

## Research Article

# An Improved Deep Learning Model for Online Tool Condition Monitoring Using Output Power Signals

Lang Dai <sup>1</sup>, Tianyu Liu <sup>1</sup>, Zhongyong Liu <sup>1</sup>, Lisa Jackson <sup>2</sup>, Paul Goodall <sup>3</sup>,  
Changqing Shen <sup>4</sup> and Lei Mao <sup>1,5</sup>

<sup>1</sup>Department of Precision Machinery and Precision Instrumentation, University of Science and Technology of China, Hefei 230031, China

<sup>2</sup>Department of Aeronautical and Automotive Engineering, Loughborough University, Loughborough LE11 3TU, UK

<sup>3</sup>Department of Mechanical, Electrical and Manufacturing Engineering, Loughborough University, Loughborough LE11 3TU, UK

<sup>4</sup>School of Rail Transportation, Soochow University, Suzhou 215131, China

<sup>5</sup>Key Laboratory of Precision Scientific Instrumentation of Anhui Higher Education Institutes, University of Science and Technology of China, Hefei 230031, China

Correspondence should be addressed to Lei Mao; leimao82@ustc.edu.cn

Received 20 July 2020; Revised 3 November 2020; Accepted 10 November 2020; Published 23 November 2020

Academic Editor: Selda Oterkus

Copyright © 2020 Lang Dai et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Something like normal functionality of tools in a manufacturing process is typically designed to ensure reliability, where fast and accurate identification of tool abnormal operation plays a vital role in intelligent manufacturing. In this study, a novel method is proposed to assess the cutting tool condition, which consists of a convolutional neural network with wider first-layer kernels (W-CONV), and long short-term memory (LSTM). The analysis benefits from the use of output power signals from the cutting tool, since they can be obtained easily and efficiently, enabling the proposed method to be applicable in practical operation for online condition monitoring. Moreover, effectiveness of the proposed method is investigated, using test data from cutting tools at various tool wear conditions. Results demonstrate that with the proposed method, tool wear condition can be identified accurately and efficiently. Furthermore, with test data collected at cutting tools with different sizes, the robustness of the proposed method can be further clarified.

## 1. Introduction

Cutting tools are widely used in the manufacturing process and play an important role for accuracy of shape/position and surface quality. During the cutting process to the workpiece, the cutting tool is bound to have light or heavy wear or even damage, and this inevitably leads to the change of surface roughness and dimensional tolerance. With severe tool wear, large amounts of chatter will be produced. However, in current manufacturing processes, identification of cutting tool condition heavily relies on experts' experience [1]. Therefore, an efficient and accurate tool condition monitoring (TCM) technique is urgently required, especially for intelligent manufacturing. Since it can improve manufacturing quality, system reliability and

productivity, while processing costs and downtime, can be reduced [2].

On the above basis, a set of TCM techniques have been proposed, which can be divided into two groups, including direct measurement method and indirect measurement technique [3]. In the direct measurement method, cutting tool condition can be measured directly, such as by measuring surface roughness and side wear. However, the manufacturing process should be stopped in order to perform such measurements, which hinders the use of such techniques for online TCM. On the other side, with indirect measurement techniques, various signals can be collected during cutting tool operation, from which distinguishable features can be extracted. A mapping relationship is then established between features and wear conditions, where

tool condition can be assessed. As online monitoring can be achieved using indirect measurement techniques, they are more applicable for industrial applications.

In indirect measurement methods, various signals can be collected from tools, such as monitoring cutting force, acoustic emission, vibration, and power [4, 5]. Among various signals, cutting force is most commonly used for TCM. Huang et al. [6] investigated the relationship between cutting force and tool wear condition, where the change of cutting force during the cutting process of a milling cutter was monitored. Saglam and Unuvar [7] proposed a neural network model for assessing wear condition of a milling cutter. In the analysis, cutting force signals were selected as model inputs, and tool wear condition was then identified using the trained model. However, expensive sensor and data acquisition device restrict the use of cutting force for many practical applications for TCM. As a possible solution, vibration signals including acceleration are used for online TCM. The backpropagation neural network (BPNN) was used to classify tool wear conditions using vibration signals [8]. However, it is noted that vibration signals are easily affected; thus, influence of tool self-excited vibration and extra noise must be resolved in TCM using vibration signals. From previous studies, acoustic emission (AE) has also been proposed for TCM. Sundaram et al. [9] used AE signals to monitor the tool wear during the cutting process, where tool wear had a significant effect above 200 kHz. Machine learning methods, such as support vector machine (SVM) and artificial neural network (ANN), were used to detect tool wear condition using AE [10]. However, additional sensors should be installed to capture acoustic signals, which increase the cost and complexity of the monitoring system [11].

Recently, power signals have been proposed and applied in TCM studies. A cutting power model was developed for TCM [12]. Moreover, cutting energy coefficients were extracted from power signals and used in TCM analysis [13]. In these studies, time series analysis is usually performed using power signals; thus, a predefined threshold between normal and abnormal tool conditions is required. Furthermore, other techniques have been proposed to extract features from tool power signals, including wavelet-based transform techniques. A novel model based on continuous wavelet transform and blind source separation was proposed for tool wear condition monitoring [14]. In [15], features were extracted from tool power signals using wavelet packet transform, from which different tool wear states were discriminated. Moreover, some machine learning methods have been used to deal with power signals. In [16], a model was established to monitor the tool life, where features extracted from power were updated by Bayesian inference. In addition, tool life was predicted in milling using power signals with neural network technique [17]. An ANN was used to evaluate real-time tool wear with spindle power signals, where features extracted from power signals were used as ANN inputs for training and validation [18].

It can be concluded from above studies that compared to other signals, tool power signals have great potential in TCM, since they can represent actual tool wear condition and can be measured easily and efficiently. However, in existing studies using tool power signals for TCM, extra and complex signal

processing techniques are needed to extract features, which restrict its application for online analysis. Moreover, extraction of distinguishable features still relies on experts' knowledge; thus, automatic TCM cannot be achieved.

With the increase of computing capacity, more advanced neural network models, including deep learning, are widely used in different fields such as image processing, speech recognition, fault diagnosis, and also TCM [19–21]. In the analysis, deep learning methods tend to be used in end-to-end model, combining feature extraction, dimensionality reduction, and monitoring. In [22], multiple stacked sparse autoencoders were used for TCM using unsupervised learning. A deep stacked autoencoder network was applied to classify spindle vibration data of different tool wear states [23]. In [24], a framework based on LSTM was proposed, where force, vibration, and AE signals were selected as model inputs. To reduce the computation time of the LSTM network, a convolutional bidirectional long short-term memory network was proposed in [25], which combined advantages of both deep convolutional neural network (CNN) and long short-term memory network (LSTM).

However, although several studies have been devoted to TCM with deep learning networks, the applications of deep learning network using tool power signals are still limited. Therefore, an accurate and online TCM technique is still urgently required, especially in intelligent manufacturing applications.

In this study, a novel method is proposed for accurate and online TCM using tool power signals, where both W-CONV and LSTM techniques are combined. With the proposed method, feature extraction and state classification can be included into one model, thus allowing the raw power signal as model input, without any preprocessing techniques. The effectiveness of proposed method is investigated using test data from cutting tools at different wear conditions. Moreover, test data from cutting tools with different sizes are used to further clarify the robustness of the proposed method.

