Journal of Intelligent Manufacturing (2020) 31:953–966
https://doi.org/10.1007/s10845-019-01488-7

Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations

Zhiwen Huang[1] · Jianmin Zhu[1] · Jingtao Lei[2] · Xiaoru Li[1] · Fengqing Tian[1]

Received: 2 February 2019 / Accepted: 8 August 2019 / Published online: 14 August 2019

Abstract

Tool wear monitoring has been increasingly important in intelligent manufacturing to increase machining efficiency. Multi-domain features can effectively characterize tool wear condition, but manual feature fusion lowers monitoring efficiency and hinders the further improvement of predicting accuracy. In order to overcome these deficiencies, a new tool wear predicting method based on multi-domain feature fusion by deep convolutional neural network (DCNN) is proposed in this paper. In this method, multi-domain (including time-domain, frequency domain and time–frequency domain) features are respectively extracted from multisensory signals (e.g. three-dimensional cutting force and vibration) as health indictors of tool wear condition, then the relationship between these features and real-time tool wear is directly established based on the designed DCNN model to combine adaptive feature fusion with automatic continuous prediction. The performance of the proposed tool wear predicting method is experimentally validated by using three tool run-to-failure datasets measured from three-flute ball nose tungsten carbide cutter of high-speed CNC machine under dry milling operations. The experimental results show that the predicting accuracy of the proposed method is significantly higher than other advanced methods.

Keywords Tool wear predicting · Multi-domain · Feature fusion · Convolutional neural network · Milling

Introduction

During the process of metal cutting, dull or damaged cutters cause a loss of surface quality in the finished workpiece and increase manufacturing costs (Javed et al. 2016; Rehorn et al. 2005), thus real-time tool wear monitoring is extremely important in intelligent manufacturing. In order to avoid the interruption in the cutting process, researchers usually collected cutting force, vibration, acoustic emission, spindle motor current and power, etc. signals for online tool wear monitoring (Zhou and Xue 2018; Dimla Snr. 2000).

Relevant researches show that time domain, frequency domain and time–frequency domain features can effectively characterize tool wear condition. Morgan et al. (2018) extracted time domain features (including mean, standard deviation, maximum and sum) and frequency domain features (e.g. power and energy spectrum) from cutting force, spindle vibration and current signals to characterize tool wear condition. Zhang et al. (2015) used 41 generalized fractal dimensional features of acoustic emission signal, 4 energy features based on empirical mode decomposition and 18 energy features based on Hilbert transform of cutting acoustic signal for tool condition monitoring. Benkedjouh et al. (2015) extracted time–frequency features based on wavelet packet decomposition from cutting force, vibration and acoustic emission signals for health assessment and life prediction of cutting tools. The time domain features for tool wear diagnostic and prognostic used by Tobon-Mejia et al. (2012) included root mean square (RMS), peak and standard deviation of the force signal, RMS and kurtosis of vibration signal, and mean and standard deviation of acoustic emission signal. However, these extracted features are only for specific signals or domains, and it is obviously difficult to achieve a universal characterization of tool wear. Wang et al. (2017) comprehensively utilized time domain, frequency domain and time–frequency domain features (including RMS, variance, maximum, skewness, kurtosis, peak-to-peak, spectral skewness, spectral kurtosis and wavelet energy) of force and vibration signals for virtual tool wear sensing, which

✉ Jianmin Zhu
jmzhu_usst@163.com

[1] College of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China

[2] School of Mechatronic Engineering and Automation, Shanghai University, Shanghai 200072, China

provided certain references for universal characterization of tool wear.

As the widely used techniques for tool wear monitoring, artificial intelligence methods can establish nonlinear mapping between multi-domain features and tool wear. Pandiyan et al. (2018) used support vector machine (SVM) to monitor tool condition in grinding process and genetic algorithm to obtain features subset selection, and achieved a prediction accuracy of 94.7% for predicting abrasive belt condition states. Kong et al. (2018) utilized radial basis function kernel principal component analysis to fuse extracted features and Gaussian process regression to predict flank wear width of turning tool with a mean absolute error (MAE) of 7 μm. Karandikar et al. (2015) described the naive Bayes classifier method for tool condition monitoring and obtained good prediction accuracy of tool wear. Yu et al. (2017) proposed a weighted hidden Markov model for continuous-state tool wear monitoring and tool life prediction with a mean square error (MSE) of 150.9098 μm. Wang et al. (2014) developed a modified fuzzy ARTMAP network for online incremental learning of tool condition classification and achieved a recognition accuracy of over 95%. Wu et al. (2018) used an adaptive network fuzzy inference system to predict remaining useful life of machining tools with root mean square error (RMSE) of 1.8 μm. Ghosh et al. (2007) estimated milling tool wear using neural network-based sensor fusion technology with feature space filtering, and the prediction error was less than 25 μm. Purushothaman (2010) used artificial neural network (ANN) based on extended Kalman filter weight updation for tool wear monitoring and transformed input patterns to reduce dimensions and obtained a classification accuracy of 95.18%. Li et al. (2009) presented fuzzy neural network to estimate tool wear during milling operation and selected features through genetic algorithms, and the MSE of prediction performances is 17.4 μm. Nevertheless, due to the fact that the ability to learn high dimensional data and complex nonlinear relationships are limited by shallow network structure in aforementioned artificial intelligence methods, researchers have to rely on their prior experience or make multiple attempts to artificially reduce the multi-domain feature dimension, which easily causes the loss of potential effective features, and hinders the further improvement in accuracy of tool wear prediction.

Recently, deep learning (Hinton and Salakhutdinov 2006) has made great achievements in the classification and recognition of large dataset images (Krizhevsky et al. 2012), which has attracted the attention of researchers in the field of tool wear monitoring. Fu et al. (2017) transformed vibration signals collected during the drilling process into two-dimensional images and input them into the designed DCNN model, which obtained better overall performance than SVM and ANN model. However, they only considered the DCNN model as classifier for machine vibration state, which was unable to perform the real-time continuous prediction of tool wear. Chen et al. (2018) selected four time domain features (i.e. maximum, minimum, mean and standard deviation) from signals and applied deep belief network (DBN) to predict tool wear and obtained lower prediction error than ANN and support vector regression (SVR). Aghazadeh et al. (2018) employed wavelet transform to extract time–frequency domain features from signals and proposed a tool wear monitoring method based on spectral subtraction and convolutional neural network (CNN) in milling operation, which obtained higher prediction accuracy than SVM, Bayesian rigid network and K nearest neighbor method. Zhao et al. (2017) selected max and mean values of divided signals as health indicators, and then designed a convolutional bidirectional long-term memory network model to predict milling tool wear and obtained better performance than other recurrent models. Although these deep learning methods work more effectively than traditional methods of machine learning, extracted features fed into them are only for specific signals or domains, thus, cannot accomplish a generic solution that can be used for any collected signals. Besides, the feature datasets used in these methods are of low dimensionality and small capacity, which limited the improvement in accuracy of tool wear prediction.

In order to improve prediction accuracy, this paper proposes a new tool wear predicting method, which respectively extracts universal time domain, frequency domain and time–frequency domain features from multisensory signals, and then designs the DCNN model to map the relationship between these multi-domain features and real-time tool wear. Concretely, the rest of this paper is organized as follows: Sect. Convolutional neural network briefly introduces CNN. The principle of proposed method is described in Sect. Proposed methodology. Section Experimental datasets introduces experimental datasets. Results are discussed and comparisons are presented in Sect. Results and discussions. Finally, Sect. Conclusions concludes this paper.

Convolutional neural network

CNN has recently gained great attention in diagnostic and prognostic fields due to their data mining and feature mapping capabilities, which can automatically extract highly distinguishing features of input data by using different kernel filters. Therefore, high performance in different tasks can be achieved even with less pre-processing. The basic structure of CNN is composed of an input layer, series of alternated convolution layers and pooling layers, few fully connected layers, and an output layer.

Convolutional layer

The convolutional layer is to extract features of the input local regions through different convolution kernels, which are regarded as the feature filters. In the forward propagation process, the convolutional layer mainly performs the convolution operation between the feature filters and the input local regions, being followed by activation function. The output of convolutional layer can be regarded as the feature map obtained by feature extraction of feature matrix and the general formulation is as follows:

$$\mathbf{X}_j^l = f\left(\sum_{i\in M_j}\left(\mathbf{X}_i^{l-1} * \mathbf{K}_{ij}^l\right) + \mathbf{b}_j^l\right) \tag{1}$$

where $\mathbf{X}_j^l$ is the jth feature map of the lth layer, $f(\cdot)$ is the activation function, M_j is the group of the input to calculate the jth output, $\mathbf{X}_i^{l-1}$ is the ith input of the lth layer, $\mathbf{K}_{ij}^l$ is the kernel used in the lth layer, $\mathbf{b}_j^l$ is the jth bias of the lth layer.

