RESEARCH ARTICLE

Research on Tool Remaining Life Prediction Method Based on CNN-LSTM-PSO

SHUO WANG, ZHENLIANG YU, GUANGCHEN XU, AND FENGQIN ZHAO

School of Mechanical and Power Engineering, Yingkou Institute of Technology, Yingkou, Liaoning 115114, China

Corresponding author: Zhenliang Yu (yuzhenliang_neu@163.com)

ABSTRACT Efficient and accurate prediction of tool Remaining Useful Life (RUL) is the key to improve product accuracy, improve work efficiency and reduce machining costs. Aiming at the problems of weak tool wear state features, difficult extraction, and low prediction precision and accuracy, this research proposes a CNN-LSTM-PSO tool remaining life prediction method based on multi-channel feature fusion. Firstly, based on computer vision, feature extraction, information fusion technology, the multi-source sensor signals collected during the tool life cycle are effectively processed and analyzed, and a sample data set of spatio-temporal correlation of traffic flow is constructed. Secondly, the sample data set was input into the CNN-LSTM-PSO model, the CNN network obtained the sequence feature vector by extracting the spatial characteristics of traffic flow data, and the feature vector was input into the multi-layer LSTM network to extract the time-dependent features, and the PSO algorithm optimized the hyperparameters in the CNN-LSTM model. The accuracy of tool RUL prediction model and the efficiency of model fitting are further improved. The results show that the CNN-LSTM-PSO model can effectively predict tool wear, with the mean absolute error (MAE) value of 1.0892, the root mean square error (RMSE) value of 1.3520, and the determination coefficient R^2 value of 0.9961; Through the comparative analysis of ablation experiments, it is found that the method proposed in the research has the highest efficiency in fitting the tool RUL prediction model, the lowest values of MAE value and root mean square error RMSE, and the value of determination coefficient R^2 is closest to 1, which has certain advantages.The proposed method has reference value and engineering practical significance for the related research of tool wear residual life prediction.

INDEX TERMS CNC machining, CNN-LSTM-PSO, feature fusion, remaining useful life, tool wear.

I. INTRODUCTION

Whether the tool wear is serious in the process of CNC machining plays a decisive role in the processing accuracy of the product. When the tool wear is serious, the quality of the product will be reduced, resulting in the increase of the rejection rate, and even cause the occurrenceof machine tool accidents [1]. Therefore, tool RUL prediction has become a basic and prerequisite work in the field of tool health maintenance and intelligent tool change in recent years [2], [3]. The effective management of tool remaining

The associate editor coordinating the review of this manuscript and approving it for publication was Yiming Tang.

useful life can improve production efficiency and reduce operation and maintenance costs, but the value of the tool can not be fully utilized if excessive protection strategies are adopted, and unnecessary tool change downtime will also lead to a waste of time. It can be seen that how to accurately predict the remaining useful life of the tool and make the maximum use of the tool will be an important research direction in the future high-end manufacturing industry.

In early studies, mostly data-driven methods combining sensor monitoring data with machine learning algorithms were mostly used for tool life prediction [4], [5]. Monitoring data refers to the acquisition of raw signals using sensor technology and the extraction of tool wear characteristics in the

time domain, frequency domain, and time-frequency domain. At present, the most commonly used feature extraction methods are: Empirical Mode Decomposition (EMD) [6], Fourier transform [7] and wavelet packet analysis [8]. EMD method can effectively extract tool wear state features from time domain and frequency domain, but it has high requirements for signal frequency processing, and extremely serious end effects and mode aliasingmay occur in the use process [9]. Fourier transform has independent adaptability, so that the time domain features can be better revealed in the frequency domain, so it is widely used to extract the frequency domain features of the sample signal [10]. Wavelet packet analysis is the use of different types of filters to decomple the time domain features into different frequency band signals, in order to achieve the purpose of refining the signal, so it is often used to extract the time-frequency domain features of the sample signal [11]. The machine learning algorithm takes the extracted tool wear features as the input of the model, simulates the whole process of tool life degradation, and compares the current working state with historical data to complete the prediction of tool remaining life [12], [13]. Wei et al. [14] used genetic algorithm to optimize BP neural network, which improved the optimization ability and learning ability of the model, and ensuredthe efficiency and accuracy of tool wear recognition. Cao et al. [15] used RBF neural network and D-S evidence theory for information fusion to diagnose tool wear faults. Experiments show that the model can effectively diagnose tool wear faults, and its prediction accuracy has been improved. Soufiane et al. [16] used continuous wavelet transform for feature extraction, and proposed an Improved Extreme Learning Machine (IELM), which mapped the input data through a nonlinear function to generate a degradation model, so as to obtain health indicators to complete the prediction of tool remaining life. Yu et al. [17] combined time domain, frequency domain, and wavelet analysis to extract the features of force and vibration signals, and constructed the IHDGWO-SVM model for tool wear prediction. The experimental results show that the prediction accuracy of the proposed model is 92 %, which is significantly higher than other models. Although the use of machine learning technology to predict tool residual life has achieved certain results, the prediction model can only mine the shallow features of sample data, and the generalization ability is insufficient. The network fitting speed is slow, and the prediction accuracy and accuracy are not high, and it is more dependent on signal processing technology and expert experience. With the development of Deep Learning theory (DL), especially Recurrent Neural Network (RNN) [18], Long Short-Term Memory network (LSTM) [19] and Convolutional Neural Network (CNN) [20], its prediction effect is significantly higher than that of machine learning technology, and it has been widely used in the research of tool wear life prediction. Compared with machine learning methods, these prediction models have more powerful feature learning and mapping capabilities, and can automatically mine deep features for prediction without prior knowledge or the help of human experts. RNN is prone to gradient disappearance or gradient explosion when processing time-series data of tool wear. To solve this problem, researchers mostly use LSTM network to predict tool wear time-series. In recent years, the work of tool life prediction based on Long LSTM has been gradually carried out, Ma et al. [21] by analyzing the singularity of the original vibration signal, eliminated the influence of the milling path, and constructed a stacked LSTM model to predict tool wear. It is found that the accuracy of tool wear prediction is improved by this method. Although the LSTM network can perfectly process the time series information of tool wear, it is difficult to extract the deep features hidden in the sample, which leads to the incomplete feature extraction of tool wear prediction, and there is still room for improvement. Compared with LSTM network, CNN Network has strong feature extraction ability and low computational complexity, which can mine the deep features hidden in the sample. Lim et al. [22] crop the surface contour image of the machined parts and input it into the CNN network for tool wear prediction. The results show that the CNN model can meet the requirements of tool wear prediction, and the accuracy rate reaches 98.9%. Although these methods have successfully predicted tool wear, it is still challenging to fully reveal the effective features existing in the monitoring signal due to the defects in the network structure [23]. It is well known that the tool wear law is fast in the early stage, gentle in the middle stage, and the most rapid and intense in the later stage. The tool wear process is gradual, random, nonlinear and has a serious dependence on timing characteristics [24]. It can be seen that using only one model to predict the tool life will lead to a single extracted feature and easy to produce overfitting. Therefore, the combination of CNN Network and LSTM network has become an inevitable trend. Zhang et al. [25] used the CNN-LSTM model to predict the remaining useful life of the tool. The results show that the prediction accuracy of the proposed method reaches 87.3 %, which is significantly better than other single algorithms, but there is still room for improvement. The CNN model is used to mine the potential deep features in space, and the LSTM model is used to capture the time series information in time, so that the time series features and spatial features of the data are fully utilized, so as to make up for the shortcomings of the above single prediction model. To further improve the prediction effect of the model, it is necessary to optimize the hyperparameters in the CNN-LSTM model. At present, the more common hyperparameter optimization methods include random optimization [26], gradient-based optimization [27], genetic algorithm optimization [28], particle swarm optimization [29], etc. Particle Swarm Optimization (PSO) algorithm [30] can perform global optimization with fewer parameters, and its powerful search performance and individual optimization ability can speed up the convergence speed of the model, so it has been widely used and studiedby scholars in recent years.

Therefore, based on computer vision, feature extraction, deep learning, hyperparameter optimization and other technologies, the research proposes a CNN-LSTM-PSO tool remaining useful life prediction model with multi-channel feature fusion. The model uses CNN Network as the feature learner to complete the deep mining of tool wear features in space and realize data dimension reduction. The LSTM network was used as the trainer to capture the time dependence of tool wear sequence in time, so that the temporal and spatial characteristics of the data were fully utilized, so as to construct the nonlinear mapping relationship between tool wear characteristics and tool wear value, and the PSO was used to optimize the hyperparameters in the CNN-LSTM prediction model. Thus, the performance of RUL prediction model can be improved. This method can monitor the tool wear value in real time. According to the prediction results, the machine tool can make intelligent judgment and make corresponding treatment, so as to improve the product processing quality and reduce the rejection rate, which has certain practical significance.

The rest of the paper is structured as follows. Section II mainly discusses the tool remaining life prediction method and process based on CNN-LSTM-PSO; Section III introduces the method of multi-channel information feature fusion in detail, and constructs the tool wear sample data set. Section IV uses the proposed method to conduct experimental test and evaluation of tool wear, and verifies the superiority of CNN-LSTM-PSO model. Section V presents some important conclusions of the paper.

II. TOOL REMAINING LIFE PREDICTION METHOD

The section introduces the Remaining Tool Life (RUL) prediction method based on CNN-LSTM-PSO model, mainly including the improved scheme of RUL prediction model, the operation process of RUL prediction model and the construction principle of CNN-LSTM-PSO hybrid prediction model.

A. IMPROVEMENT SCHEME OF RUL PREDICTION MODEL

In the automated production process, a high-precision life prediction model can very effectively predict the future degree of tool wear. It is of great significance to study the intelligent tool change when tool wear is at the critical threshold. Due to the limitations of a single algorithm prediction model, at present, different algorithms are integrated and optimized to predict the life of complex systems, in order to achieve higher prediction effect. Based on the above background, the research proposes a CNN-LSTM-PSO tool RUL prediction model based on multi-channel feature fusion. By monitoring the vibration signals of three channels, the cutting force signals of three channels and the acoustic emission (AE) signals of one channel, the wear value of the tool is output, so as to realize the predictive maintenance of the CNC machining tool. The tool can be intelligently changed before the tool wear is at the critical threshold. The improvements are as follows:

(1) The prediction of remaining tool life using a single signal will not contain all the tool wear information. So the data processing part of the RUL prediction model uses multi-channel feature fusion technology to extract vibration signals, cutting force signals, AE signals for feature extraction and fusion, and batch normalization and dimension reduction processing. The generalization ability of the model is improved, the over-fitting phenomenon is avoided, and the prediction accuracy of the model is improved.

(2) The CNN is used as the feature learner of RUL prediction model. The unique structure of CNN model with local connection and weight sharing reduces the complexity of the network, and the spatial continuity of sample features is maintained after convolution and pooling operations.

