Image-based tool condition monitoring based on convolution neural network in turning process

Rui Kou1 · Shi-wei Lian2 · Nan Xie2 · Bei-er Lu2 · Xue-mei Liu1

Received: 5 May 2021 / Accepted: 24 October 2021 / Published online: 6 January 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract
Tool wear has a significant impact on machining quality, efficiency, and cost, so it is vitally important for manufacturing systems. The current work of Tool Condition Monitoring (TCM) mainly processes the time series signals from multisensory using intelligent algorithms. However, the limits of these methods are as follows: (1) the image information is not integrated into the time series signals, and (2) the traditional methods face the problems of poor generalization and fast convergence. Thus, a novel integrated model based on the multisensory feature fusion and neural network is presented. The sensor data is first pre-processed using Piecewise Aggregate Approximation (PAA) and then recoded into images using Gramian Angular Field (GAF). The images, together with the tool infrared images, are inputs to the Convolutional Neural Network (CNN) model, which realizes the output of flank wear value. Both time series signals and tool infrared images are used to achieve the classification, and the final classification accuracy in the test set is 91%. The results show the high computation efficiency and the good generalization performance of the presented methodology.

Keywords Tool condition monitoring · Deep learning · Turning

1 Introduction
Cutting is currently the most commonly used machining method. During the machining process, the tool is prone to wear and failure due to complex temperature and stress changes on the contact surface of the tool and the workpiece [1]. Tool wear and failure makes the workpiece unable to achieve the desired shape, size, and accuracy. In severe cases, it may also cause damage to the equipment, resulting in machine failure and thus the inability to process. Data shows that 20% of machine downtime is due to tool damage [2]. In order to avoid harmful effects on the part and the machine, it is necessary to monitor the tool condition in real time, so that it can be replaced before it wears out.

The Prognostics and Health Management (PHM) tries to detect the current state of the equipment and its failure in advance through prediction. It overcomes the drawbacks of the traditional reactive maintenance policy (fail and fix) and helps decision-making with the prediction information, thereby extending the lifetime of the equipment [3]. Tool Condition Monitoring (TCM) is an important part of PHM. The development of TCM technology plays an important role in promoting intelligent production. By collecting and analyzing data reflecting the tool condition during machining process, it can lower machining costs and improve machine utilization [4, 5].

The commonly used indirect measurements of TCM are divided into three stages: (1) signal acquisition, (2) feature extraction and selection, and (3) classification [6]. In the feature extraction and selection stage, the main features that characterize the original data are extracted from a large amount of data, and methods such as Principal Component Analysis (PCA) and Kernel Principal Component Analysis (KPCA) are commonly used. In the classification stage, algorithms such as K-Nearest Neighbor (KNN), Support Vector Machines (SVM), and fuzzy are used.

Elgargni et al. [7] proposed the method of combining PCA and Discrete Wavelet Transform (DWT) with neural network and established an effective and reliable tool tracking and health identification software program. Barreiro et al. [8] combined acoustic emission and vibration signals to
provide complementary signals in different spectral bands. The tool condition and the transition between wear conditions were identified using frequency band analysis. This method is applied to a TCM system of especial king-size multiple cutting edge tools for machining in the milling process of super long and thick steel plate. Painul et al. [9] proposed a TCM algorithm based on K-star, which extracts a set of statistical features from vibration signals to form the input of the algorithm. Experimental results show that the algorithm can achieve 78% classification accuracy.

In recent years, the rapid development of deep learning in various fields has improved solving the problems of difficult feature extraction, poor model generalization, and local optimization and significantly improved the recognition ability of industrial big data. Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN) are the most widely used deep learning networks.

Xu et al. [10] proposed a Gated Regression Unit (GRU)-based neural network for tool condition monitoring, which can analyze time series data on multiple time scales. Li et al. [11] established a mapping model between vibration and sound signals and boring tool state by collecting vibration and sound signals from the boring process and proposed a deep-hole boring tool state monitoring method based on Long Short-Term Memory (LSTM) network. Zhou et al. [12] extracted wear features from machining signals and developed a model for predicting the remaining useful life of a tool under variable working conditions by using LSTM.

The research related to condition monitoring and fault diagnosis based on CNN [13] is still in the exploratory stage. Ambadear et al. [14] carried out dry cutting experiments on low carbon steel parts with carbide blade. The images of cutting tools and turning parts were taken at regular intervals by microscope and used as inputs of CNN model to classify tools into three wear grades: low, medium, and high. The experimental results show that the accuracy of this method is 87.26%. Janssens et al. [15] proposed a CNN-based feature learning model for condition monitoring, and the results show that the model performs significantly better than traditional feature-engineering-based methods.