The novelty of this work lies in two aspects. Firstly, with combination of W-CONV and LSTM, advantages of both models can be achieved, such as ability of expanding receptive field using W-CONV [26], and time-related feature extraction capacity with LSTM. Secondly, tool power signals are used in this analysis, due to their characteristics of easy and efficient collection, the proposed method can be used for practical applications for online TCM, without adding extra cost and complexity to the monitoring system.

The paper is organized as follows. Section 2 presents the methodologies, including W-CONV, LSTM, and proposed framework. In Section 3, the testing system and corresponding test data are described. In Section 4, effectiveness of the proposed method in TCM is investigated, in terms of both accuracy and required identification time. From the findings, conclusion is made in Section 5.

## 2. Methodologies

*2.1. Convolutional Neural Network.* As a feedforward neural network, the convolutional neural network (CNN) is widely

used in the field of image processing and fault diagnosis. Figure 1 depicts a typical structure of CNN. It can be seen that its hidden layers include several convolutional layers and pooling layers.

The convolutional layer uses convolution kernels to perform feature extraction with the inputs, and maps the features extracted from the previous layer to the next layer. An important characteristic of the convolutional layer is weight sharing, which can reduce computational complexity in the process of nonlinear convolution operations.

According to the dimension of input data, the convolution operation can be divided into one-dimensional convolution and two-dimensional convolution. Two-dimensional convolution is often used for image processing [27]. Since power signal is one-dimensional data, the one-dimensional CNN is used in this study. For time series data, convolution kernels move along the time axis to perform convolution operations. After that, the activation function is used to process the information by adding the bias to obtain a feature map that retains useful information. This can be expressed as follows:

$$U_i^{l+1}(x) = f \left\{ \sum_{i=1}^l \omega_{ij}^l * U_i^l(x) + b_i^l \right\}, \quad (1)$$

where  $U_i^l(x)$  is the  $i$ -th feature map in the  $l$ -th layer,  $f$  represents the activation function,  $\omega_{ij}^l$  represents the convolution kernel, and  $b_i^l$  is the bias.

Moreover, as shown in Figure 1, the pooling layer is located after the convolutional layer. The main function of this pooling layer is to reduce the number of parameters, where downsampling is widely applied to reduce the feature map dimension. In this study, the maximum pooling method is used in the pooling layer, which selects the largest statistical value of the region to represent its characteristics. The result of pooling the  $i$ -th channel of  $l+1$ -th layer can be expressed as

$$P_i^{l+1}(x) = \max_{(x-1)T \leq t \leq xT} \{v_i^l(t)\}, \quad (2)$$

where  $v$  represents the  $t$ -th neuron in the  $i$ -th channel of the  $l$ -th layer and  $T$  represents the pooling step.

**2.2. Long Short-Term Memory.** As the hidden layer structure in the conventional convolutional neural network is relatively simple, gradient disappearance or gradient explosion may be experienced when dealing with large amounts of data.

To address the above issues, the LSTM network is usually used, which is a special kind of recurrent neural network (RNN). Figure 2 shows the internal structure of LSTM neuron. It can be observed that the LSTM network has a three-state gate structure, including input gate, forget gate, and output gate, which provides selective memory in the network.

In the LSTM network, the composition of gates includes a sigmoid network and a bit multiplier. The sigmoid layer

can output a value between 0 and 1, and its value determines whether the input value can pass through the gate.

The forget gate determines what information from the previous moment is retained to the current moment and can play a role of forgetting useless information. This can be written as

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f), \quad (3)$$

where  $f_t$  is the forget gate,  $\sigma$  is a logical function with an output of 0~1,  $x_t$  represents the current input, and  $h_{t-1}$  represents the output at the previous moment,  $W_f$  is the weight of the forget gate, and  $b_f$  is the biases for the forget gate.

The input gate determines the current network input  $x_t$  and what is reserved to the current unit state  $c_t$ :

$$\begin{aligned} i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i), \\ C_t &= f_t \cdot C_{t-1} + i_t \cdot \tanh(W_c \cdot [h_{t-1}, x_t] + b_c), \end{aligned} \quad (4)$$

where  $i_t$  is the input gate,  $\tanh$  is the activation function,  $W_i$  is the weight of the input gate, and  $b_i$  is the biases for the input gate.

The output gate controls what information the unit state  $c_t$  outputs to the current value  $h_t$  of LSTM:

$$\begin{aligned} O_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o), \\ h_t &= O_t \cdot \tanh(C_t), \end{aligned} \quad (5)$$

where  $O_t$  is the output gate,  $W_o$  is the weight of the output gate, and  $b_o$  is the biases for the output gate.

**2.3. Proposed Model.** In this study, a deep learning network depicted as Figure 3 is proposed for TCM, where tool wear condition can be determined automatically by extracting features from tool power signals.

It can be seen from Figure 3 that wide convolution kernels are used at the 1st convolutional layer, with which the network receptive field can be improved and high-frequency noise can be reduced. Moreover, the use of wide convolution kernels can reduce the number of convolutional layers, thus reducing the complexity of the network structure and parameter calculation.

It should be noted that except for the 1st convolutional layer, small convolution kernels are used at other convolutional layers, since the small convolution kernels can optimize local features and deepen the network depth to improve the performance of the network. Furthermore, outputs from the last pooling layer are injected to the LSTM layer, which can compensate shortcomings of CNN's incomplete feature extraction in time series, such that the model generalization ability can be improved. Finally, the output from LSTM layer is regarded as inputs of softmax function, which can convert input neuron (0-1) probability distribution in order to classify different wear states. The softmax function can be expressed as follows:

$$\text{softmax}(a_i) = \frac{e^{a_i}}{\sum e^{a_i}}, \quad (6)$$

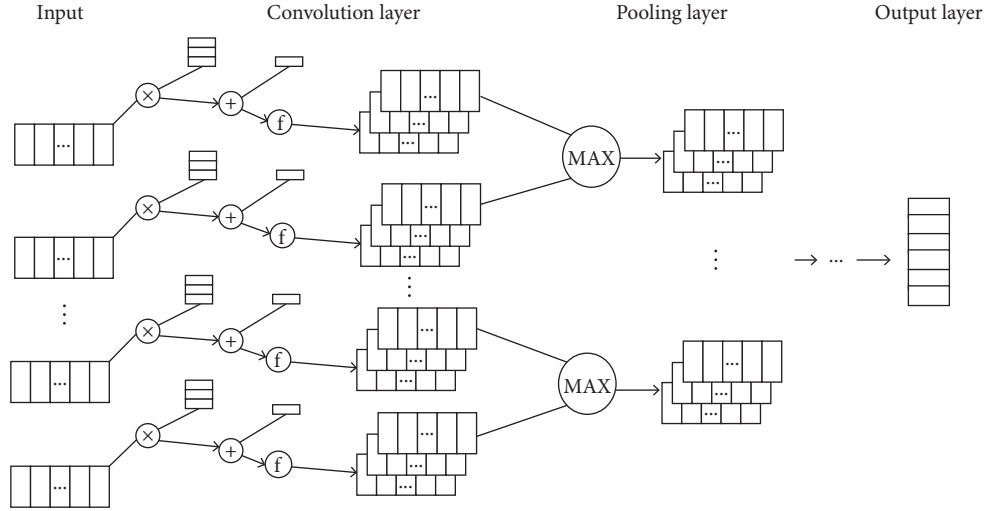


FIGURE 1: The typical structure of CNN.