In the backpropagation process, parameters of convolutional layer are updated by back-propagation algorithm (Lecun et al. 1998a), and the general formulation is as follows:

$$\frac{\partial L}{\partial \mathbf{K}_{ij}^l} = \sum_{u,v}\left(\left(\delta_j^l\right)_{u,v}\cdot\left(\mathbf{P}_i^{l-1}\right)_{u,v}\right), \quad \frac{\partial L}{\partial \mathbf{b}_j^l} = \sum_{u,v}\left(\delta_j^l\right)_{u,v} \tag{2}$$

where is the loss function, $(\mathbf{P}_i^{l-1})_{u,v}$ is the patch in $\mathbf{X}_i^{l-1}$ that was multiplied element-wise by $\mathbf{K}_{ij}^{l+1}$ during convolution in order to compute the element at (,) in the output convolution result $\mathbf{X}_j^l$, and δ_j^l is the jth element of the sensitivities in the th layer.

Pooling layer

The convolutional layer can obviously reduce the number of connections between layers, but the number of neurons in each feature map is not significantly reduced. A higher dimensionality is prone to over-fit, so the pooling layer is introduced to perform dimensionality reduction of the feature map to avoid overfitting and ensure the scale-invariant characteristics of the feature map to a certain extent (Lecun et al. 1998b). The calculated formulation for forward propagation of the pooling layer is as follows:

$$\mathbf{X}_j^l = f\left(\beta_j^l \cdot down(\mathbf{X}_j^{l-1}) + \mathbf{b}_j^l\right) \tag{3}$$

where $down(\cdot)$ represents a sub-sampling function,β_j^l is multiplicative bias and $\mathbf{b}_j^l$ is an additive bias of the lth layer.

Similarly, the back-propagation of the pooling layer is calculated as follows:

$$\frac{\partial L}{\partial \beta_j^l} = \sum_{u,v}\left(\delta_j^l \cdot down\left(X_j^{l-1}\right)\right)_{u,v}, \quad \frac{\partial L}{\partial \mathbf{b}_j^l} = \sum_{u,v}\left(\delta_j^l\right)_{u,v} \tag{4}$$

Fully connected layer

Fully connected layer is traditional multilayer perception, and the neurons in which are all connected to the neurons of the previous layer. The output is as follows:

$$\mathbf{X}^l = \sigma\left(\mathbf{W}^l\mathbf{X}^{l-1} + \mathbf{b}^l\right) \tag{5}$$

where $\sigma(\cdot)$ is the activation function, $\mathbf{W}^l$ is the weight and $\mathbf{b}^l$ is the bias of the lth layer.

The parameters of fully connected layer are also updated by back-propagation algorithm, and the formulation is as follows:

$$\frac{\partial L}{\partial \mathbf{W}^l} = \delta^l\left(\mathbf{X}^{l-1}\right)^T, \quad \frac{\partial L}{\partial \mathbf{b}^l} = \delta^l \tag{6}$$

Proposed methodology

The framework of proposed tool wear predicting method based on multi-domain fusion by DCNN in milling operation is shown in Fig. 1. This method mainly includes multi-sensor data acquisition, multi-domain feature extraction, model designing based on DCNN, and performance evaluation, which involves offline modeling and online predicting.

In the offline modeling process, multi-sensors are installed near workpiece on the CNC milling machine to collect the real-time signals in the cutting process, meanwhile microscope is used to measure the actual tool wear (i.e. flank wear width) as the target value for training the subsequent DCNN model. Then the multi-domain (including time domain, frequency domain and time–frequency domain) features of acquired signals are extracted and divided into training datasets and validating datasets. Particularly, the extracted features and measured tool wear are normalized before training the subsequent DCNN model. Next, the model is designed based on DCNN for learning the relationship between multi-domain feature and real-time tool wear, using the training datasets, validating datasets, and the target value as input. Finally, after continuous iterative training and performance evaluation, an excellent-performance DCNN model is obtained.

In the online predicting process, the monitoring signals of real-time tool wear are acquired by multi-sensor data acquisition, and then multi-domain features are extracted as the testing datasets, which are input to the trained DCNN model for real-time tool wear prediction. Then the normalized predicting values of tool wear are obtained through the

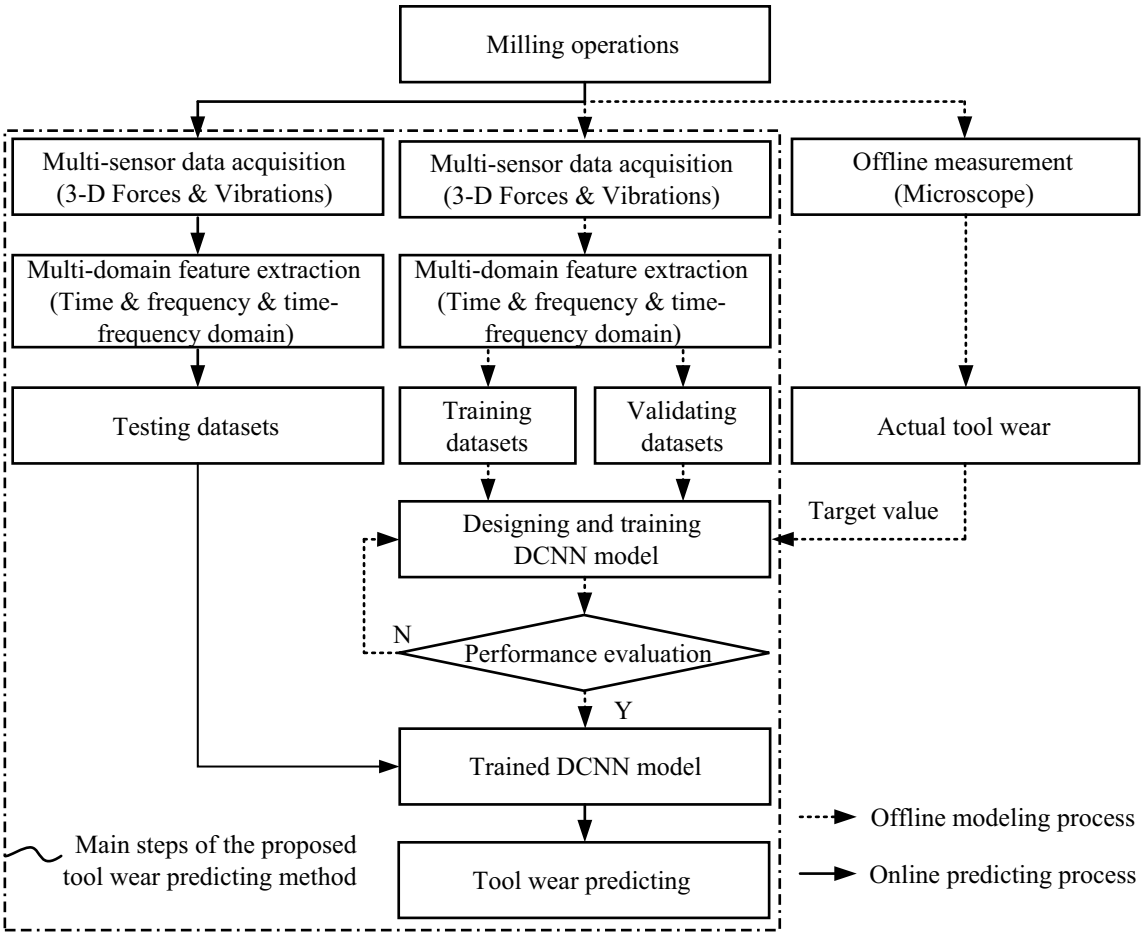

Fig. 1 The framework of the proposed tool wear predicting method in milling operations

trained model, and finally the prediction results are reversely normalized to obtain the final predicting value of tool wear. Therefore, the proposed method in this paper realizes adaptive multi-domain feature fusion and continuous tool wear prediction in the milling process through offline modeling and online predicting.

Multi-sensor data acquisition

Data acquisition for monitoring tool wear during machining process is inevitably disturbed by the noise of operating environment. Due to the limited applicability of its application, a single sensor can hardly collect signals contained multi-dimensional information of tool wear (Mali et al. 2017). Therefore, multi-sensor data acquisition is adopted to predict tool wear. The vibration signal accompanying the cutting process contains a wealth of information related to cutting wear and is considered to be a monitoring signal with high sensitivity to tool wear (El-Wardany et al. 1996), but the vibration signal is susceptible to the machining environment and sensor installation location.

