

# A multi-source information fusion method for tool life prediction based on CNN-SVM

Shuo WANG, Zhenliang YU\*, Peng LIU, Man Tong WANG

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

\*Corresponding Author: Zhenliang YU, email address: yuzhenliang\_neu@163.com

## Abstract:

For milling tool life prediction and health management, accurate extraction and dimensionality reduction of its tool wear features are the key to reduce prediction errors. In this paper, we adopt multi-source information fusion technology to extract and fuse the features of cutting vibration signal, cutting force signal and acoustic emission signal in time domain, frequency domain and time-frequency domain, and downscale the sample features by Pearson correlation coefficient to construct a sample data set; then we propose a tool life prediction model based on CNN-SVM optimized by genetic algorithm (GA), which uses CNN convolutional neural network as the feature learner and SVM support vector machine as the trainer for regression prediction. The results show that the improved model in this paper can effectively predict the tool life with better generalization ability, faster network fitting, and 99.85% prediction accuracy. And compared with the BP model, CNN model, SVM model and CNN-SVM model, the performance of the coefficient of determination R2 metric improved by 4.88%, 2.96%, 2.53% and 1.34%, respectively.

**Keywords:** CNN-SVM; tool wear; life prediction; multi-source information fusion

## 1 Introduction

CNC machining center is a set of high-tech, high precision, high efficiency in one of the high precision end equipment, specifically for processing complex curved parts, its key technology to improve the level of equipment manufacturing industry is of great significance. CNC machining center due to the complexity of the processing object tool wear more serious, when the tool wear exceeds a given threshold will greatly affect the accuracy of the workpiece processing, resulting in the processing of product quality is not up to standard, not only waste processing input time and economic losses, and even lead to machine accidents<sup>[1]</sup>. For complex curved parts with high precision machining requirements, how to make the tool wear before the critical threshold for intelligent tool change will be an important research direction for the future high-end manufacturing industry.

Currently, data-driven methods combining sensor monitoring data with machine learning algorithms are widely used for tool life prediction<sup>[2]</sup>. Monitoring data refers to the extraction of tool wear features in the time domain, frequency domain, and time-frequency domain using sensor technology to collect raw signals. However, most experts and scholars predict tool wear for only one

signal<sup>[3]</sup>, which often has a low prediction accuracy, so in this paper, three signals, cutting vibration signal, cutting force signal and acoustic emission signal, are collected in real time, and a multi-source information fusion strategy is adopted to fuse the features extracted from each signal and construct a sample feature matrix, so as to improve the accuracy of tool life prediction. Machine learning algorithm is to use the extracted tool wear features as the input of the model, simulate the whole process of tool life degradation, and compare the current working state with the historical data to complete the prediction of the remaining tool life<sup>[4]</sup>. Common machine learning algorithms include BP neural network, RBF neural network, Support vector machine (SVM) etc.

Wei Weihua<sup>[5]</sup> et al. optimized the BP neural network by genetic algorithm, which improved the optimization and learning ability of the model and ensured the efficiency and accuracy of tool wear recognition. Weiqing Cao<sup>[6]</sup> et al. diagnosed the tool wear fault by fusing the information of RBF neural network and D-S evidence theory, and the experiment showed that the model could effectively diagnose the tool wear fault and its prediction accuracy was improved. The model can effectively diagnose the tool wear fault and its prediction accuracy is improved. Zhang Kun<sup>[7]</sup> et al. constructed DCM-SVR model to predict the tool wear value of

machining process, which can correct the systematic error online and compensate the predicted value, and the results of comparison with other methods show that the prediction performance of DCM-SVM is improved by 28.7% and the root mean square error is decreased by 64.7%. Although the traditional tool life prediction methods have achieved certain results, the prediction model can only tap the shallow features of the sample data, the generalization ability is insufficient, the network fitting speed is slow, and it relies more on signal processing techniques and expert experience.

In 2006, Hinton<sup>[8]</sup> et al. proposed the theory of deep learning, and convolutional neural network (CNN) is a typical representative of deep learning. Convolutional neural networks (CNNs) have powerful feature extraction ability, can adaptively mine the deep features of the input data, and get rid of the model's over-reliance on signal processing techniques and expert experience, so they have been widely used and researched by scholars in recent years. P. K. Ambadekar<sup>[9]</sup> et al. established a tool life prediction system using CNN convolutional neural networks, where an inverted microscope regularly takes The images of the tool were used as input and different categories of tool wear were used as output, and the results showed that the accuracy of the prediction using CNN model reached 87.26%, which can meet the practical needs of production. However, the output layer of the convolutional neural network (CNN) generally consists of a fully-connected layer and a Softmax layer. When dealing with data with a high degree of nonlinearity, the number of features in the output of the fully-connected layer increases proportionally, which can cause the overfitting phenomenon<sup>[10]</sup>; moreover, the prediction performance of the Softmax layer is not as good as that of the support vector machine (SVM) in dealing with regression problems. Therefore, the combination of convolutional neural network (CNN) and support vector machine (SVM) can make up for the shortcomings of the above CNN model.

CNN-SVM is a tool life regression prediction model proposed by combining convolutional neural network (CNN) and support vector machine (SVM) methods, but if we want to continue to improve the model prediction performance we need to optimize the model hyperparameters, such as penalty parameter  $p$  and kernel function width  $g$ , etc. Currently, the more common parameter optimization methods include manual parameter tuning, random optimization<sup>[11]</sup>, gradient-based optimization<sup>[12]</sup>, and genetic algorithm optimization<sup>[13]</sup>. The genetic algorithm is scalable and easy to combine with other algorithms, and it can achieve fast optimization with less computation time and high robustness when computational accuracy is required, so it has attracted a lot of attention from scholars in the field of hyperparametric optimization in recent years<sup>[14]</sup>.

Therefore, this paper uses a CNC machining center as a platform to collect cutting vibration signals, cutting force signals and acoustic emission signals of tools under

different wear states in real time using sensor technology, and proposes a tool life prediction model based on CNN-SVM optimized by genetic algorithm (GA). The model uses CNN convolutional neural network as a feature learner and SVM support vector machine as a trainer for regression prediction. The powerful computational capabilities of the convolutional and pooling layers of the CNN convolutional neural network model are utilized to reduce the loss rate of tool wear features during translation and effectively control the fitting ability of the model; meanwhile, the powerful depth search and global search capability of the genetic algorithm is utilized to optimize two parameters, penalty factor  $c$  and kernel function radius  $g$ , in the SVM support vector machine to improve the tool life prediction accuracy.

