.

To analyze whether different weight combinations have an impact on model training, an orthogonal experiment of $L3^3$ was established. The experimental design is shown in Table 1. Three factors are the ratio of three weight values (Due to the constraint that the sum of the three weight values must equal 1, it is not feasible to directly generate an orthogonal test table by assigning weight values. Therefore, an orthogonal test is designed by setting the weight ratios instead.). The CNN-AFEADF converges when Epoch = 7-10. The average loss of Epoch = 7-10 is

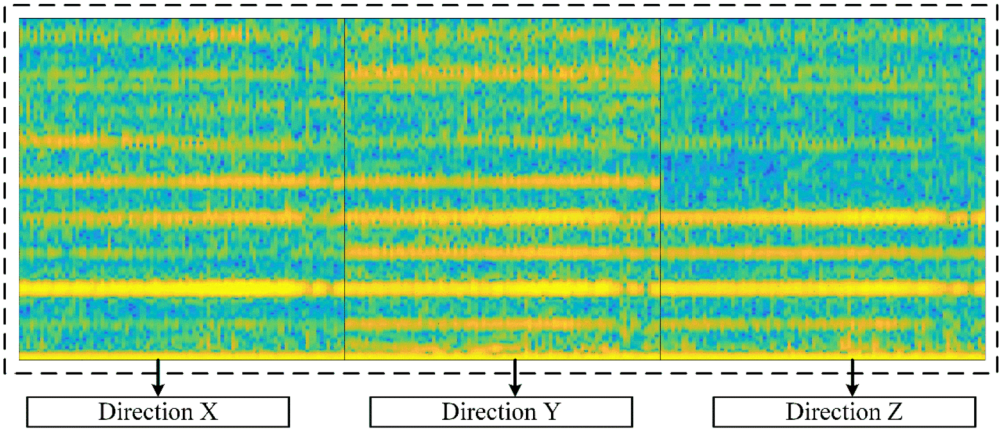


Figure 14 Traditional fusion

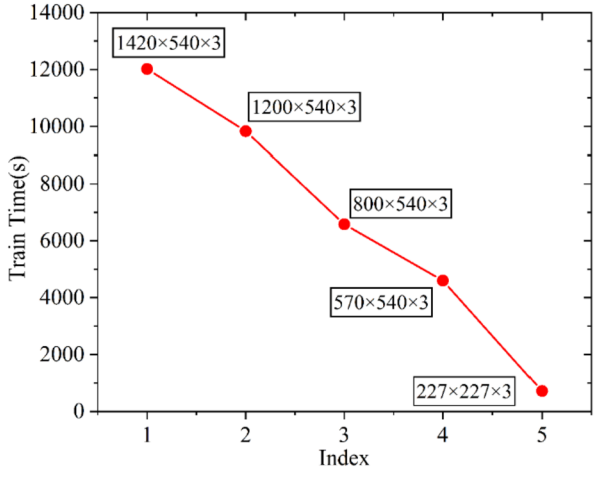


Figure 15 Training time

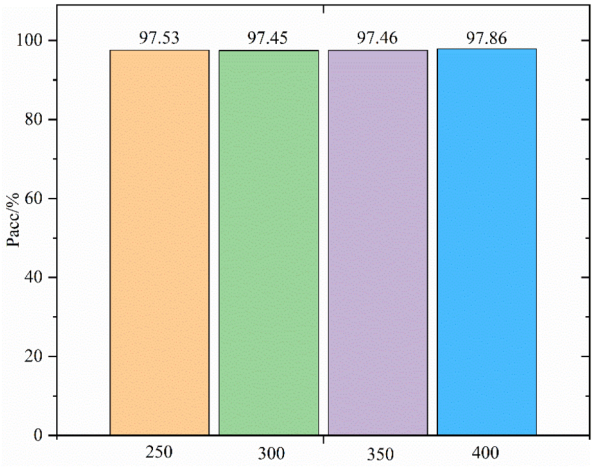


Figure 16 Comparison of accuracy under different input sizes

taken. In order to avoid accidental results in the experiment, each set of parameters is trained five times. The average of five times was taken as the Loss value of the orthogonal experiment. Based on the range analysis, it can be concluded that B has the greatest influencing factor, while A has the smallest influencing factor. According to range analysis, the three fusion weight parameters have different effects on the CNN-AFEADF. Among the data analyzed, the Y-direction data has the greatest impact on the model training fit, while the X-direction data has the least impact on the model fit.