(3) To capture the time series information in the tool wear process, the LSTM is used as the trainer. The LSTM network is a further optimization of the traditional RNN network, which can deal with longer time series data and avoid the phenomenon of gradient disappearance or gradient explosion.

(4) The powerful search and global optimization ability of PSO algorithm is used to iteratively optimize the hyperparameters in CNN-LSTM network, which reduces the subjective influence of manual selection parameters, thereby improving the prediction accuracy of tool wear model.The CNN-LSTM-PSO tool remaining life prediction model is shown in Fig. 1.

B. OPERATION PROCESS OF RUL PREDICTION MODEL

Regression prediction of tool RUL mainly includes four main modules, which are data preprocessing, CNN module, LSTM module, and PSO optimization module. As shown in Figure1. The tool life prediction process of its CNN-LSTM-PSO model is shown in Fig. 2, and the specific steps are as follows:

Step 1: The data preprocessing module denoises the original signals of multiple channels, extracts and fuses the features in the time domain, frequency domain and time-frequency domain respectively, and uses the Pearson correlation coefficient formula to reduce the dimension of the feature data to construct a sample data set of spatio-temporal correlation of traffic flow.

Step 2: Use the PSO module to initialize the initial learning rate, the number of hidden layer units and other hyperparameters, in preparation for finding the best hyperparameters of the CNN-LSTM model.

Step 3: The tool wear sample data set formed in step 1 is divided into training set and test set with a ratio of 2:1, while the CNN-LSTM model is constructed according to the initialized hyperparameters.

Step 4: Input the training set into the CNN module for convolution and pooling operations, extract the deep features of the traffic flow data, and obtain a compressed tool wear feature mapping vector after PCA dimension reduction.

Step 5: The compressed feature vector is input into the LSTM module, and the multi-layer LSTM network further captures the time-dependent characteristics of tool wear data.
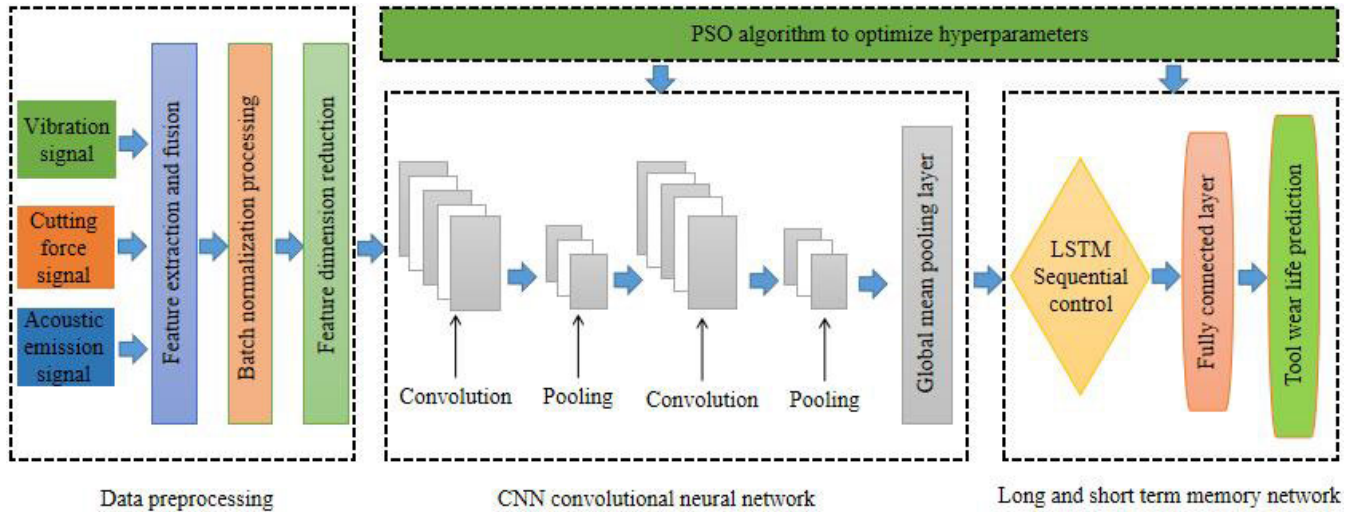
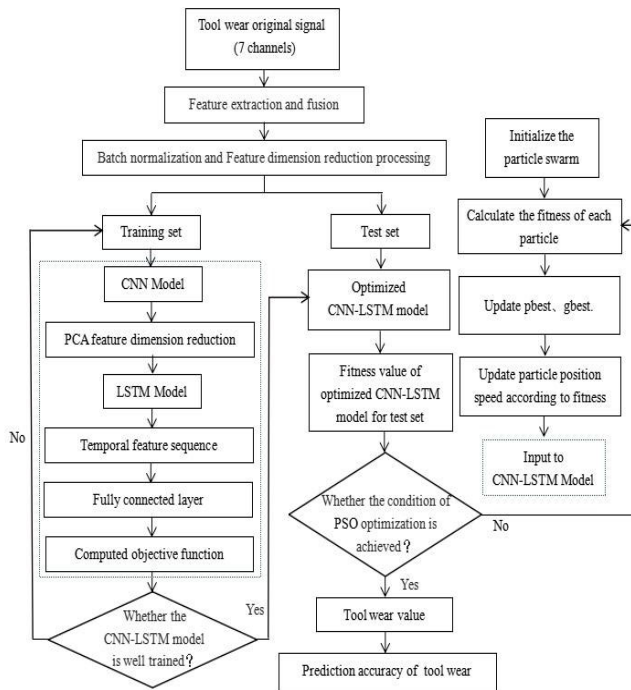**FIGURE 1.** CNN-LSTM-PSO tool remaining life prediction model.

FIGURE 2. CNN-LSTM-PSO model prediction flow.

Step 6: Use the fully connected layer and Softmax layer to obtain the probability sequence and calculate the objective function. If the training termination condition is not reached, return to step 4 and continue the above calculation until the maximum number of iterations is reached or the objective function converges.

Step 7: Input the test set formed in step 3 into the trained CNN-LSTM model and calculate the tool wear prediction accuracy. If the termination condition is not reached, the PSO module iteratively updates the position of the particle in the

multi-dimensional search space until the optimal network structure parameters are found.

Step 8: Retrain the CNN-LSTM model with the new hyperparameters and training set, and repeat the operation process of step 7 until the PSO module reaches the maximum number of iterations or finds the best hyperparameters.

C. PREDICTION PRINCIPLE OF CNN-LSTM-PSO MODEL

With the explosive development of industrial big data, compared with traditional machine learning prediction methods, deep learning methods have shown more and more significant capabilities and advantages in dealing with nonlinear and large amounts of data [31]. However, due to the limitations of a single algorithm prediction model, the CNN model and LSTM model are fused and optimized to monitor the condition of complex tool system, in order to achieve higher prediction effect. At the same time, the PSO method is used to optimize the hyperparameters of the CNN-LSTM hybrid model to avoid memory loss and gradient dispersion caused by too long step size, so as to improve the performance of the tool RUL prediction model.

1) IMPROVED CNN-LSTM PREDICTION MODEL

The CNN-LSTM prediction model proposed in this paper can not only deeply mine the spatial features of vibration signals, cutting force signals and AE signals, but also retain the effective information in the time dimension. The model outputs the tool wear value in real time by constructing a nonlinear mapping relationship with the wear thickness of the tool, and generates a failure alarm when the total wear thickness exceeds the wear threshold. Its specific principle is as follows:

(1) The sample data set after multi-channel feature fusion was input into the CNN Network, and the convolution layer convolved the data of the input layer with multiple

convolution kernels to reduce the number of training parameters of the model. The sample features obtain the weight value of each layer through the convolution operation, which indirectly represents the local features of the sample. The higher the level, the more detailed the local feature extraction, and the spatial continuity of the sample can be maintained. The convolution operation is shown in Equation (1):

$$X_i^k = \sum_{j=1}^{n} W_i^{kj} \otimes X_{i-1}^j + b_i^k \tag{1}$$

where X_i^k represents the feature matrix of the kth neuron output by the ith layer, W_i^{kj} represents the weight value of the Kth neuron output by the ith layer, $\otimes$ represents the convolution operator, X_{i-1}^j represents the feature matrix of the Jth neuron output by the i-1 layer, and b_i^k is the bias coefficient of the kth neuron output by the ith layer.

(2) To improve the prediction accuracy of the tool RUL model, the CNN network uses the ReLU function for non-linear activation, which has good unsaturated characteristics and avoids the phenomenon of gradient disappearance. The activation function is shown in Equation (2):

$$V_i^k = \text{Relu}\left(X_i^k\right) = \begin{cases} 0, & x_i^k < 0 \\ x_i^k, & x_i^k > 0 \end{cases} \tag{2}$$

where x_i^k is each eigenvalue in the feature matrix X_i^k.

(3) Each tool wear feature data is input into the pooling layer after convolution operation, and the pooling type is selected as the maximum pooling, which can not only retain the original features but also reduce the size of the feature layer, simplify the complexity of the neural network, and improve the robustness of the sample features. Max pooling is shown in Equation (3):

$$C_i^k(s, t) = \underset{\substack{1 + (s-1)\,Q \le d \le sQ \\ 1 + (t-1)\,P \le h \le tP}}{Max} \left\{ V_i^k(d, h) \right\} \tag{3}$$

where $V_i^k(d, h)$ is the eigenvalue of row d, column h of the ith feature matrix input by the pooling layer, $C_i^k(s, t)$ is the eigenvalue of row s, column t of the ith feature matrix obtained after the pooling process, P and Q are the length and width of the pooling region, respectively.

(4) The n feature matrices with dimension $S \times T$ obtained from each row of the sample data set after convolution and pooling operation are input into the global average pooling layer. The dimension of the pooling kernel of the global average pooling layer is consistent with the dimension of the feature matrix, and the dimension of the n feature matrices is reduced to reduce the collinearity of the sample features and avoid the influence of redundant features, thereby reducing the training time of the LSTM network. Therefore, the entire CNN model finally outputs a feature vector $X_t = \{x_1, x_2, \ldots x_i, \ldots, x_j, \}$. The calculation method of x_i is shown

in Equation (4):

$$x_i = \frac{1}{ST} \sum_{s=1}^{S} \sum_{t=1}^{T} C_i^k(s, t) \tag{4}$$

(5) The LSTM is used to further process the feature vector X_t output by the CNN model, and the relationship between sample features and time series is constructed. Firstly, the memory features of the input of the previous layer unit are forgotten or memorized through the forgetting gate f_t in the LSTM network. The calculation method of the forgetting gate f_t is shown in Equation (5)

$$f_t = \sigma\left(W_f \cdot [H_{t-1}, X_t] + b_f\right) \tag{5}$$

where, H_{t-1} is the time series information output at the previous time, X_t is the sample feature input at this time, σ is the sigmoid function, W_f is the weight value of the forgetting gate training, b_f is the bias coefficient of the forgetting gate training.