## 2 Construction of CNN-SVM-GA prediction model

### 2.1 Convolutional Neural Network (CNN)

Convolutional neural network (CNN)<sup>[15]</sup> is a kind of neural network, a typical representative of deep learning, fundamentally it is a further extension of BP neural network, its main difference is the convolutional operation and pooling operation, which can realize local connection and weight sharing and greatly shorten the training time. CNN network structure contains not only the input layer, fully connected layer and output layer in BP network, but also its unique convolutional, pooling and RELU layers, the training model parameters still use gradient descent method to finally complete the regression prediction task. The principle is as follows:

(1) The sample feature matrix is input to the CNN convolutional neural network for convolutional operation. The sample information is indirectly characterized by the local features of the sample through the weight value of each layer derived from the convolutional operation, and the higher the layer is, the more detailed the local features are extracted, and also the spatial continuity of the sample is maintained:

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

Where  $X_i^k$  denotes the feature matrix of the  $k$ th neuron at the output of the  $i$ th layer, and  $W_i^{kj}$  denotes the weight value of the  $k$ th neuron in the  $i$ th layer, and  $\otimes$  denotes the convolution operator, and  $X_{i-1}^j$  denotes the feature matrix of the  $j$ th neuron at the output of layer  $i-1$ , and  $b_i^k$  is the bias coefficient of the  $k$ th neuron in layer  $i$ .

(2) In order to improve the prediction accuracy of the tool wear life model, the CNN network uses ReLU function for nonlinear activation, which has good non-saturation characteristics to avoid the gradient

disappearance phenomenon. The activation function is shown in equation (2):

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

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

(3) Each tool wear feature data is input to the pooling layer after convolution operation, and the pooling type is chosen as maximum pooling, which can retain the original features and reduce the parameters of network training, and improve the robustness of the extracted features. The maximum pooling is shown in equation (3):

$$C_i^k(s, t) = \text{Max}_{\substack{1+(s-1)Q \leq d \leq sQ \\ 1+(t-1)P \leq h \leq tP}} \{V_i^k(d, h)\} \quad (3)$$

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

(4) The  $n$  feature matrices of dimension  $S \times T$ , which are derived from each row of the sample feature matrix after two convolution and pooling operations, are input to the global average pooling layer. The dimensionality of the pooling kernel of the global average pooling layer is kept consistent with the dimensionality of the feature matrix, and the  $n$  feature matrices are dimensionality reduced to reduce the covariance of the sample features and avoid the influence of redundant features, thus reducing the training time of the LSTM long and short term memory network, so the whole CNN model finally outputs a feature vector  $X_t = \{x_1, x_2, \dots, x_i, \dots, x_j, \dots\}$  where  $x_i$  is calculated as shown in equation (4):

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

According to the above, CNN networks also have shortcomings, such as overfitting when encountering data sets with a small number of features or high nonlinearity, which affects the accuracy of prediction. To address this problem, the SVM classifier needs to be used instead of the Softmax classifier in the CNN model to compensate for this disadvantage.

## 2.2 Support vector machine (SVM)

Support vector machine (SVM) [16] was proposed in 1995 by Cortes and Vapnik et al. Based on statistical theory, this learning model has a supervised mechanism that can perform tasks such as pattern recognition, classification, and regression analysis. In this paper, the feature vector output from the global average pooling layer is used as the input of the SVM support vector machine model. The biggest advantage of the SVM algorithm is that it can handle data with high nonlinearity, and the number of features in the data set has basically no effect on its model complexity, so it can accomplish

regression prediction for data sets with relatively large number of features. The mathematical model of SVM is shown in equation (5):

$$\begin{cases} \min \frac{1}{2} \|w\|^2 + \rho \sum_{r=1}^L \xi_r \\ \text{s. t. } y_r(wX_r + b) + \xi_r \geq 1, r = 1, 2, \dots, L \end{cases} \quad (5)$$

Where  $w$  is the normal vector of the hyperplane, and  $\rho$  is the penalty parameter, the  $\xi_r$  is the relaxation factor,  $b$  is the offset coefficient, and  $X_r$  is the feature vector of the  $r$ th sample, and  $y_r$  is the tool wear value,  $L$  is the total number of feature samples, and the total number of samples in this paper is 315.

The model of Eq. (5) is mostly used to deal with linearly divisible sample feature data, but the tool life sample data is linearly indivisible, so it is necessary to introduce the kernel function to up-dimension each labeled sample data. In this paper, the Gaussian radial basis kernel function is used to transform the nonlinear data of each label state into linear data in high-dimensional space, so that the analysis is possible, and then the optimal classification hyperplane is constructed based on the principle of maximizing the classification interval to complete the prediction of tool life, and the Gaussian radial basis kernel function is shown in Eq:

$$K(X) = \text{sgn} \left( \sum_{r=1}^L a_r^* y_r \exp \left( -\frac{\|X_r - X\|^2}{2g^2} \right) + \theta^* \right) \quad (6)$$

where  $\text{sgn}$  is the sign function,  $a_r^*$  is the Lagrangian multiplier,  $g$  is the kernel function width, and  $X$  is the sample label data, and  $\theta^*$  is the configuration factor.

The width parameter  $g$  and the penalty coefficient  $c$  of the radial basis kernel function are the focus of the SVM algorithm tuning, which directly affect the training speed and prediction accuracy of the model, so how to find the optimal  $c$  and  $g$  parameter matching is the key of SVM model regression analysis.

