IEEE Access
Multidisciplinary Rapid Review Open Access Journal

Received November 19, 2019, accepted December 3, 2019, date of publication December 9, 2019, date of current version December 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2958330

Tool Wear Predicting Based on Multisensory Raw Signals Fusion by Reshaped Time Series Convolutional Neural Network in Manufacturing

ZHIWEN HUANG[1], JIANMIN ZHU[1], JINGTAO LEI[2], (Member, IEEE), XIAORU LI[1], AND FENGQING TIAN[1]

[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

Corresponding author: Jianmin Zhu (jmzhu_usst@163.com)

This work was supported by the National Natural Science Foundation of China under Grant 50975179, Grant 51375289, and Grant 51775323.

ABSTRACT Tool wear monitoring is a typical multi-sensor information fusion task. The handcrafted features may be a suboptimal choice that will lower the monitoring accuracy and require significant computational costs that hinder the real-time applications. In order to solve these problems, this paper proposed a new multisensory data-driven tool wear predicting method based on reshaped time series convolutional neural network (RTSCNN). In this method, the reshaped time series layer is introduced to represent the multisensory raw signals, the alternately convolutional and pooling layers is employed to adaptively learn distinctive characteristics of tool wear directly from multisensory raw signals while the multi-layer perceptron with regression layer performs automatic tool wear prediction. In addition, three tool run-to-failure datasets measured from three-flute ball nose tungsten carbide cutter of high-speed CNC machine under milling operations are used to experimentally demonstrate the performance of the proposed RTSCNN-based multisensory data-driven tool wear predicting method. The experimental results show that the prediction error of the RTSCNN-based data-driven method is observably lower than other state-of-art methods.

INDEX TERMS Tool wear predicting, multi-sensor, raw signals, convolutional neural network, reshaped time series, manufacturing.

I. INTRODUCTION

As an important part of smart manufacturing, the tool condition monitoring and prognostic techniques have been increasingly investigated to ensure high surface quality of the workpiece and increase machining efficiency [1], [2]. Generally, these techniques can be divided into model-driven and data-driven methods. Model-driven methods use mathematical models to describe the wear conditions of the cutting tool [3]. However, owing to the uncertainty of machining process and the complexity of wear mechanisms, it is difficult to construct accurate models and ensure high precision that meet the practical needs [4]. Therefore, model-driven methods have significant limitations in practical applications.

Alternatively, data-driven methods establish the model about the tool wear based on a large volume of historical measured data, and make decisions upon the online data collected from various monitored sensors [5], which are very suitable for industrial field application of tool wear monitoring. In the data-driven methods, researchers have used measurements of forces [6], vibrations [7], acoustic emission (AE) [8], and spindle powers [9] to estimate tool wear state. Compared to a single signal, multisensory signals make the result of tool wear monitoring more reliable [10]. Consequently, in order to get a more accurate, multi-sensor information fusion techniques have been widely used in tool wear monitoring. Moreover, thanks to the development of Industry 4.0, the massive

The associate editor coordinating the review of this manuscript and approving it for publication was Varuna De Silva.

data in manufacturing systems can be acquired much larger than ever before [11], [12], which provides new opportunities for tool wear monitoring and prognostic field.

Machine learning is one of the main methods to efficiently deal with massive data [13] and automatically provide accurate results in the data-driven tool wear monitoring fields. Wang *et al.* [14] utilized different dimension reduction techniques to fuse statistical, frequency and time-frequency domain features and employed support vector regression (SVR) to predict tool wear. Pandiyan *et al.* [15] used k-nearest neighbor (KNN) classifier-based genetic algorithm (GA) to optimize time and frequency domain features and employed support vector machine (SVM) to monitor tool wear condition. Kong *et al.* [16] adopted radial basis function based kernel principal component analysis (KPCA) to fuse time, frequency and wavelet domain features and carried out Gaussian process regression (GPR) to calculate turning tool wear. Yu *et al.* [17] chose root mean square (RMS) of the vibration signals as the health indicator and utilized a weighted hidden Markov model (HMM) to monitor tool wear. Wu *et al.* [18] selected features from the time and frequency domain based on Pearson correlation coefficient (PCC), monotonicity and auto-correlation, and then input them into an adaptive network fuzzy inference system to predict remaining useful life of machining tools.

Through literature review, it can be found that although a certain accuracy have achieved in tool condition monitoring, these shallow methods of machine learning had to select different handcrafted features for various raw signals, largely depending on the researchers' prior knowledge [19]. Such handcrafted features not only cannot implement a generic solution for any acquisition signals, but also often makes raw signals lose a certain part, which are limited to improve the monitoring accuracy. Furthermore, these manual feature extractions and dimensionality reductions usually turns out to be computationally costs [20], which eventually hinder the usage of such methods in real-time monitoring applications.

Recently, with the strong data mining ability, deep learning [21] provides advanced analytics and offers great potentials to smart manufacturing in the age of big data [22] and has gradually become one of the most effective way to address the above drawbacks [23], [24]. Chen *et al.* [25] utilized deep belief network (DBN) to predict tool wear using multi-sensor data and achieved lower predicting error than artificial neural network and SVR method. Zhao *et al.* [26] proposed a convolutional bi-directional long-term memory network (CBLSTM) to predict tool wear in milling operation and obtained lower errors than other regression models, such as linear regression (LR), SVR and multi-layer perceptron (MLP), and recurrent models, such as recurrent neural network (RNN), long-term memory network (LSTM) and bi-directional long-term memory network (BLSTM). Although these deep learning methods work effectively than traditional methods of machine learning without manual feature optimizations, raw features fed into these deep networks are still extracted manually based on specific problems.

As one of the most effective deep learning methods [27], deep convolutional neural network (DCNN) has also been widely applied on the data-driven tool wear monitoring models. Cheng *et al.* [28] developed the DCNN model to identify the wear state of an abrasive belt based on sound signals and achieve higher accuracy than traditional methods, such as Bayes, SVM and back-propagation neural network. Aghazadeh *et al.* [29] used spectral subtraction and DCNN to monitor tool wear condition in milling operation, which got higher prediction accuracy than Bayesian rigid network, SVM and KNN method. Fu *et al.* [30] created image representation of vibration signals and fed them into the proposed DCNN model for machine states monitoring during drilling process, which obtained better overall performance than ANN and SVM model. Obviously, they merely applied DCNNs as classifiers for wear state estimation, which are difficult to realize the real-time prediction of continuous tool wear. Huang *et al.* [31] employed DCNNs to automatic fuse raw multi-domain feature from force and vibration signals, which achieved real-time tool wear prediction in machine process. However, the features are still extracted manually from acquisition signals base on prior knowledge about signal processing, which could not implement end-to-end prediction while limited to effectiveness of these proposed method. Therefore, although these deep learning methods has achieved better results than the traditional shadow machine learning methods, the application of deep learning directly based on multisensory raw signals fusion for tool wear predicting is still developing.

In order to reduce the prediction error and improve the prediction efficiency, this paper proposes a new multisensory data-driven tool wear predicting method based on RTSCNN. Concretely, the main contributions of this paper are summarized as follows.

(1) An end-to-end data-driven tool wear predicting method is proposed based on multisensory raw signals fusion by using the designed RTSCNN. Significantly, unlike traditional methods relying on handcrafted features, the proposed method combines adaptive multisensory feature extraction with automatic continuous tool wear prediction, which skips the steps of manual feature extraction and dimensionality reduction.