However, the above methods still have many drawbacks. Firstly, most of these methods do not take into account both the change of vibration signal caused by the change of cutting force during machining process of the tool and the wear due to bonding and oxidation caused by high temperature. Secondly, most of these methods focus on the study of the tool state, what have not been refined to the determination of the VB value of the tool wear. Thirdly, in the feature extraction and selection stage, the method for feature extraction and selection may not be suitable for other working conditions.

In addition, CNN is usually for the extraction of image information, but rarely for the extraction of features of time series signals. The processing of time series signals is also mostly done by one-dimensional CNN (1D-CNN). For instances, some studies applied 1D-CNN to process vibrational signals for structure health monitoring [16, 17] and bearing fault diagnosis [18]. But compared with signal input method in 1D-CNN, two-dimensional CNN can do fine tuning with large database, achieving higher accuracy and robustness [19]. Converting time series signals to images can take advantage of 2-dimensional CNN for large data processing and can incorporate additional image information to achieve TCM.

In this paper, a CNN-based tool condition monitoring model built on the Software as a Service (SaaS) layer of the cloud platform is proposed. It combines vibration sensor and current sensor data with infrared images of the tool to achieve TCM and realizes data transmission and interaction between the physical entity of the tool and the virtual model. Furthermore, the monitoring of the tool state using deep learning models is presented in the virtual digital system.

2 Architecture of tool condition monitoring model based on CNN

In machine tool process, the number of signals is of high volume and complex, including analog signal and image signal. GAF can transform signal curve into image information. CNN can learn features from massive data and generalize the learned features to other data of the same type. It is suitable for image information processing. Therefore, in the mechanism model and algorithm module of Platform as a Service (PaaS) layer of cloud platform, the cloud tool condition monitoring model is constructed, and the mapping relationship between tool condition edge and cloud side is established, which is used for TCM service at SaaS layer.

The architecture of tool condition monitoring model is shown in Fig. 1. The input of the model is the vibration signal and current signal collected by the sensors in the process of machine tool processing and the infrared image of the tool collected by the infrared sensor. After Gramian Angular Difference Field (GADF) image processing, all the images are taken as the input of CNN, and the output is the VB value range of the tool. In the initial case, the weight of CNN is not determined, so the original data is used for training and reducing the distance between the predicted value and the real value, and dropout is added to prevent the overfitting of the model. The image data is processed by the convolution layer and pooling layer and classified by the Softmax layer, and finally the optimal model of tool condition monitoring is obtained.
2.1 Time series data process using Gramian Angular Field (GAF)

Signal processing technology is widely used in various fields; the quality of signal processing often has a very important impact on the experimental results. Machine tool processing time series data needs smoothing, noise reduction, and other operations, and then the time-domain and frequency-domain features need to be extracted for analysis to characterize the signal state. In this paper, multiple sensors are used to implement TCM. In order to combine vibration and current time signals with tool infrared images for classification, time series signals need to be converted to images. Inspired by the great success of deep learning in computer vision, Wang et al. [20] proposed three temporal data processing algorithms Gramian Angular Summary Field (GASF), Gramian Angular Difference Field (GADF), and the (MTF). These algorithms can encode time series data into images, realize the visualization of time series data, and can be applied to the field of computer vision to realize the functions of image classification and recognition.

Given the vibration data sequence of machining process, normalize the value of vibration data within the interval $[-1, 1]$ using Equation (1) or $[0, 1]$ using Equation (2),

$$
\hat{x}_i = \frac{x_i - \min(X)}{\max(X) - \min(X)}
$$  \hspace{1cm} (1)

$$
\hat{x}_i = \frac{x_i - \min(X)}{\max(X) - \min(X)}
$$  \hspace{1cm} (2)

In this paper, Equation (1) is used. The new vibration data sequence is transformed into polar coordinate system. Among them, the value of time series and its corresponding time stamp are expressed by angle and radius, respectively:

$$
\phi = \arccos(\hat{x}_i), -1 \leq \hat{x}_i \leq 1, \hat{x}_i \in \hat{X}
$$  \hspace{1cm} (3)

$$
r = \frac{t_i}{N}, t_i \in \{0...N\}
$$  \hspace{1cm} (4)

where $t_i$ is the time stamp and $N$ is the constant factor of regularized polar coordinate span; this method based on polar coordinates analyzes the time series data from another point of view. The normalized vibration signal of X-axis is shown in Fig.2. With the increase of time, the corresponding value of polar coordinates will twist between different angles on the span circle. The advantage of GAF is that the whole coding method is bijective that given a time series, there is only one result in the polar coordinate system. At the same time, compared with the Cartesian coordinate system, GAF maintains the absolute time relationship through coordinates. After the time series data are transformed into polar coordinate system, the angle can be used to identify the temporal correlation in different time intervals by considering the triangular relationship between the points. The temporal correlation is expressed as Gram matrix, as shown in Equation (5):
Then the definitions of GASF and GADF are as follows (Eqs. (6) and (7)):

\[
GASF = \cos(\phi_i + \phi_j) = \bar{X}' \cdot \bar{X} - \sqrt{I - \bar{X}'^2} \cdot \sqrt{I - \bar{X}^2} \quad (6)
\]

\[
GADF = \sin(\phi_i + \phi_j) = \sqrt{I - \bar{X}'^2} \cdot \bar{X} - \bar{X}' \cdot \sqrt{I - \bar{X}^2} \quad (7)
\]

where \( I \) is the unit row vector \([1, 1, ..., 1]\).