In response to the above analysis, this paper introduces the PSO for optimization. The commonly used values for the STFT window length are 128, 256, and 512. The current window length values are set to the following seven groups according to the interval of 64, as shown in Table 2.

To mitigate the impact of small variations in model fitting training results, 140 sets of weight values are randomly generated through the computer code, and the 140 sets of weight values are randomly assigned to 7 sets of window lengths. Then assign different weight values and window length values to the model for training. The epoch is set to 10, and the mean square error MSE is used as the prediction evaluation index of model training. The model completely converges when Epoch = 7 and the fitting is completely stable. The training results of the combination of different weights and window lengths are shown in the blue bar chart in Fig. 18.

$$MSE = \frac{1}{N}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2. \tag{15}$$

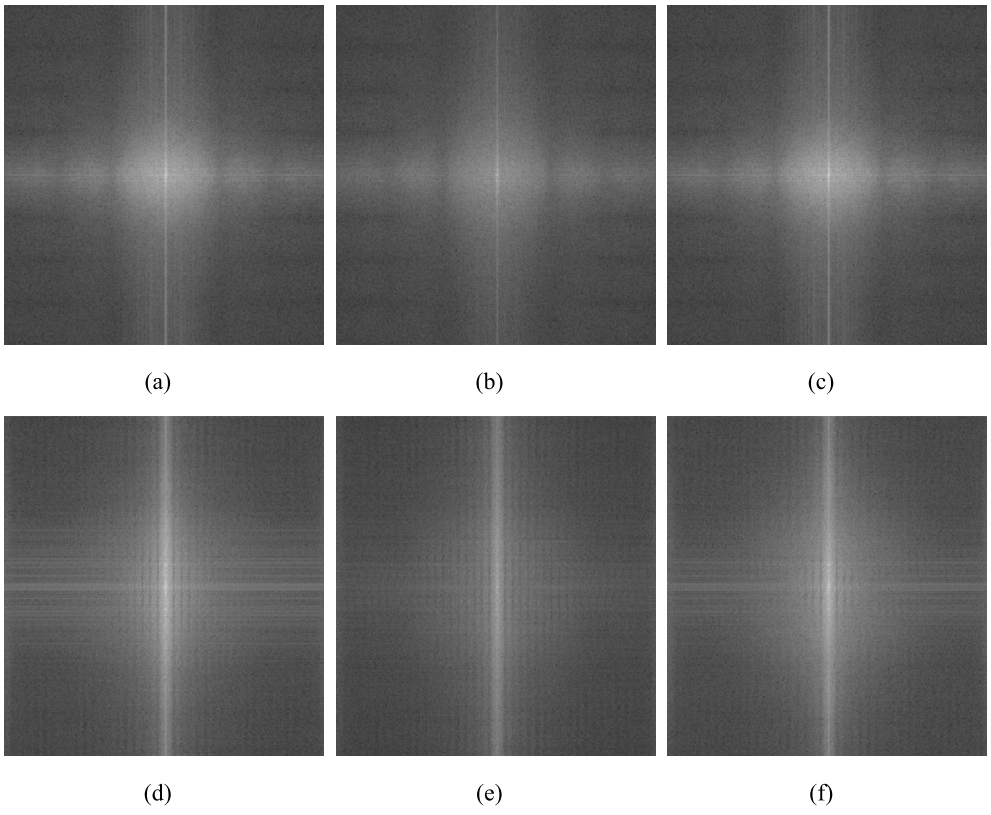


Figure 17 Image frequency feature map. (a) Red channel of 128. (b) Green channel of 128. (c) Blue channel of 128. (d) Red channel of 512. (e) Green channel of 512. (f) Blue channel of 512