(2) A new RTSCNN architecture is designed by introducing a reshaped time series layer into the 2-D CNN combined with the MLP. Moreover, this RTSCNN architecture can effectively mine highly distinctive features from massive multisensory raw signals for tool wear prediction.

(3) The proposed method is validated by using multisensory raw signals (i.e. 3-D forces, 3-D vibrations and AE) collected form three milling cutters during complete wear process with a comprehensive performance evaluation. Compared with other traditional intelligent methods, the proposed RTSCNN-based method achieves better predicting performance.

The rest of this paper is organized as follows: Section II briefly introduces the theoretical background. The framework of proposed method and the architecture of proposed RTSCNN are explained in Section III. Section IV describes the experimental datasets. Results are discussed and comparisons are presented in Section V. Finally, Section VI draws the conclusions of this paper.

II. THEORETICAL BACKGROUND

A. CONVOLUTIONAL NEURAL NETWORK

CNNs have been successfully applied in intelligent diagnostic and prognostic fields due to its powerful feature learning capabilities [32]. The basic architecture of CNN consists of a series of alternated convolutional layer and pooling layer, which is used to adaptively extract highly distinguishing features from input data by adopting a series of kernel filters. Besides, CNN contains a traditional multi-layer perceptron (MLP). For continuous predicting tasks, the MLP is composed of several fully connected layers including a regression layer used as an output layer.

In the convolutional layer, the convolution operation between the different feature filters and the local regions of the input matrix is followed by activation function, and the output can be regarded as a series of feature maps obtained by feature extraction of input data [31]. Concretely, Let $\boldsymbol{X}_i^{l-1}$ represent the ith input of the lth layer, $\boldsymbol{X}_i^l$ represent the jth feature map of the lth layer, $\boldsymbol{K}_{ij}^l$ denote the kernel used in the lth layer, $\boldsymbol{b}_i^l$ denote the jth bias of the lth layer. Thus, the general output of convolutional layer is as follows:

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

where $f(\cdot)$ denote the activation function, M_j is the group of the input to calculate the jth output.

The pooling layer is usually followed after the convolutional layer to significantly reduce the spatial size of the feature map and the parameters of the network. The max-pooling operation is commonly used in pooling layer, which performs dimensionality reduction to avoid overfitting and ensure the scale-invariant characteristics of the feature map. The calculated formulation of the pooling layer is as follows:

$$\boldsymbol{X}^l = f\left(\beta_j^l \cdot down(\boldsymbol{X}_j^{l-1}) + \boldsymbol{b}_j^l\right) \tag{2}$$

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

MLP is fully connected layer, which the neurons are all connected between the adjacent layers. The calculated formulation is as follows:

$$\boldsymbol{X}^l = \sigma\left(\boldsymbol{W}^l\boldsymbol{X}^{l-1} + \boldsymbol{b}^l\right) \tag{3}$$

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

B. DROPOUT

Although CNNs have achieved many successes in machine learning, deep networks with a huge number of parameters still suffer from a serious overfitting problem, which results in excellent network performance on the training dataset and poor performance on the testing dataset [33]. However, dropout is an effective technique that can help to reduce overfitting, which was originally proposed by Hinton *et al.* [34]. What's more, this technique was found to improve the performance of neural networks in various application fields, such as object classification, digit recognition, speech recognition, etc. [35].

In practice, dropout means randomly and temporarily removing neurons from the network with a fixed probability, along with all its incoming and outgoing connections in the training process. This prevents complex co-adaptations on the training data. Nevertheless, the neurons from the network are used without dropout in the testing process, while the outgoing weights of those neurons are multiplied by the same probability. In this way, dropout can not only help to improve the feature extraction capability of the network, but also help to overcome overfitting in the training of the network. In addition, dropout can be interpreted as a way of regularizing a neural network by adding noise to its hidden units [35]. Thus, the robustness of deep networks are enhanced. To sum up, the dropout technique is employed on the proposed network.

III. PROPOSED METHODOLOGY

A. OVERALL FRAMEWORK

To directly take advantage of multisensory raw signals for predicting tool wear, the new multisensory data-driven tool wear predicting method based on reshaped time series convolutional neural network (RTSCNN) is proposed in this paper. The overall framework of the proposed method is shown in Fig.1, which involves two processes including offline modeling and online predicting, and the general procedures are summarized respectively as follows.

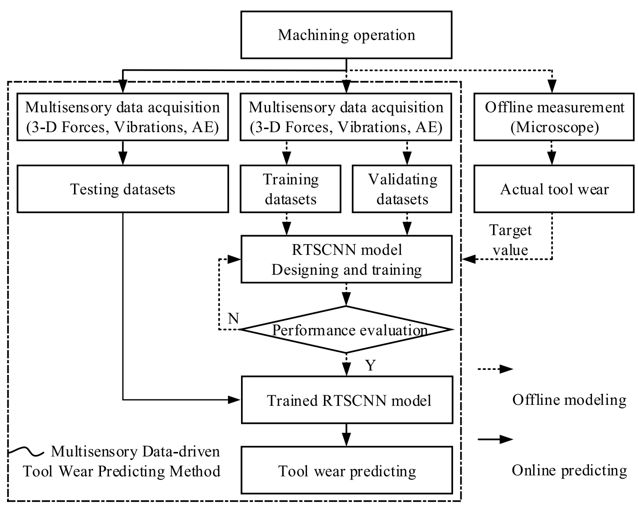


FIGURE 1. The overall framework of the proposed new multisensory data-driven tool wear predicting method based on RTSCNN.

For offline modeling, the multisensory raw signals (3-D forces, 3-D vibrations, and AE) are firstly acquired in the

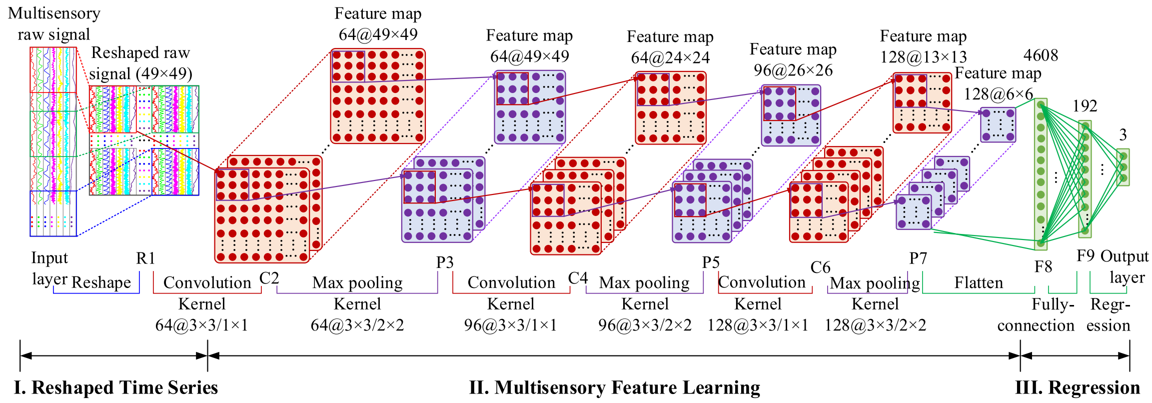


FIGURE 2. The illustration of proposed RTSCNN architecture for tool wear predicting using multisensory raw signals.

cutting process by different kinds of sensor and are divided into training and validating datasets respectively, meanwhile microscope is adopted to measure the actual flank wear width of cutter after each cutting, which is used as the target value for training and validating the subsequent model. Then, the RTSCNN model is designed to directly learn the relationship between multisensory raw signals and real-time tool wear, using the obtained datasets and the target value as input. Finally, after continuous iterative training, an end-to-end tool wear prediction method based on the designed RTSCNN is obtained through performance evaluation.