In order to overcome the deficiency, vibration sensor is usually combined with other sensors for tool wear monitoring in practical applications (Salehi et al. 2015). The cutting force increases with the aggravation of tool wear during the cutting process, which is generally considered to be one of the most important monitoring signals of tool wear (Huang et al. 2007). Moreover, the vibration and cutting force signals in the x, y, and z directions are different in sensitivity to the change of tool wear (Dimla Sr. et al. 2000). Therefore, this paper uses the vibration and cutting force signals in three directions to predict tool wear in milling process.

Multi-domain feature extraction

The amount of raw signal data measured by multi-sensors for tool wear monitoring is large and inevitably involves various types of interference information such as environmental noise, so it is necessary to extract features that can reflect tool wear during the cutting process (Duro et al. 2016). As the aggravation of tool wear increases, the time domain features as important characterizing parameters of

signals collected by multi-sensors will inevitably change (Kuljanic and Sortino 2005). In addition, the increase of tool wear usually causes the change of frequency structure, and the frequency domain features is extracted to monitor tool wear of measured signals (Gierlak et al. 2016). The vibration and cutting force caused by tool wear during the machining process are non-stationary signals, containing a wealth of information of tool wear (García and Núñez López 2018). To this end, the researchers used wavelet packet transform to extract the time–frequency domain features of collected signals (Zhu et al. 2009; García and Núñez López 2018), which is as an effective complement to time domain and frequency domain features for monitoring tool wear. As a result, this paper extracts respectively six time domain features, two frequency domain features and one time–frequency domain feature of cutting force and vibration signals in x, y and z directions to characterize the tool wear information during the cutting process. The details of these features are as shown in Table 1 (Wang et al. 2017).

In order to facilitate the frequency and time–frequency domain feature calculation, 1024 is used as the sample data point to extract respectively nine features from six channel signals per time according to Table 1. Then, for each sample batch, a total of 54 multi-domain features are obtained to form a column of original feature matrix, which is regarded as feature axis of multi-domain feature matrix. Next, as mentioned earlier, DCNNs have achieved brilliant achievements in image recognition, which can adaptively extract highly distinctive features from the square pixel matrixes of input images. Motivated by the similarity between the square pixel matrix of high-dimensional image and multi-domain feature matrix of collected signals, we design time axis to have the same dimension as the feature axis of multi-domain feature matrix. Therefore, feature extraction is carried out successively to form a whole multi-domain feature matrix with a size of 54×54,

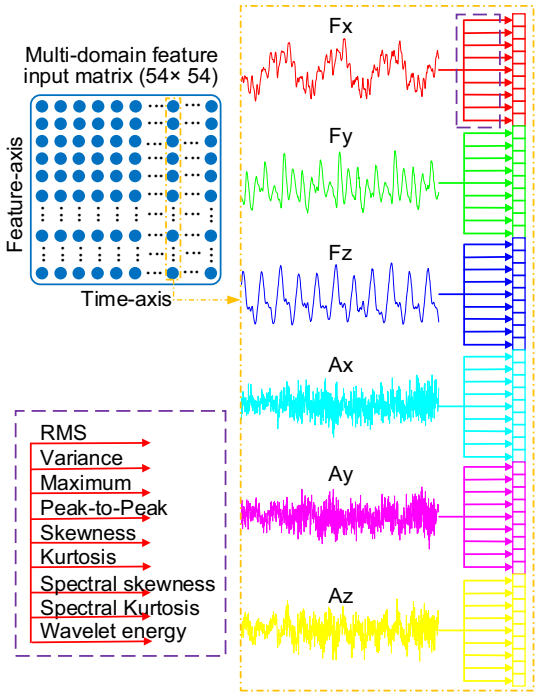

Fig. 2 The illustration of input matrix of multi-domain feature

as is shown in Fig. 2, which combines the feature axis and time axis. Finally, based on time sequence of collected signals, a series of multi-domain feature matrixes are extracted as the input of the DCNN model for tool wear predicting.

Designing and training DCNN model

Model designing is devoted to learning the complicated relationship between multi-domain features and real-time tool wear, which is the core of the proposed tool wear predicting method. Because of its powerful feature extracting and

Table 1 List of extracted features

Domain	Features	Expression
Time	RMS	$z_{rms} = \sqrt{\sum_{i=1}^{n} z_i^2 / n}$
	Variance	$z_{var} = \sum_{i=1}^{n} \left(z_i - \bar{z}\right)^2 / n$
	Maximum	$z_{max} = max(z)$
	Skewness	$z_{skew} = E\left[((z-\mu)/\sigma)^3\right]$
	Kurtosis	$z_{kurt} = E\left[((z-\mu)/\sigma)^4\right]$
	Peak-to-peak	$z_{p-p} = max(z) - min(z)$
Frequency	Spectral skewness	$f_{skew} = \sum_{i=1} k\left((f_i - \bar{f})/\sigma\right)^3 \cdot S(f_i)$
	Spectral kurtosis	$f_{kurt} = \sum_{i=1} k\left((f_i - \bar{f})/\sigma\right)^4 \cdot S(f_i)$
Time–frequency	Wavelet energy	$E_{WT} = \sum_{i=1}^{N} wt_{\phi}^2(i)/N$

data mapping ability, DCNN is widely used in solving high dimensional and intricate nonlinear problems. To this end, we designed a tool wear predicting model based on DCNN in this paper.

Concretely, after the input of multi-domain feature matrix, three convolutional layers and three pooling layers alternate with each other for adaptive mining sensitive features of tool wear. The convolutional and pooling kernel size are respectively 5×5 and 3×3, and their kernel stride are respectively 1×1 and 2×2. In this way, the high-level representation for each raw feature is obtained. Finally, the regression layer with three neurons are connected to the fully connected layer to perform automatic wear prediction of three flutes.

In mechanical intelligence diagnostic and prognostic fields, many methods still use the sigmoid function as the activation function of CNN, but the sigmoid function has the deficiencies of gradient disappearance and slow convergence in the training process. However, the rectified linear units (ReLU) function can effectively overcome these deficiencies. Hence, the ReLU function is applied as the activation function of the convolutional, pooling and fully connected layers. The details of the designed DCNN model are shown in Table 2.

The Euclidean distance between the actual tool wear value $\boldsymbol{y}$ and the predicted tool wear value $\hat{\mathbf{y}}$ is used to train the designed DCNN model and provide the error feedback, which is taken as the loss function as follows:

$$L = \frac{1}{2N}\sum_{i=1}^{N}\left\|\hat{\mathbf{y}}_i - \mathbf{y}_i\right\|_2^2 \tag{7}$$

where N is the number of samples.

As a widely used strategy, cross-validation is used to effectively improve generalization ability of the model (Arlot and Celisse 2010). Therefore, in order to avoid overfitting, cross-validation strategy is used for training the designed DCNN model. Besides, as an effective method to reduce the shift of internal covariance, normalization is widely used to accelerate the training process of deep neural network (Zhang et al. 2018), thus we utilize normalization in the training processing of the proposed model. To further improve the predicting performance, the training process uses the back-propagation algorithm to minimize the loss function, which can be divided into the following steps. First, use the forward propagation algorithm to calculate the output of each layer until the last layer. Then, use back propagation algorithm to update the parameters of each layer. Finally, repeat this cycle until the model is converged.

Performance evaluation

In order to quantitatively evaluate the overall performance of the proposed tool wear predicting method in this paper, the following four criteria are adopted, including Pearson correlation coefficient (PCC), mean absolute percentage error (MAPE), mean absolute error (MAE) and root mean squared error (RMSE). The computational formula of these criteria is shown in Table 3.

Among these criteria for performance evaluation, the larger the value of PCC, the closer the predicted tool wear amount gets to the actual tool wear amount, then the model performance is better and the prediction accuracy is higher. On the contrary, the smaller the value of MAPE, MAE and

Table 3 List of different criteria

Criterion	Expression
Pearson correlation coefficient	$f_{\mathrm{PCC}}=\frac{\sum_{i=1}^{N}(\mathbf{y}_i-\bar{\mathbf{y}})(\hat{\mathbf{y}}_i-\bar{\hat{\mathbf{y}}})}{\sqrt{\sum_{i=1}^{N}(\mathbf{y}_i-\bar{\mathbf{y}})^2\sum_{i=1}^{N}(\hat{\mathbf{y}}_i-\bar{\hat{\mathbf{y}}})^2}}$
Mean absolute percentage error	$f_{\mathrm{MAPE}}=\sum_{i=1}^{N}(\lvert\hat{\mathbf{y}}_i-\mathbf{y}_i\rvert/\mathbf{y}_i)/N$
Mean absolute error	$f_{\mathrm{MAE}}=\sum_{i=1}^{N}\lvert\hat{\mathbf{y}}_i-\mathbf{y}_i\rvert/N$
Root mean squared error	$f_{RMSE}=\sqrt{\sum_{i=1}^{N}(\hat{\mathbf{y}}_i-\mathbf{y}_i)^2/N}$

Table 2 Details of designed DCNN model used in proposed method

No.	Layer type	Kernel size/stride	Kernel channel size	Output size	Padding	Activation function
1	Input			54×54		
2	Convolution1	5×5/1×1	64	50×50	No	ReLU
3	Pooling1	3×3/2×2	64	25×25	No	ReLU
4	Convolution2	5×5/1×1	96	21×21	No	ReLU
5	Pooling2	3×3/2×2	96	10×10	No	ReLU
6	Convolution3	5×5/1×1	128	6×6	No	ReLU
7	Pooling3	3×3/2×2	128	3×3	No	ReLU
8	Fully-connected			192×1		ReLU
9	Output			3×1		Sigmoid

RMSE, the closer the predicted tool wear amount gets to the actual tool wear amount, then the model performance is better and the prediction accuracy is higher. Particularly, PCC and MAPE are dimensionless criteria, while MAE and RMSE are dimension criteria.