Table 1 Orthogonal experiment

Experiment number	A	B	C	Loss
1	4	4	2	1435.2915
2	4	6	4	1420.4937
3	4	8	6	1503.7995
4	6	4	4	1418.0161
5	6	6	6	1536.0786
6	6	8	2	1654.8592
7	8	4	6	1388.3474
8	8	6	4	1542.8412
9	8	8	2	1736.5331
Range analysis	102.712331	217.84558	148.44429	

Table 2 Different window length values

Window length value						
128	194	256	320	384	448	512

Put all the data into the PSO for optimization. The optimized weights and STFT window lengths are shown in Table 3. The PSO iterative optimization process is shown in Fig. 19. The particles are constantly searching for the global optimum. Since their initial positions are close to the global optimum, the final results after completing the iterations remain near the initial positions. Optimization results are shown in Table 4, and average Loss after training of this set of parameters is 1195. The comparison before and after optimization is shown in Fig. 18 and Table 3. In Fig. 18, the optimized green histogram is lower than the blue histogram, and the yellow histogram represents the average value. In Table 3, the maximum optimization (MaxOpt), minimum optimization (MinOpt), and average optimization (AvgOpt) optimization rates are listed. The parameters are determined by PSO, which can improve the optimization effect by at least 8.95%. The training results are shown in Fig. 20.

4.3 Model predictions and comparisons

To verify the performance of the surface roughness prediction method proposed in this paper, two groups of experiments under different working conditions were set up for verification. The machining parameters for the two sets of working conditions are shown in Table 5. Owing to the variations in machining parameters, the surface roughness measured by the roughness measuring instrument also differs. The trend in roughness variation is shown in Fig. 21.