(6) The input sample feature X_t is logically calculated through the input gate i_t, the memory of the whole system is updated, and the new memory feature C_t is generated according to the path set by the system. The calculation methods of the input gate i_t and the memory feature C_t are shown in Equations (6) and (7)

$$i_t = \sigma\left(W_i \cdot [H_{t-1}, X_t] + b_i\right) \tag{6}$$
$$C_t = f_t \otimes C_{t-1} + i_t \otimes \tanh\left(W_c \cdot [H_{t-1}, X_t] + b_c\right) \tag{7}$$

where, W_i is the weight value of the input gate training, b_i is the bias coefficient of the input gate training, C_{t-1} is the memory feature output at the last moment, tanh is the activation function, W_c is the weight value of the memory feature training, and b_c is the bias coefficient of the memory feature training.

(7) The output gate o_t controls the memory feature C_t and outputs the timing feature H_t, which is transmitted to the next layer unit. The calculation methods of output gate o_t and timing feature H_t are shown in Equations (8) and (9)

$$o_t = \sigma\left(W_o \cdot [H_{t-1}, X_t] + b_o\right) \tag{8}$$
$$H_t = o_t \otimes \tanh\left(C_t\right) \tag{9}$$

where, W_o is the weight value trained for the output gate and b_o is the bias coefficient trained for the output gate.

(8) Construct tool wear feature vector $M_t = \{H_1, H_2, H_3, \ldots, H_t, \ldots H_n\}$, input to the fully connected layer to complete the prediction of tool RUL, and obtain the tool wear value y_t. The calculation method is shown in Equation (10):

$$y_t = W_s M_t + b_s \tag{10}$$

where, y_t is the predicted value of tool wear, W_s and b_s are the weight and bias values of the fully connected layer, respectively.

2) PARTICLE SWARM OPTIMIZATION (PSO)

PSO is the way to complete the evolution of the flock is the mutual assistance and information sharing between individuals. The PSO algorithm is similar to the process of bird predation. Due to its simple principle and convenient operation, it is widely used in the global optimization process of hyperparameters [32]. Each individual in the population is called a particle, and each particle is a possible solution of the parameter optimized in the global search space. Its characteristic index mainly includes three aspects: position, speed and fitness value. Firstly, the fitness value of each particle is calculated by the fitness function, and the optimal position and velocity of all particles are memorized. In each iteration, the particle adjusts any dimension component of the velocity to make it reach a new position and calculate it, and repeat until the particle finds the optimal position or reaches the iteration number, so as to complete the optimization process of the particle in the multi-dimensional search space. The optimization process of the PSO algorithm is shown in Fig.3. In the research, the PSO algorithm is used to optimize the hyperparameters in the LSTM model, and the optimal solution is obtained to avoid the overfitting phenomenon during model training.
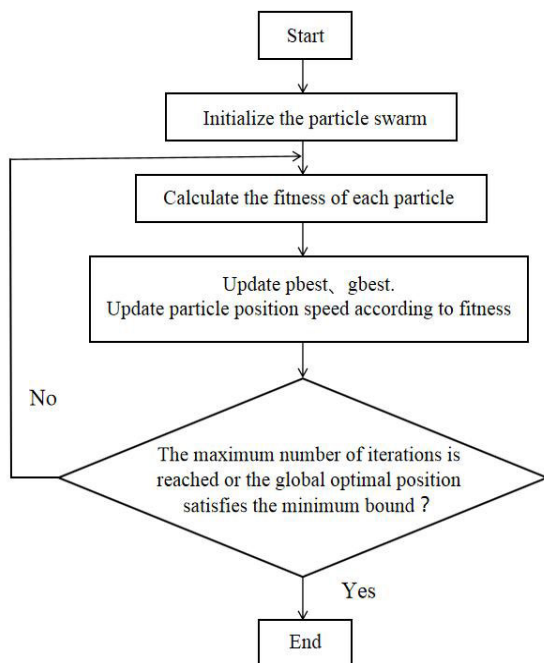

FIGURE 3. PSO algorithm optimization process.

III. MULTI-CHANNEL INFORMATION FEATURE FUSION METHOD

At present, most experts and scholars only predict the remaining life of the tool for one signal [33], which cannot contain all the wear information of the tool, so the monitoring accuracy is low. For the problem of single cutting signal, the algorithm can be used to fuse multiple signals and generate fusion samples containing all wear information, but attention should be paid to the way and method of signal fusion. The section mainly uses sensor technology to collect signals related to tool wear, and introduces the method of multi-channel information feature fusion, which mainly includes the monitoring and collection of original signals, the extraction and fusion of multi-channel features, and the construction of RUL prediction model sample data set.

A. RAW SIGNAL MONITORING AND ACQUISITION

Tool residual life monitoring has been recognized as an important method to prevent excessive tool wear and maintain part tolerance and surface quality during milling [34]. Its essence is to use sensor technology to collect the signals related to tool wear in real time, and use data-driven technology to capture the associated features of tool wear, so as to construct a reference model for feature monitoring. With the development of sensor technology, various sensors are used to monitor tool wear, such as vibration, AE, current, cutting force, sound, power and so on are widely used in tool wear monitoring, and the research mainly collects vibration, cutting force, AE signals for tool wear state monitoring.

(1) The vibration signal is caused by the periodic vibration of the cutting system composed of the workpiece or the tool, and the strength of the vibration between the systems is closely related to the wear state of the tool. The cutting vibration signal is collected by Kistler 8636C piezoelectric accelerometer adsorbed on the surface of the part to be processed. The original signal collected includes three channels of X axis, Y axis and Z axis vibration signal.

(2) The cutting force gradually increases with the increase of tool wear, so the cutting process can be monitored by monitoring the amplitude and change of the cutting force signal. And the sensitivity of the cutting force sensor can be made higher to perceive small changes in cutting force, so the use of cutting force monitoring tool wear has great practical significance. In this paper, the cutting force signal is collected by Kistler 8152 three-direction platform dynamometer and 5070 A charge amplifier. The original signal collected contains three channels of cutting force signal of X axis, Y axis and Z axis.

(3) As an online non-destructive monitoring technology, AE monitoring technology can effectively avoid the low frequency noise of machine tools in the processing environment, and its anti-interference ability is strong, so it is widely used in the field of tool wear monitoring [36]. In the research, the AE signal is collected by Kistler 9265B acoustic transmitter, and the original AE signal collected has only one channel.

To sum up, the signal acquisition platform for tool residual life monitoring is shown in Fig.4. The vibration signals of 3 channels, the cutting force signals of 3 channels and the original AE signals of 1 channel are collected respectively, with a total of 7 channels. In this experiment, the tool is walked once every Δt time, and the original signals of 7 channels can be collected each time. The number of collected
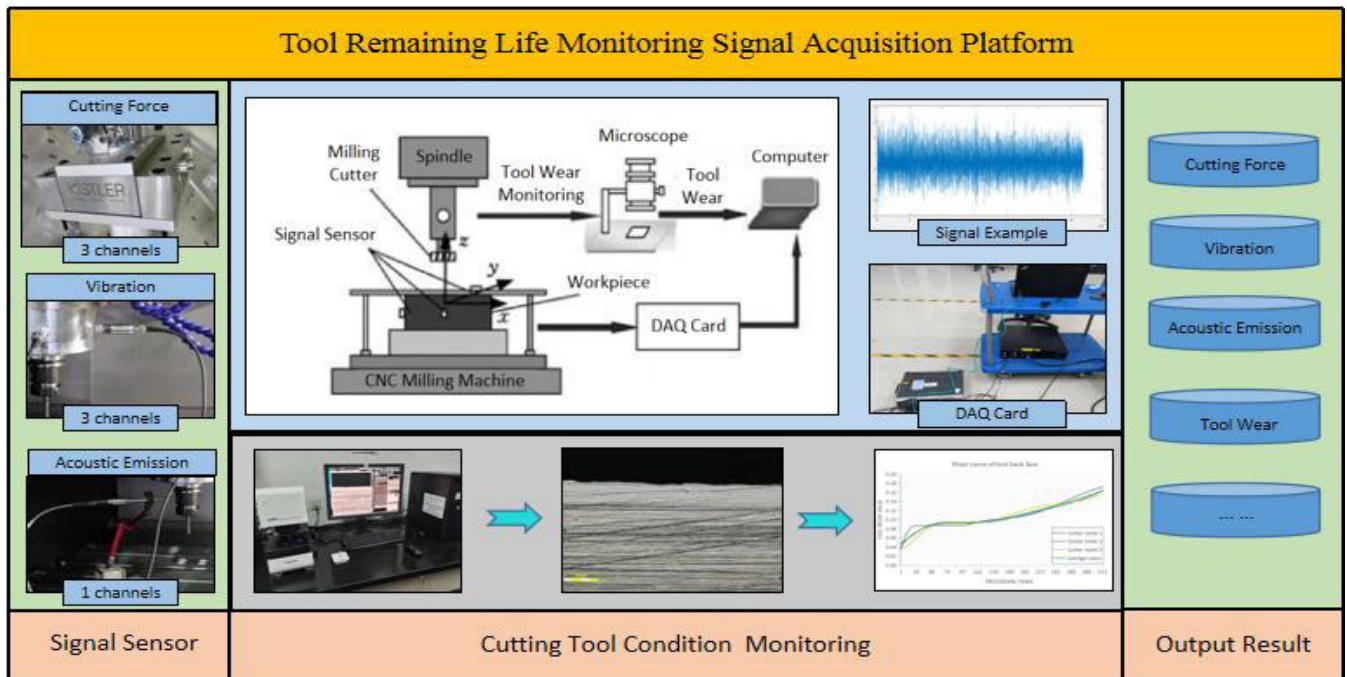
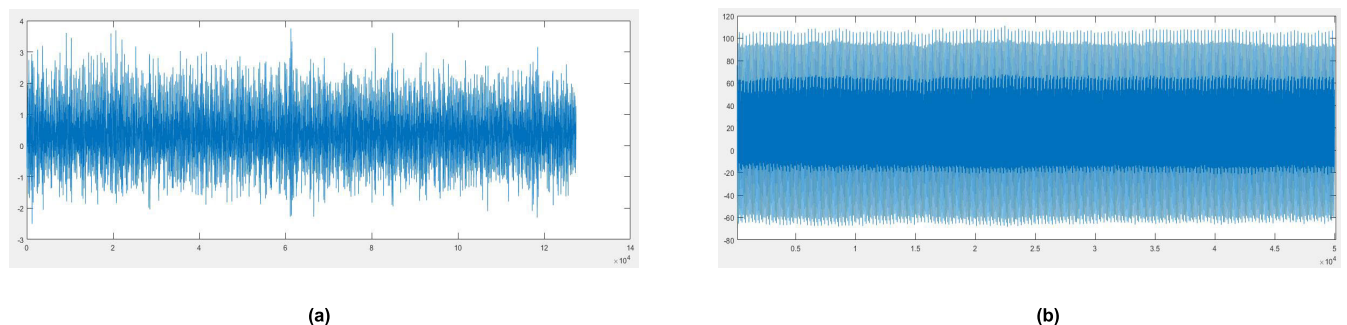**FIGURE 4.** Tool remaining life monitoring signal acquisition platform.