For online predicting, the raw signals are collected through multisensory data acquisition module as the testing datasets, which are directly fed into the well-trained RTSCNN model for real-time flank wear width prediction of cutter. As a result, through offline modeling and online predicting, the proposed method in this paper simultaneously achieves adaptive feature learning of multisensory raw signals and automatic tool wear predicting in the machining process.

B. ARCHITECTURE OF RTSCNN

To adaptively capture characteristics of tool wear from multisensory raw signals and automatically predict tool wear in real-time, this paper develops a new RTSCNN architecture combined multisensory feature learning based on the typical 2-D CNN with regression prediction based on the traditional MLP. The key innovation of the proposed architecture is to directly establish the relationship between complicated multisensory raw signals and real-time tool wear without any signal process technique and feature selection method. Concretely, as is illustrated in Fig. 2, the proposed RTSCNN consists of three sequential stages: a reshaped time series stage, a multisensory feature learning stage, and a regression stage.

1) RESHAPED TIME SERIES

With the multisensory raw signals massively collected, there is a necessary method for processing them. Although various signal representation techniques are investigated for better feature learning of the DCNN model by transforming the time series signal into different domains, its performance is highly affected by the learning of input signal representation [36]. Therefore, it is necessary to investigate feature learning automatically from multisensory raw signals in machining processes.

As is well-known, DCNNs have been applied successfully in image recognition and classification, which can automatically extract highly distinguishing features from the pixel matrixes of input images. Motivated by the similarity between the pixel matrix of high-dimensional image and the raw data matrix of multisensory time series signal, in this paper, we introduce a reshaped time series stage to represent the multisensory raw signals. The procedure of the reshaped time series are summarized in Table 1. Different from traditional data-driven methods extracted the features from the raw signals; this stage is an effective method to handle directly the multisensory raw signals in time series, which avoids time-consuming and exhausted work of various data processing methods.

TABLE 1. The algorithm of the reshaped time series.

Algorithm: Reshaped time series

Input: The raw signals $\boldsymbol{X}^0$ with size LxN collected form the multisensory sensors, which L is length of the sampling signals to reshape every time, N is the number of signal channels.
Output: The reshaped multisensory matrix $\boldsymbol{X}^1$ with size KxK.

For $j = 0$ to K-1 **do**
  For $i = 0$ to K-1 **do**
    $\boldsymbol{X}^1(i, j) \leftarrow \boldsymbol{X}^0((j*N+i)\%L, (j*N+i)/L)$
Return $\boldsymbol{X}^1$

2) MULTISENSORY FEATURE LEARNING

The main function of this stage is to mine sensitive features of tool wear from multisensory raw signals through multiple pairs of convolutional layers and pooling layers. In details, as shown in Fig.2, after the reshaped time series layer, three convolutional layers and three pooling layers are alternated with each other to learn highly distinguishing and robust features. The kernel size of both convolutional layer and pooling layer is 3×3, and their kernel stride respectively are 1×1 and 2×2.

In mechanical intelligence diagnostic and prognostic fields, lots of methods still adopt the sigmoid function as the activation function of convolutional and pooling layer, but this function has the deficiencies of slow convergence and gradient disappearance in the process of DCNN training [31]. However, the rectified linear units (ReLU) function can effectively overcome these deficiencies, because it makes the weights in the shallow layer more trainable when using back-propagation learning method to adjust the parameters [37]. Hence, the ReLU function is applied as the activation function of the convolutional and pooling layer.

3) REGRESSION

The tool wear predicting studied in this paper is a multi-objective regressive task. The high-level characteristic representation for raw signals obtained in the multisensory feature learning stage is directly fed into a traditional MLP with two layers to perform automatic predicting of three flutes wear of milling tool. The first layer is a fully connected layer with ReLU activation function. In the output layer, we use a sigmoid function that outputs a continuous prediction for each flute wear of milling tool.

C. TRAINING OF RTSCNN

In the training process of RTSCN model, the Euclidean distance between the actual tool wear value $\mathbf{y}$ and the predicted tool wear value $\hat{\mathbf{y}}$ is taken as the loss function, and the calculated formulation is as follows:

$$L = \frac{1}{2N}\sum_{i=1}^{N}\left\|\hat{\mathbf{y}}_i - \mathbf{y}_i\right\|_2^2 \tag{4}$$

where N is the number of samples.

As a widely used strategy, cross-validation is employed to effectively improve generalization ability of the model [38]. Therefore, cross-validation strategy is employed to avoid over-fitting in the training process of the proposed RTSCNN model. In addition, as an effective strategy to reduce the shift of internal covariance, normalization is widely applied to accelerate the training process of deep neural network. Thus, we utilized normalization in the RTSCNN network. Furthermore, the back-propagation algorithm is utilized to reduce the Euclidean distance in the training process of RTSCNN model.

IV. EXPERIMENTAL DATASETS

A. EXPERIMENTAL SETUP

In this paper, a set of experimental data measured from three-flute ball nose tungsten carbide cutter of high-speed computerized numerical control (CNC) machine under dry milling operations [39] was utilized to experimentally demonstrate the performance of the RTSCNN-based data driven tool wear prediction method. During the wear process of cutter in milling operations, a Kistler quartz 3-component dynamometer was mounted between the machining table and workpiece to measure cutting forces in the form of charges, and three Kistler piezo accelerometers were mounted on the workpiece to measure the vibration signals in x, y, z directions respectively. Besides, a Kistler acoustic emission (AE) sensor was mounted on the workpiece to monitor the high frequency stress wave generated by the cutting process. Meanwhile, NI DAQ PCI1200 was applied to collect multisensory signals with a continuous sampling frequency of 50 kHz.

These acquired multisensory raw signals were stored and used for training, validating and testing the RTSCNN-based data driven method of tool wear predicting. Besides, after each cutting process is completed, the actual flank wear width of each flute edge is measured offline by using a LEICA MZ12 microscope, which is regarded as the target value for training and validating the RTSCNN model. The experimental platform is shown in Fig.3; correspondingly, the operation parameters are shown in Table 2 [39].

TABLE 2. The operation parameters of the experimental platform.

Parameters	Value
The running speed of the spindle	10400 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

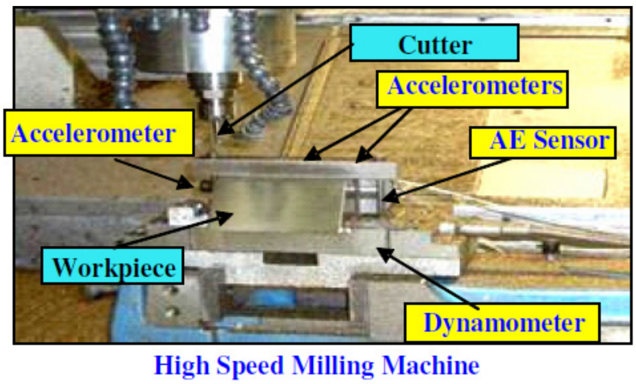


FIGURE 3. Illustrations for the experimental platform.