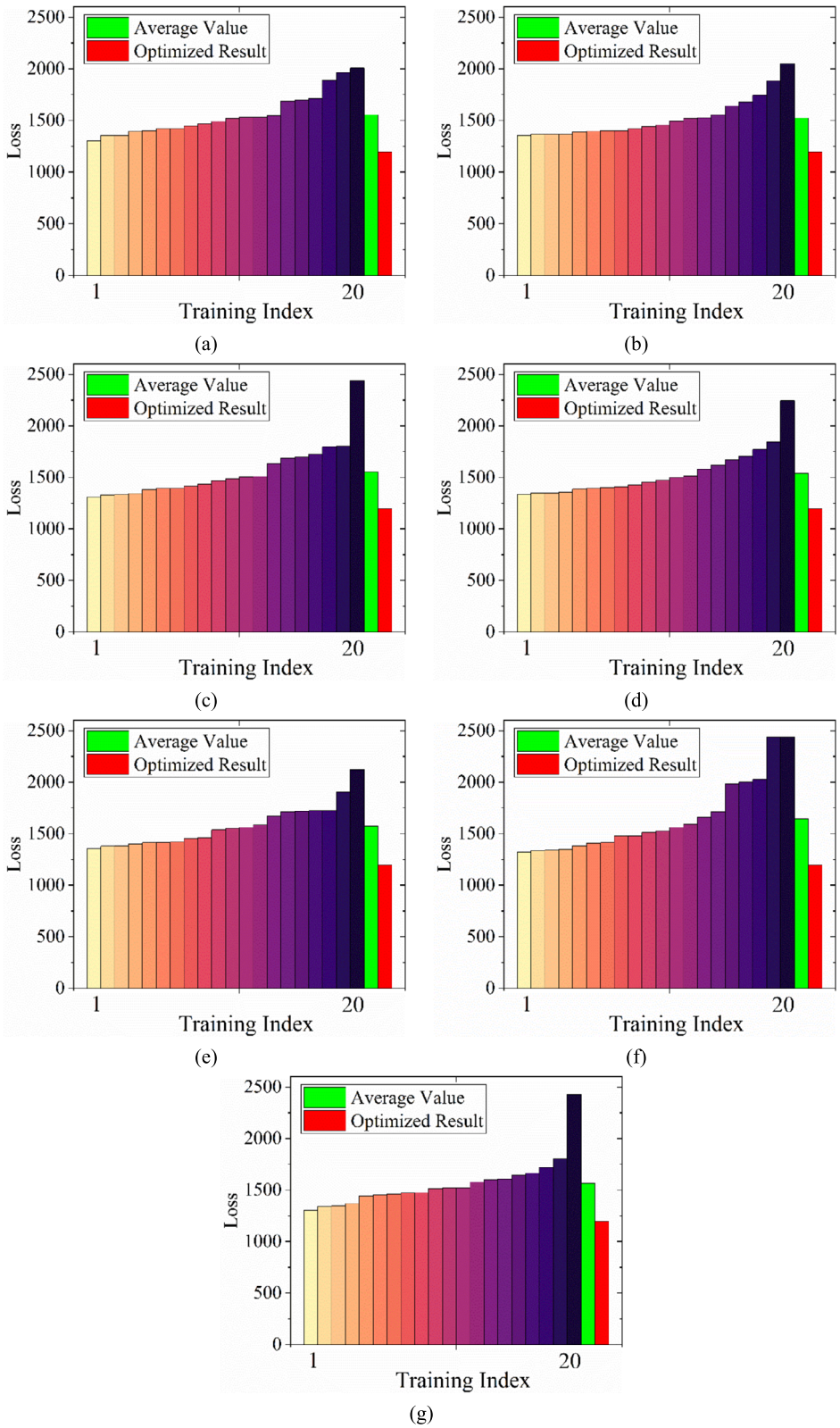


Figure 18 Optimized result. (a) The window length is 128 training groups. (b) The window length is 192 training groups. (c) The window length is 256 training groups. (d) The window length is 320 training groups. (e) The window length is 384 training groups. (f) The window length is 448 training groups. (g) The window length is 512 training groups

Table 3 Optimization Result

X	Y	Z	Windows	Loss after optimization
0.453	0.153	0.394	155	1195

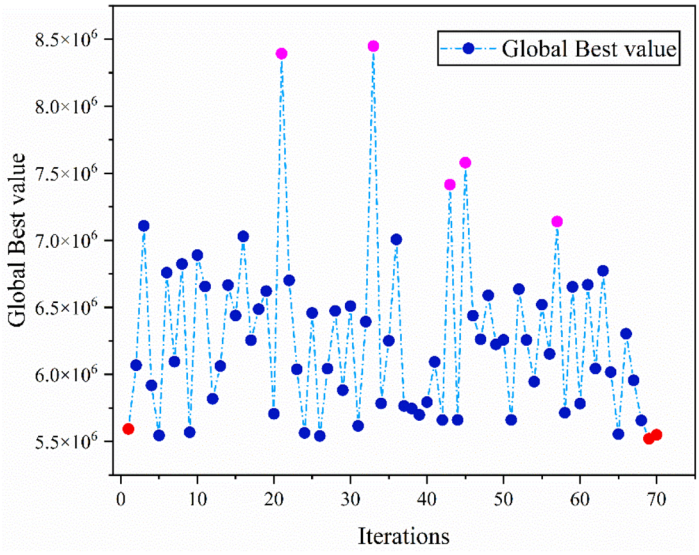


Figure 19 PSO iteration

Table 4 Optimization effect

Category	128	192	256	320	384	448	512
Avg Loss	1556	1523	1554	1538	1574	1648	1565
MinOpt	8.95%	13.52%	9.51%	11.50%	13.08%	10.60%	9.03%
MaxOpt	67.95%	71.67%	104.08%	87.87%	77.48%	103.99%	103.06%
AvgOpt	30.23%	27.42%	30.06%	28.74%	31.71%	37.89%	30.95%