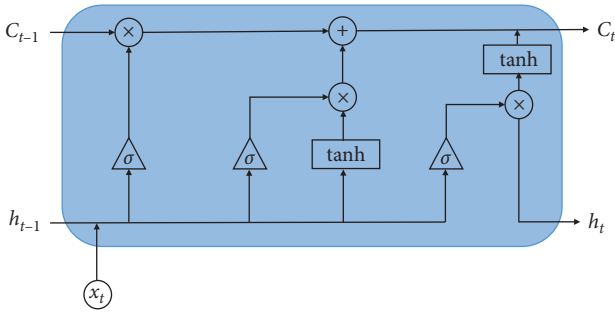


FIGURE 2: Internal structure of LSTM neuron.

where  $a_i$  represents the output of the  $i$ -th neuron.

In the analysis, the batch normalization (BN) method is used to reduce variance transfer between layers. It can normalize all hidden layers and reduce differences between samples. With BN, the network training time can be reduced effectively, and gradient disappearance or gradient explosion can also be minimized. It should be mentioned that feature maps are batch normalized between convolution operation and activation processing at convolutional layers. This process can be expressed as

$$\mu_{\text{BN}} = \frac{1}{n} \sum_{m=1}^n x_m, \quad (7)$$

$$\sigma_{\text{BN}}^2 = \frac{1}{n} \sum_{m=1}^n (x_m - \mu_{\text{BN}})^2, \quad (8)$$

$$\hat{x}_m = \frac{x_m - \mu_{\text{BN}}}{\sqrt{\sigma_{\text{BN}}^2 + \varepsilon}}, \quad (9)$$

$$y_m = \gamma \hat{x}_m + \beta = \text{BN}_{\gamma\beta}(x_m), \quad (10)$$

where  $x_m$  represents the value of  $x$  in a batch,  $n$  represents the batch size, and  $\gamma$  and  $\beta$  are the scaling parameters and translation parameters, respectively.

Moreover, the reason of adding LSTM in the CNN is further clarified. As power signals from the tool cutting process are used in the analysis, which are periodic signals containing both global and local features, the proposed model should have the capability of spatial-temporal feature extractions. Therefore, local correlation of time series signals should be obtained to integrate local features before recognizing global features. With long-term memory of LSTM, current inputs can be processed based on previous state information; thus, local features can be fused and global features can be processed. Therefore, in the proposed model, the CNN can extract local features and reduce the noise influence, while LSTM can further extract time-related global features.

### 3. Description of Testing System and Corresponding Test Data

In this study, high speed steel containing 8% cobalt (HSS-Co8) with 4 flutes is used as the end milling tool, as the wear of HSS-Co8 can be measured easily. Moreover, two end milling tools with diameters of 8 mm and 10 mm are used in the test, respectively; thus, the robustness of proposed model can be illustrated. Table 1 lists technical parameters of two end milling tools.

In the test, the cutting process is conducted using a Hurco 3-axis VN1 CNC, where three phase current and voltage are collected with sample frequency of 50 kHz. It should be mentioned that since the sensors used in the tests will not add complexity to the monitoring system, the proposed model can be applied for practical machining applications. Figure 4 depicts the test bench used in the analysis.

Moreover, in the experiment, each end mill is assigned a work piece with a dimension of 30 mm × 150 mm × 120 mm. The plate material is selected as commercial aluminium grade 6082 T651, as it is a common alloy used in manufacturing for its machinability and material properties.

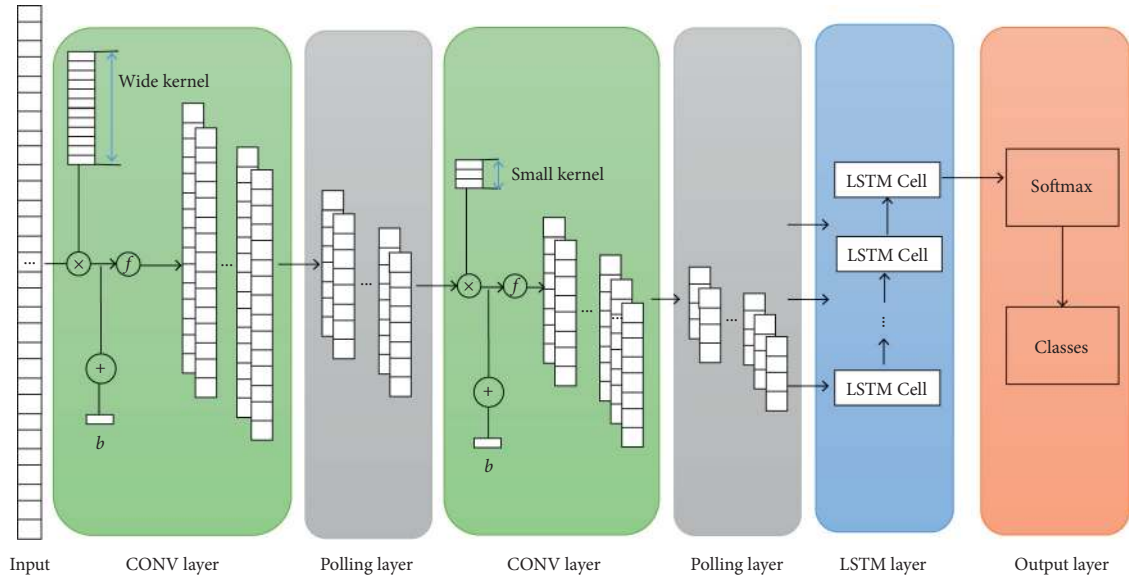


FIGURE 3: The structure of the proposed model.

TABLE 1: Technical parameters of end mills.

Mill dia. (mm)	Shank dia. (mm)	Overall length (mm)	No. of teeth/flutes
8	10	69	4
10	10	72	4



FIGURE 4: Illustration of test bench.

During the cutting process, the milling tools are required to perform several milling sessions. It should be mentioned that at each cutting, the tools are used for a total of 40 minutes in the first session, as there is little sign of flank wear growth after only 20 minutes cutting at the first session. Except for the 1st milling session, each subsequent session is carried out for 20 minutes. After each cutting session, the tools are inspected optically, and flank wear land size is measured and recorded. Each test contains 4 milling sessions, which represents a total of 100 minutes of machining.

In the test, three phase voltage and current can be collected from end milling tools at 5 different tool wear conditions (0 min, 40 min, 60 min, 80 min, and 100 min). With collected voltages and currents, instantaneous power can be calculated as follows:

$$P_{\text{ins}} = \sum_{n=1}^3 v(t)_n i(t)_n, \quad (11)$$

where  $v(t)_n$  and  $i(t)_n$  are the collected voltage and current measurements at time  $t$ , and  $n$  represents the phase of  $v$  and  $i$ .

Based on above test configuration, 5 sets of power signals can be obtained from each test. Figure 5 depicts a typical calculated power signal from a cutting test using equation (11), where the start of spindle rotation, arrival and leave of cutting tool to the work piece, and cutting end are highlighted.