B. DATA PREPARATION

In the experiment, three tools named C1, C4 and C6 are used as the datasets for verify the proposed method. Each tool contains 300 data files of collected multisensory signals and flank wear, corresponding to 300 milling operations. And each data file contains 7 channels of multisensory signals, that is, force signal in the x direction (Fx), force signal in the y direction (Fy), force signal in the z direction (Fz), acceleration signal in the x direction (Ax), acceleration signal in the y direction (Ay), acceleration signal in the z direction (Az) and AE signal. Then, for each sample batch, the 7 multi-signals are obtained to form a column of multisensory raw signal matrix, which is regarded as multi-signals axis. Through a series of attempts to determine the 343 continuous sampling data regarded as time-series axis. Thus, a whole multisensory raw signal matrix with a size of 343 × 7 is obtained, which combines the multi-signals axis and time-series axis.

TABLE 3. Description of experimental datasets.

Tool Symbol	Training Datasets	Validating Datasets	Testing Datasets	Total Datasets
C1	108,000	36,000	36,000	180,000
C4	108,000	36,000	36,000	180,000
C6	108,000	36,000	36,000	180,000

TABLE 4. List of different criteria.

Criterion	Expression
Mean absolute error	$f_{\mathrm{MAE}} = \sum_{i=1}^{N} \lvert \hat{\boldsymbol{y}}_i - \boldsymbol{y}_i \rvert / N$
Root mean squared error	$f_{\mathrm{RMSE}} = \sqrt{\sum_{i=1}^{N} (\hat{\boldsymbol{y}}_i - \boldsymbol{y}_i)^2 / N}$

Finally, with the continuous signal acquisition, a series of raw multisensory signals matrixes are obtained.

In order to verify the RTSCNN model, the raw multisensory signals matrixes are divided into training datasets, validating datasets and testing datasets respectively, and the description of these datasets is shown in Table 3. The performance of the proposed RTSCNN-based multisensory data-driven tool wear predicting method is evaluated based on leave-one-out cross-validation, in which one dataset is chosen for testing while the rest datasets are used for model training and validating.

The proposed RTSCNN model was calculated on Alibaba elastic compute service (ECS) platform with 8-core CPU and NVIDIA Tesla P100 GPU in Ubuntu16.04 operating system. Besides, CUDA8 is used to accelerate the calculation in the platform. The training time for one epoch is about 32.8s, and the testing time for every sample is only 0.01s. Therefore, the RTSCNN model can be an effective solution for tool wear predicting.

V. RESULTS AND DISCUSSIONS

A. PERFORMANCE METRICS

In order to quantitatively evaluate the overall performance of the RTSCNN-based data driven multisensory tool wear prediction method, this paper makes use of the following two criteria, namely mean absolute error (MAE) and root mean squared error (RMSE). The computation formula of two criteria is as shown in Table 4.

Among the above dimensional metrics for performance evaluation, the smaller the value of MAE and RMSE, the predicted value of flank wear width gets closer to the actual value of flank wear width, then the prediction error is lower and the RTSCNN model performance is superior.

B. MULTI-SENSOR FUSION

To verify its effectiveness on the multi-sensor information fusion, the proposed RTSCNN model is firstly utilized for the tool wear prediction based on different single signals in this paper. The corresponding results are shown in Fig. 4.

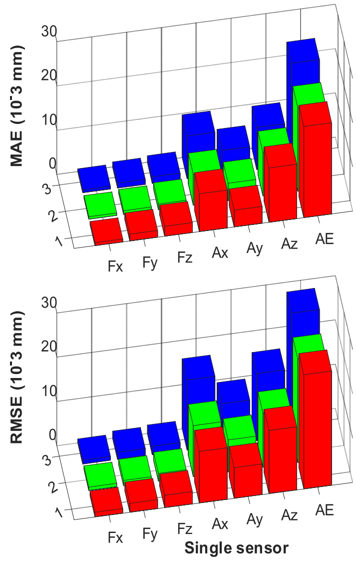


FIGURE 4. The performance comparison under single sensor.

It is clearly that the force signal is more sensitive with the tool wear and the vibration signal is ordinary sensitive during the cutting process. On the contrary, the AE signal is insensitive with tool flank wear during the cutting process. In addition, the vibration signals and cutting force signals in the x, y, and z directions are different in sensitivity to tool flank wear. Therefore, it is necessary to utilize the proposed RTSCNN model to adaptively fuse the multisensory raw signals to predict the tool wear during cutting process.

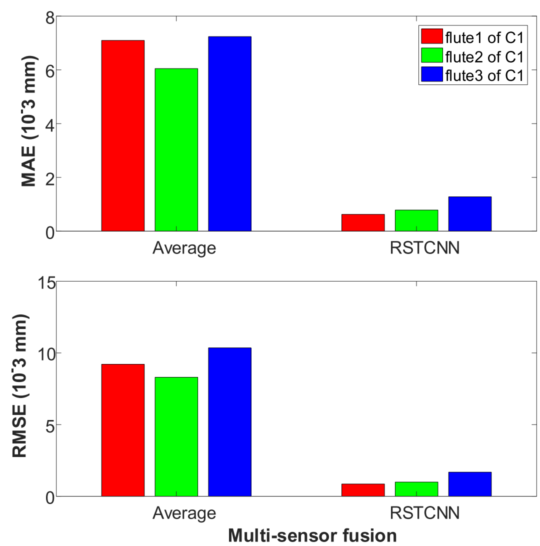


FIGURE 5. The performance comparison under multi-sensor fusion.

Furthermore, the average fusion strategy is adopted to fuse the above seven single signals for tool wear prediction, which is compared with the proposed multisensory data-driven tool wear predicting method based on RSTCNN. Correspondingly, the predicted tool wear under different multi-sensor fusion strategy is illustrated in Fig. 5. It is obviously found that both MAE and RMSE of the multisensory fusion method based directly on the RSTCNN model are significantly lower than average fusion strategy of single signals. Moreover, the RSTCNN model can not only automatically fuse the multi-information from the raw multisensory signals, but also adaptively learn the deep intrinsic features hidden in the raw multisensory signals during cutting process. Consequently, for better feature learning and information fusion, applying the RSTCNN model to fuse multisensory signals is an effective strategy for monitoring tool wear.

C. PARAMETRIC OPTIMIZATION

When applying the multisensory raw signals for tool wear prediction by using the RTSCNN-based model, parameter optimization can enhance the model performance. However, due to the fact that the parameters vary with the experimental datasets, therefore optimizing the parameters to select the appropriate ones of the corresponding experimental datasets is a necessary part in the usage of the RTSCNN model. Concretely, the parameters to be optimized include gradient descent algorithm, batch size, learning rate, number of epochs and dropout.

1) GRADIENT DESCENT ALGORITHM

The gradient descent algorithms are most popular way to optimize neural networks, but due to theoretical explanations of their strengths and weaknesses are difficulty to come by, the algorithms are usually regarded as black box optimizers [40]. Therefore, in order to accelerate training process of RTSCNN model, the Adadelta, Adagrad, adaptive moment estimation (Adam), Nestrov accelerated gradient (NAG) and stochastic gradient descent (SGD) algorithms are compared experimentally to choose the appropriate gradient descent optimization algorithm. Correspondingly, the results of performance evaluation are shown in Fig. 6. It can be seen from Fig. 6 that the Nesterov algorithms not only have a lower MAE, but also have a lower RMSE. Finally, the NAG algorithm is chosen to accelerated train the RTSCNN model.