## 2.3 Genetic Algorithm (GA)

Genetic Algorithm (GA) [17] is an intelligent algorithm that originates from the laws of nature and the mechanism of superiority and inferiority among living organisms. Using genetic algorithm, global search for superiority can be achieved, usually with three most important steps of selection, crossover and mutation, which are similar to the genetic laws of individual biological chromosomes. Therefore, this algorithm is widely used to solve search problems or to optimize some hyperparameters. Firstly, through coding, the set of strings of problem solutions is transformed into individuals that can be recognized by the genetic algorithm. Therefore, individuals with high adaptation values will survive and generate the next generation,

while individuals with low adaptation values will be eliminated; secondly, individuals with medium adaptation values will be "crossed over" to generate new individuals, which will form a new population with the original adaptive individuals; finally, the new population will be "mutated", i.e. Finally, the new population is subjected to "variation", i.e., the adaptation value of some individuals in the population is changed; so on and so forth, the whole population develops to a higher level and finally evolves the most adaptive individuals, i.e., the optimal solution, to complete the task of global optimization, etc. The optimization process of the genetic algorithm (GA) is shown in Figure 1. In this paper, it is the genetic algorithm (GA) that is used to complete the selection of hyperparameters in the SVM model, so as to improve the prediction accuracy and precision of the model.

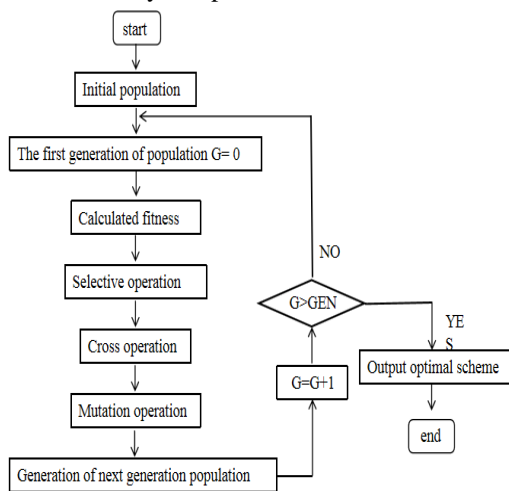


Figure 1 Flow chart of Genetic algorithm (GA)

## 2.4 CNN-SVM model

The essence of CNN-SVM convolutional support vector machine multi-input single-output regression prediction model is to use the CNN convolutional neural network model as a feature fuser and the SVM support vector machine as a trainer for regression prediction. The principle is firstly based on CNN convolutional neural network structure, using its convolutional layer in the network to obtain the weight parameters, pooling layer for dimensionality reduction, the sample set can be automatically feature mining and extraction from the input information without doing complex pre-processing, and fusion of features from shallow to deep as the network is continuously passed backwards. Its fusion pattern framework diagram is shown in Figure 2. Then the output feature vector (fusion value) is directly used as the input of SVM support vector machine for training, and the SVM model transforms these fusion values from low-dimensional space to high-dimensional space after CNN model processing, and then constructs an optimal decision function with the principle of maximizing classification interval to complete the regression prediction problem of data in low-dimensional space,

which can realize the tool life prediction in the milling process using this method. Intelligent prediction of tool life in milling. The structure diagram of the CNN-SVM model is shown in Figure 3.

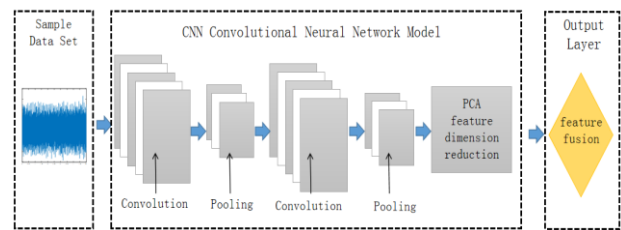


Figure 2 CNN model fusion model framework diagram

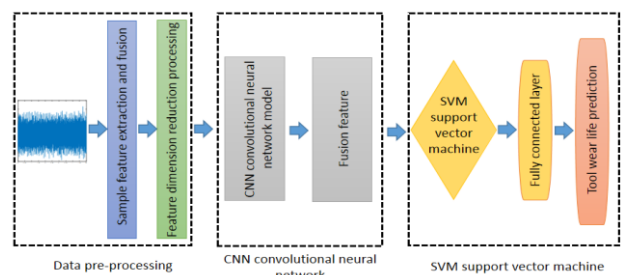


Figure 3 Structure of CNN-SVM model

## 2.5 CNN-SVM-GA hybrid model

The hybrid CNN-SVM model constructed in this paper uses genetic algorithm (GA) to optimize the two parameters of penalty factor  $c$  and kernel function radius  $g$  in the tool life prediction model of CNN-SVM described above. The resulting optimal solution is decoded as a parameter of the support vector machine to improve its generalization ability, speed up the network fitting, and make the tool wear prediction more accurate. The algorithmic flow of the tool life prediction technique based on CNN-SVM optimized by genetic algorithm is shown in Figure 4, and the specific steps are as follows:

Step 1: The original signal (7 channels) related to tool wear is processed for noise reduction and feature extraction and fusion in the time domain, frequency domain and time-frequency domain, respectively.

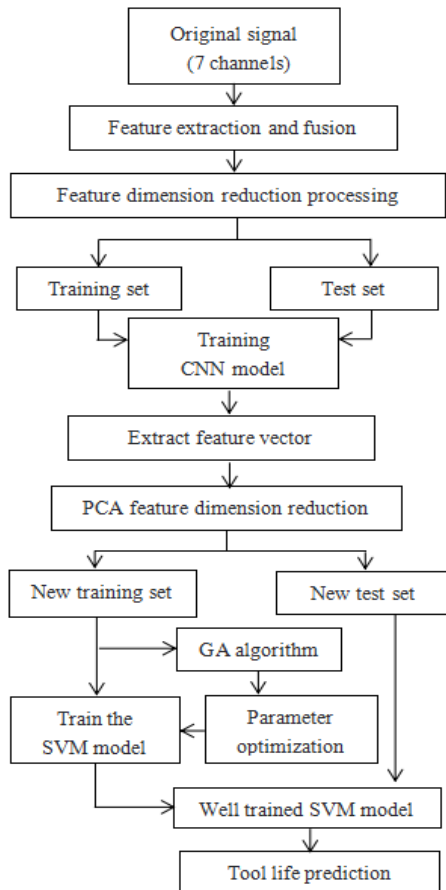
Step 2: Using Pearson's correlation coefficient formula for the above feature data to perform dimensionality reduction and random division of them to construct the training set and test set of the model.