Each group verified the experimental data of the test set of 48 groups, each group took 5 images, a total of 240 images. The model prediction error is measured as a percentage of the average relative prediction error. The first group of prediction results is shown in Fig. 21(a), with an average relative prediction error of 2.74%. The second group of prediction results is shown in Fig. 21(b), with an average relative prediction error of 3.15%. The average relative error between the two groups is 2.95%.

$$\bar{P}_{acc}(\%) = 1 - \frac{1}{n}\sum_{i=1}^{n}\left|\frac{Ra_i^{\exp} - Ra_i^{pre}}{Ra_i^{\exp}}\right| \times 100\%. \tag{16}$$

In addition, SVM, Artificial Neural Networks (ANN), and General Regression Neural Network (GRNN) were used for training prediction. And the SVM kernel function (KernelFunction) was set to linear kernel, the GRNN was configured with 10 hidden neurons, and the ANN featured two hidden layers, each containing 10 neurons. The kurtosis value of the vibration signal corresponding to each surface roughness is calculated. The kurtosis value is input into SVM, ANN, and GRNN training as features, as shown in Fig. 22, CNN-AFEADF has a significantly better prediction error than SVM, ANN, and GRNN.

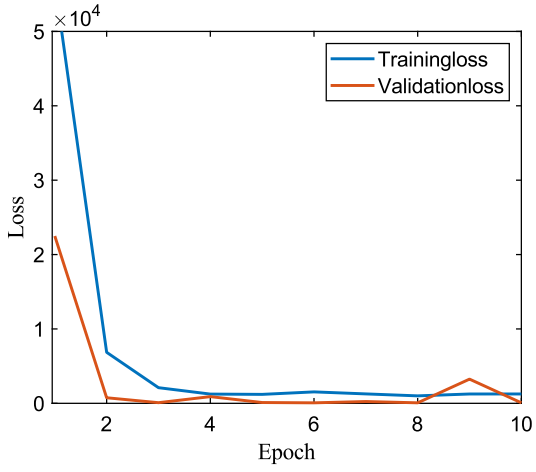
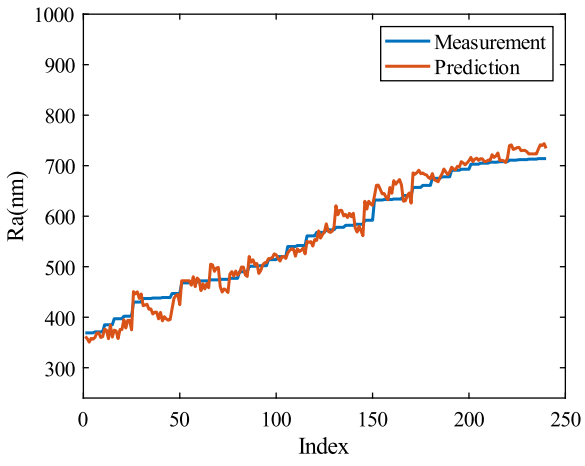


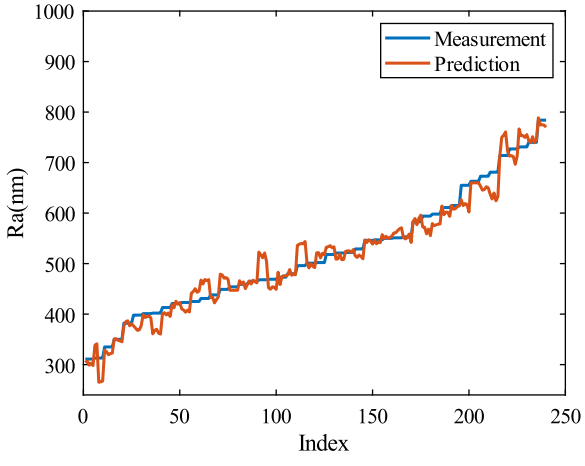
Figure 20 Optimized training results

Table 5 Processing parameters

Category	Speed	Feed	Depth of cut (finishing)
Condition 1	15,000 r/min	1500 mm/min	0.2 mm
Condition 2	16,000 r/min	1700 mm/min	0.2 mm