Fig. 3 is the result of GADF transformation of the time series data in Fig. 2. GADF provides a time-dependent method. When the position in the image moves from the lower left corner to the upper right corner, the time increases. \( G_{(i,j)} \) represents the superposition relationship of direction relative to time interval, when \( k=0 \), the diagonal \( G_{(i,i)} \) contains the original value/angle value information. So, from the diagonal, the high-dimensional features of time series data could be learned through convolution neural network.

In order to reduce the size of the GADF image, Piecewise Aggregate Approximation (PAA) [21] is used to smooth the curve while maintaining the trend of the original data. Supposed the sensor signal \( X = \{x_1, x_2, ..., x_n\} \) be a sequence of \( n \) length, and convert it into a sequence of \( m \) length \( S = \{s_1, s_2, ..., s_m\} \), where the relation \( n > m \) is satisfied and \( k = n/m \) is an integer, and then any element \( s_i \) in the sequence \( S \) satisfies the following equation (Eq. (8)):

\[
s_i = \frac{1}{k} \sum_{j=k(i-1)+1}^{k(i)} x_j, \quad 1 \leq i \leq m \quad (8)
\]

This method represents the characteristics of the original data by calculating the mean value. When analyzing the time series data, it pays more attention to the integrity and change trend of the data. The data processed by PAA can reflect the shape of the data and reduce the size of the data at the same time, effectively reducing the amount of calculation, improving the calculation efficiency, and laying a foundation for the visualization of time series data.

### 2.2 TCM feature identification based on CNN

CNN can learn features from massive data and generalize features to other data of the same type, so it is often used in the field of computer vision with multi-pixel information. The default input format of CNN is image, and the neuron is designed into three dimensions: width, height, and depth. The input of tool condition monitoring method based on CNN is the image after GADF and the infrared image. The CNN network includes six layers of convolution layer, three layers of pooling layer, and one layer of full connection layer. The result is classified by Softmax, as shown in Fig. 4.

The basic idea of CNN network is convolution. Each convolution layer is composed of several neuron filters convolution cores. The parameters of each convolution core are optimized by Backpropagation Through Time (BPTT) algorithm, and the image features can be extracted. These convolution cores are connected with the previous layer of neurons according to the information they feel. In the first convolution layer, each convolution core is not connected to each pixel of the whole picture but only connected to the pixels in its own region. The convolution core in the second convolution layer is connected to the convolution core in the first layer, and so on. Therefore, convolution neural network
often contains multiple convolution layers. The first convolution layer only extracts some low-dimensional features. With the increase of the number of layers, the subsequent convolution layer can further convolute the previous features into higher-level features and extract more and more complex features. Set the sensor signal image \( X \) after GADF as the network input and \( W \) as the convolution kernel function, and then the convolution of the two is (Eq. (9)):

\[
s(i, j) = (X * W)(i, j) = \sum_{m} \sum_{n} x(i+m, j+n)w(m, n)
\]  

(9)

where * is the convolution operator, \( s(i, j) \) is the matrix of the image and convolution kernel after convolution operation, \( m, n \) are the number of pixels of the image, and \( i, j \) indicate the size of the convolution matrix, which is called a convolution feature of the image.

The convolution operation on the image is equivalent to a convolution core sliding on the image matrix until the whole image is scanned. Each sliding will produce a corresponding node; the specific process is shown in Fig. 5.

In the convolution process, each image has different characteristics because of the difference of the image of sensor signal GADF. Each convolution unit has more than one set of weight values and filters. In the process of training, CNN can constantly find the filter that can classify a specific task, that is, CNN can accurately find the features that can represent the tool state; regardless of their specific position in the image, higher-level convolution can classify the tool state considering all the features.

CNN can get weight or feature through training, and it has the ability of learning and clustering, but it cannot classify different kinds of signals. Therefore, it is necessary to add multi-classification structure to CNN structure and solve multi-classification problem by Softmax regression model.