Table 2 lists the control parameters during the cutting process and measured tool wear size after different milling sessions.

#### 4. Effectiveness of Proposed Method in TCM

**4.1. Segmentation of Tool Output Power Signal.** Figure 6 shows instantaneous power of the 8 mm tool at five different wear conditions (0 min, 40 min, 60 min, 80 min, and 100 min as listed in Table 2). It should be mentioned that only the power signal when the tool cuts at work piece is used; that is, power signals from 10 passes indicated in Figure 5 are used in the following analysis, since tool wear

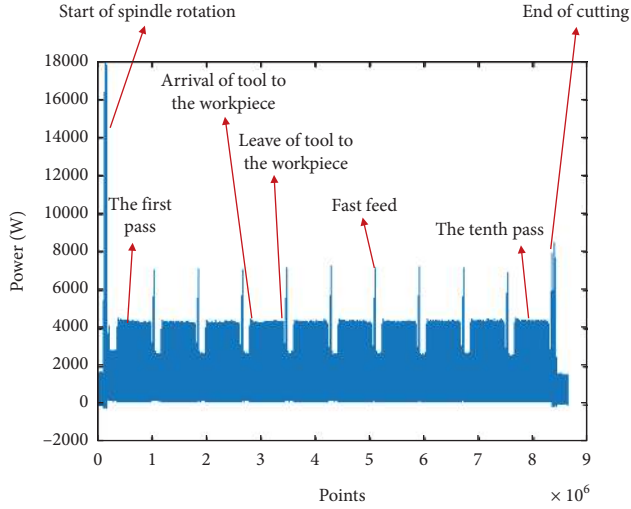


FIGURE 5: Tool power signal from a cutting test.

TABLE 2: Controlling parameters during the cutting process.

Mill dia. (mm)	Time	RPM (r/min)	Feed rate (m/min)
8	0	4000	580
8	40	4000	580
8	60	4000	580
8	80	4000	580
8	100	4000	580
10	0	3100	600
10	40	3100	600
10	60	3100	600
10	80	3100	600
10	100	3100	600

condition can be expressed when performing the cutting process.

As can be seen from Figure 6, the output power signal increases with increased tool wear, but it is difficult to distinguish different tool wear conditions directly from power signals, since clearly features from different tool wear conditions cannot be observed.

In order to apply the proposed model, power signals are configured and divided into different samples. The rotation speeds of 8 mm and 10 mm tools are 4000 r/min and 3100 r/min, respectively, corresponding to 750 and 968 points at each rotation cycle. Therefore, in this analysis, each sample contains 2048 points to ensure that at least two rotation cycles of data are included, from which distinguishable features can be extracted. At each tool wear condition, 2000 samples are randomly selected from power signals. Moreover, 70% of samples are used for training, and the rest are used for testing. It means that there are 7000 samples for training and 3000 samples for testing in total.

To verify the model, a classification accuracy  $a_r$  is introduced as follows, which has been widely used in previous studies for accuracy evaluation [28]:

$$a_r = \frac{M_C}{M_T}, \quad (12)$$

where  $M_C$  represents the correctly predicted samples and  $M_T$  represents the total samples.

**4.2. Configuration of Proposed Model.** In this analysis, the proposed model shown in Figure 3 is used for tool wear condition assessment, which consists of two convolutional layers, two pooling layers, an LSTM layer, and an output layer. In the model, mean square error is used as loss function, based on which the Adam algorithm is selected to update the learning rate automatically, and weights are updated with back propagation technique. Moreover, the dropout method is used in the model to prevent overfitting.

The analysis is implemented is performed in kears framework with python3.7, on a desktop with i7-6700K CPU and 32G RAM. Moreover, the number of proposed model training iterations is set to 20, as convergence is observed with this iteration.

As described in Section 2, wide convolution kernels are used at the 1st convolution layer. Based on previous studies [26], the size of convolution kernel in the 1st convolutional layer is selected as 64, in order to obtain larger receptive field. Moreover, the number of convolution kernel at the 1st convolution layer is set as 16, to accommodate the input with 2048 points. In the second convolutional layer, the size of the convolution kernel is 3, which is commonly used size in previous studies [26, 29]. However, there is no protocol for the determination of kernel number at the 2nd convolutional layer, and this will be investigated later. Regarding parameters at the LSTM layer, 32 units are selected with trail-and-error analysis, as too many units will increase the network computing time, while too little will reduce the accuracy.

From above configuration, the only uncertain parameter is the number of kernels at the 2nd convolution layer. In this study, five candidate values, including 16, 32, 64, 128, and 256, are used, and their performance is evaluated in terms of both accuracy and training time. Figure 7 shows the results with different kernel numbers at the 2nd convolution layer.

It can be seen from Figure 7 that the classification accuracy is lower when the number of convolution kernels is too small. However, with too large kernel number, the accuracy will be reduced due to overfitting. Moreover, training time will increase with an increased kernel number. Considering both the accuracy and training time, 64 convolution kernels are selected in this study, where test accuracy can reach 96.8%, and the training time is 42 s.

Table 3 lists the configured model parameters used in this study.

**4.3. Performance of Proposed Model in TCM.** In this section, the performance of the proposed model in TCM is firstly investigated using test data from an 8 mm tool. Moreover, in order to further clarify the benefit of the proposed model, its classification results are compared to those with the conventional CNN. Furthermore, the effectiveness of the proposed model in TCM for a 10 mm tool is also investigated, such that the robustness of the proposed model can be illustrated.

