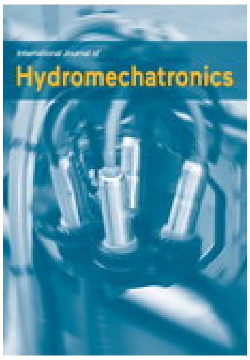
INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research



International Journal of Hydromechatronics

ISSN online: 2515-0472 - ISSN print: 2515-0464
https://www.inderscience.com/ijhm

TCM in milling processes based on attention mechanism-combined long short-term memory using a sound sensor under different working conditions

Guoxiao Zheng, Wei Chen, Qijia Qian, Anil Kumar, Weifang Sun, Yuqing Zhou

DOI: 10.1504/IJHM.2022.10047740

Article History:

Received:	14 December 2021
Last revised:	24 May 2022
Accepted:	06 April 2022
Published online:	25 August 2022

Copyright © Inderscience Enterprises Ltd.

TCM in milling processes based on attention mechanism-combined long short-term memory using a sound sensor under different working conditions

Guoxiao Zheng

College of Mechanical and Electrical Engineering,
Wenzhou University,
Wenzhou, China
Email: 909936695@qq.com

Wei Chen

Manufacturing Department,
Zhejiang Keteng Precision Machinery Co., Ltd,
Wenzhou, China
Email: 1550302411@qq.com

Qijia Qian

Technology Center,
Wenzhou Ruiming Industrial Co., Ltd,
Wenzhou, China
Email: 77648750@qq.com

Anil Kumar and Weifang Sun

College of Mechanical and Electrical Engineering,
Wenzhou University,
Wenzhou, China
Email: anil_taneja86@yahoo.com
Email: vincent_suen@126.com

Yuqing Zhou*

College of Mechanical and Electrical Engineering,
Wenzhou University,
Wenzhou, China
and
College of Mechanical and Electrical Engineering,
Jiaxing Nanhu University,
Jiaxing, China
Email: 609123725@qq.com
*Corresponding author

Copyright © 2022 Inderscience Enterprises Ltd.

Abstract: Tool condition monitoring (TCM) is essential for the milling process to ensure machining quality, and several deep learning (DL)-based methods have been proposed to obtain good regression accuracy for TCM, such as RNN and LSTM. Unfortunately, the performances of these DL-based methods are not good enough under different working conditions. A novel method combining attention mechanism and long short-term memory (LSTM) is proposed. Firstly, sound time series signal obtained from a machining process is converted into several feature sequences, and these feature sequences are input to the attention mechanism-combined LSTM (AMLSTM) to train the weight of the feature sequences. Finally, the trained AMLSTM model with the optimal weight of the feature can be used to estimate the tool wear value. The application of the proposed method in milling TCM experiments shows that the AM-LSTM-based method is significantly better than SVR-based, RNN-based, and LSTM-based methods under different working conditions. Moreover, skewness and kurtosis are two important features for TCM.

Keywords: tool condition monitoring; TCM; long short-term memory network; attention mechanism.

Reference to this paper should be made as follows: Zheng, G., Chen, W., Qian, Q., Kumar, A., Sun, W. and Zhou, Y. (2022) 'TCM in milling processes based on attention mechanism-combined long short-term memory using a sound sensor under different working conditions', *Int. J. Hydromechatronics*, Vol. 5, No. 3, pp.243–259.

Biographical notes: Guoxiao Zheng received his BS in Mechanical Design and Manufacturing from Zhejiang University of Technology, Hangzhou, China, in 2019. He is currently working toward his MS in Mechanical Engineering from Wenzhou University. His research interests include condition monitoring and fault diagnosis.

Wei Chen is currently the Director of the Manufacturing Department, Zhejiang Keteng Precision Machinery, Wenzhou, China. His research interests include intelligent manufacturing and intelligent operation and maintenance.

Qijia Qian is currently the Director of the Technology Department, Wenzhou Ruiming Industrial Co., Ltd., Wenzhou, China. His research interests include tool condition monitoring and intelligent manufacturing.

Anil Kumar is currently an Associate Professor with the College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou, China. His research interests include fault diagnosis and deep learning.

Weifang Sun is currently an Associate Professor with the College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou, China. His research interests include digital information analysis, and artificial intelligence methods.