Step 3: building a convolutional neural network, trained using the training and test sets from step 2, the output of which is a feature vector.

Step 4: Perform PCA feature dimensionality reduction on the extracted feature vectors to reduce the training time of the SVM and form a new training and test set.

Step 5: The SVM model is trained with the training set formed in step 4, and the  $g$  and  $c$  parameters of the support vector machine are optimized using a genetic algorithm.

Step 6: Input the test set formed in step 4 to the improved CNN-SVM model to test the model diagnostic effect.



**Figure 4** Improved CNN-SVM lifetime prediction model

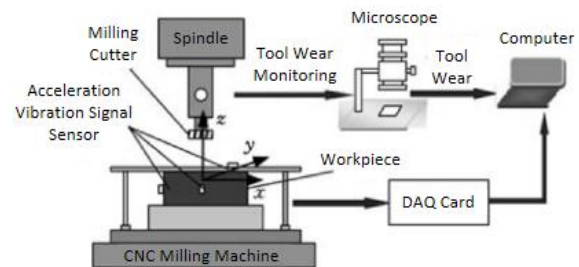
### 3 Tool wear experiment process

The experimental data were obtained from the open data of the 2010 High Speed CNC Machine Tool Health Prediction Competition of the Prediction and Health Management Society (PHM), New York, USA<sup>[18]</sup>, whose tool wear experimental conditions are shown in Table 1.

**Table 1** Experimental conditions for tool wear

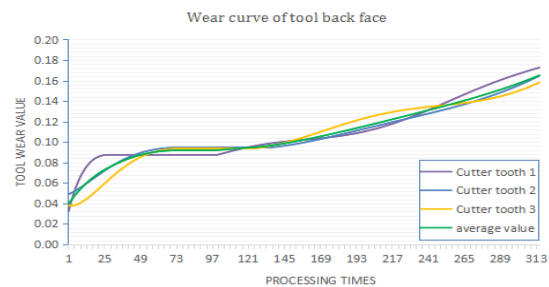
Hardware Conditions	Model and main parameters	Cutting Conditions	Parameters
CNC Milling Machine	CNC Milling Machine Roders Tech RFM760	Spindle speed	10400
Force Gauge	Three-way force gauge Kistler 9265B	Feeding speed	1555
Charge amplifier	Multi-channel charge amplifier Kistler 5019A	Back draft	0.2
Milling Material	cube Inconel 718	Side-draft amount	0.125
Tools	Ball end carbide milling cutter 3 teeth	Feed amount	0.001
Data Acquisition Cards	Data Acquisition Cards NIDAQ	Sampling frequency	50
Wear Gauge	Microscope LEICA MZ12	Cooling conditions	Dry cutting

In the process of machining, the spindle speed was 10400 RPM, the feed was 0.001 mm, the feed speed was set to 1555 mm/min, the tool side draft was 0.125 mm, and the tool back draft was 0.2 mm. The shape of the milled part was square, and the end face was milled by round-trip milling, and the length of the milled part was about 108 mm. The surface length is about 108 mm, and the machining process does not use cutting fluid.  $\Delta t$  The wear value of the rear face of the three teeth of the ball end mill was checked after each time. In this paper, the experimental data set of the first tool of C1 group is selected, and the data set collects and monitors the data of X, Y and Z axes cutting force signals, X, Y and Z axes vibration signals and acoustic emission signals, with a total of 7 channels, each channel walking 315 times, the acquisition frequency is 50 KHz per channel, and the number of sampling points is above 200000 each time, and its related specific data acquisition system is shown in Figure The specific data acquisition system is shown in Figure 5.



**Figure 5** Tool wear data acquisition system

Since the milling cutter used in the experiment has three teeth, the wear of the three teeth was measured after every  $\Delta t$ . The wear of the three teeth was measured after each time. Figure 6 shows the wear curve of the first group of test tools, the purple curve is the wear of the first tooth, the blue curve is the wear of the second tooth and the yellow curve is the wear of the third tooth. In this paper, the average value of the wear of these three tool teeth is taken to represent the actual wear of the tool, and this average value is the sample target value of the improved CNN-SVM convolutional support vector machine model, i.e., the output data. From the figure, it can be seen that the tool wear is faster at the beginning of the tool wear period, flatter when it enters the middle period, and faster at the later period, which is consistent with the theory related to tool wear, which indirectly verifies the accuracy of the data set.

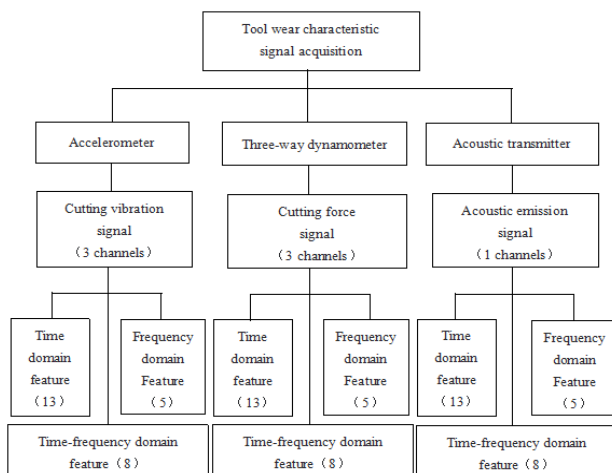


**Figure 6** Test tool wear variation curve

## 4 Pre-processing of tool wear characteristics

### 4.1 Feature extraction and fusion

During CNC machining, sensor technology is used to collect real-time signals related to tool wear. In this paper, cutting vibration signals are collected using a Kistler 8636C piezoelectric accelerometer, cutting force signals are collected using a Kistler 8152 three-way platform dynamometer, and acoustic emission signals are collected using a Kistler 9265B acoustic transmitter. The number of signal data collected above 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, so it is necessary to perform noise reduction processing on the various types of raw signals collected above. The number of times the tool is walked in this experiment is 315, and the number of acquisition points for each knife walk is about 200000 or more. In order to avoid adverse effects during model training, the sampling points with data labels of 50001 to 100000 in each acquisition signal are extracted for feature extraction and fusion in this paper.