Experimental datasets

Experimental setup

In order to experimentally verify the performance of the tool wear predicting method proposed in this paper, a set of experimental data measured from three-flute ball nose tungsten carbide cutter of high-speed CNC machine under dry milling operations was used (Li et al. 2009). During the tool wear process in milling operations, a Kistler 3-component dynamometer was mounted between the workbench and the workpiece to measure cutting forces, and three Kistler piezo accelerometers were mounted respectively on the workpiece in x, y, z directions to measure the vibration signals.

Table 4 The operation parameters of the experimental platform

Parameters	Value
The running speed of the spindle	10,400 rpm
The feed rate in x direction	1555 mm/min
The depth of cut (radial) in y direction	0.125 mm
The depth of cut (axial) in z direction	0.2 mm

Meanwhile, DAQ NI PCI1200 was adopted to acquire signals of these multi-sensors with a continuous sampling frequency of 50 kHz. These acquired data was stored and processed, then used for training, validating and testing of the DCNN model for tool wear prediction. Correspondingly, after each milling process was completed, the actual wear (i.e. flank wear width) of each cutting edge was measured offline using a LEICA MZ12 microscope, which was stored and set as the target value for training and validating of the DCNN model for tool wear prediction. The operation parameters of the experimental platform are shown in Table 4.

Data preparation

In this paper, three tools (C1, C4 and C6) with complete wear data are selected as the datasets. Each tool contains 300 data files, corresponding to 300 cutting operations. Here, the collected data of C1 is taken as an example for data analysis. The sample signals of force and vibration in x, y, and z directions under different wear conditions are shown in Fig. 3. It is clear that the amplitude of both force and vibration signals in three directions increase significantly with the aggravation of tool wear during the cutting process. Therefore, applying multi-sensor data acquisition to collect force and vibration signals in x, y, and z directions is an effective strategy to monitor tool wear.

Using multi-domain feature extraction to obtain respectively nine features described in Table 1 from three-dimensional force and vibration signals. Visually, the normalized

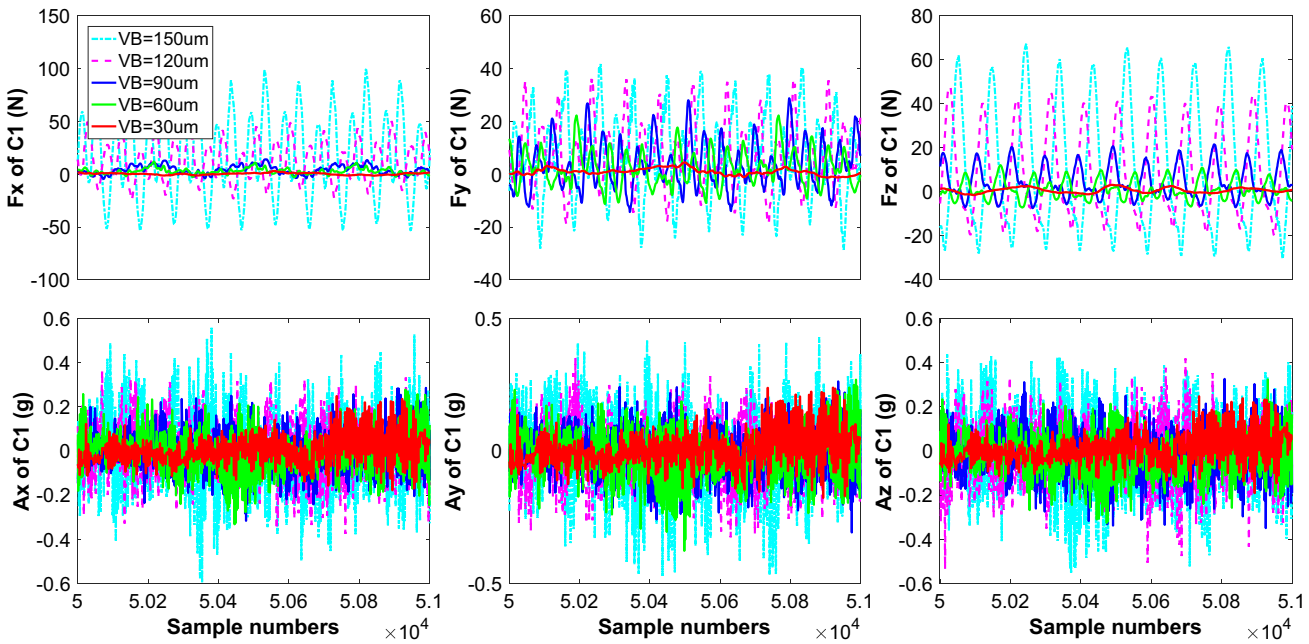

Fig. 3 The sample signals of force and vibration of C1 under different tool wear conditions in three directions

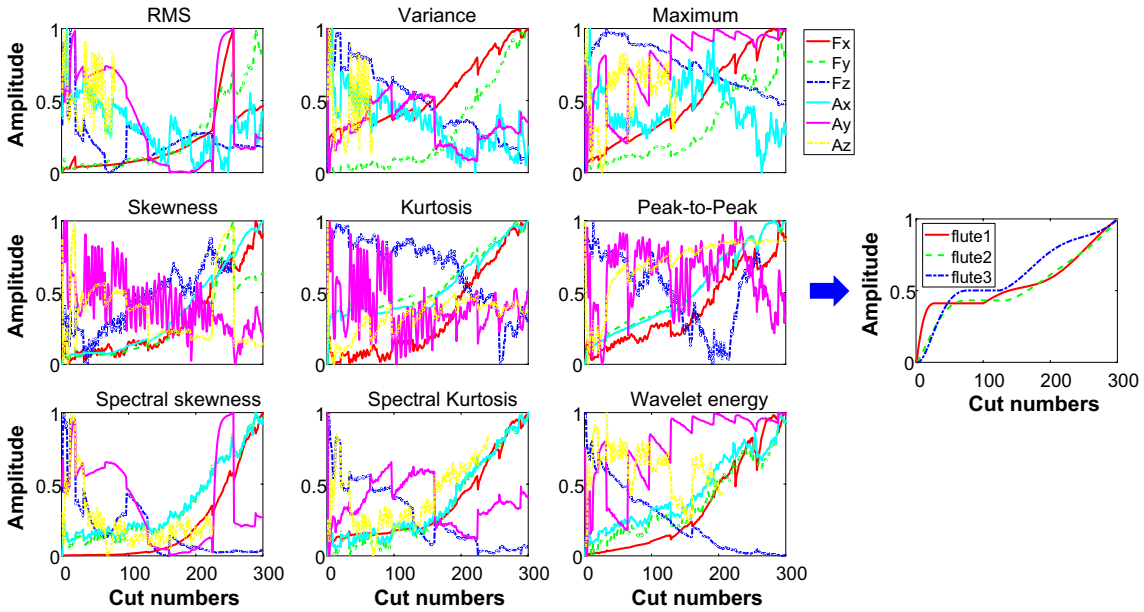

Fig. 4 The extracted features of force and vibration of C1 in three directions

extracted features and normalized actual flank wear measurements are shown in Fig. 4.

It is found that some of the original features are positively or negatively correlated with the tool wear during the cutting process. Thus, it is a feasible strategy to predict tool wear by multi-domain features. However, the relationship between the other original features of the force and vibration signals and the tool wear trend is not obvious. Therefore, it is necessary to use the designed DCNN model to mine adaptively the deep intrinsic features hidden in the original features to monitor the tool wear.

Computing platform

According to the multi-domain feature extraction method described in Sect. Multi-domain feature extraction, training datasets, validating datasets and testing datasets with different original feature matrixes are obtained to verify the DCNN model. Concretely, the description of experimental datasets is shown in Table 5.