FIGURE 5. Comparison results between the original signal and the noise reduction signal Raw (a) signal data Signal data after noise reduction.

points of the original signals of a single walk in each channel is about more than 200000. It can be seen that the amount of signal data is huge, and there is a lot of noise, which is often caused by the instability of the system at the moment of cutting in and out of the tool. Therefore, it is necessary to denoise all kinds of raw signals collected above, so as to avoid adverse effects during model training.

The sampling points with data labels of 50001-100000 in the original signals of each acquisition are extracted as the research objects. Taking a single knife walk of a certain channel as an example, the comparison results between the original signal and the denoised signal are shown in Fig.5. It can be seen that the signal after noise reduction has uniform fluctuation and no noise. In the research, the full life cycle experiment of three ball-end milling cutters is carried out, and the number of cutter removal in each experiment is 315 times, with a total of 7 channels. Therefore, the original signal of

each tool can form a 315×7 original signal matrix of tool wear after noise reduction.

B. EXTRACTION AND FUSION OF MULTI-CHANNEL FEATURES

The time domain feature of the signal is to unlimitlessly expand a certain period of the milling process, and find and analyze the law of the related variables changing with time. Although the collected signals have continuous waveforms, it is difficult to directly extract the tool wear related features from the original signals due to the high sampling frequency and frequent noise interference. Therefore, time domain analysis is necessary. Time domain analysis is to calculate the relevant parameters of the original signal and analyze the data, so as to make the extracted time domain features more representative. In the research, to realize the intelligent prediction of tool life, the time domain characteristics of the

TABLE 1. Extraction of time-domain features of the original signal.

Serial number	Time domain feature	Calculation method
1	Average value	$\bar{x} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i$
2	Standard deviation	$x_{std} = \sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$
3	Mximum value	$X_{\max} = \max\{x_i\}$
4	Minimum value	$X_{\min} = \min\{x_i\}$
5	Peak-to-peak value	$X_p = X_{\max} - X_{\min}$
6	Root mean square,	$x_{RMS} = \sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n} x_i^2}$
7	Skewness	$X_s = \dfrac{1}{n}\sum\limits_{i=1}^{n}\dfrac{\left(\lvert x_i\rvert - \bar{x}\right)^3}{\sigma^3}$
8	Kurtosis	$X_k = \dfrac{1}{n}\sum\limits_{i=1}^{n}\dfrac{\left(\lvert x_i\rvert - \bar{x}\right)^4}{\sigma^4}$
9	Amplitude factor	$C = \dfrac{X_{max}}{X_{RMS}}$
10	Waveform factor	$S = \dfrac{X_{RMS}}{\dfrac{1}{n}\sum\limits_{i=1}^{n}\lvert x_i\rvert}$
11	Impact factor	$I = \dfrac{X_{\max}}{\dfrac{1}{n}\sum\limits_{i=1}^{n}\lvert x_i\rvert}$
12	Margin factor	$L = \dfrac{X_{\max}}{\left(\dfrac{1}{n}\sum\limits_{i=1}^{n}\sqrt{\lvert x_i\rvert}\right)^2}$
13	Energy	$E = \sum\limits_{i=1}^{n}\lvert x_i\rvert$

TABLE 2. Extraction of frequency domain features of the original signal.

Serial number	Frequency domain feature	Calculation method
1	The frequency domain amplitude average	$\bar{P} = \dfrac{1}{n}\sum\limits_{i=1}^{n} P(f_i)$
2	The center of gravity frequency	$P_c = \dfrac{\sum\limits_{i=1}^{n} f_i \cdot P(f_i)}{\sum\limits_{i=1}^{n} P(f_i)}$
3	The mean square frequency	$P_m = \dfrac{\sum\limits_{i=1}^{n} f_i^2 \cdot P(f_i)}{\sum\limits_{i=1}^{n} P(f_i)}$
4	The variance frequency	$P_{RMS} = \sqrt{P_m}$
5	The frequency variance	$P_V = \dfrac{\sum\limits_{i=1}^{n}(f_i - P_c)^2 \cdot P(f_i)}{\sum\limits_{i=1}^{n} P(f_i)}$

original signal extracted mainly include 13 kinds, which are the average value, standard deviation, skewness, kurtosis, maximum value, minimum value, peak-to-peak value, root mean square, amplitude factor, waveform factor, impact factor, margin factor, energy, and its calculation method is shown in Table 1.

The frequency domain feature of the signal is to describe the rules between the related variables of the observed signal in terms of frequency, which is more profound and convenient than time domain analysis. Fourier Transform is the most commonly used method of frequency domain analysis, its essence is to convert the signal in the time domain to the frequency domain, by extracting the spectrum characteristics of the sample signal for tool life prediction. When the

wear degree of the cutter changes in the milling process, the frequency component of the signal spectrum will change accordingly. Therefore, by analyzing the frequency domain characteristics, the signal spectrum information can be accurately characterized, so as to know whether the cutter is in a healthy state. The frequency domain features extracted in this experiment mainly include 5 kinds, namely, the frequency domain amplitude average, the center of gravity frequency, the mean square frequency, the variance frequency, and the frequency variance. Let f_i be the frequency spectrum converted from the time domain signal (i.e., the original signal) by fast Fourier transform (FFT), and $P(f_i)$ be the power spectral density, then the calculation methods of the 5 frequency domain features are shown in Table 2.

Due to the change of geometric characteristics or process parameters of machining parts, the signal collected by the sensor will have instantaneous mutation in the process of tool wear signal monitoring, so it is also necessary to analyze the signal in the time-frequency domain. In the research, wavelet packet analysis is used to sample the high frequency signal and low frequency signal in the process of layer by layer decomposition. After high and low frequency signal decomposition, the low frequency and high frequency part have the same resolution, and the signal is subdivided into different frequency bands. With the change of tool wear state, the frequency band structure of the monitoring signal will also change, resulting in the change of energy parameters in different frequency bands. Therefore, the energy of each frequency band can accurately characterize the wear degree

of the tool [36]. The energy value of the frequency band is calculated as shown in Equation (11):

$$E_m = \frac{1}{M} \sum_{t=1}^{M} W_t^2 \tag{11}$$

where, E_m is the energy value of the frequency band; M is the number of frequency bands; W_t is the wavelet packet coefficient of the tth frequency band.
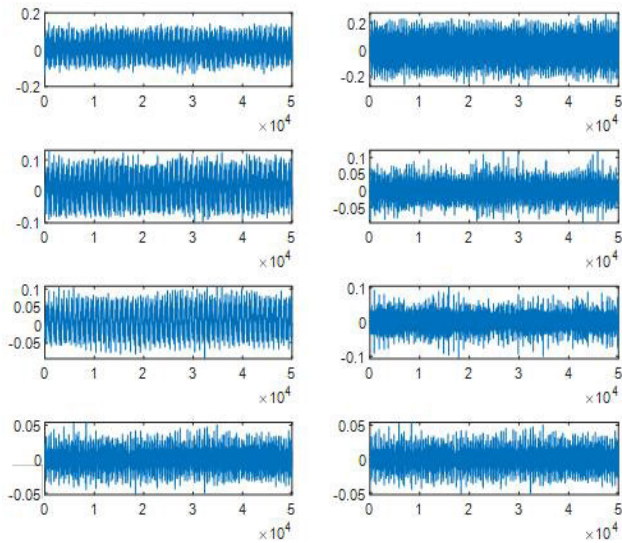

FIGURE 6. Frequency bands for wavelet packet decomposition.

The tool wear signal is decomposed by wavelet packet according to the above principle, and the number of decomposed layers is set to three layers, which are completed by db5 wavelet basis. Then the decomposed signals of each layer are reconstructed according to wavelet coefficients for more accurate analysis. Due to the orthogonality of the wavelet packet basis, the energy of the frequency band can be characterized by the wavelet packet coefficients of each frequency band. As shown in Fig.6, the wear signal of a single tool walk in a channel is decomposed by three levels, and the frequency domain is divided into 8 frequency bands, so that 8 time-frequency domain features are extracted.

In summary, the feature quantities related to the tool wear state are extracted from the 315 × 7 tool wear signal matrix in the time domain, frequency domain and time-frequency domain, respectively. As shown in Fig.7, 13 time domain features, 5 frequency domain features, and 8 time-frequency domain features were extracted from each channel signal respectively, and a total of 26 feature values were extracted. The features of 7 channels of tool wear information were extracted and fused respectively, and 182 features related to tool wear state were finally obtained. After matrix recombination, a 315 × 182 fusion feature matrix was output.

C. CONSTRUCTION OF SAMPLE DATA SET FOR RUL PREDICTION MODEL

The ball-end milling cutter used in the experiment has three teeth. To improve the precision and accuracy of tool wear prediction, the wear of the tool face behind the three teeth is measured after each tool move, and the average value is taken to characterize the actual wear of the tool. Since the number of tool moves is 315 times for each tool during signal acquisition, the average value of the measured wear amount after each tool move constitutes a sample target matrix, and the dimension of the matrix is 315 × 1. Each value in the matrix is the output data of the CNN-LSTM-PSO prediction model of the current tool.

The fitting operation speed of the tool remaining life prediction model is closely related to the number of features. The more features the model is, the more complex the model is, and the slower the operation is. According to the above mentioned, the feature matrix of 315 × 182 is obtained after the extracted features are fused by multi-channel features, but not all the features can characterize the wear amount of the tool face. To find the correlation between the feature matrix and the target matrix more clearly, the above multi-channel fusion feature matrix and tool wear values are normalized, and the normalization processing formula is shown in Equation (12):

$$X_n = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{12}$$

where, X is the sample of each feature, $X_{\min}$ is the minimum value of the sample feature, and $X_{\max}$ is the maximum value of the sample feature.

Taking the X-axis cutting force signal of a tool as an example, the correlation between the sample data after normalization and the tool wear curve is analyzed. It can be seen from Fig.8 that there are many features that are not correlated with the tool wear value or the features with weak correlation will interfere with the tool wear prediction model and should be deleted. The Pearson correlation coefficient is the most widely used correlation coefficient analysis method, which can be used to measure the correlation between the extracted feature values and the tool wear [37]. Its calculation formula is shown in Equation (13):

$$P_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - \left(\sum x_i\right)^2} \sqrt{n \sum y_i^2 - \left(\sum y_i\right)^2}} \tag{13}$$

where P_{xy} represents the Pearson correlation coefficient of signal feature x and tool wear value y. Where n represents n groups of signal values, x_i represents the ith value of signal feature value, and y_i represents the ith value of tool wear.