**Softmax regression model** is a generalization of logistic regression model in multi-classification. It transforms the score value of linear classification into probability, so the sum of these probability values is 1. In the multi-classification problem, the class label can take more than two values. Suppose that there is a picture collection \( X^{(i)}, y^{(i)} \), for each \( x^{(i)} \), the model calculates its probability value \( p(y^{(i)} = k | x^{(i)}), k = 1, 2, 3, ..., K \). The formula is as follows:

\[
h_{\theta}(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1 | x^{(i)}; \theta) \\ p(y^{(i)} = 2 | x^{(i)}; \theta) \\ \vdots \\ p(y^{(i)} = K | x^{(i)}; \theta) \end{bmatrix} = \frac{1}{\sum_{j=1}^{K} \exp(\theta_j^T x^{(i)})} \begin{bmatrix} \exp(\theta_1^T x^{(i)}) \\ \exp(\theta_2^T x^{(i)}) \\ \vdots \\ \exp(\theta_K^T x^{(i)}) \end{bmatrix}
\]

(10)
where $\theta = [\theta_1, \theta_2, ..., \theta_K]$ is the parameter of Softmax, the cost function of Softmax regression $J(\theta)$:

$$J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^{m} \sum_{j=1}^{K} 1[y(i) = j] \log \frac{\exp(\theta_j^T x(i))}{\sum_{l=1}^{K} \exp(\theta_l^T x(i))} \right] + \frac{\lambda}{2} \sum_{i=1}^{K} \sum_{j=0}^{N} \theta_{ij}^2 \tag{11}$$

Therefore, Softmax regression can be used to calculate the probability of each tool condition and classify the tool conditions.

### 3 Experiment

#### 3.1 Experimental design

The experiments are performed on a CNC lathe (TX30). The vibration data and current data are collected by CT1010SLFP vibration sensor and JLB50A current sensor and transmitted to the computer by NI USB-6009 data acquisition cards. At the same time, the infrared image of the tool in the machining process was captured by using the infrared camera FLIR A35. After each machining, the micro-measuring instrument AMT3114 was used to measure and record the tool wear until the tool was damaged. The process of the experiment is as follows: Machining $30 \text{ mm}$ workpiece along the axial direction with the cutter; after machining, remove the cutter from the tool holder and measure the tool wear VB. Fig. 6 shows the installation of sensors.

For each machining process, the corresponding data acquisition card signal channel is shown in Table 1. The whole tool wear process is divided into six stages, corresponding to different tool wear VB values, and corresponding labels are given, as shown in Table 2, infrared image as shown in Fig. 7, and some tool wear VB values and pictures as shown in Fig. 8. The experimental design adopts the way of orthogonal experiment, which can balance sampling in the range of multi-factor and multi-level, making each experimental combination more representative and meeting the requirements of the experiment. Considering the

![Fig. 6 The installation of sensors](image)

| Signal channel | Measurement data |
|----------------|------------------|
| Channel 1      | $V_x$ —— Vibration X |
| Channel 2      | $V_y$ —— Vibration Y |
| Channel 3      | $V_z$ —— Vibration Z |
| Channel 4      | $I_A$ —— Current A |
| Channel 5      | $I_B$ —— Current B |
| Channel 6      | $I_C$ —— Current C |

![Table 1 Signal channel of measurement data](image)

| Label | Flank wear | Tool condition |
|-------|------------|----------------|
| 0     | 0–0.1mm    | Initial wear   |
| 1     | 0.1–0.2mm  | Initial wear   |
| 2     | 0.2–0.3mm  | Stable wear    |
| 3     | 0.3–0.5mm  | Stable wear    |
| 4     | 0.5–0.7mm  | Sharp wear     |
| 5     | Broken     | Broken         |

![Table 2 Corresponding table of tool condition status](image)

![Fig. 7 Infrared image of cutting tool in cutting process](image)
three cutting parameters and tool conditions, the orthogonal table L9 (3^4) is used in this experiment. Each experiment includes about 200000 sample points. The text size of all data sets converted to txt format is about 2.67GB. The specific experimental processing parameters are shown in Table 3.

### Table 3 The experimental design

| No. | Spindle speed (rpm) | Feed rate (mm/min) | Cutting depth (mm) | Axial cutting length (mm) | Sampling rate (kHz) | material | Tool       |
|-----|---------------------|--------------------|-------------------|--------------------------|---------------------|----------|------------|
| 1   | 800                 | 70                 | 0.2               | 30                       | 10                  | 45#steel | TNMG160404 |
| 2   | 800                 | 90                 | 0.3               |                          |                     |          |            |
| 3   | 800                 | 50                 | 0.4               |                          |                     |          |            |
| 4   | 1000                | 90                 | 0.2               |                          |                     |          |            |
| 5   | 1000                | 50                 | 0.3               |                          |                     |          |            |
| 6   | 1000                | 70                 | 0.4               |                          |                     |          |            |
| 7   | 1200                | 50                 | 0.2               |                          |                     |          |            |
| 8   | 1200                | 70                 | 0.3               |                          |                     |          |            |
| 9   | 1200                | 90                 | 0.4               |                          |                     |          |            |


### 3.2 Model training, testing, and evaluation

The model uses Keras open-source framework to build convolutional neural network and completes the corresponding training and testing verification. A small number of data sets (3/4) are selected for training to verify the availability of the model and avoid the waste of computing resources.