The proposed DCNN model for tool wear prediction in milling operations is conducted using a NVIDIA Tesla P100 GPU on Alibaba elastic compute service (ECS) with 8-core CPU in Ubuntu16 04 operating system, and CUDA8 is used to accelerate the calculation. The training time for each epoch is about 6.27 s, and the testing time of a single sample is only 0.0025 s. Therefore, the proposed DCNN model is a highly effective tool wear predicting method.

Results and discussions

When using the designed DCNN model for tool wear prediction, it is crucial to find the right parameters. However, the parameters of the DCNN model vary with the experimental datasets, thus adjusting the parameters to choose the appropriate parameters of the corresponding datasets is an important link in the application of the designed DCNN model. In the experimental studies, the parameters includes gradient descent algorithm, batch size, learning rate and number of epochs. Considering the similarity of the wearing process, only datasets of C1 are used as an example for parameters selection.

Table 5 Description of experimental datasets

Tool symbol	Training datasets	Validating datasets	Testing datasets	Total datasets
C1	27,000	9000	9000	45,000
C4	27,000	9000	9000	45,000
C6	27,000	9000	9000	45,000

Selection of gradient algorithm

As a popular optimization algorithm, the gradient descent algorithm is frequently used to optimize neural networks. However, due to the difficulty in achieving theoretical explanations of their advantages and disadvantages, the algorithm is usually regarded as black box optimizers (Ruder 2016). To this end, the following common gradient descent optimization algorithms (including Adadelta, Adagrad, Nestrov, RSMprop, and SGD) are compared experimentally to select the appropriate gradient descent algorithm. The training and validating loss under different gradient descent algorithms are shown in Fig. 5. Correspondingly, the results of performance evaluation are shown in Fig. 6.

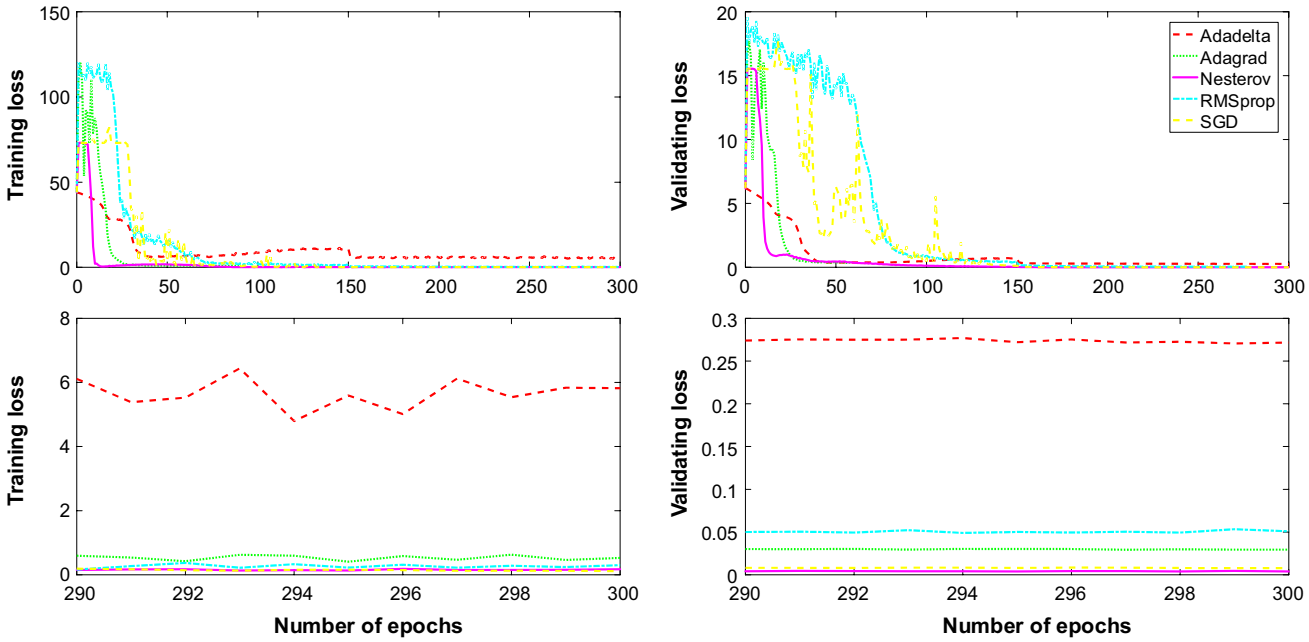

Fig. 5 The training and validating loss of C1 under different gradient descent algorithms

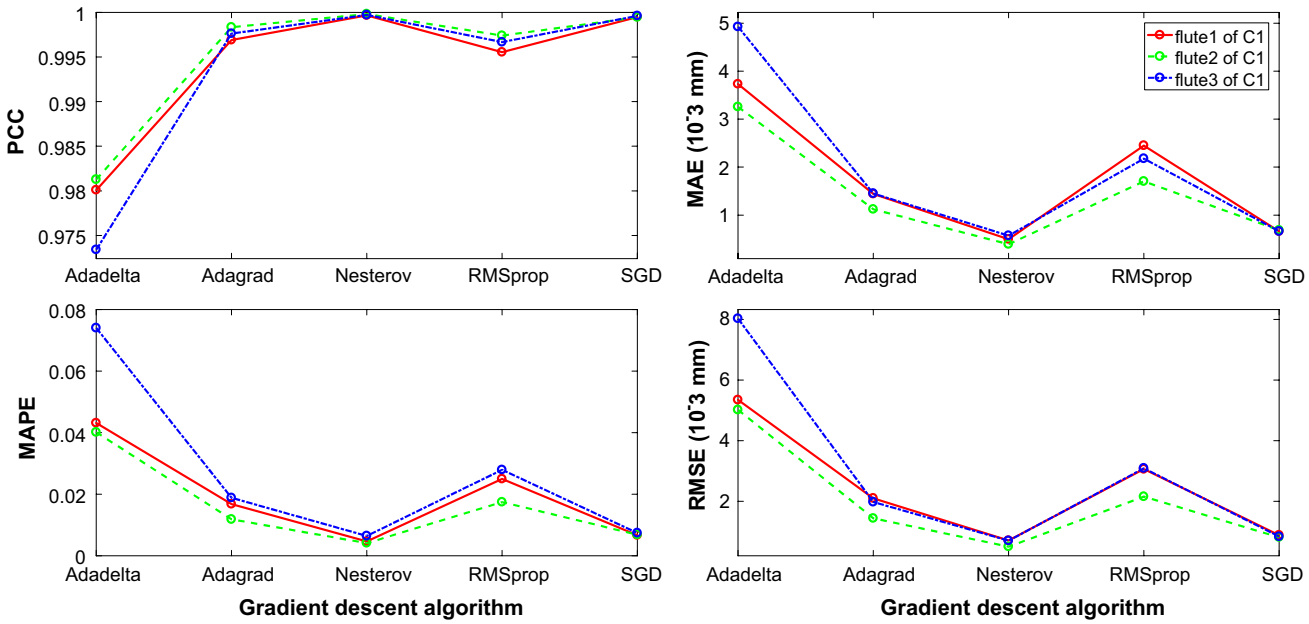

Fig. 6 The performance comparison of C1 under different gradient descent algorithms

It can be seen from Figs. 5 and 6 that the Adadelta and RMSProp algorithms not only have a slow convergence, but also a big training error. Although Adagrad algorithm converges faster, the training error is bigger. Nesterov and SGD obtain a smaller training error, but the Nesterov algorithm has faster converge speed and lower error. Therefore, the Nesterov algorithm is chosen to train the DCNN model.

Selection of batch size

Due to the large capacity of datasets and the limitation of computing resources, it is difficult to simultaneously calculate the gradient and update the parameters of all samples during network training process. It is usually preferable to divide the dataset into suitable batches (Keskar et al. 2016). The batch size is an important parameter in network training. Thus, different batch sizes in the experiment are selected for DCNN training. The training time and loss are shown in Fig. 7. The results of performance evaluation under different batch sizes are shown in Fig. 8.

As is shown in Figs. 7 and 8, when batch size is smaller, the training loss is smaller, moreover, the PCC is larger, the MAPE, MAE and RMSE are smaller, the predicted tool wear is closer to the actual wear, but the training time per epoch is longer. Conversely, when batch size is larger (especially over 50), the training time per epoch is significantly reduced and tended to be stable, but the PCC is smaller, the MAPE, MAE and RMSE increase significantly. To sum up, when batch size equals to 50, not only can the model prediction accuracy be ensured, but also the network training time can be reduced.