Pearson correlation coefficient formula was used to calculate the correlation between the above 182 features and the tool wear value. Fig.9 shows the correlation of each feature's Pearson coefficient. In the research, the feature less than 0.5 is counted as weak correlation, and there are 48 feature values in total. The features between 0.5 and 0.9 is counted as moderate correlation, with a total of 87 features. And greater
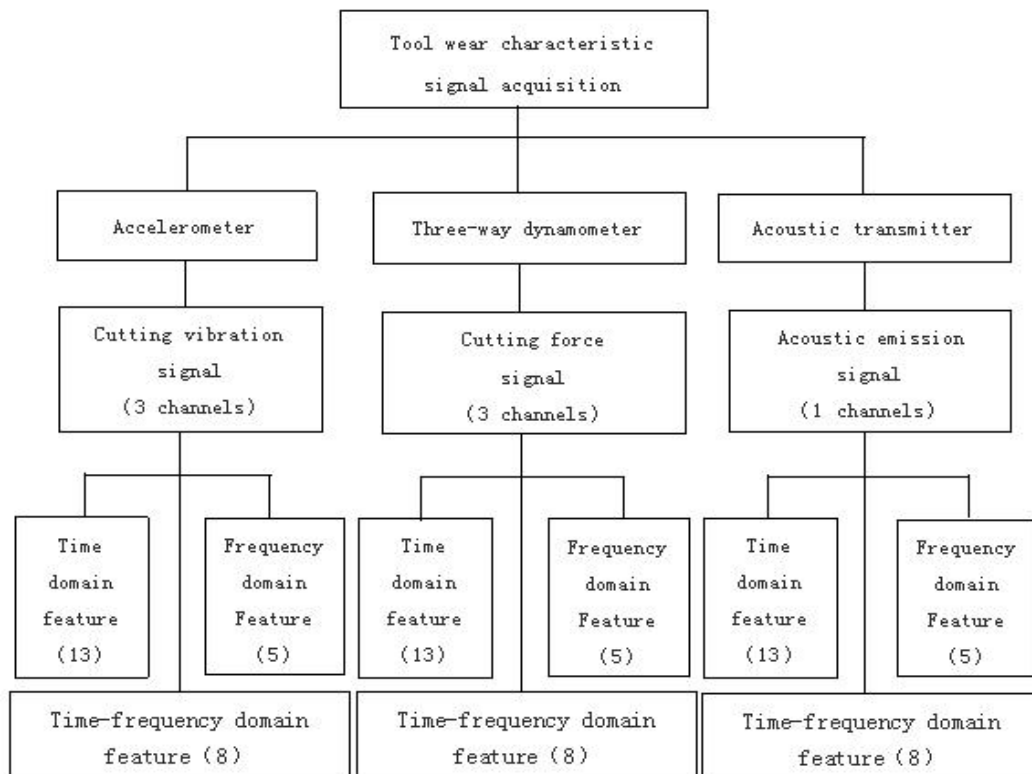
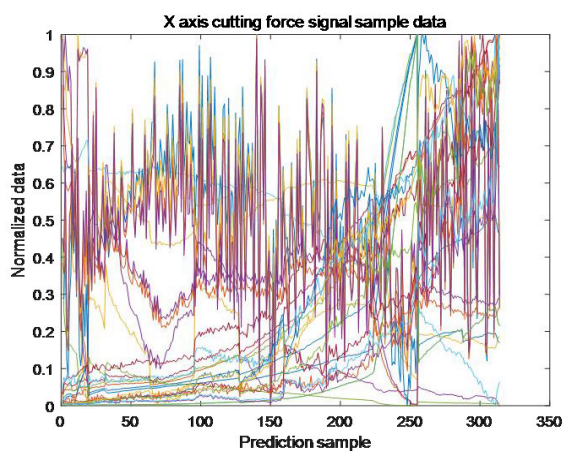**FIGURE 7.** Wear feature extraction and fusion scheme.

FIGURE 8. Normalised sample dataset.

than 0.9 are considered to be strongly correlation, with a total of 47 features. It can be found from Fig.9 that the correlation between the extracted features of AE signals and tool wear values is not high, which indicates that the signal-to-noise separation of AE signals is difficult and feature extraction is difficult. This conclusion indirectly indicates that it is necessary to use multi-sensor information as the monitoring of tool remaining useful life prediction to avoid the influence of information outliers of a single sensor on the prediction results.

Taking a tool as an example, the multi-channel fusion feature matrix of 315×182 is filtered and reduced. Irrelevant features or features with weak correlation are deleted, and only 47 strong correlation features are retained, so as to improve the speed of model operation and avoid overfitting. Fig.10 shows the sample data of X-axis cutting force signal after dimension reduction. Compared with the original sample data set in Fig.8, the uncorrelated or weakly correlated features are deleted, and the 7 strongly correlated features are retained, which shows that the effect is remarkable. Therefore, 47 strongly correlated features are fused and reconstituted, and the dimension of the sample feature matrix is 315×47. The sample feature matrix is the input data of the current CNN-LSTM-PSO prediction model for cutting tools. Because there are three tools in this experiment, a tool wear sample data set with spatial-temporal correlation of traffic flow can be constructed by the above method, and its dimension is $315 \times 48 \times 3$.

IV. EXPERIMENTAL VERIFICATION AND ANALYSIS OF TOOL RESIDUAL LIFE

In the section, the RUL prediction method of CNN-LSTM-PSO tool based on multi-channel feature fusion proposed in the research is verified and analyzed by taking the five-axis CNC machining center as the experimental platform. It mainly includes the experimental conditions of tool remaining life, the selection of RUL prediction model
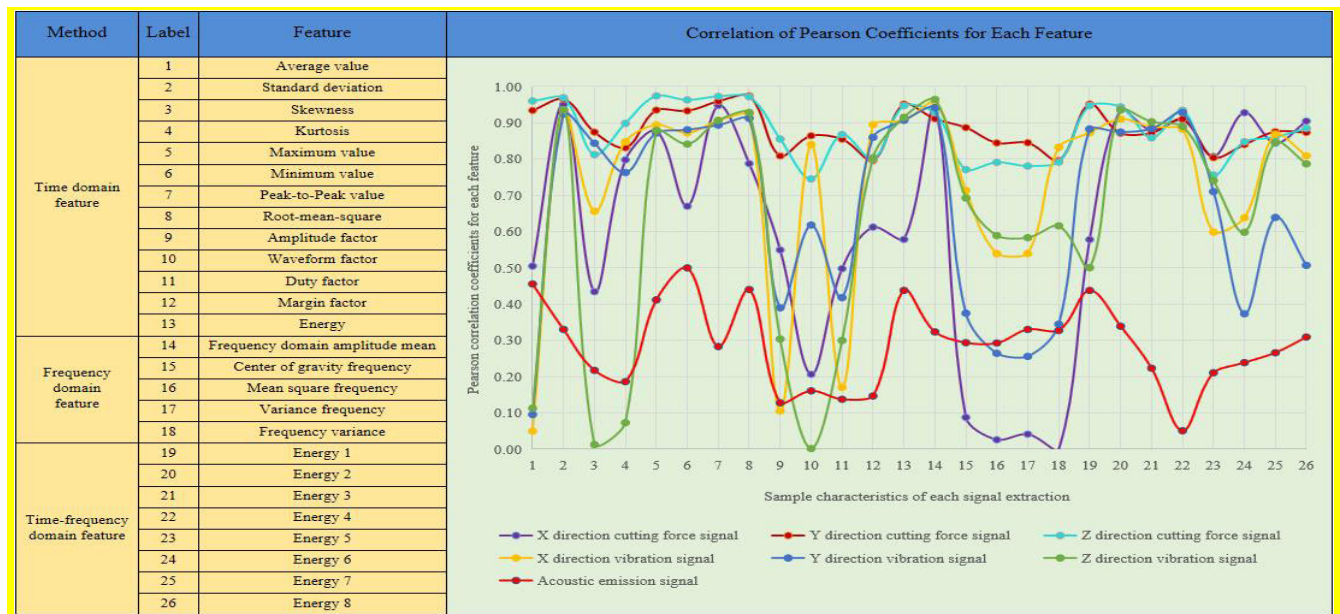
Method	Label	Feature	Correlation of Pearson Coefficients for Each Feature
Time domain feature	1	Average value	
	2	Standard deviation	
	3	Skewness	
	4	Kurtosis	
	5	Maximum value	
	6	Minimum value	
	7	Peak-to-Peak value	
	8	Root-mean-square	
	9	Amplitude factor	
	10	Waveform factor	
	11	Duty factor	
	12	Margin factor	
	13	Energy	
Frequency domain feature	14	Frequency domain amplitude mean	
	15	Center of gravity frequency	
	16	Mean square frequency	
	17	Variance frequency	
	18	Frequency variance	
Time-frequency domain feature	19	Energy 1	
	20	Energy 2	
	21	Energy 3	
	22	Energy 4	
	23	Energy 5	
	24	Energy 6	
	25	Energy 7	
	26	Energy 8	

FIGURE 9. Correlation of pearson coefficients for each feature.
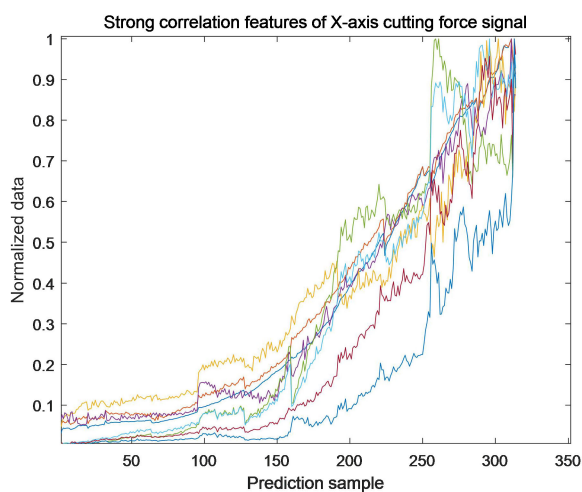
FIGURE 10. Sample data after dimensionality reduction.

FIGURE 11. Experimental conditions for tool wear.

parameters and the prediction results of CNN-LSTM-PSO model.

A. EXOERIMENTAL CONDITIONS FOR TOOL REMAINING LIFE

The experimental conditions of tool wear are shown in Fig.11. The cutting vibration signal is collected by Kistler 8636C piezoelectric accelerometer, the cutting force signal is collected by Kistler 8152 three-direction platform dynamometer, and the AE signal is collected by Kistler 9265B acoustic transmitter.