The training data set of CNN is composed of vibration signal and three-phase current encoded image and infrared image. Because the time series of a machining process is too long (about 20W data points at a time) and contains irrelevant data caused by start-up or stop when the machine is started or stopped, the representative part of the complete sensor data series can be intercepted, and the relevant images of vibration data and current data can be obtained by using the GADF method. After data normalization, PAA is used to reduce the number of data to 1500 without affecting the consistency of time series; GADF is used to convert the time series data after PAA into three channel image of 350 × 350 pixels. The label of the image is consistent with the label of the tool in this case.

After data preprocessing, a total of seven images, including six 3-channel images and one infrared image, were generated in each cutting process. A total of 1120 images were generated in 160 experiments. Seventy-five percent of these pictures are used as training set and 25% as test set. The test set does not participate in the whole process of training, so there are 840 training set samples and 280 test set samples. The following will elaborate the generation of CNN model from model training, model testing, and evaluation.

#### 3.2.1 Model training

In the training phase, the CNN network uses Softmax for regression classification, uses cross entropy loss function to calculate the distance between the actual output of the neural network and the label, and updates the model parameters and weights through continuous training. In the training stage, the data sets and tags are loaded into the CNN neural network model. The model is iteratively optimized in the continuous training process and constantly updates the weights in the neurons to reduce the loss value. Finally, the neural network model with classification ability is obtained. Table 4 shows the training parameters of CNN model.

The curve of accuracy and loss value of convolutional neural network under above training parameters is shown in Fig.9. The horizontal axis is the training round epoch, and the vertical axis is the accuracy and loss value. It can be seen that after about 2000 steps of epoch, the accuracy of the model gradually increased to about 80%, loss gradually decreased to less than 0.25, and then the accuracy of the model quickly reached more than 95% and gradually converged. Although the model oscillates slightly, with the increase of training times, the accuracy and loss of the model tend to be stable, which shows that the model can mine the tool state-related features in the signal image and achieve good training effect.

#### 3.2.2 Model testing and evaluation

When the model training is completed, the trained model with tool state classification function is applied to the test set to evaluate the performance of the model. This section will evaluate the advantages and disadvantages of the model using F1 score. The F1 score is the weighted average sum of the precision and the recall, i.e., the F1 score is high when both the precision and the recall are high, and the formulas are as follows:

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

\[
F1\text{Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}
\]

where \(TP, FP,\) and \(FN\) represent the number of true positive samples, false positive samples, and false negative samples, respectively.

Based on a total of 280 pictures in the test set, the evaluation indicators are shown in Table 5. It can be seen that the accuracy of the test set is 91%, the macro-average F1 score value of six kinds of Tool States is 0.91, and the weighted average F1 score value is 0.92. It shows that the model also shows a satisfactory classification effect in the test set and has a good performance in predicting the “1” initial wear state and “5” damage state, that is, the model is more sensitive to the damaged state of the tool and can quickly locate the damaged tool in practical application.

At the same time, in terms of training time, the average training time of each epoch is 850 MS, and the whole training process takes about 3 h, but the time consumption on the test set becomes shorter. This is because only the model parameters need to be loaded in the test set to classify the
tool state, so the monitoring model can also achieve real-time monitoring. Because the training network needs to consume a lot of computing resources, the twin data transmission of sensors depends on bandwidth, and it also takes a certain amount of time to convert the data into images. Therefore, this algorithm is built in the cloud with strong computing power to achieve the effect of near real-time monitoring. In addition, although training consumes a lot of time, we can speed up the model calculation and testing by using more powerful CPU or GPU, optimizing code, and introducing parallel computing and other methods.

Fig. 10 shows the confusion matrix of the model. Combined with Table 5, it can be seen that the model performs better in classification “0,” “1,” and “2” which are initial wear and stable wear and worse in classification “4” which is sharp wear. The model is difficult to classify severe wear. The possible reason is that at this stage, the tool will quickly enter the damage stage, resulting in less available data. At the same time, due to the instability of the tool at this stage, the form, temperature, weak point, and angle of the tool are not the same, and the vibration curve, current curve, and infrared image are not the same. It is hard to summarize a general feature. When the tool goes to this stage, the data

![Training Loss and Accuracy](image)

![Confusion matrix](image)