**Figure 7** Wear feature extraction and fusion scheme

The feature quantities related to the tool wear state are extracted in the time domain, frequency domain, and time-frequency domain for the above three types of signals, respectively. In order to realize the intelligent tool wear prediction, the time domain features of the original signal are extracted, including 13 kinds, namely, mean value, standard deviation, skewness, cliffness, maximum value, minimum value, peak-to-peak value, root mean square, amplitude factor, waveform factor, impact factor, margin factor, and energy; the frequency domain features are extracted, including 5 kinds, namely, frequency domain amplitude mean, center of gravity frequency, mean square frequency, variance frequency, and frequency variance; the time The extraction of frequency domain features mainly uses wavelet packet analysis to subdivide the original signal into different frequency bands, when the tool wear state changes the energy parameters of different frequency bands will also change, so the energy of each frequency band is the extracted

time-frequency domain features. The wavelet packet decomposition is performed on the original signal, and the number of decomposed layers is set to 3, all of which are completed by db5 wavelet base, and the frequency domain is divided into 8 frequency bands, so that 8 time-frequency domain features are extracted. In this experiment, the original signals of cutting vibration signal (3 channels), cutting force signal (3 channels) and acoustic emission signal (1 channel) are extracted and fused every  $\Delta t$  time, as shown in Figure 7, 13 time domain features, 5 frequency domain features and 8 time-frequency domain features are extracted from each channel signal, so 26 features can be extracted from each channel signal, for a total of 7 channels and 182 features in total. The total number of features is 182.

### 4.2 Feature dimensionality reduction processing

The speed of the tool wear prediction model fitting operation is closely related to the number of features, the more features the more complex the model is, the slower the operation is, so it is necessary to filter and optimize all the features. The best way is to find the correlation between the above mentioned 182 features and the tool wear, and to delete the uncorrelated or weakly correlated features, thus optimizing the extraction of the tool wear signal features and making the model operation speed increase. The Pearson correlation coefficient is the most widely used correlation coefficient analysis method, which can be used to measure the correlation between the extracted eigenvalues and the tool wear amount<sup>[19]</sup>. It is calculated as shown in equation (7):

$$P_{xy} = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}} \quad (7)$$

where  $P_{xy}$  denotes the Pearson correlation coefficient of the signal feature  $x$  and the tool wear value  $y$ . where  $n$  denotes that there are  $n$  sets of signal values and  $x_i$  denotes the  $i$ th value of the signal characteristic value, and  $y_i$  denotes the  $i$ th value of tool wear. The Pearson correlation coefficient formula is used to calculate the correlation between the above 182 features and the tool wear values. Figure 8 shows the correlation of Pearson coefficients for each feature, the red area is  $|P_{xy}| < 0.5$  the features that are weakly correlated, with a total of 48 feature values; the yellow area is  $0.5 \leq |P_{xy}| < 0.9$  The yellow area is for the features that are moderately correlated, with a total of 87 feature values; the green area is for the features that are strongly correlated, with a total of 87 feature values.  $|P_{xy}| \geq 0.9$  The green area is for the features that are strongly correlated, with a total of 47 eigenvalues. In this paper, the 47 strongly correlated features are used as the input data for the training and

prediction of CNN-SVM model, so as to improve the computational speed and accuracy of tool wear prediction.

Feature extraction	Feature	X direction cutting force signal	Y direction cutting force signal	Z direction cutting force signal	X direction vibration signal	Y direction vibration signal	Z direction vibration signal	Acoustic emission signal
Time domain feature	Average value	0.50	0.95	0.96	0.98	0.98	0.91	0.44
	Standard deviation	0.95	0.95	0.95	0.95	0.95	0.94	0.93
	Skewness	-0.87	-0.87	-0.79	0.88	-0.84	0.91	0.91
	Kurtosis	-0.50	-0.50	-0.50	-0.50	-0.72	0.87	0.82
	Median value	0.88	0.94	0.97	0.89	0.87	0.88	0.41
	Median value	-0.67	-0.90	-0.96	-0.87	-0.88	-0.84	0.50
	Peak-to-Peak value	0.95	0.96	0.97	0.95	0.89	0.91	0.28
	Root-mean-square	0.79	0.87	0.87	0.82	0.81	0.82	0.44
	Amplitude factor	-0.05	-0.81	-0.86	-0.81	-0.83	0.50	-0.13
	Waveform factor	-0.21	-0.86	-0.75	-0.84	-0.62	0.88	-0.13
Frequency domain feature	Duty factor	-0.50	-0.85	-0.87	-0.87	-0.82	0.88	-0.14
	Harsh factor	-0.81	-0.80	-0.80	-0.89	-0.88	-0.89	-0.10
	Energy	0.99	0.96	0.96	0.96	0.91	0.92	0.44
	Frequency domain amplitude mean	-0.98	-0.85	-0.85	-0.88	-0.88	-0.88	-0.21
	Center of gravity frequency	-0.03	0.89	0.77	0.71	0.80	0.89	-0.20
	Mean square frequency	-0.03	0.84	0.79	0.54	0.58	0.89	-0.20
	Variance frequency	-0.04	0.85	0.78	0.54	0.58	0.88	-0.21
	Frequency variance	0.00	0.80	0.79	-0.83	-0.84	-0.82	-0.21
	Energy 1	0.58	0.95	0.95	0.87	0.88	0.81	0.41
	Energy 2	0.94	0.87	0.94	0.85	0.87	0.94	0.24
Time-frequency domain feature	Energy 3	0.88	0.87	0.96	0.89	0.88	0.85	-0.22
	Energy 4	0.80	0.80	0.80	0.88	0.89	0.89	0.28
	Energy 5	0.80	0.80	0.75	0.80	0.79	0.79	-0.21
	Energy 6	0.80	0.84	0.85	0.84	0.87	0.80	-0.21
	Energy 7	0.84	0.88	0.86	0.87	0.84	0.85	-0.27
	Energy 8	0.80	0.87	0.89	0.81	0.81	0.79	-0.21

Figure 8 Correlation of Pearson coefficients for each feature

## 5 Tool wear experiment results and analysis

### 5.1 Construction of the sample set data

In this paper, firstly, sensor technology is used to collect the signals related to tool wear (7 channels), secondly, the original signals are subjected to noise reduction processing, feature extraction, feature fusion, dimensionality reduction processing and other operations respectively, and 47 strongly correlated features are derived to form a feature matrix as the input data for the training and prediction of the life prediction model, and its sample feature matrix dimension is  $315 \times 47$ ; the wear of the three teeth of the milling tool is Finally, the above feature matrix is randomly sampled and feature coded, and then the training set and test set are divided, and the first 200 data are taken as the training set and the remaining data are taken as the test set.