According to the above experimental conditions, the tool wear measurement test was carried out on three ball end milling cutters. In the pro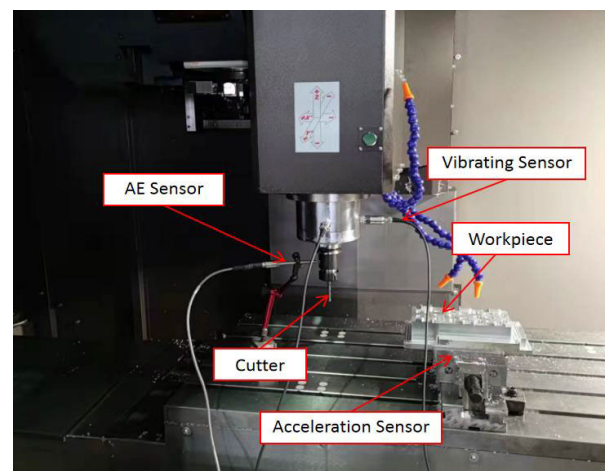cess of machining, the spindle speed was 10400RPM, the feed rate was 0.001mm, the feed rate was set to 1555 mm/min, the tool side feed rate was 0.125mm, and the tool back feed rate was 0.2mm. The shape of the milling part is cuboid, the round-trip milling end face is used, the surface length of the milling part is about 108mm, and the processing process does not use cutting fluid, the specific CNC machining cutting parameters are shown in Table 3.

Since the milling cutter used in the experiment has three teeth, LEICA MZ12 microscope was used to measure the wear value of the three teeth in real time after each knife move. The tool wear change curve is shown in Fig.12, the purple curve is the wear amount of the first tooth, the blue curve is the wear amount of the second tooth, and the yellow curve is the wear amount of the third tooth. It can be seen from the figure that the tool wear is fast in the early stage,
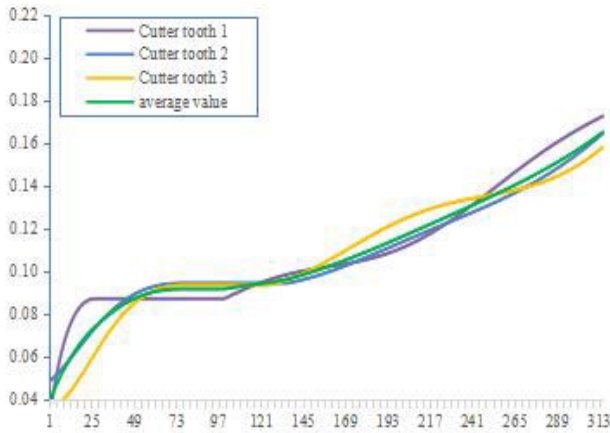
FIGURE 12. Test tool wear variation curve.

TABLE 3. CNC machining cutting parameter.

Hardware Conditions	Models and main parameters	Cutting Conditions	Parameters
CNC Milling Machines	CNC milling machines Roders Tech RFM760	Spindle speed	10400RPM
Force Gauges	Three-way force gauges Kistler 9265B	Feed rate	1555 mm/min
Charge amplifier	Multi-channel charge amplifier Kistler 5019A	Back draft	0.2 mm
Milling Materials	Cube Inconel 718	Side draft	0.125 mm
Knives	Ball end carbide milling cutters 3 teeth	Feeds	0.001 mm
Data Acquisition Cards	Data Acquisition Cards NI DAQ	Sampling frequency	50 KHz
Wear Gauges	Microscope LEICA MZ12	Cooling conditions	Dry cutting

gentle in the middle stage, and accelerated in the later stage, which is consistent with the related theory of tool wear, which indirectly verifies the accuracy of the data set.

B. SELECTION OF RUL PREDICTION MODEL PARAMETERS
In the research, the initial learning rate, the number of hidden layer units and the L2 regularization coefficient in the CNN-LSTM model are selected as the optimization objects of the PSO algorithm, to improve the nonlinear fitting performance and prediction accuracy of the RUL prediction model. To avoid the influence of external factors, the number of particle swarm individuals in PSO algorithm was set to 10, the minimum Root Mean Square Error was used as the fitness function, and the maximum number of iterations was
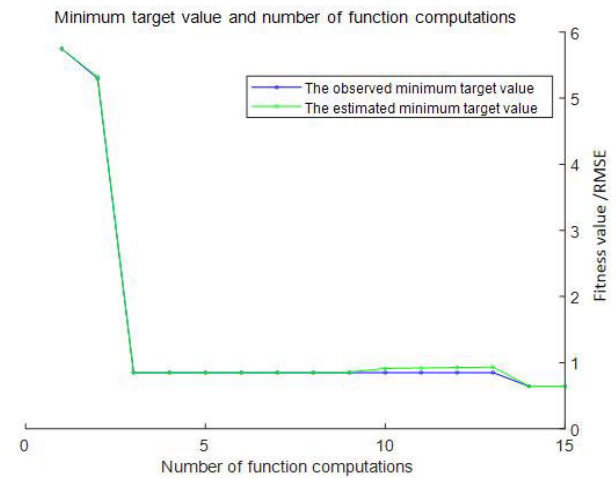

FIGURE 13. The process of PSO algorithm optimizing the hyperparameters of CNN-LSTM model.

set to 15. The initial learning rate parameter of the optimized CNN-LSTM model was set between 0.001 and 1, the number of hidden layer units was set between 1 and 100, and the L2 regularization coefficient was set between 1E-10 and 1E-2. The hyperparameter optimization process of its PSO algorithm is shown in Fig. 13. It can be found that with the increase of the number of iterations, the Root Mean Square Error value of CNN-LSTM model decreases rapidly in the first three iterations and tends to be stable in the later stage, and the global optimal hyperparameter collocation appears in the 14th iteration. The optimal hyperparameter collocation is the initial learning rate value of 0.0091, the number of hidden layer units is 46, and the L2 regularization coefficient value is 1.04E-10.

To quantify the prediction performance of tool life model, it is necessary to develop evaluation indexes. In the research, three objective evaluation indexes are selected, which are MAE value, RMSE valueand determination coefficient R^2. The MAE value can obtain an evaluation value, but only through the comparison between different models can reflect the quality of the model. Both mean square error (RMSE) and coefficient of determination R^2 can directly characterize the quality of the model. The smaller the RMSE value is, the closer the coefficient of determination R^2 value is to 1, the higher the precision and accuracy of the prediction model. The calculation formula of the three evaluation indicators is as follows:

$$MAE = \frac{\sum_{t=1}^{m} |y_t - \hat{y}_t|}{m} \tag{14}$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^{m} (y_t - \hat{y}_t)^2}{m}} \tag{15}$$

$$R^2 = 1 - \frac{\sum_{t=1}^{m} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{m} (y_t - \bar{y}_t)^2} \tag{16}$$

**TABLE 4.** Influence of the PSO algorithm on the prediction model.

Algorithm	Number of Units	Initial Learn Rate	L2 Regularization	Test set prediction results		
				MAE	RMSE	R^2
CNN-LSTM	10	0.001	1.0E-02	8.3887	10.8422	0.7378
	20	0.01	1.0E-04	4.1106	4.8654	0.9499
	30	0.01	1.0E-06	1.8552	2.8124	0.9833
	40	0.1	1.0E-08	4.9347	5.9053	0.9261
CNN-LSTM-PSO	46	0.0091	1.04E-10	1.0892	1.3520	0.9961

where m is the number of samples output by the fully connected layer, the number of samples in the research is 315, $\hat{y}_t$ is the predicted value of tool wear, and y_t is the actual value of tool wear.

Table 4 shows the influence results of PSO algorithm on tool RUL prediction model, in which the number of hidden layer units, initial learning rate, L2 regularization coefficient and other hyperparameters of CNN-LSTM model are selected manually at random, and it can be seen that the CNN-LSTM model optimized by PSO algorithm has the best prediction effect. Compared with the CNN-LSTM model, the MAE value and RMSE valueare reduced, and the determination coefficient R^2 is improved. The R^2 performance index of the tool RUL prediction model based on PSO algorithm to optimize hyperparameters is more than 0.99, while the R^2 performance index of the CNN-LSTM model with manually selected parameters is between 0.7378 and 0.9833, which shows that its fluctuation range is wide and the prediction results are unstable. This is mainly because after optimizing the hyperparameters of the CNN-LSTM model, the PSO algorithm obtains more accurate hyperparameter collocation, finds the most critical attributes that affect the accuracy of tool wear prediction, avoids the blindness of parameter setting, and thus improves the prediction effect. Other structural parameters of the tool RUL prediction model are shown in Table 5.

C. PREDICTION RESULTS OF CNN-LSTM-PSO MODEL

In the research, the CNN-LSTM-PSO model with multichannel feature fusion is used for tool residual life regression prediction. The fitting effect of all samples is shown in Fig. 14, which shows that the CNN-LSTM model optimized by PSO algorithm has strong generalization ability and robustness. Fig.15 shows the test error of CNN-LSTM-PSO model test set samples. The determination coefficient R^2 value of tool RUL prediction model is 0.9961, the RMSE value is 1.3520, and the MAE value is 1.0892. In summary, the CNN-LSTM-PSO model can effectively predict
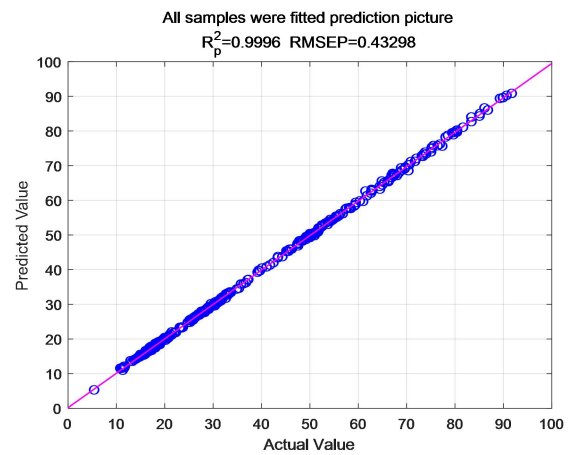
FIGURE 14. Fitting effect diagram of the sample.
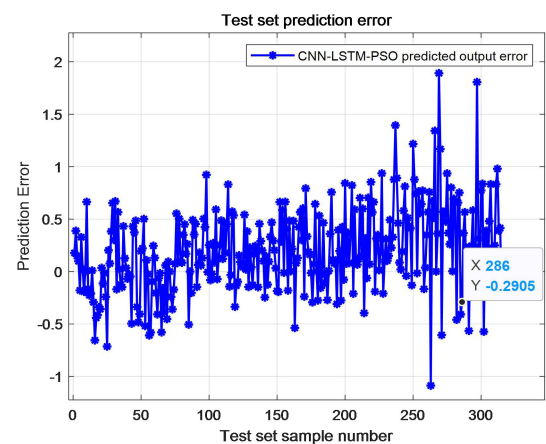
FIGURE 15. Test error of test set sample.

the remaining useful life of the tool, and has achieved good results.