### Table 5 The performance of the CNN model

| Label            | Precision | Recall | F1 score |
|------------------|-----------|--------|----------|
| Predicted as “0” class | 0.92      | 0.94   | 0.93     |
| Predicted as “1” class | 0.92      | 1.00   | 0.96     |
| Predicted as “2” class | 0.96      | 0.81   | 0.88     |
| Predicted as “3” class | 0.98      | 0.86   | 0.92     |
| Predicted as “4” class | 0.76      | 0.95   | 0.85     |
| Predicted as “5” class | 0.92      | 0.92   | 0.92     |
| Macro average    | 0.91      | 0.91   | 0.91     |
| Weighted average | 0.92      | 0.91   | 0.92     |
| Accuracy         | /         | /      | 0.91     |
set of the tool in this state will be reduced if the machine tool is shut down artificially to protect the workpiece and the machine tool. The information and features extracted in the model training process are less than those in other tool states. In the stage of tool breakage, the tool vibration and temperature rise sharply, which are different from the other states, and can be better classified in this stage.

### 3.3 Results comparison and analysis

The model classification accuracy and its possible improvements will be discussed first in this section. Then, in order to verify the effectiveness of the model, the monitoring performance of the model will be compared from two aspects of different model inputs and different monitoring algorithms.

#### 3.3.1 Discussion of classification results

The model performs relatively poorly in class 4, i.e., “sharp wear.” The possible reason is that the training set samples for class 4 are too small. Because at that stage the tool is about to enter the damage stage, the state is unstable, and it is difficult to collect a sufficient number of samples. To improve the classification accuracy for class 4, a possible improvement is to adjust the structure of the CNN, such as changing the convolution kernel size and the number of convolution layers. One of the possible improvement directions is to introduce inception layer [22] in the CNN. The inception module is shown in Fig. 11. The advantages of this algorithm are that the features can be abstracted from different scales simultaneously and more nonlinearity are introduced to improve the generalization ability. Thus, it may be possible to improve the classification accuracy of category 4.

#### 3.3.2 Comparison between single sensor image input and multi-sensor image input

With the development of sensing technology and communication technology, there are more and more data collected by sensors of different scales and types. Simultaneous interpreting different data from different sensors reveals the state of the tool.

In this experiment, only a single vibration sensor data is used as the input of CNN neural network, not all the sensor images. After the model training is completed, it is applied to the test set. The confusion matrix is shown in Fig. 12, and the evaluation index is shown in Table 6. It can be seen from Table 6 that the prediction result of this model is better.

![Confusion matrix](image)

**Fig. 12** Confusion matrix under single input

| Label          | Precision | Recall | F1 Score |
|----------------|-----------|--------|----------|
| Predicted as “0” class | 0.98      | 0.75   | 0.85     |
| Predicted as “1” class | 0.75      | 0.98   | 0.85     |
| Predicted as “2” class | 0.91      | 0.79   | 0.85     |
| Predicted as “3” class | 0.78      | 0.81   | 0.79     |
| Predicted as “4” class | 0.71      | 0.73   | 0.72     |
| Predicted as “5” class | 0.80      | 1.00   | 0.89     |
| Macro average   | 0.82      | 0.84   | 0.82     |
| Weighted average| 0.83      | 0.82   | 0.82     |
| Accuracy        | /         | /      | 0.82     |

**Table 6** The performance of the CNN model under single input
only in “5” category. F1 score is 0.89, and the classification effectiveness of other categories is far from the expected.

Comparing Table 5 and Table 6, Fig.10 and Fig.12, it can be seen that the classification of multi-sensor model is better than that of single sensor model. When all kinds of sensor information are comprehensively used, the ability of multi-source information synthesis can be trained, and the classification ability of the model can be enhanced. Multiple types of input can make neural network extract more effective and sensitive features from different data.

3.3.3 Comparison with related algorithms

Xie et al. [23] proposed a tool condition monitoring method based on LSTM. The whole LSTM network consists of one input layer, two hidden layers, and one output layer: the input layer uses tool twin data as input, the hidden layer includes one LSTM layer, and one full connection layer and the output layer uses regression to classify signals as well. After repeated iterative training of labeled data sets, LSTM neural network with tool condition monitoring ability is finally obtained. The number of LSTM neurons is 64, dropout value is 0.4, and batch size value is 128. The trained model is applied to the test set, and the accuracy of the test set is 89.7%. At the same time, we also built a simple RNN model, the learning rate is $1 \times 10^{-5}$, the number of RNN neurons is 64, under the above parameters for training and testing, and the accuracy is 84.6%.

Ref [24] proposed a tool wear condition monitoring method based on PCA and C-SVM. In the same data set, the algorithm extracts six features, the penalty factor C and the parameter $g$ of RBF kernel function are 0.82932 and 0.15604, respectively. The data of the three tool states are classified under above parameters. A total of 10 tests are carried out, and the average accuracy is 73.3%.