(a) Condition 1 prediction result



(b) Condition 2 prediction result

Figure 21 CNN-AFEADF test sets results

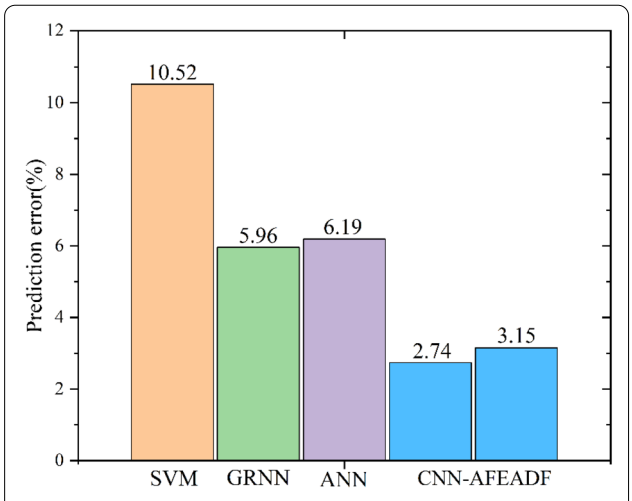


Figure 22 Prediction comparison of different model

5 Conclusion

In this paper, a feature extraction layer based on STFT and a Data Fusion layer are established. The weight parameters and STFT window length parameters are determined by the PSO. The input layer, feature extraction layer, fusion layer, convolution layer, pooling layer, fully connected layer, output layer, and parameter optimization are integrated. This surface roughness intelligent prediction method can realize automatic extraction of data features and adaptive Data Fusion. Parameters are automatically optimized based on the effect of data training. Compared to traditional prediction methods, the model proposed in this paper maintains a lower prediction error. The main findings of this study are:

(1) Recognizing that convolutional neural networks necessitate a substantial number of features to effectively characterize processing quality and thereby minimize prediction errors, this paper proposes a time-frequency domain processing method for signal feature extraction. The obtained image possesses rich time-frequency domain features. Analysis indicates that the feature variations derived from the STFT are more stable and regular. Additionally, the time-frequency domain signal processing method facilitates automatic feature extraction while maintaining a low prediction error. Compared to traditional model data preprocessing, STFT extraction of time-frequency images eliminates the need for preprocessing steps such as signal noise reduction and principal component analysis. This approach integrates traditional neural network models with signal processing, enhancing the efficiency of modeling and application in processing quality assessment.

(2) This paper reduces the data volume through data fusion, effectively shortening model fitting time by a factor of 16.74. This improvement enhances model prediction efficiency and underscores the significance of data fusion processing. Range analysis reveals that the influence weight of the dataset under varying parameter fusions exhibits different sensitivities to neural network training. Consequently, the fusion layer weight parameters and STFT window length parameters are optimized using PSO. This approach effectively reduces the loss value of model training, with the average loss decreasing by approximately 30%. Thus, adaptive data fusion is achieved.

(3) Two test sets with different working conditions were evaluated, yielding prediction errors of 2.74% and 3.15%.

Compared with traditional models, including the SVM (10.52%), GRNN (5.96%), and ANN (6.19%), the proposed model demonstrates a lower prediction error compared to traditional neural network fitting models. This model has demonstrated the ability to maintain low prediction error under the feature extraction and data fusion methods presented in this paper. A modeling scheme is provided for online real-time evaluation of machining quality.

Acknowledgements
Thanks to the State Key Laboratory of Mechanical Transmission for Advanced Equipments of Chongqing University for the support of this study

Author contributions
XZ: Technical implementation, Data Curation, Experiments.
SW: Methodology, Review & Editing, Funding acquisition.
FG: Technical support for machine tools, Programming support.
HW: Support in academic writing, Results discussion.
HW: Data acquisition, Collation and analysis.
YL: Proofreading.
All authors read and approved the final manuscript.