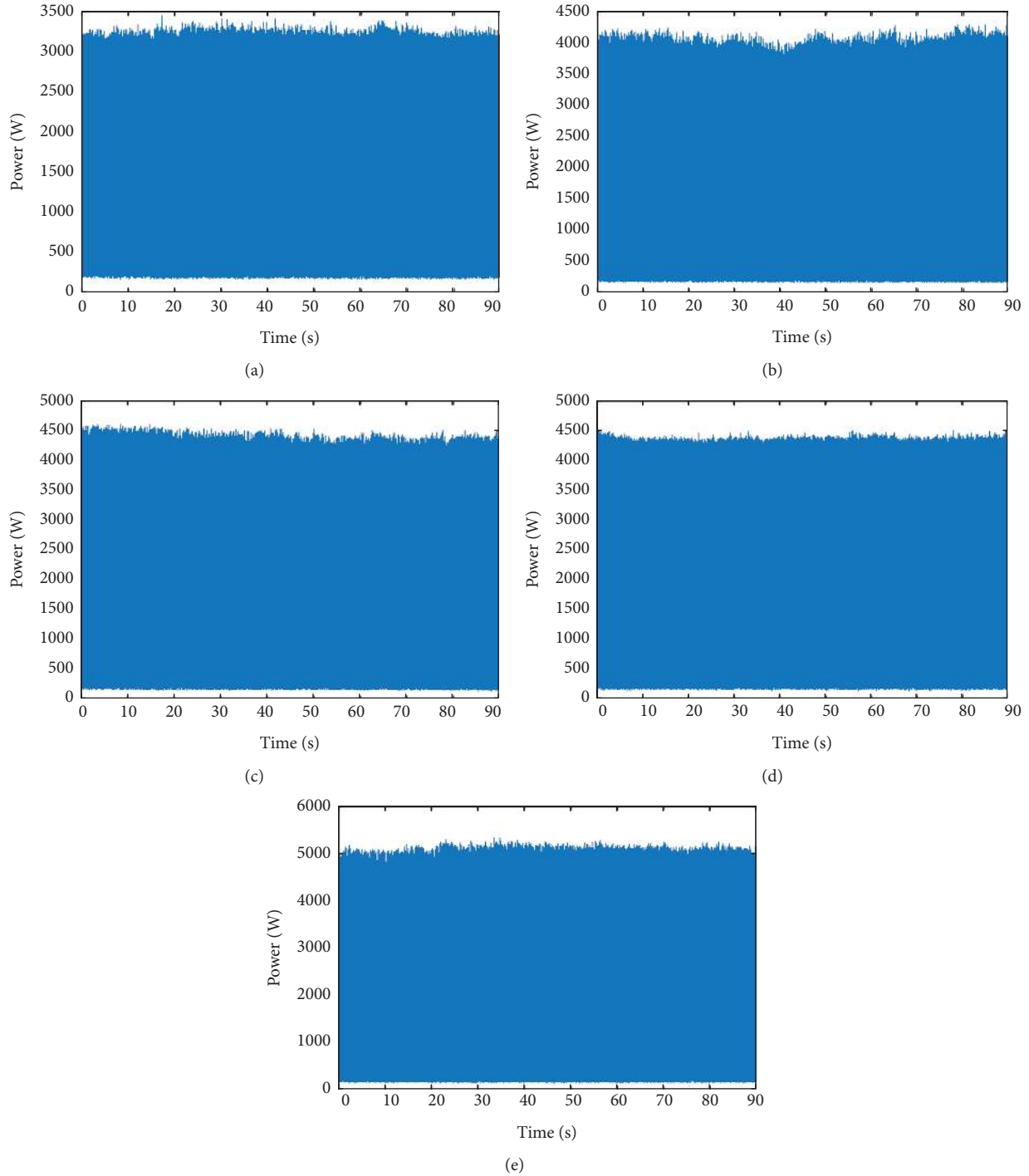


FIGURE 6: Instantaneous power of 8 mm tool at different wear conditions. (a) 00 min. (b) 40 min. (c) 60 min. (d) 80 min. (e) 100 min.

#### 4.3.1. Performance of Proposed Model in TCM for 8 mm Tool.

In this case, 10 trials are performed, where training and test samples are randomly selected from the test data. Classification results are shown in Figure 8 and Table 4. All samples are randomly collected from the power data for all 10 passes cutting.

It can be found from above results that with the proposed model, all training accuracy can achieve 99% and the average test accuracy achieves 95.9%, while the highest test accuracy

and the lowest test accuracy are 98.6% and 90.6%, respectively. Figure 9 depicts the confusion matrixes for the lowest accuracy and the highest accuracy. It can be seen that the reason for the low classification accuracy is mainly due to the misclassification of the wear state for the tool at 100 min and 80 min. This situation may be due to the influence of noise on the test sample at that experiment, resulting in insignificant signal features in various states. However, the overall classification test results indicate that accurate tool wear

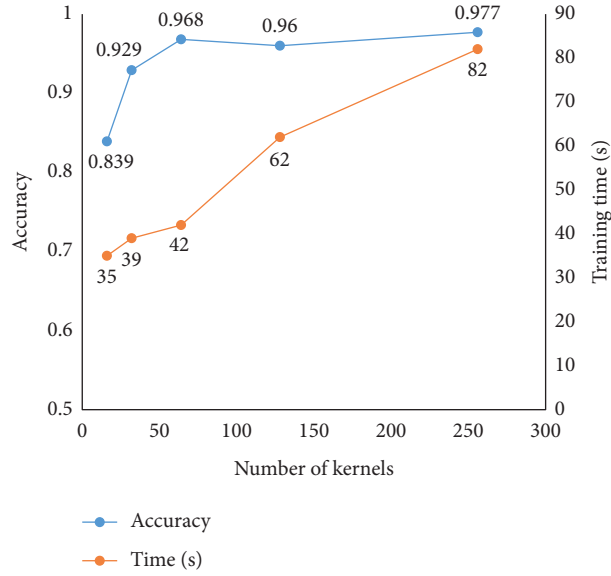


FIGURE 7: Performance using different kernel numbers at the 2nd convolution layer.

TABLE 3: Parameters of the proposed model.

Name	Filters	Size/stride	Units	Input	Output	Function	Parameters
Conv-1	16	64/16	—	1 × 2048	16 × 128	Relu	1040
Pooling1	—	2/2	—	16 × 128	16 × 64	—	0
Conv-2	64	3/1	—	16 × 64	64 × 64	Relu	3136
Pooling2	—	2/2	—	64 × 64	64 × 32	—	0
LSTM	—	—	32	64 × 32	32 × 32	Tanh	12416
Dense	—	—	—	32 × 32	1 × 5	Softmax	5125

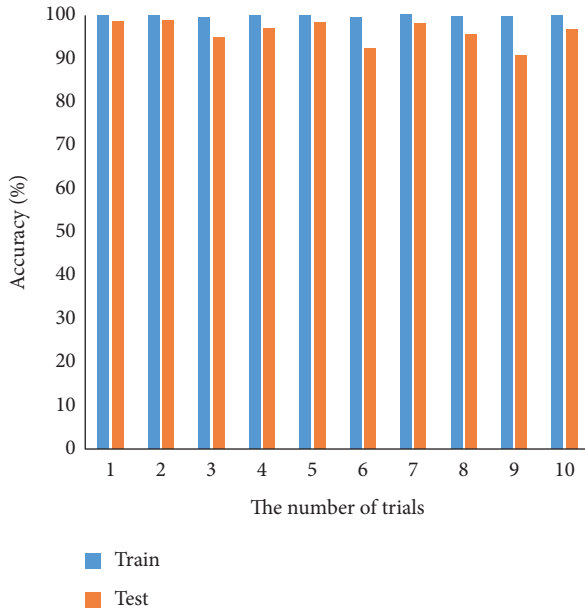


FIGURE 8: Performance of the proposed model in TCM for the 8 mm tool.

condition, especially the minor tool wear state, can be identified efficiently using the proposed model.

TABLE 4: Results of 10 trials of the proposed model.

Trial	1	2	3	4	5	6	7	8	9	10
Train $a_r$ (%)	99.7	99.8	99.3	99.8	99.7	99.4	99.9	99.6	99.6	99.7
Testing $a_r$ (%)	98.4	98.6	94.6	96.8	98.2	92.1	97.8	95.4	90.6	96.5

To illustrate the effectiveness of the proposed model, classification results using W-CONV are obtained with 10 trials, which are shown in Figure 10. It can be seen that the average test accuracy of 10 trials is only 86.2%, while the highest test accuracy and the lowest test accuracy are 98.4% and 68.9%, respectively. The reason is that with W-CONV, some global features related to time characteristics would be ignored during training process in the big data case, thus leading to worse classification accuracy.