The same data set is also applied to the tool condition monitoring model based on Support Vector Regression (SVR). After 10 tests in the test set, the average accuracy of the test set is 70.2%. At the same time, we also build a simple CNN structure, which has two layers of convolution layer: the basic learning rate is 0.0001, and the accuracy of test set is 80.7%. The specific comparison of the above algorithms is shown in Table 7.

| Methods       | Iterations | Number of tool condition | Testing accuracy (%) |
|---------------|------------|--------------------------|----------------------|
| SVR           | 10         | 6                        | 70.2                 |
| PCA and C-SVM | 10         | 6                        | 73.3                 |
| Simple RNN    | 300        | 6                        | 84.6                 |
| LSTM          | 300        | 6                        | 89.7                 |
| Simple CNN    | 4000       | 6                        | 80.7                 |
| CNN           | 4000       | 6                        | 91.0                 |

As can be seen from Table 7, compared with the traditional monitoring algorithm, the neural network algorithms based on machine learning provide better accuracy in the same data set, and the accuracy in the test set is more than 80%. Although compared with the traditional monitoring algorithm, the training time of neural network is generally longer, and the number of epoch rounds is also larger. According to the size of training set, the training time varies from several hours to dozens of hours and consumes more computing resources. However, when it is applied to tool condition monitoring, it tends to be more accurate. In terms of monitoring time, the neural network only takes seconds to test. At the same time, convolutional neural network can extract more and higher dimensional features from the original data, and it can show better generalization when processing image information. Even if it is applied to data sets under other working conditions, it only needs to retrain the network to get a new model.

It can be seen that the performance of RNN, CNN, and LSTM with gating unit is better than the traditional “feature extraction + classification” monitoring method. In the traditional tool condition monitoring methods, it often takes a lot of time to preprocess the data in the feature extraction stage to capture sensitive features, such as using wavelet analysis, Fourier transform, and other methods. Although the extracted features can represent the tool state, they still contain some redundant information, and more sensitive features may be omitted, which is inevitable for manual feature extraction. This method is time-consuming and needs to rely on empirical knowledge. With the explosion of industrial data, the disadvantages of this kind of methods are gradually significant. On the contrary, the method based on deep learning can adaptively learn features from the original data and mine deep-seated features without relying on empirical knowledge and manpower, effectively avoiding the limitations of human feature extraction. Because dropout layer is added to the network, the model does not rely on local features, which improves the generalization of the model.

4 Conclusion

In this paper, the tool condition monitoring system framework on cloud platform is constructed, and the tool condition monitoring model based on convolution neural network is proposed to realize the tool condition monitoring using by sensors. In the training phase of convolutional neural network, the sensor data is preprocessed, the data curve is smoothed by PAA algorithm, and then the data is imaged
by GADF, and the sensor data is converted into graphics for CNN to use. At the same time, the infrared sensor image is used as the neural network data set for training and testing. Both time series signals and image information are used for classification purpose. In the test phase, the model is evaluated by confusion matrix, accuracy rate, F1 score, and other indicators. The results show that the algorithm can achieve satisfactory monitoring performance:

1) Compared with single sensor model, multiple sensor model can improve classification accuracy from 82% to 91%. It shows the necessity of multi-sensor to improve the accuracy of the model.

2) Compared with related algorithms, the algorithm performs best in classification accuracy. While the classification accuracy of traditional algorithms can only achieve 73.3% in the test set, the algorithm can achieve 91% accuracy rate. Moreover, the results also show that it has better goodness of fit and generalization performance than other neural network algorithms.

**Author contribution**  Kou Rui: experiments and writing—algorithm and original draft preparation. Lian Shi-wei: data processing and writing—reviewing. Xie Nan: scheme development and overall experimental design. Lu Bei-er: experimental platform construction. Liu Xue-mei: validation.

**Funding**  This work was supported by National Key R&D Program of China (2018YFB1700902).

**Availability of data and materials**  The authors confirm that the data supporting the findings of this study are available within the article.

**Declarations**

**Ethics approval**  Not applicable.

**Consent to participate**  Consent to participate in this study was obtained from all the authors.

**Consent for publication**  Not applicable.

**Conflict of interest**  The authors declare no competing interests.

**References**

1. Wang G, Li Z, Dong Y (2018) Advances in intelligent tool condition monitoring research. Aeronautical Manufacturing Technology 61(06):16–23. https://doi.org/10.16080/j.issn1671-833x.2018.06.016

2. Kurada S, Bradley C (1997) A review of machine vision sensors for tool condition monitoring. Comput Ind 34:55–72. https://doi.org/10.1016/s0166-3615(96)00075-9

3. Baur M, Albertelli P, Monno M (2020) A review of prognostics and health management of machine tools. Int J Adv Manuf Technol 107:2843–2863. https://doi.org/10.1007/s00170-020-05202-3