Funding
This work is supported by the National Key Research and Development Program Young Scientist Program under Grant No. 2022YFB3402400, the National Natural Science Foundation of China under Grant No. 52375407, Chongqing graduate research innovation project under Grant No. CYS23141, Chongqing Talent Program under Grant No. CQYC202105002.

Availability of data and material
The data that support the findings of this study are available on request from the corresponding author upon reasonable request.

Code availability
The code that support the findings of this study are available on request from the corresponding author upon reasonable request.

Declarations

Competing interests
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.

Author details
[1]State Key Laboratory of Mechanical Transmission for Advanced Equipments, Chongqing University, Chongqing 400044, China. [2]College of Mechanical and Vehicle Engineering, Chongqing University, Chongqing 400044, China. [3]Department of Mechanical Engineering, School of Engineering, Cardiff University, Cardiff, CF24 3AA, UK.

Received: 4 August 2024 Revised: 7 October 2024
Accepted: 31 October 2024 Published online: 12 December 2024

References

1. C. Chen, D. Wu, Y. Liu, Recent advances of AI for engineering service and maintenance. Auton. Intell. Syst. **2**(1), 19 (2022)
2. Y. Guo, Y. Sun, K. Wu, Research and development of monitoring system and data monitoring system and data acquisition of CNC machine tool in intelligent manufacturing. Int. J. Adv. Robot. Syst. **17**(2), 1729881419898017 (2020)
3. E.C. Ani, K.A. Olu-lawal, O.K. Olajiga, et al., Intelligent monitoring systems in manufacturing: current state and future perspectives. J. Eng. Sci. Technol. **5**(3), 750–759 (2024)
4. L. Yao, Z. Ge, Big data quality prediction in the process industry: a distributed parallel modeling framework. J. Process Control **68**, 1–13 (2018)
5. V. Nasir, M. Kooshkbaghi, J. Cool, et al., Cutting tool temperature monitoring in circular sawing: measurement and multi-sensor feature fusion-based prediction. Int. J. Adv. Manuf. Technol. **112**, 2413–2424 (2021)
6. W. Yan, X. Wang, Q. Gong, et al., A service-oriented energy assessment system based on BPMN and machine learning. Auton. Intell. Syst. **2**(1), 18 (2022)
7. D. Wang, Robust data-driven modeling approach for real-time final product quality prediction in batch process operation. IEEE Trans. Ind. Inform. **7**(2), 371–377 (2011)
8. V. Nasir, F. Sassani, A review on deep learning in machining and tool monitoring: methods, opportunities, and challenges. Int. J. Adv. Manuf. Technol. **115**(9), 2683–2709 (2021)
9. T. Wuest, C. Irgens, K.D. Thoben, An approach to monitoring quality in manufacturing using supervised machine learning on product state data. J. Intell. Manuf. **25**, 1167–1180 (2014)
10. M. Kovačič, B. Šarler, Application of the genetic programming for increasing the soft annealing productivity in steel industry. Mater. Manuf. Process. **24**(3), 369–374 (2009)
11. N. Liu, et al., A novel approach to predicting surface roughness based on specific cutting energy consumption when slot milling Al-7075. Int. J. Mech. Sci. **118**, 13–20 (2016)
12. C. Zhao, A quality-relevant sequential phase partition approach for regression modeling and quality prediction analysis in manufacturing processes. IEEE Trans. Autom. Sci. Eng. **11**(4), 983–991 (2013)
13. Q. Xu, C. Xu, J. Wang, Forecasting the yield of wafer by using improved genetic algorithm, high dimensional alternating feature selection and SVM with uneven distribution and high-dimensional data. Auton. Intell. Syst. **2**(1), 24 (2022)