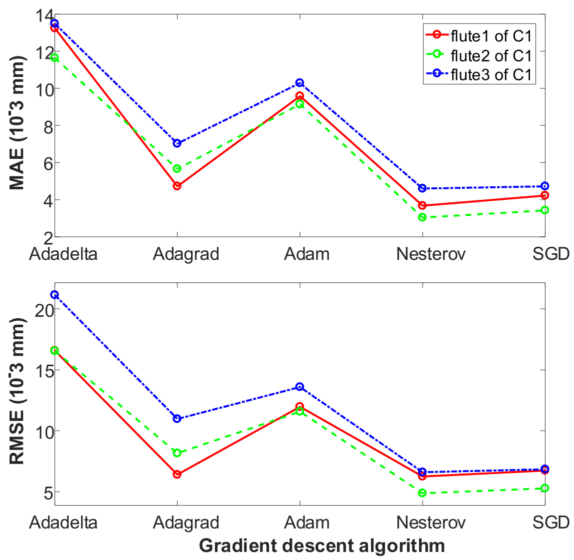


FIGURE 6. The performance comparison under different gradient descent algorithms.

2) BATCH SIZE

In the training process of deep neural network by using gradient descent optimization algorithm, the batch size is a crucial parameter to balance computing resources and model performance. It is often reported that when increasing the batch size for a training task, there exists a threshold after which there is a deterioration in the quality of the model [41]. To this end, to find the appropriate threshold for enhancing model generalization, different batch sizes are compared experimentally in the training process of the RTSCNN model. Correspondingly, both training time and training loss are shown in Fig. 7. Besides, the results of performance evaluation under different batch sizes are shown in Fig. 8.

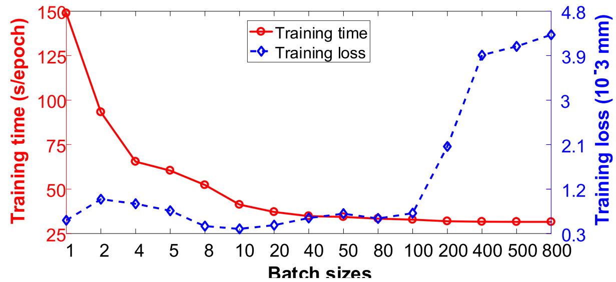


FIGURE 7. The training time and loss under different batch sizes.

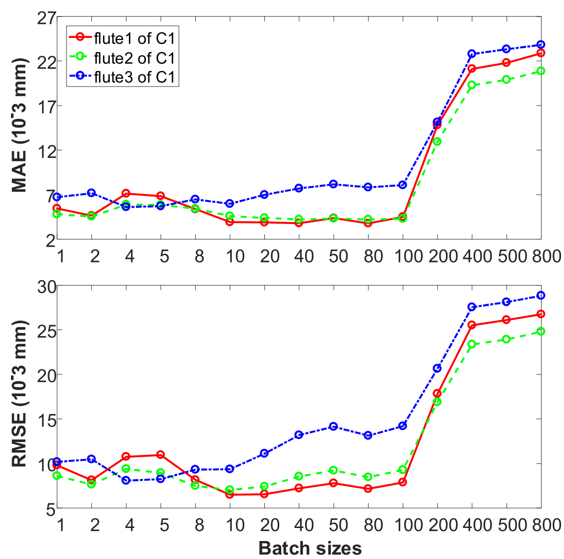


FIGURE 8. The performance comparison under different batch sizes.

As is shown from Fig. 7 and Fig. 8, it is obviously found that when the batch size is very small, both network training loss and performance metrics (including MAE and RMSE) are very small, which means the predicted value of tool flank wear is closer to the actual tool flank wear, but it takes a longer training time to conduct one epoch. Conversely, as the batch size becomes larger (especially when the batch size is larger than 100), the training time required for one epoch is progressively reduced and tended to be stable, but both training error and performance metrics are increased progressively. Therefore, during the process of comprehensive comparison, we found that when the batch size is determined as 100, not only the model prediction accuracy can be improved, but also the network training time is reduced.

3) LEARNING RATE

When using gradient descent algorithm for training the RTSCNN model, learning rate is a vital parameter for adjusting network weights and model convergence. To this end, suitable learning rate is necessarily selected to improve the efficiency of the proposed model. In this experiment, different learning rates were adopted to train RTSCNN network, and the performance metrics including MAE and RMSE under different learning rates are obtained as shown in Fig. 9.

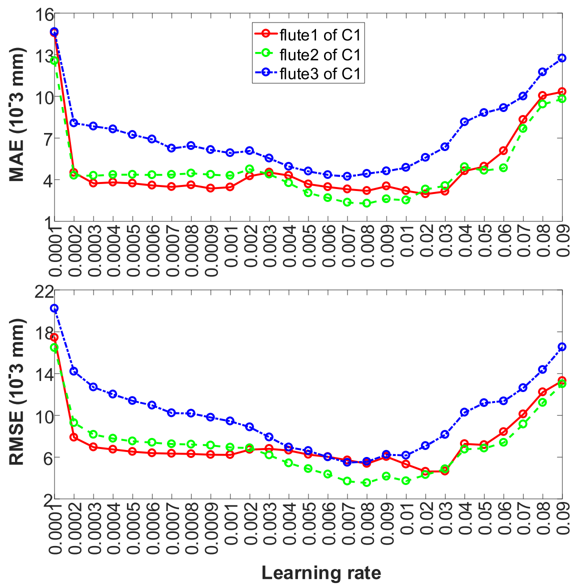


FIGURE 9. The performance comparison under different learning rates.

It can be seen from Fig. 9 that if learning rate is smaller or larger than 0.007, as the learning rate decreases or increases, the MAE and RMSE basically becomes larger and larger, that is, the predicted value of tool flank wear is farther and farther to the actual value of tool flank wear. In other words, too smaller or larger learning rates can increase the prediction error, and moderate ones can speed up error convergence. Therefore, by considering the overall performance between convergence error and prediction accuracy, the appropriate learning rate was selected as 0.007.

4) NUMBER OF EPOCHS

The number of epochs is an important parameter for training the RTSCNN network, and that can affect the prediction error and generalization. To this end, we adopted different number of epochs for training the RTSCNN network in this experiment, and the corresponding results of performance evaluation metrics are obtained as shown in Fig. 10.

As is know from Fig. 10, we can easily find that as the number of epochs is larger and larger, the network training is more and more sufficient and the prediction error are lower and lower to a certain extent. However, when the number of

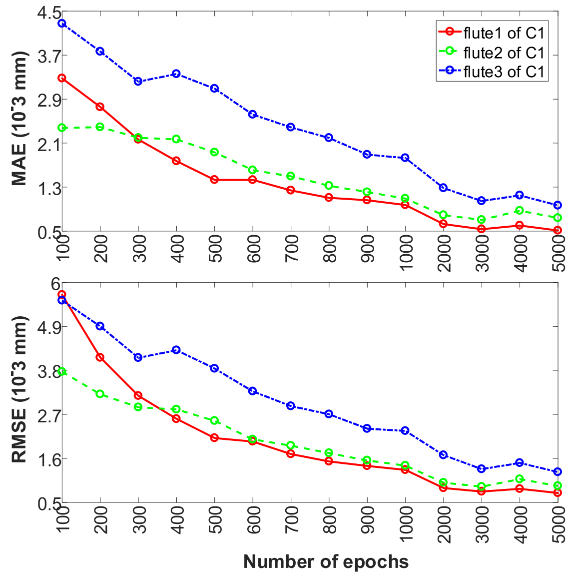


FIGURE 10. The performance comparison under different number of epochs.