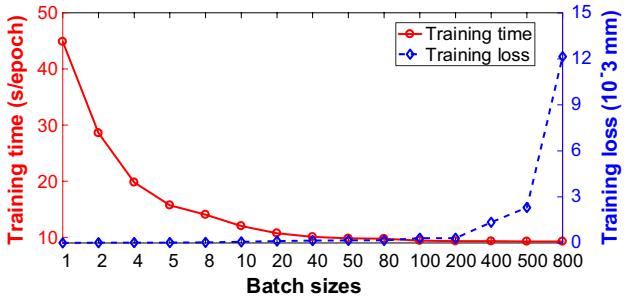

Fig. 7 The training time and loss of C1 under different batch sizes

Selection of learning rate

In the model training process, the gradient descent algorithm is used for optimization. The learning rate is an important parameter, which not only affects the weight of each layer, but also the error convergence. In order to improve the efficiency of the network, it is very important to choose the appropriate learning rate. In the experiment, the DCNN network training is carried out using different learning rates. The results of performance evaluation under different learning rates are shown in Fig. 9.

It is obviously seen from Fig. 9 that when learning rate is too large or too small, the PCC is smaller, the MAPE, MAE and RMSE are larger, and namely the error between the predicted tool wear and actual tool wear is larger.

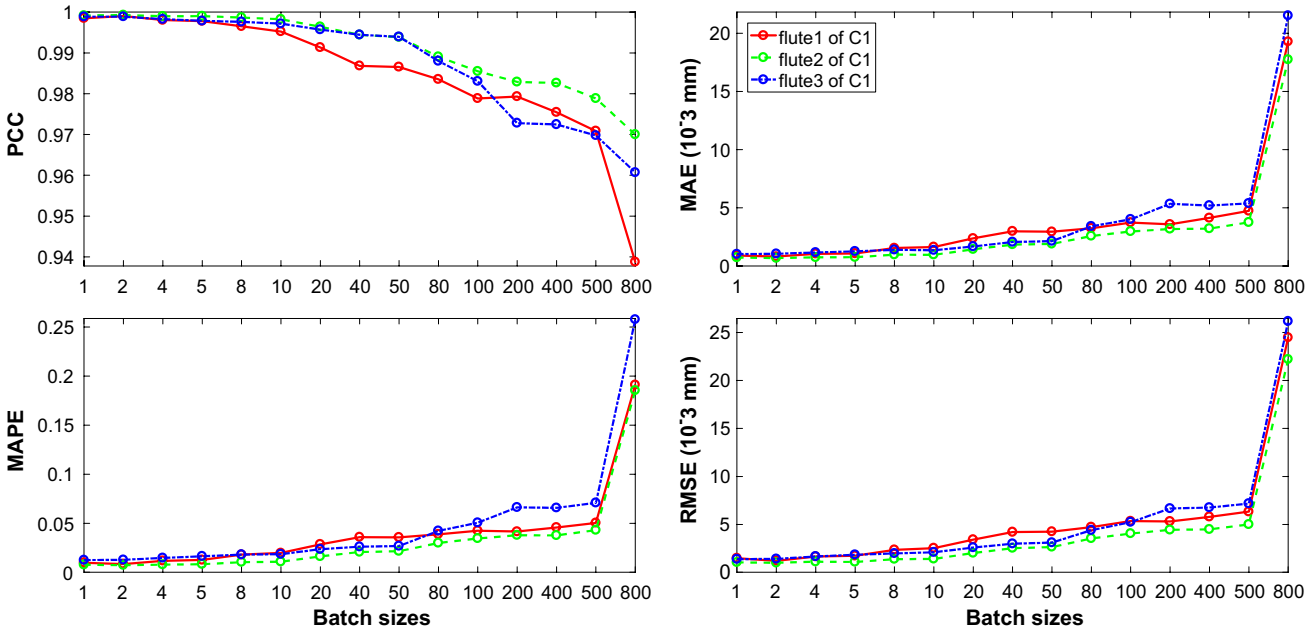

Fig. 8 The performance comparison of C1 under different batch sizes

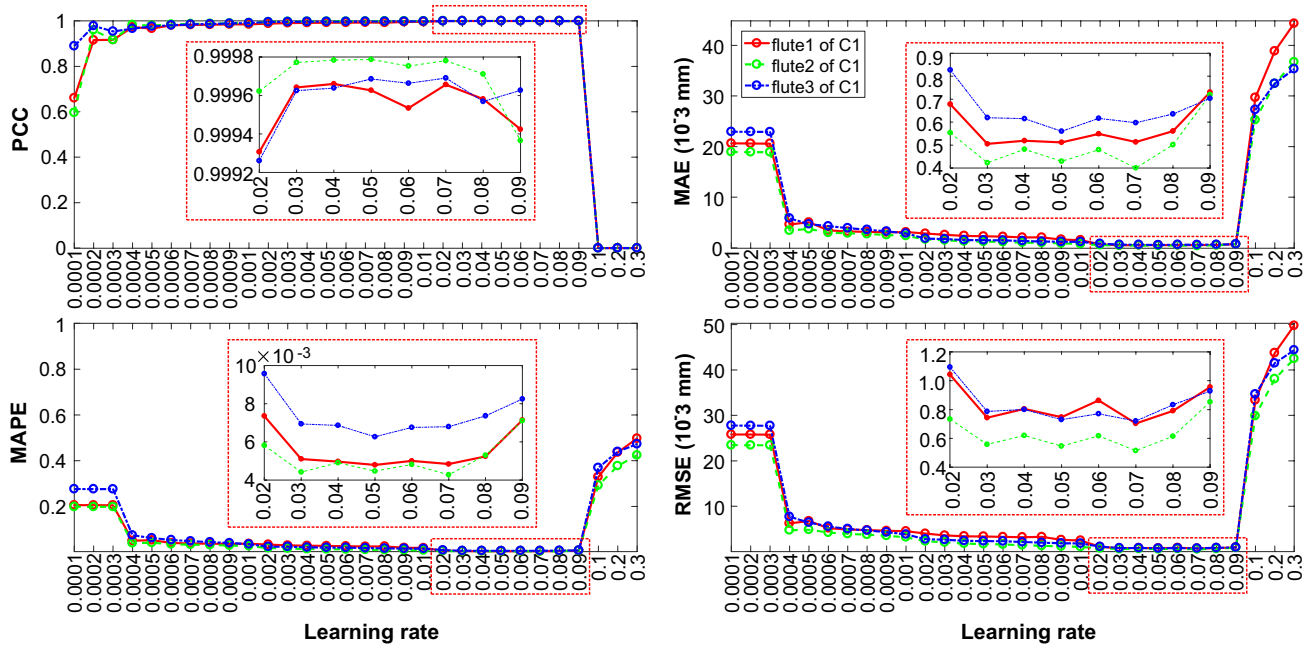

Fig. 9 The performance comparison of C1 under different learning rates

Nevertheless, when learning rate is moderate, the PCC is larger, the MAPE, MAE and RMSE are smaller, and therefore the error between the predicted tool wear and the actual tool wear is smaller. Considering the overall performance of the tool wear prediction, this experiment selects 0.06 as the appropriate learning rate.

Selection of epoch

The dataset is fully trained once called an epoch, number of epochs is crucial for the DCNN model training, which affects the convergence error of the model. In this experiment, different number of epochs are used for the DCNN