### 5.2 Setting of prediction model parameters

In this paper, the sample set data are input to a CNN-SVM model based on genetic algorithm (GA) optimization for tool life prediction, where the initial learning rate parameter of CNN convolutional neural network is set to 0.001, the cross-entropy function is used as the loss function of the whole model, and the Adam optimizer is selected to optimize the hyperparameters, which is set to make the model generalization ability stronger. Second, the Softmax classifier on the fully connected layer is replaced with the SVM algorithm to better handle data with high nonlinearity, and an optimal decision function is constructed to complete the regression prediction of tool wear.

The CNN-SVM-GA model selects the penalty parameter in the SVM model  $\rho$  and the kernel function width  $g$ , which are both set between 0 and 3, as the 2 parameters for the optimization search process. The genetic algorithm (GA) adopts the strategy of superiority selection, crossover and variation to find the optimal hyperparameter pairing, with the crossover rate set to 0.35,

the variation rate set to 0.1, the population size set to 20, and the evolutionary generation set to 3000. The specific parameters are shown in Table 2. Ten optimization operations were performed according to the parameters in Table 2, and the average value was taken as the final result, where the penalty parameter  $\rho$  The optimized kernel function  $g$  is 1.421.  $\rho$  The optimized parameters,  $g$ , are migrated to the CNN-SVM model to complete the tool life prediction.

Table 2 Genetic algorithm (GA) parameter settings

GA algorithm parameters	Parameter Value
Evolutionary Algebra	3000
Population size	20
Crossover Rate	0.5
Variation rate	0.1

In order to quantify the prediction performance of the tool life model, three objective evaluation indicators are selected, namely the mean absolute error MAE, the root mean square error RMSE and the coefficient of determination R2. Among them, the mean absolute error MAE can obtain an evaluation value, but the comparison between different models is necessary to reflect the model's merit; the mean square error RMSE can measure the deviation between the observed value and the true value, the smaller the RMSE value, the better our model is. The smaller the RMSE value is, the better the model is; the coefficient of determination R2 can directly characterize the merit of the model, and the closer the value of the coefficient of determination R2 is to 1, the higher the accuracy and precision of the prediction model is. The three evaluation indicators are calculated as shown in equations (8) to (10):

$$MAE = \frac{\sum_{t=1}^m |y_t - \hat{y}_t|}{m} \quad (8)$$

$$RMSE = \sqrt{\frac{\sum_{t=1}^m (y_t - \hat{y}_t)^2}{m}} \quad (9)$$

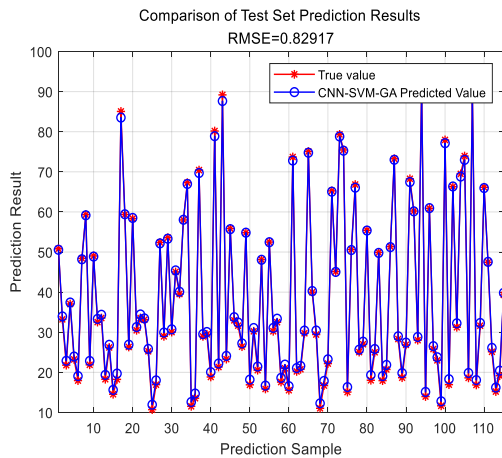
$$R^2 = 1 - \frac{\sum_{t=1}^m (y_t - \hat{y}_t)^2}{\sum_{t=1}^m (y_t - \bar{y})^2} \quad (10)$$

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

### 5.3 Tool life prediction results

Based on the open data of the CNC machining center tool health prediction contest, the CNN-SVM algorithm optimized by genetic algorithm (GA) was used for tool wear regression prediction, and its test set prediction results are shown in Figure 9. The mean absolute error MAE value of the model was calculated to be 0.7231, the root mean square error RMSE value was 0.8292, and the coefficient of determination R2 value was 0.9985. The

results show that the regression prediction of tool life can be effectively performed using the CNN-SVM-GA-based model with good results.



**Figure 9** CNN-SVM-GA test set prediction results

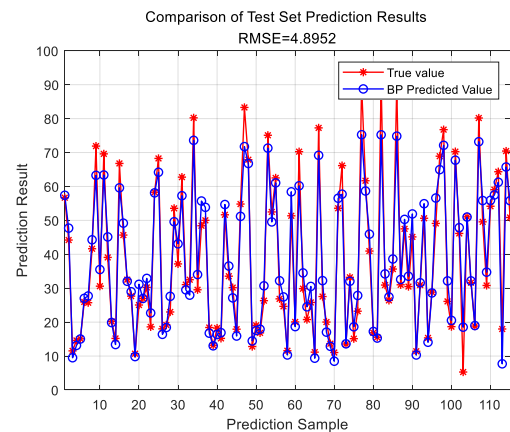
Table 3 shows the effect of genetic algorithm (GA) on the tool life regression prediction model, where the penalty parameters of the CNN-SVM model  $\rho$  and hyperparameters such as kernel function width  $g$  are chosen randomly by relying on manual, it can be seen that the CNN-SVM model optimized using genetic algorithm (GA) has the best tool life prediction. Compared with the CNN-SVM model, its mean absolute error MAE and root mean square error RMSE are reduced and the coefficient of determination  $R^2$  is improved, and its performance index reaches 0.99, while the performance index of the CNN-SVM model with manually selected parameters is maintained at a maximum of about 0.98. This is mainly because the hyperparameter optimization of the CNN-SVM model by Genetic Algorithm (GA) has obtained more accurate hyperparameter pairings, found the most critical attributes affecting the accuracy of tool life prediction, and avoided the blindness of setting parameters, thus improving the prediction effect.