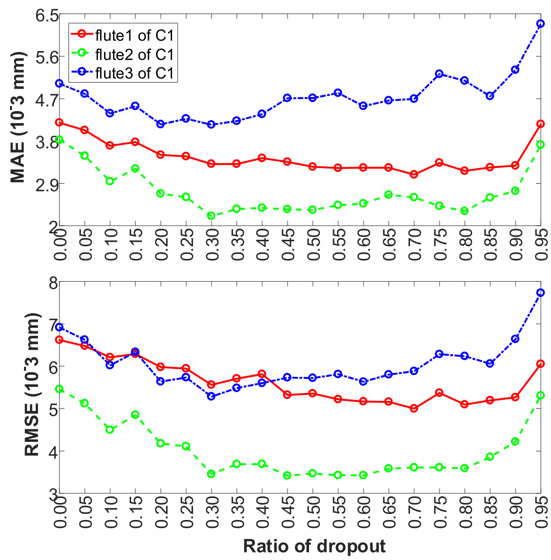


FIGURE 11. The performance comparison under different ratios of dropout.

epochs is greater than 2000, the training of RTSCNN network is over-fitted, and the prediction error tends to be stable. Finally, in this experiment, the optimized number of epochs was determined as 2000.

5) RATIO OF DROPOUT

As a widely used technique for overcoming the model overfitting, dropout is an effective way to improve the overall performance of deep networks. To this end, different ratios of dropout are used for training the RTSCNN network in the experiment. The corresponding results of performance evaluation metrics under different ratios of dropout are obtained as shown in Fig. 11.

It can be obviously seen from Fig. 11 that if ratio of dropout is too large, namely neurons for training wear randomly

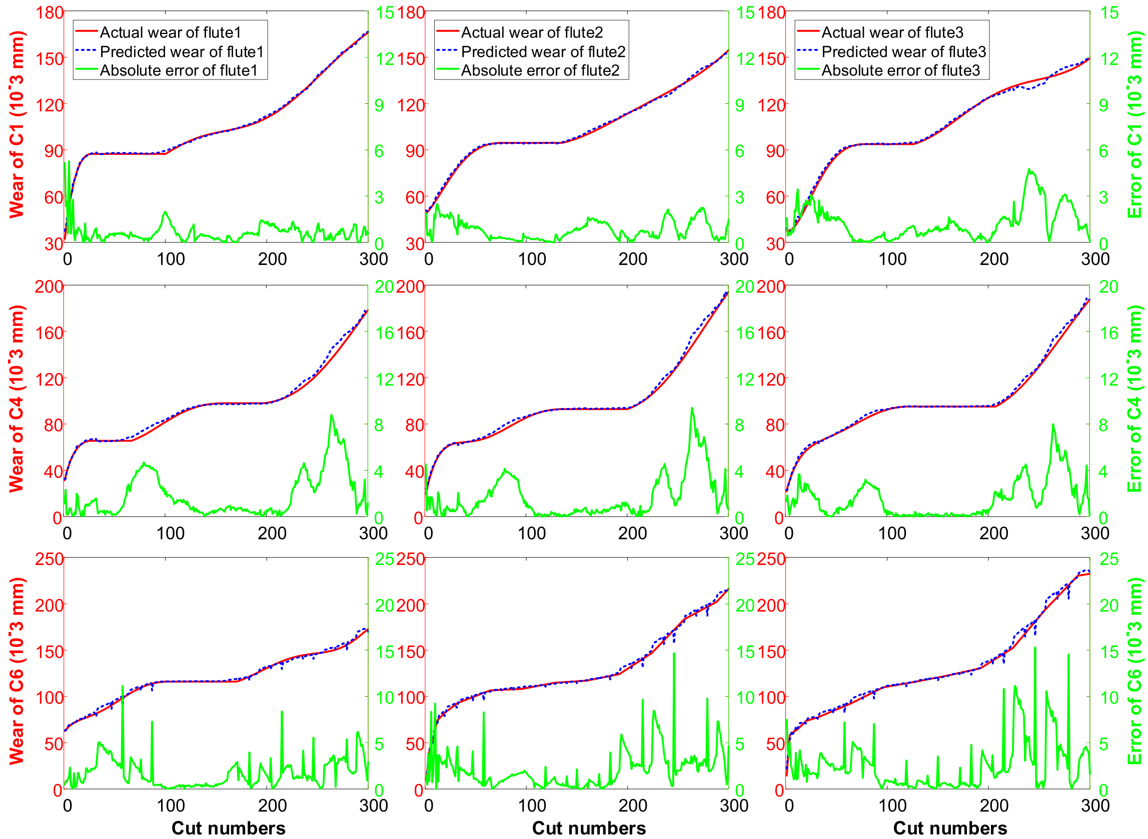


FIGURE 12. The tool wear predicting result of the proposed method.

TABLE 5. The optimized parameters of proposed method.

Gradient descent algorithm	Batch size	Learning rate	Number of epoch	Ratio of dropout
Nesterov	100	0.007	2000	0.030

removed too many, the prediction error is larger. In addition, if ratio of dropout is too small, namely almost no neurons to be randomly removed in the training process. Nevertheless, when ratio of dropout is moderate, the prediction error is smaller, and therefore the predicted value of tool flank wear is closer to the actual value of tool flank wear. As a result, by comprehensive considering the ability to prevent model overfitting and the prediction accuracy, the optimized ratio of dropout was determined as 0.30 in this experiment.

D. METHOD COMPARISON

According to a series of parameter optimization experiments, the parameters of the RTSCNN network applied in the proposed multisensory data-driven tool wear predicting method are optimized as shown in Table 5.

The corresponding predicted values of tool flank wear under different datasets are illustrated in Fig. 12. Obviously, the tool flank wear can be effectively predicted from multisensory raw signals by using the proposed RTSCNN-based data-driven method.

In order to demonstrate the effectiveness and advancement, the RTSCNN-based method is compared with other methods using the original published multisensory data. Concretely, we employed traditional intelligent methods such as SVR and SVR+KPCA [14] for tool wear predicting. Moreover, the deep learning methods such as RNN, LSTM, CBLSTM [26], CNN [32] and DCNN [31] are also adopted for performance comparison. Particularly, for DCNN [31], the AE signal is added for feature extraction and the input of DCNN is 63×63, while other steps are still stayed the same according to the referenced publication. Correspondingly, the compared results under different evaluation criteria are given in Table 6.

TABLE 6. The results of the approaches in tool wear prediction.

Method	MAE (10-3mm)	RMSE (10-3mm)
SVR[14]	9.3770±2.0422	11.9681±3.3337
SVR+KPCA[14]	3.9583±0.9371	5.4428±1.5894
RNN[26]	12.1667±6.2292	15.7333±6.2164
LSTM[26]	10.7333±3.8734	13.7333±4.5742
CBLSTM[26]	7.2333±1.0263	9.2333±1.9140
CNN[32]	11.0000±1.3000	14.0428±5.5588
DCNN+AE[31]	1.5708±0.5485	2.2385±0.7105
RTSCNN	1.5658±0.6037	2.2009±0.9012

As shown in Table 6, as a traditional nonlinear regression method, SVR algorithm is weak in dealing with large-scale samples, which obtained the higher predicted error by using multisensory extracted features for tool wear predicting. Besides, as a feature dimension reduction method, KPCA algorithm can improve the predicted accuracy of SVR, but the predicted error is still higher than the proposed RTSCNN method. However, the deep learning methods such as RNN, LSTM, CBLSTM and CNN can implement nonlinear regression without features dimension reduction, but the predicted accuracy of these deep methods is still lower than the proposed RTSCNN method. In addition, the DCNN based method achieved a close prediction performance with the proposed RTSCNN-based method, but multi-domain features are still extracted manually from collected signals, which could not be implemented in an end-to-end tool wear prediction. Overall, the proposed RTSCNN-based method can ensure not only the low deviation, but also a small variance, which shows the better predicting performance in multisensory raw signals. Therefore, the proposed RTSCNN model can correctly and effectively learn the relationship between the multisensory raw signals and the tool flank wear, and achieve higher prediction accuracy of tool wear.