Compared to the results using the proposed model, the average accuracy and the minimum accuracy from W-CONV are significantly reduced. The range of accuracy is 29.7%, while that of the proposed model is only 8%. It can be seen that the proposed model is more robust and more stable, especially when dealing with big data case. The confusion matrix for the lowest accuracy of W-CONV is shown in Figure 11. It can be seen that the tool wear conditions at 60 min and 100 min cannot be classified correctly. This further confirms that with the proposed



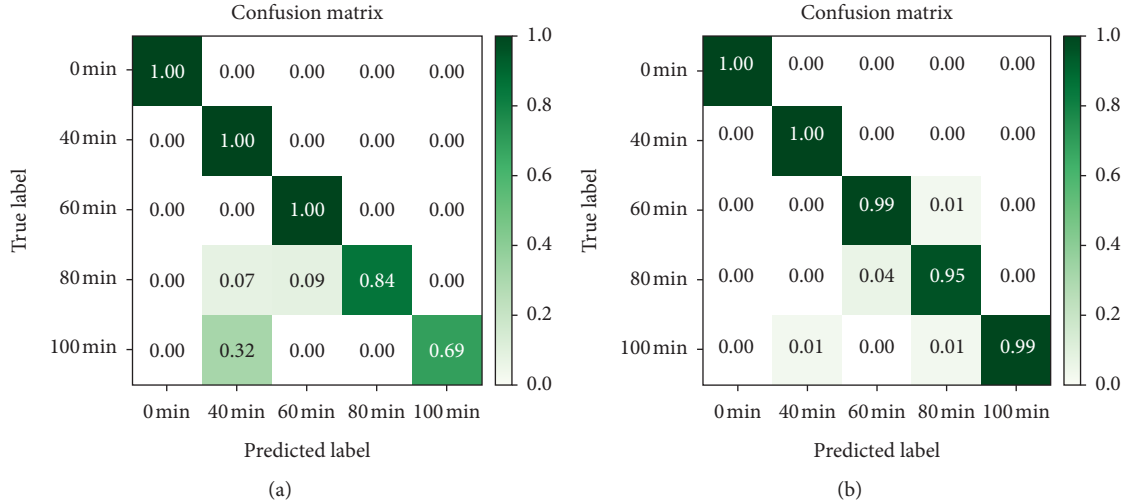


FIGURE 9: Confusion matrix of the 8 mm tool: (a) the lowest accuracy = 90.6%; (b) the highest accuracy = 98.6%.

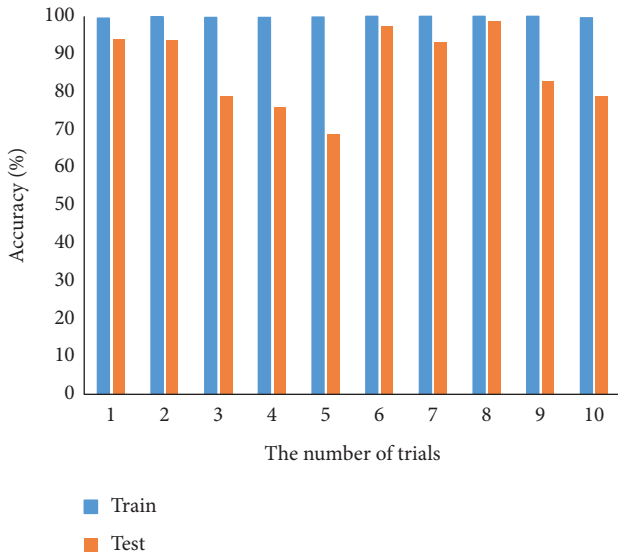


FIGURE 10: Performance of the W-CONV model.

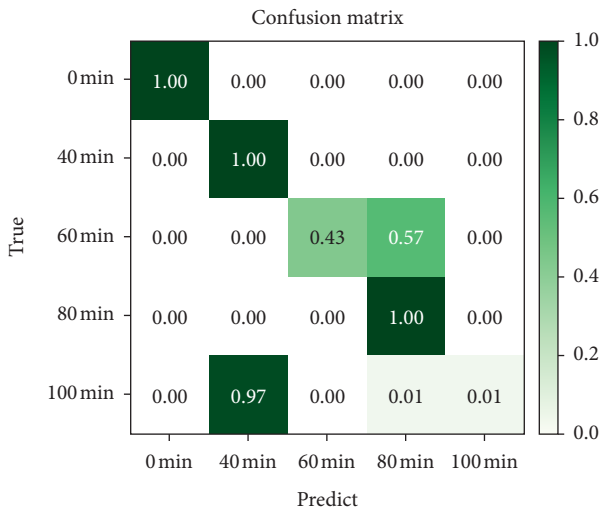


FIGURE 11: Confusion matrix of lowest accuracy using W-CONV.

model, the lack of capacity in time-related feature extraction with W-CONV can be properly addressed, thus increasing its classification accuracy in different data cases.

In addition, mean squared error (MSE), root mean squared error (RMSE), and root mean squared logarithmic error (RMSLE) are introduced to compare the models. The three accuracy criteria can be calculated as follows:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad (13)$$

$$\text{RMSLE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\log(\hat{y}_i + t1) - \log(y_i + 1))^2},$$

where  $y_i$  is the true label and  $\hat{y}_i$  is the predicted value.

For MSE, RMSE, and RMSLE, the smaller their value, the better the model performance. Tables 5–7 show the calculation results with the proposed model and W-CONV. It can be seen that the average MSEs of two model are 0.0159 and 0.0469, while the average RMSEs of two model are 0.00388 and 0.00639. Table 7 shows that the average RMSLEs of two models are 0.00288 and 0.00472. Regardless of the maximum value, minimum value or average value, MSE, RMSE, and RMSLE of the proposed model are smaller than the value of W-CONV, which is consistent with the results depicted in Figures 9–11.

For further comparison, Figure 12 shows the training process with the proposed model and W-CONV. It is found that the proposed model can reach convergence on the 2nd iteration, while W-CONV converges after the 4th iteration. Therefore, it can be concluded that compared to conventional CNN, the proposed method can provide more accurate and efficient identification for tool wear conditions.

TABLE 5: Comparison of two models using MSE of 10 trials.

Trials	1	2	3	4	5	6	7	8	9	10
Proposed model ( $10^{-3}$ )	8.3	8.1	19.9	13.1	8.9	29.4	9.4	17.1	30.6	14.5
W-CONV ( $10^{-3}$ )	23.1	23.4	72.9	84.2	111.3	11.2	24.5	8.3	51	59.4

TABLE 6: Comparison of two models using RMSE of 10 trials.

Trials	1	2	3	4	5	6	7	8	9	10
Proposed model ( $10^{-3}$ )	2.88	2.84	4.46	3.62	2.98	5.42	3.07	4.14	5.53	3.81
W-CONV ( $10^{-3}$ )	4.81	4.83	8.54	9.18	10.55	3.35	4.95	2.88	7.14	7.71

TABLE 7: Comparison of two models using RMSLE of 10 trials.