**Table 3** Effect of genetic algorithm (GA) on the prediction model

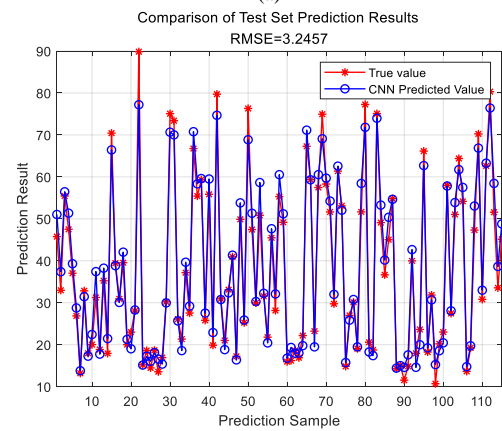
Algorithm	Hyperparameters		Test set prediction results		
	Penalty Parameter	Kernel width	MAE	RMSE	R2
CNN-SVM	0.5	0.5	2.4859	2.8570	0.9817
	1	1	1.1671	1.4557	0.9851
	2	2	3.2250	4.1678	0.9628
	3	3	4.1927	5.7604	0.9296
CNN-SVM-GA	0.511	1.421	0.7231	0.8292	0.9985

To further validate the prediction performance of CNN-SVM-GA based tool life, a comparative analysis was performed with other traditional prediction models in the past, such as BP neural network, CNN convolutional neural network, SVM support vector machine, and CNN-SVM model. Figure 10 shows the comparison

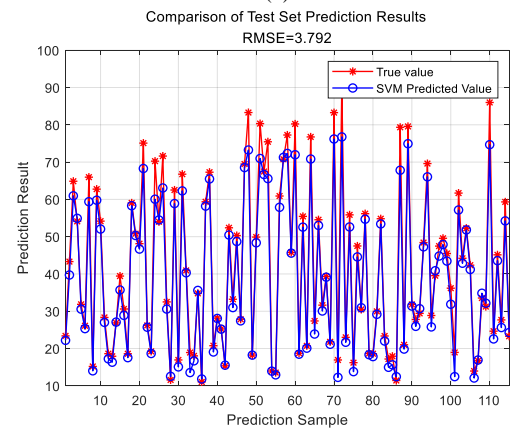
results of four traditional tool life prediction models, and it can be seen from Figure 9 and Figure 10 that the root mean square error RMSE performance of the five tool life prediction models is ranked as CNN-SVM-GA < CNN-SVM < CNN < SVM < BP, and their root mean square error is reduced by 83.06%, 78.13%, 74.45%, and 43.04%, respectively. It can be seen that the CNN-SVM model based on genetic algorithm (GA) optimization proposed in this paper has obvious advantages in tool life prediction, which is because the CNN-SVM-GA model can deeply mine the hidden layer features of the data with high nonlinearity, the feature extraction is comprehensive, and the selection of hyperparameters does not have any dependence on expert experience.



(a)

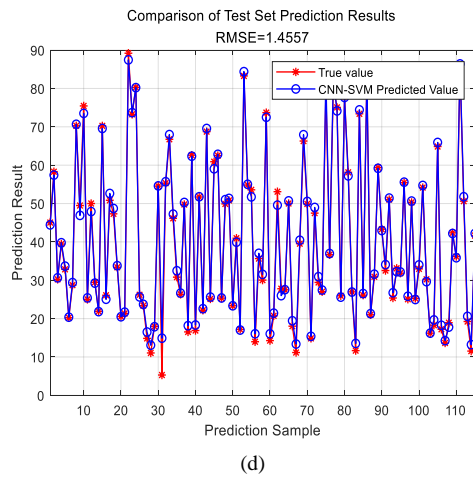


(b)



(c)





**Figure 10** Prediction results of the four traditional models

(a) BP model (b) CNN model (c) SVM model (d) CNN-SVM model

**Table 4** Comparison of prediction performance results of five models

Algorithm	Test set prediction results		
	MAE	RMSE	R2
BP Neural Network	3.6862	4.8952	0.9498
CNN Algorithms	2.5274	3.2457	0.9732
SVM Algorithms	2.5671	3.7920	0.9689
CNN-SVM Algorithm	1.1671	1.4557	0.9851
CNN-SVM-GA Algorithm	0.7231	0.8292	0.9985

Table 4 shows the comparison results of the prediction performance of the five models, and it is found that the CNN-SVM-GA model using multi-channel feature fusion for tool life prediction has the smallest mean absolute error MAE, and the index performance ranking is CNN-SVM-GA < CNN-SVM < CNN < SVM < BP, which is reduced by 80.38%, 71.83%, 71.39%, and 38.04%; the coefficient of determination R2 of the CNN-SVM-GA model proposed in this paper is 0.9985, which is closest to 1. The index performance is ranked as CNN-SVM-GA > CNN-SVM > CNN > SVM > BP, which is improved by 1.34%, 2.53%, 2.96%, and 4.88%, respectively. These two results once again prove that using the CNN-SVM-GA model proposed in this paper for tool life prediction is more effective and can achieve more effective tool life prediction and health management in the milling process.

## 6 Conclusion

This paper completes the construction of a tool life sample dataset based on machine vision, feature extraction, and information fusion, and also proposes a CNN-SVM tool life prediction model based on genetic algorithm (GA) optimization. The model uses convolutional neural network (CNN) model as the feature fusion and support vector machine as (SVM) as the

trainer for tool life regression prediction. And the prediction accuracy of the model is improved by using genetic algorithm (GA) to find the superiority of hyperparameters in the model. The results show that:

(1) The mean absolute error MAE value of 0.7231, root mean square error RMSE value of 0.8292, and coefficient of determination R2 value of 0.9985 were obtained for tool life regression prediction using CNN-SVM-GA model. This indicates that the model can effectively predict the remaining life of the tool with good results.

(2) The tool life prediction model is parameter-seeking by genetic algorithm (GA), and its decision coefficient R2 performance index reaches 0.99, which reduces the subjective influence of manual selection of parameters and avoids the blindness of setting parameters, thus improving the model prediction accuracy.

(3) Compared with the BP model, CNN model, SVM model and CNN-SVM model, the mean absolute error MAE and root mean square error RMSE values of the CNN-SVM-GA model proposed in this paper are reduced, and the value of the coefficient of determination R2 is improved to be closest to 1. This indicates that the constructed tool life prediction model has stronger generalization ability, faster network fitting and tool wear prediction is more accurate.