To further verify the prediction performance of CNN-LSTM-PSO model, the ablation experiment was carried out

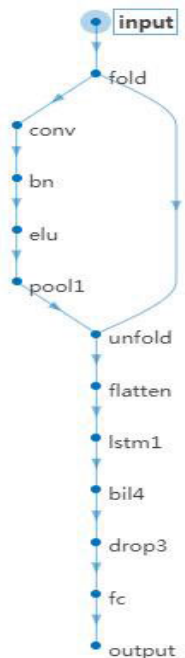**TABLE 5.** Structural parameters of CNN-LSTM-PSO model.

| Flow chart	Structural section	Network structure Name	Parameter settings
	1	Input layer（input）	Training set: 315×48×2 Test set: 315×48×1
	2	Convolutional layer （conv） Batch standardisation layer（bn） Pooling layer（pool1）	Activation function: RELU Initial Learn Rate: 0.0091 L2 Regularization: 1.04E-10 Maximum pooling
	3	LSTM layer 1（lstm1）	Number of hidden layer units: 50 Activation function: Sigmoid
	4	LSTM layer 2（bil4）	Number of hidden layer units: 46 Activation function: Sigmoid
	5	Dropout layer（drop3）	25% discard
	6	Output layer（output）	Activation function: Softmax

TABLE 6. Comparison of prediction performance of five models.

Algorithm	Model fitting time	Test set prediction results		
		MAE	RMSE	R^2
CNN-PSO Algorithms	118.26s	3.0572	3.8331	0.9689
LSTM-PSO Algorithms	114.95s	2.1770	3.0852	0.9798
CNN-LSTM Algorithm	110.13s	1.8552	2.8124	0.9833
CNN-LSTM-PSO Algorithm	94.76s	1.0892	1.3520	0.9961

to compare the prediction model with CNN-LSTM model, LSTM-PSO model and CNN-PSO model. The experimental results are shown in Fig.16, and it can be seen from Fig.16. The prediction effect of four tool RUL prediction models is CNN-LSTM-PSO > CNN-LSTM > LSTM-PSO > CNN-PSO, which indicates that CNN model, LSTM model and PSO algorithm all play a role in the process of tool RUL prediction, and none of them is dispensable. Although the prediction effect of CNN-LSTM model is better than that of CNN-PSO model and LSTM-PSO model, the hyperparameters in the model are not optimized by PSO and are not the global optimal solution, so the prediction effect of CNN-LSTM-PSO model is not as good as that of CNN-LSTM-PSO model. This result once again proves the role of PSO algorithm in the process of tool RUL prediction. The tool wear curve predicted by CNN-LSTM-PSO model is closer to the real tool wear curve than the other three models, and the fluctuation range of the error curve is the smallest. In summary, the prediction performance of the CNN-LSTM-PSO tool remaining life prediction model proposed in the research has certain advantages.

The comparison of the prediction performance of the four models is shown in Table 6. It is found that the CNN-LSTM-PSO model with multi-channel feature fusion for tool wear prediction has the shortest prediction time compared with the CNN-LSTM model, LSTM-PSO model and CNN-PSO
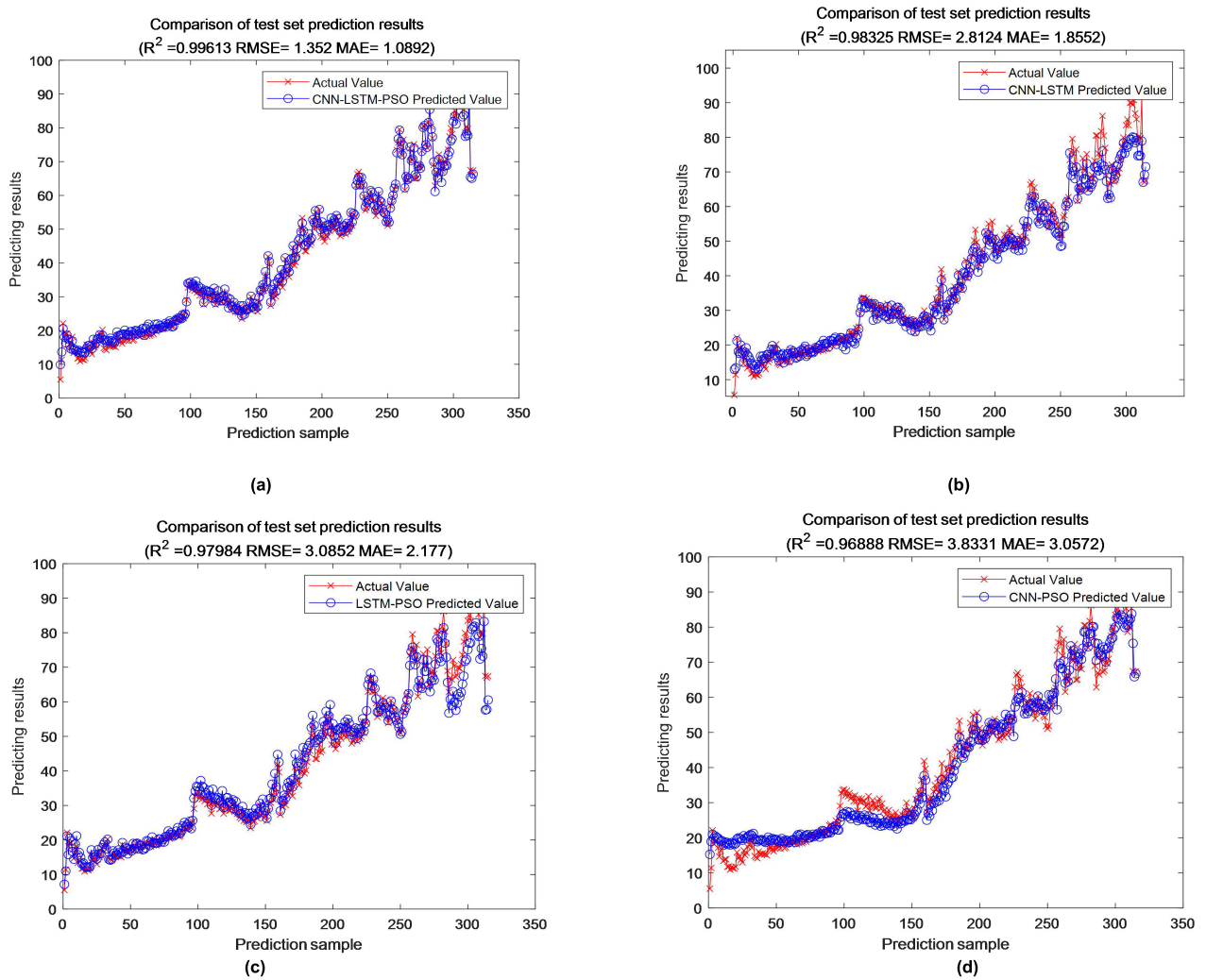
FIGURE 16. Prediction results of the four traditional models. (a) CNN-LSTM-PSO model.(b) CNN-LSTM model.(c) LSTM-PSO model.(d) CNN-PSO model.

model. Decreased by 13.96%, 17.56% and 19.87%, respectively. The RMSE value was the smallest, reduced by 41.29%, 49.97% and 64.37%, respectively. The MAE value was the smallest, reduced by 51.93%, 56.18% and 64.73%, respectively. The value of determination coefficient R^2 is closest to 1, which is increased by 1.30%, 1.66% and 2.81% respectively. These four results once again prove that the CNN-LSTM-PSO model proposed in the research has a higher accuracy in predicting the residual life of the tool, and can more effectively realize the monitoring of the residual life of the tool and intelligent tool change.

V. CONCLUSION

The original signals of seven channels including vibration signal, cutting force signal and AE signal are collected by sensor technology. At the same time, the wear value of the tool face at 315 times of each tool was measured by microscope, and the sample data set of RUL model was constructed after multi-channel feature extraction and fusion. Then a

CNN-LSTM-PSO model based on multi-channel feature fusion was proposed to regression predict tool RUL in the milling process of 5-axis machining center. The comparative study of ablation experiments is carried out, and the results show that:

(1) The multi-channel feature fusion method is used to extract, fuse, batch normalize and reduce the dimensionality of the collected signals of seven channels, which improves the generalization ability of the model and avoids the influence of the information anomaly of a single channel on the prediction results.

(2) The PSO algorithm is used to optimize the parameters of the CNN-LSTM prediction model, and the optimal hyperparameter collocation is finally obtained as the initial learning rate value of 0.0091, the number of hidden layer units is 46, and the L2 regularization coefficient value is 1.04E-10, which reduces the subjective influence of artificial selection parameters and avoids the blindness of parameter setting. As a result, the prediction accuracy is improved.

(3) The CNN-LSTM-PSO model is used to predict the remaining life of the tool, and the determination coefficient R^2 value is 0.9961, the RMSE value is 1.3520, and the MAE value is 1.0892. This shows that the model can effectively predict the remaining life of the tool, and has achieved good results.

(4) Compared with the CNN-LSTM model, LSTM-PSO model and CNN-PSO model, the CNN-LSTM-PSO model proposed in this paper has the highest fitting efficiency, the MAE value and RMSE values are reduced, and the determination coefficient R^2 value is improved, which is closest to 1. This indicates that the constructed tool life prediction model has less error, better accuracy and better effect.

The CNN-LSTM-PSO tool life prediction model can be widely used in CNC machining tool life management and intelligent operation and maintenance in various factories. By predicting the remaining useful life of the tool in real time, the predictive maintenance of the CNC machining tool is realized. When the tool wear is in the critical threshold, the intelligent tool change is carried out in advance, which is in line with the development trend of intelligent control and network interaction in the future.

REFERENCES

[1] D. Y. Pimenov, A. Bustillo, S. Wojciechowski, V. S. Sharma, M. K. Gupta, and M. Kuntoglu, "Artificial intelligence systems for tool condition monitoring in machining: Analysis and critical review," *J. Intell. Manuf.*, vol. 34, no. 5, pp. 2079–2121, Mar. 2022.

[2] N. Hu, Z. Liu, S. Jiang, Q. Li, S. Zhong, and B. Chen, "Remaining useful life prediction of milling tool based on pyramid CNN," *Shock Vibrat.*, vol. 2023, pp. 1–14, Feb. 2023.

[3] J. Gokulachandran and R. Padmana, "Prediction of remaining useful life of cutting tools: A comparative study using soft computing methods," *Process Manag. Benchmarking.*, vol. 26, no. 2, pp. 156–181, Apr. 2018.