Yuqing Zhou is currently a Professor with the College of Mechanical and Electrical Engineering, Jiaxing Nanhu University, Jiaxing, China, and Special-term Professor in Wenzhou University, Wenzhou, China. His research interests include tool condition monitoring, mechanical fault diagnosis, and machine learning.

1 Introduction

Milling processing is one of main processing method in mechanical manufacturing and it is a highly efficient processing method. Due to the complexity and efficiency of CNC milling, it has attracted much attention in the machining industry. Milling tools are regarded to be the main part in the milling process (Javed et al., 2018), and tool breakage is the main cause of unexpected downtime during milling. In milling process, tool failure take up 7–20% of the total downtime of the machine (Zhou et al., 2022a; Bhattacharyya et al., 2007), and the costs of tools and tool changes account for 3–12% of the total processing cost (Liu et al., 2015). It is a way to apply an effective tool condition monitoring (TCM) system to reduce downtime and maximise the life of the milling tools and it can reduce costs by 10–40% (Zhou and Xue, 2018).

There are usually direct and indirect means to monitor tool wear condition. The direct method can more accurately reveal the dimensional changes of tool. However, it is rarely feasible, cost-effective or reliable to directly detect physical damage for online monitoring (Bhattacharyya et al., 2007; He et al., 2012). It is necessary to stop processing and detect tool wear through machine vision technology (Szydlowski et al., 2016; Zhang and Zhang, 2013). With the progress of sensor technology, indirect TCM method has been paid more and more attention, which apply one or more sensors to acquire physical signals, e.g., cutting force (Zhu and Liu, 2018), acoustic emission (AE) (Bhuiyan et al., 2016), motor current (Xue et al., 2021), and vibration (Lei et al., 2020) signals, to extract features related to tool condition (Vashishtha et al., 2021; Chauhan et al., 2021). Subsequently, certain artificial intelligence (AI) method is employed to detect tool wear condition, e.g., support vector machine (SVM) (Gao et al., 2017), extreme learning machine (ELM) (Vashishtha and Kumar, 2022a; Zhou et al., 2022b), and artificial neural network (ANN) (Pal et al., 2011). Especially, with the development of big data and graphics processing unit (GPU) technology, deep learning (DL)-based TCM methods have been widely researched, such as CNNs (Cao et al., 2022) and RNNs (Zhao et al., 2018). DL algorithms can promote data-driven methods, which are more intelligent and autonomous in extracting features from a large number of monitoring datasets. Recently, researchers have begun to prefer long short-term memory (LSTM) methods to develop predictive models. The superior of LSTM is that it can forget about redundant information when analysing long-term sequences, so it does not take up too much memory and can capture features in long sequences. Cai et al. (2020) apply LSTM network to extract temporal features from datasets. Karim et al. (2018) apply the attention LSTM fully convolutional network to explore the effective of attention mechanism (AM) to increase the time series classification. Zhao et al. (2017) proposed convolutional bidirectional LSTM network to process original sensory data.

The above methods obtain good performance for TCM in certain situations, however, there are two problems still need to be improved:

1 the regression performances of these LSTM-based methods were poor when the working conditions of the testing samples differed from those of the training samples in several research studies (Kumar and Kumar, 2020; Cao et al., 2020; Kumar et al., 2021)

2 the influences of features selected in model are not clear.

Those features with little impact too TCM can be discarded to reduce the computation and improve the efficiency of online monitoring. Therefore, a new model combining LSTM and AM is proposed to improve the performance of TCM under different working conditions and obtain the importance of the features.

The study is organised in the following sections. Section 2 introduces related theory of the proposed method. Section 3 describes the framework of the proposed TCM method. Section 4 presents the experimental setup, the analysis of methods and experimental results. Finally, Section 5 summarises the article.

2 Related theory

2.1 LSTM

LSTM is a RNN with a special structure that can learn signal connections in short-term and long-term sequences. The LSTM network uses time back propagation for training, which overcomes the problem of vanishing gradients. The memory block constitutes the LSTM network and it is composed of successive layers (Zeng et al., 2019). In LSTM cells, the forget gate, input gate form a LSTM unit. The goal of each gate can be expressed in Table 1.

Table 1 Element control

Element	*Purpose*
Input gate (i)	Level of control for cell state change
Forget gate (f)	Level of control for cell state forget
Cell candidate (g)	Increase information
Output gate (o)	Level of control added to the cell state of the hidden state

A gate contain a cell state c_{t-1} in the unit block. The general architecture of LSTM method is shown in Figure 1 (Zheng et al., 2021).