4. Zhou Y, Sun B, Sun W (2020) A tool condition monitoring method based on two-layer angle kernel extreme learning machine and binary differential evolution for milling. Measurement 166:108186. https://doi.org/10.1016/j.measurement.2020.108186

5. Mohanraj T, Shankar S, Rajasekar R, Sakhivel NR, Pramanik A (2020) Tool condition monitoring techniques in milling process—a review. J Market Res 9:1032–1042. https://doi.org/10.1016/j.jmartr.2019.10.031

6. Lei Y, Jia F, Lin J, Xing S, Ding SX (2016) An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. IEEE Trans Industr Electron 63:3137–3147. https://doi.org/10.1109/tie.2016.2519325

7. Elgargni M, Al-Habaibeh A, Lotfi A (2015) Cutting tool tracking and recognition based on infrared and visual imaging systems using principal component analysis (PCA) and discrete wavelet transform (DWT) combined with neural networks. Int J Adv Manuf Technol 77:1965–1978. https://doi.org/10.1007/s00170-014-6576-y

8. Barreiro J, Fernández-Abia AI, González-Laguna A, Pereira O (2017) TCM system in contour milling of very thick-very large steel plates based on vibration and AE signals. J Mater Process Technol 246:144–157. https://doi.org/10.1016/j.jmatprot.2017.03.016

9. Painuli S, Elangovan M, Sugumaran V (2014) Tool condition monitoring using K-star algorithm. Expert Syst Appl 41:2638–2643. https://doi.org/10.1016/j.eswa.2013.11.005

10. Xu H, Zhang C, Hong GS, Zhou JH (2018) Gated recurrent units based neural network for tool condition monitoring. International Joint Conference on Neural Networks. https://doi.org/10.1109/ijcnn.2018.8489354

11. Li D, Shen M, Zhang H, Wang S, Chen F (2020) Wang Y (2020) Deep hole boring tools condition monitoring based on LSTM network. Modern Manufacturing Engineering 08:92–96. https://doi.org/10.16731/j.cnki.1671-3133.2020.08.015

12. Zhou JT, Zhao X, Gao J (2019) Tool remaining useful life prediction method based on LSTM under variable working conditions. Int J Adv Manuf Technol 104:4715–4726. https://doi.org/10.1007/s00170-019-04349-y

13. Lecun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. Proc IEEE 86:2278–2324. https://doi.org/10.1109/5.726791

14. Ambadekar PK, Choudhari CM (2020) CNN based tool monitoring system to predict life of cutting tool. SN Applied Sciences 2:860. https://doi.org/10.1007/s42452-020-2598-2

15. Janssens O, Slavkovikj V, Vervisch B, Stockman K, Locuiffer M, Verstockt S, Walle RVd, Hoecke SV (2016) Convolutional neural network based fault detection for rotating machinery. J Sound Vib 377:331–345. https://doi.org/10.1016/j.jsv.2016.05.027

16. Abdeljaber O, Avci O, Kiranyaz MS, Boashash B, Sodano H, Inman DJ (2018) 1-D CNNs for structural damage detection: verification on a structural health monitoring benchmark data. Neurocomputing 275:1308–1317. https://doi.org/10.1016/j.neucom.2017.09.069

17. Abdeljaber O, Avci O, Kiranyaz S, Gabbou M, Inman DJ (2017) Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. J Sound Vib 388:154–170. https://doi.org/10.1016/j.jsv.2016.10.043

18. Eren L, Inc T, Kiranyaz S (2019) A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN Classifier. Journal of Signal Processing Systems 91:179–189. https://doi.org/10.1007/s11265-018-1378-3
19. Wu Y, Yang F, Liu Y, Zha X, Yuan S (2018) A comparison of 1-D and 2-D deep convolutional neural networks in ECG classification. Proceedings of Annual International Conference of the IEEE Engineering in Medicine and Biology Society: 324-327.
20. Wang Z, Oates T (2015) Imaging time-series to improve classification and imputation. Proceedings of the 24th International Conference on Artificial Intelligence: 3939–3945.
21. Keogh EJ, Pazzani MJ (2000) Scaling up dynamic time warping for datamining applications. Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM 2000:285–289. https://doi.org/10.1145/347090.347153
22. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going Deeper with Convolutions. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition: 1-9. https://doi.org/10.1109/CVPR.2015.7298594
23. Xie N, Kou R, Yao Y (2020) Tool condition prognostic model based on digital twin system. Procedia CIRP 93:1502–1507. https://doi.org/10.1016/j.procir.2020.03.045
24. Xie N, Ma F, Duan M, Li A (2016) Tool wear condition monitoring based on principal component analysis and C-support vector machine [J]. Journal of Tongji University (Natural Science) 44:434–439. https://doi.org/10.11908/j.issn.0253-374x.2016.03.015

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.