In the future, this CNN-SVM-GA tool wear prediction model can be widely used in various factories for CNC machining tool life management and other fields. By making real-time prediction of tool life, it can realize predictive maintenance of CNC machining tools and can perform intelligent tool change before tool wear is at a critical threshold, which is in line with the future development trend of intelligent control and network interactive production.

**Author Contributions:** For Conceptualization, methodology, analysis, and writing original draft preparation, Wang Shuo; writing review and full-text editing, Yu Zhenliang; writing—original draft preparation, Liu Peng; writing—original draft preparation, Wang Mantong.

**Conflicts of interest:** The authors declare no conflict of interest. All authors have read and agreed to the published version of the manuscript.

**Acknowledgments:** The research work financed with the means of Basic Scientific Research Youth Program of Education Department of Liaoning Province, No.LJKQZ2021185; Yingkou Enterprise and Doctor Innovation Program (QB-2021-05).

## References

- [1] Danil Yu Pimenov, Andres Bustillo, Szymon Wojciechowski, et al.. Artificial intelligence systems for tool condition monitoring in machining: analysis and critical review [J]. *Journal of Intelligent Manufacturing*, 2022(1):67-68.
- [2] Dong Liang, Wang Chensheng, Yang Guang, Huang Zeyuan, Zhang Zhiyue, Li Cen. An Improved ResNet-1d with Channel

- Attention for Tool Wear Monitor in Smart Manufacturing [J]. *Sensors*, 2023, 23(3):89-90.
- [3] Shi Yuen Wong, Joon Huang Chuah, Hwa Jen Yap. Technical data-driven tool condition monitoring challenges for CNC milling: a review [J]. *The International Journal of Advanced Manufacturing Technology*, 2020,107 (prepublish).
- [4] Pu Xiaobo, Jia Lingxu, Shang Kedong, Chen Lei, Yang Tingting, Chen Liangwu, Gao Libin, Qian Linmao. A New Strategy for Disc Cutter Wear Status Perception Using Vibration Detection and Machine Learning [J]. *Sensors*, 2022,22(17).
- [5] Wei Weihua, Cong Rui, Li Yuantong, Abraham Ayodele Daniel, Yang Changyong, Chen Zengtao. Prediction of tool wear based on GA-BP neural network [J]. *Proceedings of the Institution of Mechanical Engineers*,2022, 236(12).
- [6] Cao W. The Diagnosis of Tool Wear Based on RBF Neural Networks and D-S Evidence Theory. IEEE China Council. IEEE Beijing Section, Sichuan Computer Federation. *Proceedings of 2010 3rd IEEE International Conference on Computer Science and Information Technology VOL.7* [J]. Institute of Electrical and Electronics Engineers, 2010:429-431.
- [7] Zhang Kun , Zhu Hongtao, Liu Dun, Wang Guoning, Huang Chuanzhen,Yao Peng. A dual compensation strategy based on multi-model support vector regression for tool wear monitoring [J]. *Measurement Science and Technology*, 2022,33(10).
- [8] G. E. Hinton, R. R. Salakhutdinov. Reducing the Dimensionality of Data with Neural Networks [J]. *Science*, 2006, 313(5786).
- [9] P. K. Ambadekar, C. M. Choudhari. CNN based tool monitoring system to predict the life of cutting tool [J]. *SN Applied Sciences*, 2020, 2(4).
- [10] Caesarendra Wahyu, Triwiyanto Triwiyanto, Pandiyan Vigneashwara, Glowacz Adam, Permana Silvester Dian Handy, Tjahjowidodo Tegoeh. A CNN Prediction Method for Belt Grinding Tool Wear in a Polishing Process Utilizing 3-Axes Force and Vibration Data [J]. *Electronics*, 2021, 10(12).
- [11] Stefan Droste, Thomas Jansen, Ingo Wegener. Upper and Lower Bounds for Randomized Search Heuristics in Black-Box Optimization. *Electron. Colloquium Comput. Complex*, 2003:48-48.
- [12] Weifeng Lu, Bingyu Cai, Rui Gu. Improved Particle Swarm Optimization Based on Gradient Descent Method [J]. *CSAE*, 2020: 121-126.
- [13] Salih Omran, Duffy Kevin Jan. Optimization Convolutional Neural Network for Automatic Skin Lesion Diagnosis Using a Genetic Algorithm [J]. *Applied Sciences*, 2023,13(5):87-88.
- [14] Nagra Arfan Ali, Mubarik Iqra, Asif Muhammad Mugees, Masood Khalid, Ghamdi Mohammed A. Al, Almotiri Sultan H. Hybrid GA-SVM Approach for Postoperative Life Expectancy Prediction in Lung Cancer Patients [J]. *Applied Sciences*, 2022, 12(21):655-66.
- [15] Gajera Himanshu K., Nayak Deepak Ranjan, Zaveri Mukesh A.. A comprehensive analysis of dermoscopy images for melanoma detection via deep CNN features [J]. *Biomedical Signal Processing and Control*, 2023,79(2):32-37.
- [16] Fafa Chen, Chen Fafa, Cheng Mengteng, Tang Baoping, Chen Baojia, Xiao Wenrong. Pattern recognition of a sensitive feature set based on the orthogonal neighborhood preserving embedding and adaboost\_SVM algorithm for rolling bearing early fault diagnosis [J]. *Measurement Science and Technology*,2020,31(10):89-91.
- [17] Zheng Zhang, Liang Li, Wei Zhao. Tool Life Prediction Model Based on GA-BP Neural Network [J]. *Materials Science Forum*, 2016, 3901(836):956-957.
- [18] Huimin Chen. A Multiple Model Prediction Algorithm for CNC Machine Wear PHM [J]. *International Journal of Prognostics and Health Management*, 2011,2(2):56-57.
- [19] Li Yifan, Xiang Yongyong, Pan Baisong, Shi Luojie. A hybrid remaining useful life prediction method for cutting tool considering the wear state [J]. *The International Journal of Advanced Manufacturing Technology*, 2022,121(5-6):78-90.