Trials	1	2	3	4	5	6	7	8	9	10
Proposed model ( $10^{-3}$ )	2.21	2.10	3.31	2.59	2.30	3.87	2.39	3.13	4.14	2.74
W-CONV ( $10^{-3}$ )	3.44	3.56	6.28	6.32	7.27	2.74	3.62	2.17	5.50	6.32

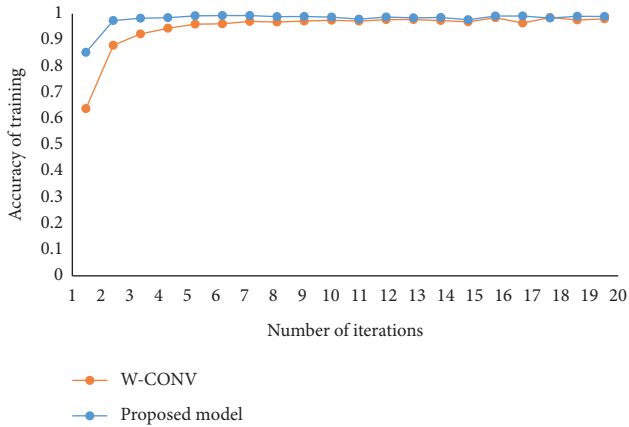


FIGURE 12: Training process of different models.

This is due to the fact that without LSTM, some global features that are related to time will be ignored during training process in the big data case.

The reason of worse classification results using W-CONV is further investigated using the smaller test dataset. In this analysis, only test data from single pass cutting are used (as shown in Figure 5). Figure 13 compares the classification accuracy using data from single pass cutting (small data case) and all 10 pass cuttings (big data case).

It can be seen that with small data case, the W-CONV average predict accuracy of 10 trials is 99.2%, while its average accuracy is only 86.2% in big data case. The reason is that with small data case, test data have been trained during the training stage with 10 trials. However, with increased data amount, this cannot be achieved, leading to the worse classification results using W-CONV.

Furthermore, the training and testing time of each model is shown in Table 8. It should be mentioned that the testing time is the time of each sample. The training time of the proposed model is slightly longer than that of W-CONV, because the LSTM layer takes a certain amount of time to extract global features. However, in terms of the testing time

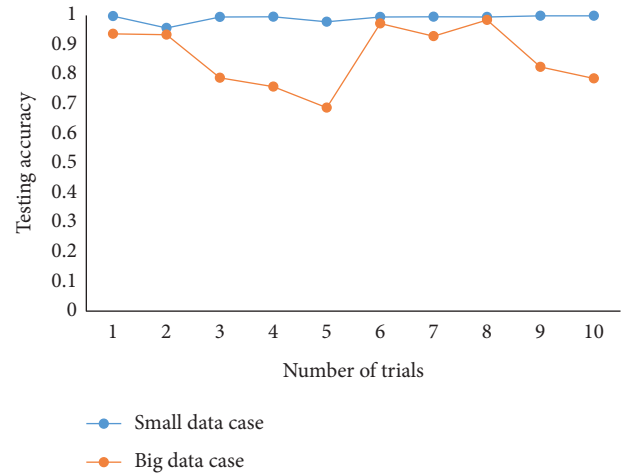


FIGURE 13: Accuracy of W-CONV with different data cases.

TABLE 8: Computational times of each model.

Model	Training time (s)	Testing time ( $\mu$ s)
Proposed	42	108
W-CONV	30.5	95

TABLE 9: Results of 10 trials of the proposed model.

Trials	1	2	3	4	5	6	7	8	9	10
Train $a_r$ (%)	99.5	99.9	100	99.7	100	99.8	100	100	100	100
Testing $a_r$ (%)	97.8	100	100	99.8	100	99.9	100	100	100	100

TABLE 10: Computational times of the proposed model of the 10 mm tool.

Training time	Testing time
40.5 (s)	108 ( $\mu$ s)

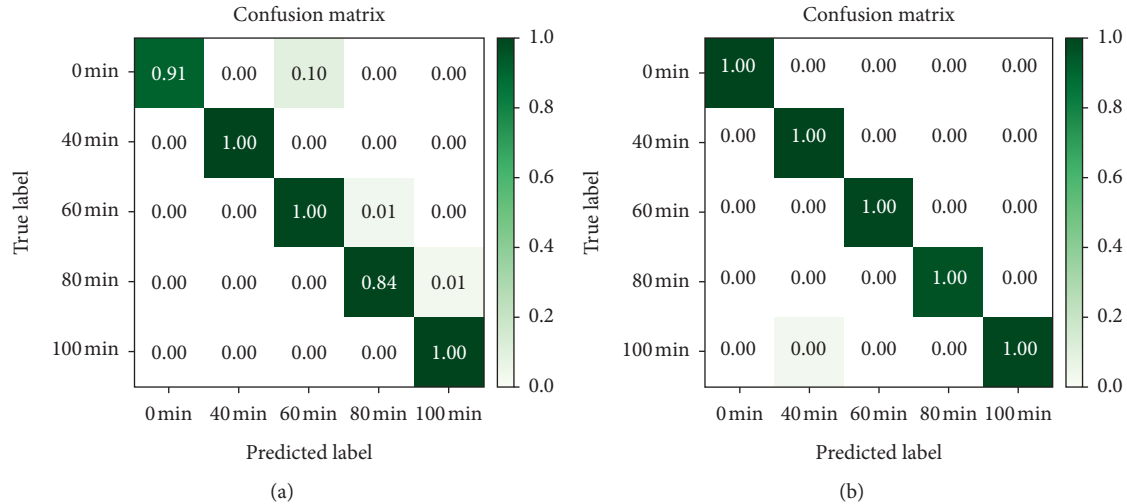


FIGURE 14: Confusion matrixes of classification for the dataset of the 10 mm tool: (a) lowest accuracy = 97.8%; (b) highest accuracy = 100%.

of each sample, both the proposed model and W-CONV can reach the microsecond level and the difference between the two models is relatively small, which will not hinder it for online monitoring purpose. However, the difference in classification accuracy from two models is significant, which is shown in Figures 8 and 10, where the proposed model can provide more accurate classification results. Therefore, with considering of both computational times and classification accuracy, the proposed model can provide accurate classification while can still meet online monitoring requirement.

**4.3.2. Performance of Proposed Model in TCM for 10 mm Tool.** To further verify the robustness of the proposed model, its effectiveness in TCM with 10 mm tool is investigated in this section. It should be mentioned that big data case is used herein; that is, training and testing samples are randomly collected from the power data for all 10 passes cutting. Results are listed in Table 9.

It can be seen from Table 9 that the proposed model can also provide accurate classification with test data from the 10 mm tool, where the average test accuracy of the 10 mm tool dataset reaches 99.75%, and the highest test accuracy and the lowest test accuracy are 100% and 97.8%, respectively.

Figure 14 shows the confusion matrix for the lowest and the highest accuracy (97.8% and 100%, respectively). It can be found that although some data are misclassified, the proposed model can still provide accurate classification, since different tool wear conditions can be discriminated correctly. Furthermore, computational times of the proposed model of the 10 mm tool are shown in Table 10. The testing time shows that the proposed model can be applied at practical applications for accurate and online TCM.

## 5. Conclusion

In this study, an improved deep learning model is proposed for tool wear condition assessment, which integrates advantages from W-CONV and LSTM, such that distinguishable features can be extracted automatically from tool test data. Moreover, output power signals from cutting tool during its operation are used in the analysis, since they can be collected easily and efficiently, without adding extra cost and complexity to the monitoring system. Therefore, the proposed model is applicable at practical applications for accurate and online tool wear condition assessment.