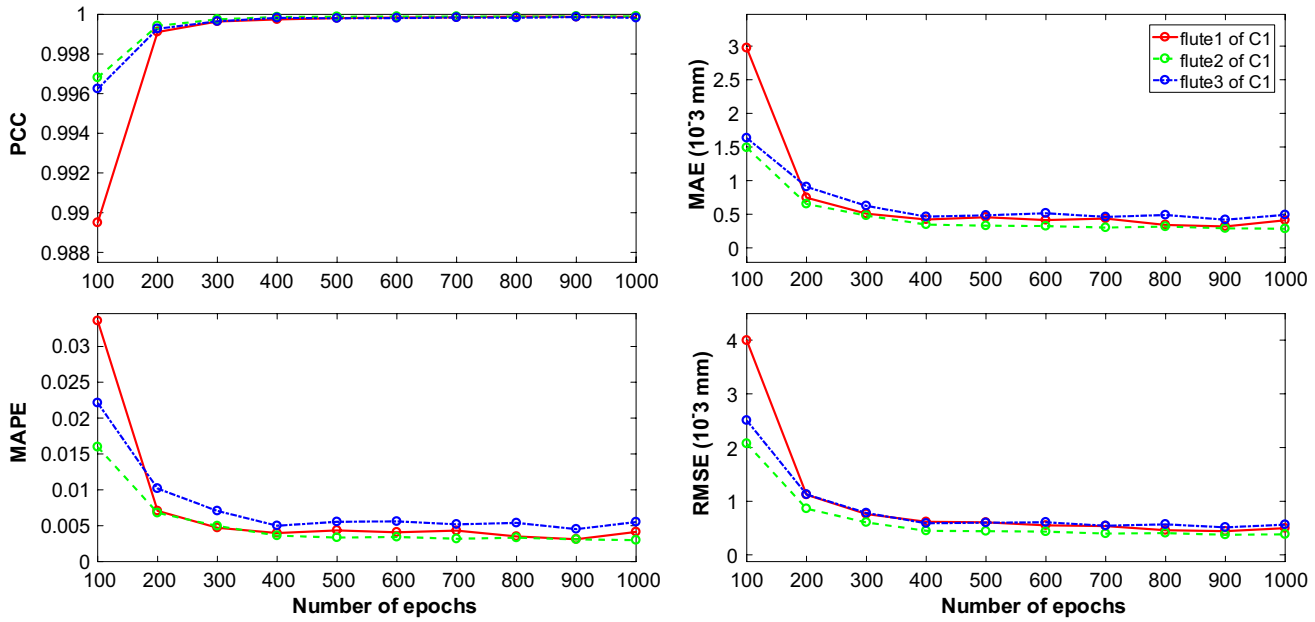

Fig. 10 The Performance comparison of C1 in different number of epochs

Table 6 The optimized parameters of proposed method

Gradient descent algorithm	Batch size	Learning rate	Number of epoch
Nesterov	50	0.06	400

model training. The corresponding results of performance evaluation under different number of epochs are shown in Fig. 10.

It is easy to know from Fig. 10 that when the number of epochs is too small, the model training is insufficient, the PCC is small, the MAPE, MAE and RMSE are large, that is, the tool prediction error is large. On the contrary, when the number of epochs is gradually increased, the model training becomes more adequate, the prediction error is gradually reduced. However, when the number of epochs is too large (especially over 400), the network training is over-fitted, and the prediction error is basically stable. As a result, the experiment determines 400 as the appropriate number of epochs.

Comparisons of other methods

Through the above series of parameter selection experiments, the appropriate parameters of the DCNN model used in the proposed method are shown in Table 6.

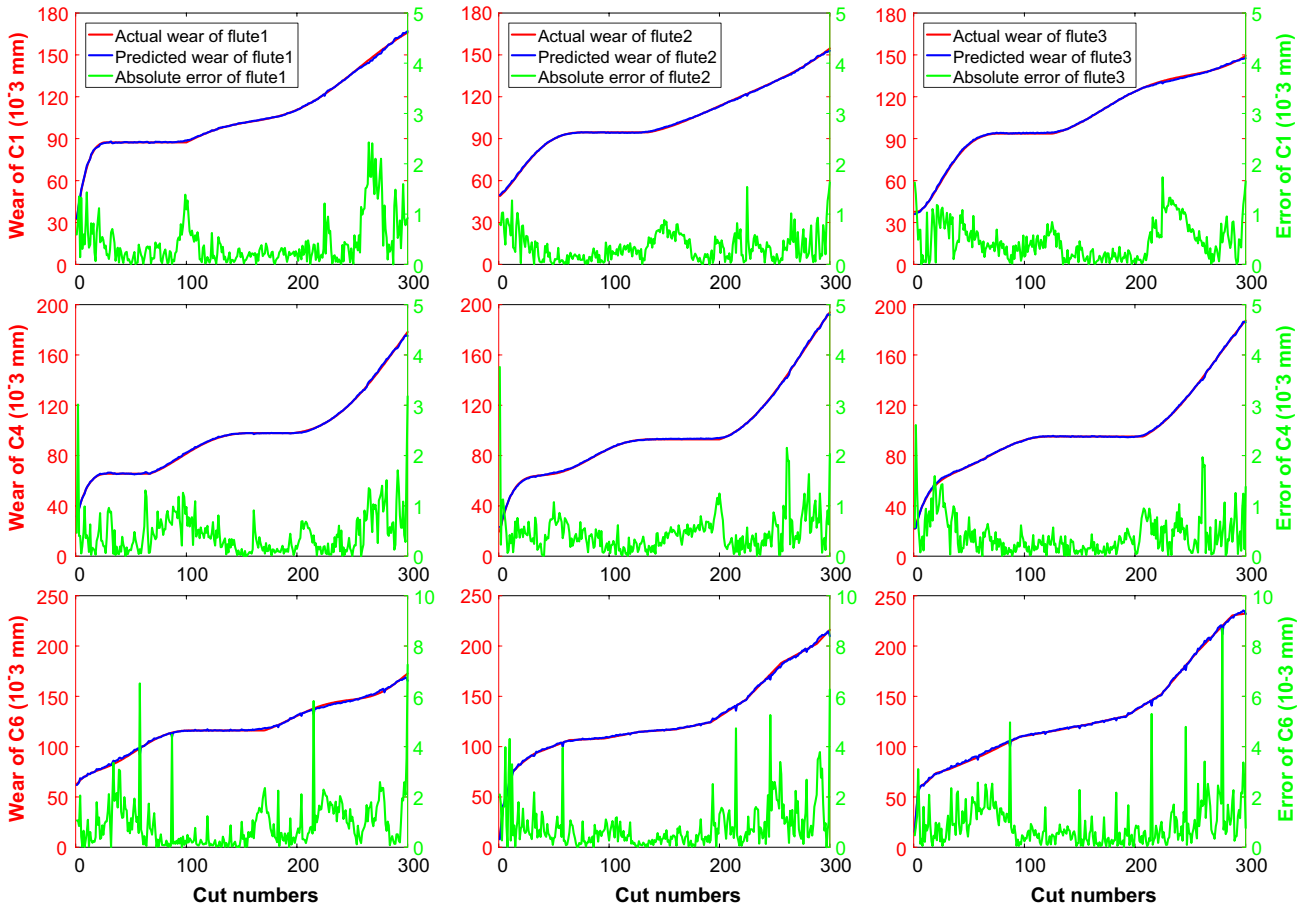

Fig. 11 The tool wear predicting result of the proposed method

Table 7 The results of the approaches in tool wear prediction under different criteria

Method	PCC	MAPE	MAE (10^{-3} mm)	RMSE (10^{-3} mm)
SVR[a]	0.9541 ± 0.0292	0.0845 ± 0.0215	9.3770 ± 2.0422	11.9681 ± 3.3337
SVR + KPCA[a]	0.9844 ± 0.0054	0.0377 ± 0.0100	3.9583 ± 0.9371	5.4428 ± 1.5894
Our proposed	0.9997 ± 0.0003	0.0057 ± 0.0016	0.5706 ± 0.2278	0.7998 ± 0.3538

[a]Results from method by Wang et al. (2017)

Corresponding, the predicted tool wear after reversed normalization under different datasets is illustrated in Fig. 11.

In order to further verify its effectiveness and advancement, the tool wear predicting method proposed in this paper is compared with other methods proposed in the literature using the originally published data, and the results are shown in Table 7.

As shown in Table 7, the mean value of the MAE of predicted tool wear in the three tools is 0.5706×10^{-3} mm and the mean value of the RMSE is 0.7998×10^{-3} mm, which is the minimum value in all comparative methods. Moreover, the variance of MAE is 0.2278×10^{-3} mm, and the variance of RMSE is 0.3538×10^{-3} mm, which is also the minimum value in all comparative methods. Therefore, the method proposed in this paper can ensure not only the low deviation of the tool wear predictions, but also a small degree of dispersion. In summary, the method can correctly and effectively establish the mapping relationship between the multi-domain features of multi-sensor and the real-time tool wear, and achieve higher tool wear prediction accuracy.

Conclusions

This paper presents a novel tool wear predicting method based on multi-domain feature fusion by the proposed DCNN model in milling operations to improve the predicting accuracy. The realization of the proposed method benefits from the following techniques. First, multi-sensor data acquisition provides comprehensive signals to monitor the tool wear. Then, multi-domain feature extraction provides original characteristics to characterize the tool wear. Next, the designed DCNN model adaptively mine sensitive features from the original characteristics and automatically predict real-time tool wear. Finally, the designed DCNN model is optimized by performance evaluation, which helps to achieve the excellent predicting result.

The main contribution of the proposed method is the continuous prediction of tool wear in milling operations with adaptive feature fusion by DCNN, omitting the step of optimizing features by dimension reduction techniques. Moreover, the innovation is that the constructed multi-domain feature matrix combines feature axis and time axis, and the designed DCNN model effectively learns the complicated relationship between obtained multi-domain feature matrixes and real-time tool wear. Ultimately, the proposed method is validated by three tool wear experimental datasets measured from three-flute ball nose tungsten carbide cutters of high-speed CNC machine under dry milling operations.