VI. CONCLUSION

In this paper, we focused on the multisensory raw signals (i.e. 3-D forces, 3-D vibrations and AE) fusion and proposed a data-driven method based on RTSCNN architecture for tool wear prediction under the machine process. The major contribution of the proposed RTSCNN-base data-driven method was the direct utilization of multisensory raw time series signals for tool wear predicting. Concretely, the reshaped time series layer is introduced to represent the multisensory raw signals, and the alternately convolutional and pooling layers of the proposed RTSCNN effectively learn characteristics of tool wear directly from multisensory raw signals while the MLP with regression layer perform automatic tool wear prediction.

Being different from those traditional tool wear predicting methods which greatly relying on the handcrafted wear features from prior knowledge, the developed RTSCNN-based data-driven tool wear predicting method can perform adaptive feature extraction and automatic prediction simultaneously without depending on prior knowledge of complex signal processing techniques and dimension reduction algorithms. In addition, experimental results of three tool run-to-failure datasets under milling operations showed that the RTSCNN-based method can be superior to the traditional intelligent method in terms of overall predicting performance.

The proposed RTSCNN-based method has presented an efficient end-to-end tool wear predicting capacity in milling operations, and the discussion may provide some references for the application in real machining processes of industrial fields. Moreover, the proposed method provides a new perspective for tool wear prediction by directly using multisensory raw signals and can be easily employed to deal with tool wear predicting in turning, drilling, etc. In future work, we will study tool wear prediction under different working conditions to improve the performance of the whole method.

REFERENCES

[1] H.-T. Yau, C.-C. Wang, J.-Y. Chang, and X.-Y. Su, "A study on the application of synchronized chaotic systems of different fractional orders for cutting tool wear diagnosis and identification," *IEEE Access*, vol. 7, pp. 15903–15911, 2019.

[2] K. Javed, R. Gouriveau, X. Li, and N. Zerhouni, "Tool wear monitoring and prognostics challenges: A comparison of connectionist methods toward an adaptive ensemble model," *J. Intell. Manuf.*, vol. 29, no. 8, pp. 1873–1890, 2016.

[3] K. Zhu and Y. Zhang, "A generic tool wear model and its application to force modeling and wear monitoring in high speed milling," *Mech. Syst. Signal Process.*, vol. 115, pp. 147–161, Jan. 2019.

[4] J. Zhu, N. Chen, and W. Peng, "Estimation of bearing remaining useful life based on multiscale convolutional neural network," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3208–3216, Apr. 2019.

[5] R. Zhao, D. Wang, R. Yan, K. Mao, F. Shen, and J. Wang, "Machine health monitoring using local feature-based gated recurrent unit networks," *IEEE Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1539–1548, Feb. 2018.

[6] M. Nouri, B. K. Fussell, B. L. Ziniti, and E. Linder, "Real-time tool wear monitoring in milling using a cutting condition independent method," *Int. J. Mach. Tool. Manuf.*, vol. 89, pp. 1–13, Feb. 2015.

[7] P. Gierlak, A. Burghardt, D. Szybicki, M. Szuster, and M. Muszyńska, "On-line manipulator tool condition monitoring based on vibration analysis," *Mech. Syst. Signal Process.*, vol. 89, pp. 14–26, May 2017.

[8] J. Chen, H. Chen, J. Xu, J. Wang, X. Zhang, and X. Chen, "Acoustic signal-based tool condition monitoring in belt grinding of nickel-based superalloys using RF classifier and MLR algorithm," *Int. J. Adv. Manuf. Technol.*, vol. 98, nos. 1–4, pp. 859–872, 2018.

[9] R. Corne, C. Nath, M. El Mansori, and T. Kurfess, "Study of spindle power data with neural network for predicting real-time tool wear/breakage during inconel drilling," *J. Manuf. Syst.*, vol. 43, pp. 287–295, Apr. 2017.

[10] R. Mali, M. T. Telsang, and T. V. K. Gupta, "Real time tool wear condition monitoring in hard turning of Inconel 718 using sensor fusion system," *Mater. Today*, vol. 4, no. 8, pp. 8605–8612, 2017.

[11] C.-L. Liu, W.-H. Hsaio, and Y.-C. Tu, "Time series classification with multivariate convolutional neural network," *IEEE Trans. Ind. Electron.*, vol. 66, no. 6, pp. 4788–4797, Jun. 2019.

[12] J. Dong, M. Wang, X. Zhang, L. Ma, and K. Peng, "Joint data-driven fault diagnosis integrating causality graph with statistical process monitoring for complex industrial processes," *IEEE Access*, vol. 5, pp. 25217–25225, 2017.

[13] C. Shi, G. Panoutsos, B. Luo, H. Liu, B. Li, and X. Lin, "Using multiple-feature-spaces-based deep learning for tool condition monitoring in ultraprecision manufacturing," *IEEE Trans. Ind. Electron.*, vol. 66, no. 5, pp. 3794–3803, May 2019.

[14] J. Wang, J. Xie, R. Zhao, L. Zhang, and L. Duan, "Multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing," *Robot. Comput.-Integr. Manuf.*, vol. 45, pp. 47–58, Jun. 2017.

[15] V. Pandiyan, W. Caesarendra, T. Tjahjowidodo, and H. H. Tan, "In-process tool condition monitoring in compliant abrasive belt grinding process using support vector machine and genetic algorithm," *J. Manuf. Process.*, vol. 31, pp. 199–213, Jan. 2018.

[16] D. Kong, Y. Chen, and N. Li, "Gaussian process regression for tool wear prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 556–574, May 2018.

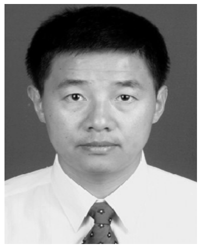
[17] J. S. Yu, S. Liang, D. Y. Tang, and H. Liu, "A weighted hidden Markov model approach for continuous-state tool wear monitoring and tool life prediction," *Int. J. Adv. Manuf. Technol.*, vol. 91, nos. 1–4, pp. 201–211, Jul. 2017.

[18] J. Wu, Y. Su, Y. Cheng, X. Shao, C. Deng, and C. Liu, "Multi-sensor information fusion for remaining useful life prediction of machining tools by adaptive network based fuzzy inference system," *Appl. Soft Comput.*, vol. 68, pp. 13–23, Jul. 2018.

[19] F. Jia, Y. G. Lei, J. Lin, X. Zhou, and N. Lu, "Deep neural networks: A promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data," *Mech. Syst. Signal Process.*, vols. 72–73, pp. 303–315, May 2016.