To validate the effectiveness of the proposed model, test data are collected from the cutting tool at different wear conditions. From 10-fold cross validation results, the proposed model can provide accurate and efficient assessment to the tool wear condition. Compared to results using conventional CNN, the proposed model shows better capability in dealing with massive data. Moreover, test data collected from cutting tools with different sizes are utilized in the analysis, such that the robustness of the proposed model in identifying wear condition of various tools can be illustrated, which will also be beneficial at practical applications. In future work, more test data will be added to investigate effectiveness of the proposed model in classification of more cutting tool states. In addition, the proposed model will be updated to provide accurate classification using the limited test to improve the robustness.

## Data Availability

Some data relevant to the article will be made available upon request for research purpose.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (NSFC) (51975549), Anhui Provincial Natural Science Foundation (1908085ME161), and State Key Laboratory of Mechanical System and Vibration (MSV202017).

## References

- [1] D. E. Dimla, "Sensor signals for tool-wear monitoring in metal cutting operations—a review of methods," *International Journal of Machine Tools and Manufacture*, vol. 40, no. 8, pp. 1073–1098, 2000.
- [2] Y. Altintas and I. Yellowley, "In-process detection of tool failure in milling using cutting force models," *Journal of Engineering for Industry*, vol. 111, no. 2, pp. 149–157, 1989.
- [3] P. Bhattacharyya and D. Sengupta, "Estimation of tool wear based on adaptive sensor fusion of force and power in face milling," *International Journal of Production Research*, vol. 47, no. 3, pp. 817–833, 2009.
- [4] T. Mohanraj, S. Shankar, R. Rajasekar, N. R. Sakthivel, and A. Pramanik, "Tool condition monitoring techniques in milling process—a review," *Journal of Materials Research and Technology*, vol. 9, no. 1, pp. 1032–1042, 2020.
- [5] X. Jiang, C. Shen, J. Shi, and Z. Zhu, "Initial center frequency-guided VMD for fault diagnosis of rotating machines," *Journal of Sound and Vibration*, vol. 435, pp. 36–55, 2018.
- [6] S. N. Huang, K. K. Tan, and Y. S. Wong, "Tool wear detection and fault diagnosis based on cutting force monitoring," *International Journal of Machine Tools & Manufacture*, vol. 47, no. 3–4, pp. 444–451, 2007.
- [7] H. Saglam and A. Unuvar, "Tool condition monitoring in milling based on cutting forces by a neural network," *International Journal of Production Research*, vol. 41, no. 7, pp. 1519–1532, 2003.
- [8] W.-H. Hsieh, M.-C. Lu, and S.-J. Chiou, "Application of backpropagation neural network for spindle vibration-based tool wear monitoring in micro-milling," *The International Journal of Advanced Manufacturing Technology*, vol. 61, no. 1–4, pp. 53–61, 2012.
- [9] S. Sundaram, P. Senthilkumar, A. Kumaravel, and N. Manoharan, "Study of flank wear in single point cutting tool using acoustic emission sensor techniques," *ARPJ Journal of Engineering and Applied Sciences*, vol. 3, pp. 32–36, 2008.
- [10] P. Krishnakumar, K. Rameshkumar, and K. I. Ramachandran, "Machine learning based tool condition classification using acoustic emission and vibration data in high speed milling process using wavelet features," *Intelligent Decision Technologies*, vol. 12, no. 2, pp. 265–282, 2018.
- [11] L. Xiaoli, "On-line detection of the breakage of small diameter drills using current signature wavelet transform," *International Journal of Machine Tools and Manufacture*, vol. 39, no. 1, pp. 157–164, 1999.
- [12] H. Shao, H. L. Wang, and X. M. Zhao, "A cutting power model for tool wear monitoring in milling," *International Journal of Machine Tools and Manufacture*, vol. 44, no. 14, pp. 1503–1509, 2004.
- [13] M. Xu, R. B. Jerard, and B. K. Fussell, "Energy based cutting force model calibration for milling," *Computer-Aided Design and Applications*, vol. 4, no. 1–4, pp. 341–351, 2007.
- [14] T. Benkedjouh, N. Zerhouni, and S. Rechak, "Tool wear condition monitoring based on continuous wavelet transform and blind source separation," *The International Journal of Advanced Manufacturing Technology*, vol. 97, no. 9–12, pp. 3311–3323, 2018.
- [15] L. Mao, L. Jackson, P. Goodall, and A. West, *Enhanced Condition Monitoring of the Machining Process Using Wavelet Packet Transform, Safety and Reliability—Safe Societies in a Changing World*, CRC Press, Boca Raton, FL, USA, 2018.
- [16] J. M. Karandikar, A. E. Abbas, and T. L. Schmitz, "Tool life prediction using Bayesian updating. part 1: milling tool life model using a discrete grid method," *Precision Engineering*, vol. 38, no. 1, pp. 9–17, 2014.
- [17] C. Drouillet, J. Karandikar, C. Nath, A.-C. Journeaux, M. El Mansori, and T. Kurfess, "Tool life predictions in milling using spindle power with the neural network technique," *Journal of Manufacturing Processes*, vol. 22, pp. 161–168, 2016.
- [18] 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," *Journal of Manufacturing Systems*, vol. 43, no. 2, pp. 287–295, 2017.
- [19] H.-J. Lee and S.-G. Lee, "Arousal-valence recognition using CNN with STFT feature-combined image," *Electronics Letters*, vol. 54, no. 3, pp. 134–136, 2018.
- [20] J. Zheng, X. Cao, B. Zhang, X. Zhen, and X. Su, "Deep ensemble machine for video classification," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 2, pp. 553–565, 2019.
- [21] C. Shen, Y. Qi, W. Jun, G. Cai, and Z. Zhu, "An automatic and robust features learning method for rotating machinery fault diagnosis based on contractive autoencoder," *Engineering Applications of Artificial Intelligence*, vol. 76, pp. 170–184, 2018.
- [22] 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 Transactions on Industrial Electronics*, vol. 66, no. 5, pp. 3794–3803, 2019.
- [23] V. Nguyen, V. Nguyen, and V. Pham, "Deep stacked auto-encoder network based tool wear monitoring in the face milling process," *Strojniški Vestnik—Journal of Mechanical Engineering*, vol. 66, no. 4, pp. 227–234, 2020.
- [24] H. B. Sun, J. D. Zhang, R. Mo, and X. Z. Zhang, "In-process tool condition forecasting based on a deep learning method," *Robotics and Computer Integrated Manufacturing*, vol. 64, Article ID 101924, 2020.
- [25] 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, p. 273, 2017.
- [26] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals," *Sensors*, vol. 17, no. 2, p. 425, 2017.
- [27] Z. Zhu, G. Peng, Y. Chen, and H. Gao, "A convolutional neural network based on a capsule network with strong generalization for bearing fault diagnosis," *Neurocomputing*, vol. 323, pp. 62–75, 2019.
- [28] H. H. Pan, X. X. He, S. Tang, and F. M. Meng, "An improved bearing fault diagnosis method using one-dimensional CNN and LSTM," *Journal of Mechanical Engineering*, vol. 64, pp. 443–452, 2018.
- [29] M. Qiao, S. Yan, X. Tang, and C. K. Xu, *Deep Convolutional and LSTM Recurrent Neural Networks for Rolling Bearing Fault Diagnosis under Strong Noises and Variable Loads*, IEEE Access, Piscataway, NJ, USA, 2020.