14. X. Yin, Z. Niu, Z. He, et al., Ensemble deep learning based semi-supervised soft sensor modeling method and its application on quality prediction for coal preparation process. Adv. Eng. Inform. **46**, 101136 (2020)
15. B.-h. Li, B.-C. Hou, W.-T. Yu, X.-B. Lu, C.-W. Yang, Applications of artificial intelligence in intelligent manufacturing: a review. Front. Inf. Technol. Electron. Eng. **18**, 86–96 (2017)
16. J. L. Wang, C.Q. Xu, J. Zhang, R. Zhong, Big data analytics for intelligent manufacturing systems: a review. J. Manuf. Syst. **62**, 738–752 (2022)
17. X. Kong, X. Li, Q. Zhou, et al., Attention recurrent autoencoder hybrid model for early fault diagnosis of rotating machinery. IEEE Trans. Instrum. Meas. **70**, 1–10 (2021)
18. Z.H. Wang, S.B. Wang, S.L. Wang, et al., An intelligent process parameters determination method based on multi-algorithm fusion: a case study in five-axis milling. Robot. Comput.-Integr. Manuf. **73**, 102244 (2022)
19. L.H. Xu, C.Z. Huang, J.H. Niu, et al., Prediction of cutting power and surface quality and optimization of cutting parameters using new inference system in high-speed milling process. Adv. Manuf. **9**, 388–402 (2021)
20. Z.Y. Zhao, S.B. Wang, Z.H. Wang, et al., Surface roughness stabilization method based on digital twin-driven machining parameters self-adaption adjustment: a case study in five-axis machining. J. Intell. Manuf. **33**, 1–10 (2022)
21. T. Meng, X.Y. Jing, Z. Yan, et al., A survey on machine learning for data fusion. Inf. Fusion **57**, 115–129 (2020)
22. H. Wu, Z. Li, Q. Tang, et al., A practical prediction method for grinding accuracy based on multi-source data fusion in manufacturing. Int. J. Adv. Manuf. Technol. **127**(3), 1407–1417 (2023)
23. S. Zeng, D. Pi, T. Xu, Milling surface roughness prediction method based on spatiotemporal ensemble learning. Int. J. Adv. Manuf. Technol. **128**(1–2), 91–119 (2023)
24. Z. Wang, S. Wang, S. Wang, et al., An intelligent prediction method of surface residual stresses based on multi-source heterogeneous data. J. Intell. Manuf., 1–17 (2023)
25. Z. He, T. Shi, J. Xuan, Milling tool wear prediction using multi-sensor feature fusion based on stacked sparse autoencoders. Measurement **190**, 110719 (2022)
26. Z.W. Huang, J.M. Zhu, J.T. Lei, et al., Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations. J. Intell. Manuf. **31**, 953–966 (2020)
27. X.W. Zhang, Z.J. Leng, Z.B. Zhao, et al., Spatial-temporal dual-channel adaptive graph convolutional network for remaining useful life prediction with multi-sensor information fusion. Adv. Eng. Inform. **57**, 102120 (2023)
28. S. Takasu, M. Masuda, T. Nishiguchi, et al., Influence of study vibration with small amplitude upon surface roughness in diamond machining. CIRP Ann. **34**(1), 463–467 (1985)

29. S. Banerji, A. Sinha, C. Liu, New image descriptors based on color, texture, shape, and wavelets for object and scene image classification. Neurocomputing **117**, 173–185 (2013)
30. J. Kennedy, R. Eberhart, Particle swarm optimization, in *Proceedings of ICNN'95-International Conference on Neural Networks*, vol. 4 (IEEE, 1995), pp. 1942–1948
31. Z. Huang, J. Zhu, J. Lei, et al., Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations. J. Intell. Manuf. **31**, 953–966 (2020)
32. D. Kinga, J.B. Adam, A method for stochastic optimization, in *International Conference on Learning Representations*, vol. 5 (ICLR, 2015), p. 6

Publisher's Note