[4] B. M. Sarat and R. T. Babu, "Multi-sensor heterogeneous data-based online tool health monitoring in milling of IN718 superalloy using OGM (1, N) model and SVM," *Measurement*, vol. 199, Jun. 2022, Art. no. 111501.

[5] K. Zhang, H. T. Zhu, D. Liu, G. N. Wang, C. Z. Huang, and P. Yao, "A dual compensation strategy based on multi-model support vector regression for tool wear monitoring," *Meas. Sci. Technology.*, vol. 33, Jul. 2022, Art. no. 105601.

[6] I. O. Olalere and O. A. Olanrewaju, "Tool and workpiece condition classification using empirical mode decomposition with Hilbert–Huang transform of vibration signals and machine learning models," *Appl. Sciences.*, vol. 13, no. 4, pp. 141–152, Feb. 2023.

[7] X. Wang, Y. Zheng, Z. Zhao, and J. Wang, "Bearing fault diagnosis based on statistical locally linear embedding," *Sensors*, vol. 15, no. 7, pp. 16225–16247, Jul. 2015.

[8] D. Kong, Y. Chen, and N. Li, "Monitoring tool wear using wavelet package decomposition and a novel gravitational search algorithm-least square support vector machine model," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, vol. 234, no. 3, pp. 822–836, Feb. 2020.

[9] W. Q. Cao, P. Fu, and X. H. Li, "The diagnosis of tool wear based on EMD and GA-B-spline network," *Sensors Transducers*, vol. 156, no. 9, pp. 159–202, Sep. 2013.

[10] A. S. Pyatykh, A. V. Savilov, and S. A. Timofeev, "Method of tool wear control during stainless steel end milling," *J. Friction Wear*, vol. 42, no. 4, pp. 263–267, Jul. 2021.

[11] Z. Wang, S. Leng, T. Min, and G. Chen, "Analysis of AE characteristics of tool wear in drilling CFRP/Ti stacked material," in *Proc. MATEC Web Conf.*, vol. 211, Oct. 2018, Art. no. 03001.

[12] X. Pu, L. Jia, K. Shang, L. Chen, T. Yang, L. Chen, L. Gao, and L. Qian, "A new strategy for disc cutter wear status perception using vibration detection and machine learning," *Sensors*, vol. 22, no. 17, p. 6686, Sep. 2022.

[13] C. W. Han, K. B. Kim, S. W. Lee, M. B. G. Jun, and Y. H. Jeong, "Thrust force-based tool wear estimation using discrete wavelet transformation and artificial neural network in CFRP drilling," *Int. J. Precis. Eng. Manuf.*, vol. 3, pp. 1–10, Jul. 2021.

[14] W. Wei, R. Cong, Y. Li, A. D. Abraham, C. Yang, and Z. Chen, "Prediction of tool wear based on GA-BP neural network," *Proc. Inst. Mech. Eng., B, J. Eng. Manuf.*, vol. 236, no. 12, pp. 1564–1573, Oct. 2022.

[15] W. Q. Cao, P. Fu, and W. L. Li, "The diagnosis of tool wear based on RBF neural networks and D-S evidence theory," in *Proc. 3rd IEEE Int. Conf. Comput. Sci. Inf. Technol.*, vol. 7, Jul. 2010, pp. 429–431.

[16] S. Laddada, M. O. Si-Chaib, T. Benkedjouh, and R. Drai, "Tool wear condition monitoring based on wavelet transform and improved extreme learning machine," *Proc. Inst. Mech. Eng., C, J. Mech. Eng. Sci.*, vol. 234, no. 5, pp. 1057–1068, Mar. 2020.

[17] Y. Liang, S. S. Hu, W. S. Guo, and H. Q. Tang, "Abrasive tool wear prediction based on an improved hybrid difference grey wolf algorithm for optimizing SVM," *Measurement*, vol. 187, Jan. 2022, Art. no. 110247.

[18] H. Lee, H. Jeong, G. Koo, J. Ban, and S. W. Kim, "Attention RNN based severity estimation method for interturn short-circuit fault in PMSMs," *IEEE Trans. Ind. Electron.*, vol. 68, no. 4, pp. 3445–3453, Apr. 2021.

[19] S. R. Han and Y. S. Kim, "A fault identification method using LSTM for a closed-loop distribution system protective relay," *Int. J. Electr. Power Energy Syst.*, vol. 148, Jun. 2023, Art. no. 108925.

[20] Y. K. Zhou, Z. Y. Wang, and X. Zuo, "Identification of wear mechanisms of main bearings of marine diesel engine using recurrence plot based on CNN model," *Wear*, vol. 520, May 2023, Art. no. 204656.

[21] K. Ma, G. Wang, K. Yang, M. Hu, and J. Li, "Tool wear monitoring for cavity milling based on vibration singularity analysis and stacked LSTM," *Int. J. Adv. Manuf. Technol.*, vol. 120, nos. 5–6, pp. 4023–4039, May 2022.

[22] M. L. Lim, M. N. Derani, M. M. Ratnam, and A. R. Yusoff, "Tool wear prediction in turning using workpiece surface profile images and deep learning neural networks," *Int. J. Adv. Manuf. Technol.*, vol. 120, nos. 11–12, pp. 8045–8062, Jun. 2022.

[23] J. Luo and X. Zhang, "Convolutional neural network based on attention mechanism and bi-LSTM for bearing remaining life prediction," *Int. J. Speech Technol.*, vol. 52, no. 1, pp. 1076–1091, Jan. 2022.

[24] J. H. Liang and H. L. Gao, "Research on tool wear morphology and mechanism during turning nickel-based alloy GH4169 with PVD-TiAlN coated carbide tool," *Wear*, vol. 508, Nov. 2022, Art. no. 204468.

[25] X. Zhang, X. Lu, W. Li, and S. Wang, "Prediction of the remaining useful life of cutting tool using the Hurst exponent and CNN-LSTM," *Int. J. Adv. Manuf. Technol.*, vol. 112, pp. 2277–2299, Jan. 2021.

[26] S. Droste, T. Jansen, and I. Wegener, "Upper and lower bounds for randomized search heuristics in black-box optimization," *Theory Comput. Syst.*, vol. 39, no. 4, pp. 525–544, Jul. 2006.

[27] W. F. Lu, B. Y. Cai, and R. Gu, "Improved particle swarm optimization based on gradient descent method," in *Proc. CSAE*, Oct. 2020, pp. 121–126.

[28] O. Salih and K. J. Duffy, "Optimization convolutional neural network for automatic skin lesion diagnosis using a genetic algorithm," *Appl. Sci.*, vol. 13, no. 5, p. 3248, Mar. 2023.

[29] X. Zhang, Y. Jiang, and W. Zhong, "Prediction research on irregularly cavitied components volume based on gray correlation and PSO-SVM," *Appl. Sci.*, vol. 13, no. 3, p. 1354, Jan. 2023.

[30] P. V. Badiger, V. Desai, M. R. Ramesh, B. K. Prajwala, and K. Raveendra, "Cutting forces, surface roughness and tool wear quality assessment using ANN and PSO approach during machining of MDN431 with TiN/AlN-coated cutting tool," *Arabian J. Sci. Eng.*, vol. 44, no. 9, pp. 7465–7477, Sep. 2019.

[31] J. V. Viisainen, F. Yu, A. Codolini, S. Chen, L. T. Harper, and M. P. F. Sutcliffe, "Rapidly predicting the effect of tool geometry on the wrinkling of biaxial NCFs during composites manufacturing using a deep learning surrogate model," *Compos. B, Eng.*, vol. 253, Jan. 2023, Art. no. 110536.

[32] J. Shi, Y. Zhang, Y. Sun, W. Cao, and L. Zhou, "Tool life prediction of dicing saw based on PSO-BP neural network," *Int. J. Adv. Manuf. Technol.*, vol. 123, pp. 4399–4412, Dec. 2022.

[33] S. Y. Wong, J. H. Chuah, and H. J. Yap, "Technical data-driven tool condition monitoring challenges for CNC milling: A review," *Int. J. Adv. Manuf. Technol.*, vol. 107, nos. 11–12, pp. 4837–4857, Apr. 2020.

[34] S. Bagri, A. Manwar, A. Varghese, S. Mujumdar, and S. S. Joshi, "Tool wear and remaining useful life prediction in micro-milling along complex tool paths using neural networks," *J. Manuf. Processes*, vol. 71, pp. 679–698, Nov. 2021.

[35] R. Zhuo, Z. Deng, B. Chen, G. Liu, and S. Bi, "Overview on development of acoustic emission monitoring technology in sawing," *Int. J. Adv. Manuf. Technol.*, vol. 116, pp. 1411–1427, Sep. 2021.

[36] S. X. Wu, P. N. Li, Z. H. Yan, L. N. Zhang, X. Y. Qiu, and J. Yang, "Wavelet packet analyses of acoustic emission signal for tool wear in high speed milling," *Key Eng. Mater.*, vols. 589–590, pp. 600–605, Oct. 2013.

[37] Y. Li, Y. Xiang, B. Pan, and L. Shi, "A hybrid remaining useful life prediction method for cutting tool considering the wear state," *Int. J. Adv. Manuf. Technol.*, vol. 121, nos. 5–6, pp. 3583–3596, Jul. 2022.

SHUO WANG received the M.Sc. degree from the College of Mechanical Engineering, Shenyang University of Science and Technology, in 2014. After graduation, he engaged in the research of tool wear related technology and obtained the engineer title, in 2018. In 2020, he began to teach courses related to metal cutting principles and tools with the Yingkou Institute of Technology. In 2022, he was a lecturer. His current research interests include data mining, deep learning, and artificial intelligence, and he has achieved a range of achievements in the field of condition monitoring and fault diagnosis.

ZHENLIANG YU received the Ph.D. degree from the School of Mechanical Engineering and Automation, Northeastern University, in 2020. He is currently a Lecturer with the School of Mechanical and Power Engineering, Yingkou Institute of Technology. His current research interests include mechanical reliability analysis, intelligent operation and maintenance, and life prediction.

GUANGCHEN XU received the M.E. degree from the College of Mechanical Engineering, Zhejiang University of Technology. He is currently an Assistant Professor with the Mechanical and Power Engineering College, Yingkou Institute of Technology. He teaches courses related to detection and control technology. His current research interests include machinery dynamics, machine design, and machine tool vibration analysis.

FENGQIN ZHAO received the Ph.D. degree in engineering from the College of Agricultural Engineering, Shenyang Agricultural University, in 2002. She is currently a Professor and a Master Supervisor with the School of Mechanical and Power Engineering, Yingkou Institute of Technology. Her current research interests include mechanical control mechanism and system research, and machine tool fault diagnosis.

● ● ●