[20] T. Ince, S. Kiranyaz, L. Eren, M. Askar, and M. Gabbouj, "Real-time motor fault detection by 1-D convolutional neural networks," *IEEE Trans. Ind. Electron.*, vol. 63, no. 11, pp. 7067–7075, Nov. 2016.
[21] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
[22] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, "Deep learning for smart manufacturing: Methods and applications," *J. Manuf. Syst.*, vol. 48, pp. 144–156, Jul. 2018.
[23] R. Huang, Y. Liao, S. Zhang, and W. Li, "Deep decoupling convolutional neural network for intelligent compound fault diagnosis," *IEEE Access*, vol. 7, pp. 1848–1858, 2019.
[24] L. Wen, X. Li, L. Gao, and Y. Zhang, "A new convolutional neural network-based data-driven fault diagnosis method," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5990–5998, Jul. 2018.
[25] Y. Chen, Y. Jin, and G. Jiri, "Predicting tool wear with multi-sensor data using deep belief networks," *Int. J. Adv. Manuf. Technol.*, vol. 99, nos. 5–8, pp. 1917–1926, 2018.
[26] R. Zhao, R. Yan, J. Wang, and K. Mao, "Learning to monitor machine health with convolutional bi-directional LSTM networks," *Sensors*, vol. 17, no. 2, pp. 273–291, 2017.
[27] G. Jiang, H. He, J. Yan, and P. Xie, "Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox," *IEEE Trans. Ind. Electron.*, vol. 66, no. 4, pp. 3196–3207, Apr. 2019.
[28] C. Cheng, J. Li, Y. Liu, M. Nie, and W. Wang, "Deep convolutional neural network-based in-process tool condition monitoring in abrasive belt grinding," *Comput. Ind.*, vol. 106, pp. 1–13, Apr. 2019.
[29] F. Aghazadeh, A. Tahan, and M. Thomas, "Tool condition monitoring using spectral subtraction and convolutional neural networks in milling process," *Int. J. Adv. Manuf. Technol.*, vol. 98, nos. 9–12, pp. 3217–3227, 2018.
[30] Y. Fu *et al.*, "Machining vibration states monitoring based on image representation using convolutional neural networks," *Eng. Appl. Artif. Intell.*, vol. 65, pp. 240–251, Oct. 2017.
[31] Z. W. Huang, J. M. Zhu, and J. T. Lei, "Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations," *J. Intell. Manuf.*, to be published, doi: 10.1007/s10845-019-01488-7.
[32] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring," *Mech. Syst. Signal Process.*, vol. 115, pp. 213–237, Jan. 2019.
[33] X. Li, Q. Ding, and J.-Q. Sun, "Remaining useful life estimation in prognostics using deep convolution neural networks," *Rel. Eng. Syst. Saf.*, vol. 172, pp. 1–11, Apr. 2018.
[34] G. Hinton, N. Srivastava, A. Krizhevsky, R. R. Salakhutdinov, and I. Sutskever, "Improving neural networks by preventing co-adaptation of feature detectors," *Comput. Sci.*, vol. 3, no. 4, pp. 212–223, Jul. 2012.
[35] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[36] J. Wang, Y. Ma, Z. Huang, R. Xue, and R. Zhao, "Performance analysis and enhancement of deep convolutional neural network: Application to gearbox condition monitoring," *Bus. Inf. Syst. Eng.*, vol. 61, no. 3, pp. 311–326, 2019.
[37] W. Zhang, C. Li, G. Peng, Y. Chen, and Z. Zhang, "A deep convolutional neural network with new training methods for bearing fault diagnosis under noisy environment and different working load," *Mech. Syst. Signal Process.*, vol. 100, pp. 439–453, Feb. 2018.
[38] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Statist. Surv.*, vol. 4, pp. 40–79, Mar. 2010.
[39] X. Li, B. S. Lim, J. H. Zhou, S. Huang, S. J. Phua, K. C. Shaw, and M. J. Er, "Fuzzy neural network modelling for tool wear estimation in dry milling operation," in *Proc. Annu. Conf. Prognostics Health Manage. Soc. (PHM)*, San Diego, CA, USA, Sep. 2009, pp. 1–11.
[40] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*. [Online]. Available: https://arxiv.org/abs/1609.04747
[41] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On large-batch training for deep learning: Generalization gap and sharp minima," 2016, *arXiv:1609.04836*. [Online]. Available: https://arxiv.org/abs/1609.04836

ZHIWEN HUANG was born in Lu'an, Anhui, China, in 1986. He received the B.S. degree in mechanical design, manufacture and automation from Anhui Polytechnic University, Wuhu, Anhui, China, in 2010, and the M.S. degree in mechatronic engineering from the University of Shanghai for Science and Technology, Shanghai, China, in 2013. He is currently pursuing the Ph.D. degree in mechanical engineering with the University of Shanghai for Science and Technology, Shanghai, China. His research interests include fault diagnosis and prognosis, process monitoring, and intelligent control.



JIANMIN ZHU was born in Rudong, Jiangsu, China, in 1968. He received the B.S. and M.S. degrees in fluid transmission and control from the Luoyang Institute of Technology, Luoyang, Henan, China, in 1991 and 1994, respectively, and the Ph.D. degree in mechanical manufacture and automation from the Huazhong University of Science and Technology, Wuhan, Hubei, China, in 2000.

From 2000 to 2008, he was a Professor with the Mechanical and Electrical Engineering College, Henan University of Science and Technology, Luoyang, Henan, China. Since 2008, he has been a Professor and Doctoral Supervisor with the School of Mechanical Engineering, University of Shanghai for Science and Technology, Shanghai, China. He has authored 5 books, more than 100 articles, and more than 20 inventions. His research interests include process monitoring and intelligent control of electromechanical systems, precision measurement technique, and complex system modeling and control.

Prof. Zhu's awards and honors include the Second Prize of National Science and Technology Progress Award, the Second Prize of the Ministry of Education Natural Science Award, and the Second Prize of Henan Province Science and Technology Progress Award.

JINGTAO LEI (M'14) was born in Luoyang, Henan, China, in 1970. She received the B.S. degree in fluid transmission and control and the M.S. degree in mechanical design and theory from the Luoyang Institute of Technology, Luoyang, Henan, China, in 1991 and 1996, respectively, and the Ph.D. degree in automotive engineering from the Beijing University of Aeronautics and Astronautics, Beijing, China, in 2007.

From 2007 to 2009, she worked as a Postdoctoral Researcher with the School of Mechanical Engineering, Beijing University of Aeronautics and Astronautics, Beijing, China. Since 2009, she has been with the School of Mechanical Engineering and Automation, Shanghai University, Shanghai, China. She is currently a Professor and Doctoral Supervisor. She has authored more than 50 articles, and more than 10 inventions. Her research interest includes intelligent robot design and artificial intelligence application.

XIAORU LI was born in Zhoukou, Henan, China, in 1980. She received the B.S. degree in mechanical design, manufacture and automation and the M.S. degree in measurement technology and instrumentation from the University of Shanghai for Science and Technology, Shanghai, China, in 2003 and 2005, respectively, where she is currently pursuing the Ph.D. degree in mechanical engineering.

Since 2005, she has been an Engineer with the Professional Laboratory Center, School of Mechanical Engineering, University of Shanghai for Science and Technology. Her research interests include the discrimination and prediction of tool wear, and mechanical dynamics.

FENGQING TIAN was born in Xinxiang, Henan, China, in 1982. He received the B.S. degree in mechanical engineering from the Henan Institute of Science and Technology, China, in 2005, and the M.S. degree in mechanical engineering from the Nanjing University of Science and Technology, China, in 2012. He is currently pursuing the Ph.D. degree in mechanical engineering with the University of Shanghai for Science and Technology, China.

From 2005 to 2015, he was a Lecturer with the Henan Institute of Science and Technology, Xinxiang, China. His research interests include mechanical dynamics, embedded systems, and its application.