The calculation of a LSTM unit can be expressed by formulas (1)–(3):

$$i_t = sigmoid\left(W_{xi}x_t + W_{hi}h_{t-1} + b_i\right) \tag{1}$$

$$f_t = sigmoid\left(W_{xf}x_t + W_{hf}h_{t-1} + b_f\right) \tag{2}$$

$$o_t = sigmoid\left(W_{xo}x_t + W_{ho}h_{t-1} + b_o\right) \tag{3}$$

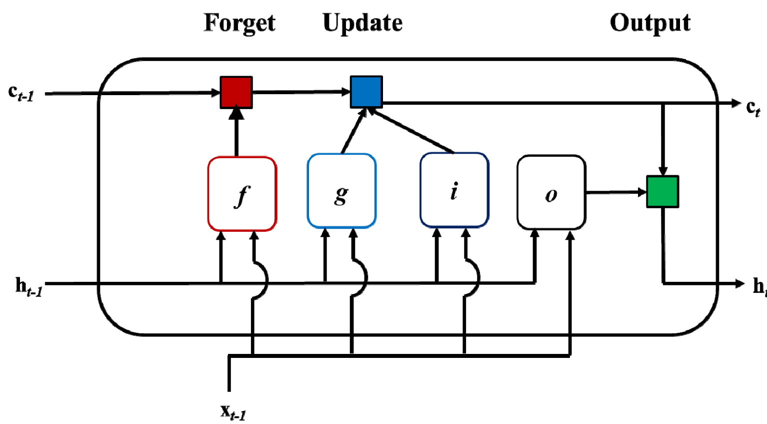
where W and B denote weight matrix and bias matrix respectively. In a LSTM unit, i denote input gate, f represent forget gate, o is output gate. The source of the information flow is the current input x_t, the hidden state h_{t-1} at the previous moment, and the cell state c_{t-1} at the previous moment. The real source of memory cell information which is the current input x_t and the hidden state h_{t-1} at the previous moment express as cell candidate in equation (4). Memory cell history information accumulation is given by equation (5):

$$g_t = \tanh\left(W_{xg}x_t + W_{hg}h_{t-1} + b_g\right) \tag{4}$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \tag{5}$$

where $\odot$ means the original vector will be multiplied by the corresponding element (element-wise).

Figure 1 LSTM unit (see online version for colours)



The accumulation of historical information does not rely on the hidden state h_t itself, but on the memory cell c_t. When the information is accumulated, the forget gate is used to limit the information of the memory cell at the previous moment, and the input gate is used to limit the new information. After updating the cell state, it is necessary to judge the state characteristics of the output cell according to the sum of the input, and pass the cell state through the tanh layer to obtain a vector with a value between –1 and 1. The vector is multiplied by the judgment condition obtained by the output gate to get the final output of this LSTM unit. The current hidden state is given by in equation (6):

$$h_t = o_t \odot \tanh(c_t) \tag{6}$$

2.2 *Attention mechanism*

The AM can extract effective information, filter out inefficient information in the signal, and eliminate redundant and irrelevant information. The attention part allocate different weights to different features to emphasise the significance of different features for prediction. LSTM can be more adapted to longer time sequence after adding the AM (Chu et al., 2017).

All time steps can be integrated into a sequence (x_1, x_2, …, x_n, …). An AM model is proposed to give prominence to the characteristics of sensitive feature analysis. In the feature sequence, an attention score β_i is distributed to every vector α_i according to the predicted significance in equations (7) and (8) (Lee et al., 2021):

$$\beta_i = \frac{\exp(\alpha_i)}{\sum_{i=1}^{k} \exp(\alpha_i)} \tag{7}$$

where

$$\alpha_i = \varphi\left(W^{(s)} h_i + b^{(s)}\right) \tag{8}$$

β is a non-standardised score that measures the quality of a vector x_n express the degree of feature.

Learnable parameters are parameters whose values are learned during the training process. With learnable parameters, we typically start out with a set of arbitrary values, and these values then get updated in an iterative fashion as the network learns. The weight matrix $W^{(s)}$ and bias term $b^{(s)}$ are learnable parameters in the network. Gloria initialisation is employed to initialise the learnable parameters. h_t is a feature sequence in the whole sequence.