The proposed method has the application in milling operations, and the discussion in this paper provides some useful suggestions for the deployment of the proposed approach in real machining processes. In addition, the method also can easily apply to tool wear prediction in other machining process, such as turning, drilling, etc. In future work, on the one hand, we will increase the type of sensors to enrich the multi-sensor fusion technology for tool wear prediction; on the other hand, try to use raw signals to predict tool wear directly.

Acknowledgements This research is supported by National Natural Science Foundation of China (Nos. 50975179, 51375289 and 51775323).

References

Aghazadeh, F., Tahan, A., & Thomas, M. (2018). Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process. *The International Journal of Advanced Manufacturing Technology, 98*(9–12), 3217–3227.

Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys, 4,* 40–79.

Benkedjouh, T., Medjaher, K., Zerhouni, N., & Rechak, S. (2015). Health assessment and life prediction of cutting tools based on support vector regression. *Journal of Intelligent Manufacturing, 26*(2), 213–223.

Chen, Y., Jin, Y., & Jiri, G. (2018). Predicting tool wear with multi-sensor data using deep belief networks. *The International Journal of Advanced Manufacturing Technology*, *99*(5–8), 1917–1926.

Dimla, D. E., Sr., & Lister, P. M. (2000). On-line metal cutting tool condition monitoring. I: Force and vibration analyses. *International Journal of Machine Tools and Manufacture, 40*(5), 739–768.

Dimla Snr, D. E. (2000). Sensor signals for tool-wear monitoring in metal cutting operations—A review of methods. *International Journal of Machine Tools and Manufacture, 40,* 1073–1098.

Duro, J. A., Padget, J. A., Bowen, C. R., Kim, H. A., & Nassehi, A. (2016). Multi-sensor data fusion framework for CNC machining monitoring. *Mechanical Systems and Signal Processing, 66–67,* 505–520.

El-Wardany, T. I., Gao, D., & Elbestawi, M. A. (1996). Tool condition monitoring in drilling using vibration signature analysis. *International Journal of Machine Tools and Manufacture, 36*(6), 687–711.

Fu, Y., Zhang, Y., Gao, Y., Gao, H., Mao, T., Zhou, H. M., et al. (2017). Machining vibration states monitoring based on image representation using convolutional neural networks. *Engineering Applications of Artificial Intelligence, 65,* 240–251.

García, P. E., & Núñez López, P. J. (2018). Application of the wavelet packet transform to vibration signals for surface roughness monitoring in CNC turning operations. *Mechanical Systems and Signal Processing, 98,* 902–919.

Ghosh, N., Ravi, Y. B., Patra, A., Mukhopadhyay, S., Paul, S., Mohanty, A. R., et al. (2007). Estimation of tool wear during CNC milling using neural network-based sensor fusion. *Mechanical Systems and Signal Processing, 21*(1), 466–479.

Gierlak, P., Burghardt, A., Szybicki, D., Szuster, M., & Muszyńska, M. (2016). On-line manipulator tool condition monitoring based on vibration analysis. *Mechanical Systems and Signal Processing, 89,* 14–26.

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science, 313,* 504–507.

Huang, S. N., Tan, K. K., Wong, Y. S., De Silva, C. W., Goh, H. L., & Tan, W. W. (2007). Tool wear detection and fault diagnosis based on cutting force monitoring. *International Journal of Machine Tools and Manufacture, 47*(3), 444–451.

Javed, K., Gouriveau, R., Li, X., & Zerhouni, N. (2016). Tool wear monitoring and prognostics challenges: A comparison of connectionist methods toward an adaptive ensemble model. *Journal of Intelligent Manufacturing, 29,* 1873–1890.

Karandikar, J., McLeay, T., Turner S., Schmitz, T. (2015). Tool wear monitoring using naive Bayes classifiers. *The International Journal of Advanced Manufacturing Technology*, *77*(9), 1613–1626.

Keskar, N. S., Mudigere, D., Nocedal, J., Smelyanskiy, M., & Tang, P. T. P. (2016). *On large-batch training for deep learning: Generalization gap and sharp minima*. arXiv preprint arXiv:1609.04836.

Kong, D., Chen, Y., & Li, N. (2018). Gaussian process regression for tool wear prediction. *Mechanical Systems and Signal Processing, 104,* 556–574.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In *21th annual conference on neural information processing systems (NIPS)*. Lake Tahoe, USA, December 3–8.

Kuljanic, E., & Sortino, M. (2005). TWEM, a method based on cutting forces-monitoring tool wear in face milling. *International Journal of Machine Tools and Manufacture, 45*(1), 29–34.

Lecun, Y. L., Bottou, L., Bengio, Y., & Haffner, P. (1998a). Gradient-based learning applied to document recognition. *Proceedings of the IEEE, 86*(11), 2278–2324.

Lecun, Y., Bottou, L., Orr, G. B., & Müller, K. R. (1998b). Efficient backprop. *Lecture Notes in Computer Science, 1524*(1), 9–50.

Li, X., Lim, B. S., Zhou, J. H., & Huang, S. (2009). Fuzzy neural network modelling for tool wear estimation in dry milling operation. In *Annual conference of the prognostics and health management society* (pp. 1–11). PHM Society.

Mali, R., Telsang, M. T., & Gupta, T. V. K. (2017). Real time tool wear condition monitoring in hard turning of Inconel 718 using sensor fusion system. *Materials Today: Proceedings, 4*(8), 8605–8612.

Morgan, J., & O'Donnell, G. E. (2018). Cyber physical process monitoring systems. *Journal of Intelligent Manufacturing, 29,* 1317–1328.

Pandiyan, V., Caesarendra, W., Tjahjowidodo, T., & Tan, H. H. (2018). In-process tool condition monitoring in compliant abrasive belt grinding process using support vector machine and genetic algorithm. *Journal of Manufacturing Processes, 31,* 199–213.

Purushothaman, S. (2010). Tool wear monitoring using artificial neural network based on extended Kalman filter weight updation with transformed input patterns. *Journal of Intelligent Manufacturing, 21*(6), 717–730.

Rehorn, A. G., Jiang, J., & Orban, P. E. (2005). State-of-the-art methods and results in tool condition monitoring: A review. *The International Journal of Advanced Manufacturing Technology, 26*(7–8), 693–710.

Ruder, S. (2016). *An overview of gradient descent optimization algorithms*. arXiv preprint arXiv:1609.04747.

Salehi, M., Albertelli, P., Goletti, M., Ripamonti, F., Tomasini, G., & Monno, M. (2015). Indirect model based estimation of cutting force and tool tip vibrational behavior in milling machines by sensor fusion. *Procedia CIRP, 33,* 239–244.

Tobon-Mejia, D. A., Medjaher, K., & Zerhouni, N. (2012). CNC machine tool's wear diagnostic and prognostic by using dynamic Bayesian networks. *Mechanical Systems and Signal Processing, 28,* 167–182.

Wang, G., Guo, Z., & Qian, L. (2014). Online incremental learning for tool condition classification using modified fuzzy ARTMAP network. *Journal of Intelligent Manufacturing, 25*(6), 1403–1411.

Wang, J., Xie, J., Zhao, R., Zhang, L., & Duan, L. (2017). Multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing. *Robotics and Computer Integrated Manufacturing, 45*(C), 47–58.

Wu, J., Su, Y., Cheng, Y., Shao, X., Deng, C., & Liu, C. (2018). Multi-sensor information fusion for remaining useful life prediction of machining tools by adaptive network based fuzzy inference system. *Applied Soft Computing, 68,* 13–23.

Yu, J., Liang, S., Tang, D., & Liu, H. (2017). A weighted hidden Markov model approach for continuous-state tool wear monitoring and tool life prediction. *International Journal of Advanced Manufacturing Technology, 91*(1–4), 1–11.

Zhang, W., Li, C. H., Peng, G. L., Chen, Y. H., & Zhang, Z. J. (2018). A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load. *Mechanical Systems and Signal Processing, 100,* 439–453.

Zhang, K. F., Yuan, H. Q., & Nie, P. (2015). A method for tool condition monitoring based on sensor fusion. *Journal of Intelligent Manufacturing, 26*(5), 1011–1026.

Zhao, R., Yan, R., Wang, J., & Mao, K. (2017). Learning to monitor machine health with convolutional bi-directional LSTM networks. *Sensors, 17*(2), 273.

Zhou, Y., & Xue, W. (2018). Review of tool condition monitoring methods in milling processes. *International Journal of Advanced Manufacturing Technology, 96*(5–8), 2509–2523.

Zhu, K. P., Wong, Y. S., & Hong, G. S. (2009). Wavelet analysis of sensor signals for tool condition monitoring: A review and some new results. *International Journal of Machine Tools and Manufacture, 49*(7), 537–553.