The participating feature coefficients and the sequence are weighted and averaged to form a new feature sequence in equation (9):

$$\gamma = \sum_{1 \leq i \leq k} \beta_i \cdot x_i \tag{9}$$

Then, this representation is used as input to build the TCM model. In detail, the fully connected (FC) layer, the LSTM layer, and the regression function layer are combined to build a tool wear predictive model. The network of the attention module is shown in Figure 2, in which the AM assigns different weights for features to reconstruct a new feature sequence.

Figure 2 Attention module (see online version for colours)

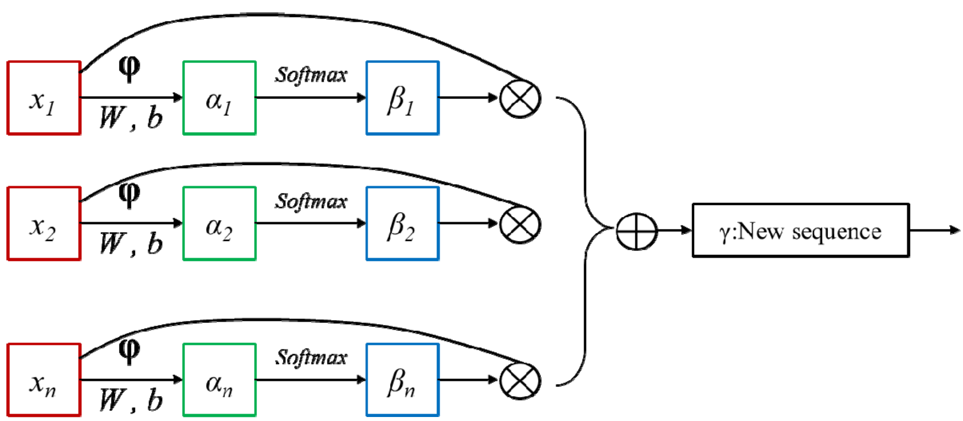
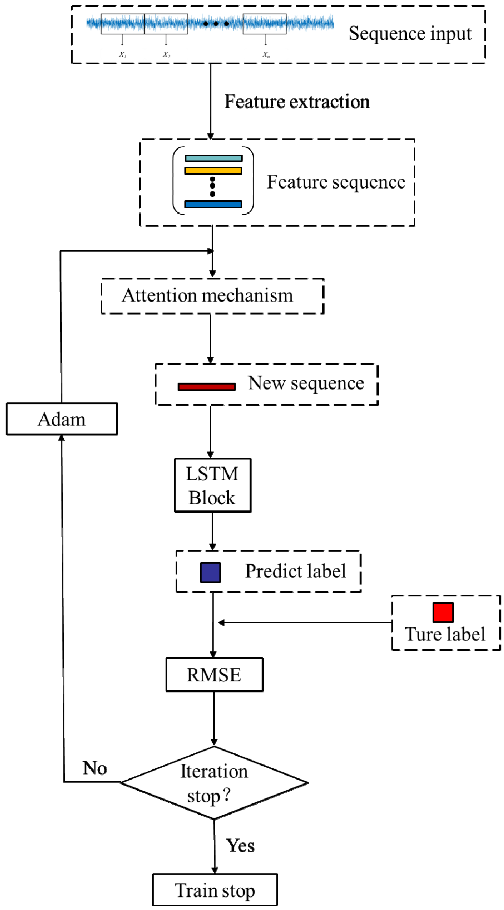


Figure 3 LSTM and attention (see online version for colours)



3 Proposed method

3.1 Framework

The proposed TCM method framework based on AM combined LSTM model is illustrated in Figure 3. Firstly, sound time series signal collected from milling process are transformed to feature space through feature extraction. New feature sequence are input iteratively into AM and LSTM network to estimate tool wear value more accurately. In the above process, feature weights are calculated and update in each iteration for improving the prediction accuracy of tool wear value.

The idea of the proposed method is utilise the most relevant parts of the input sequence in a flexible manner, by a weighted combination of all of the input vectors, with the most relevant vectors being attributed the highest weights. In this way, the importance of feature is defined as the weighted arithmetic mean of the output of LSTM, where larger weights should be set to the more relevant feature to the TCM, otherwise smaller weights should be assigned to feature not conveying useful information. Therefore, the proposed method reconstructs a new network to calculate parameters for improving the performance of TCM and obtaining the weights of selected feature parameters.

Table 2 Feature parameters

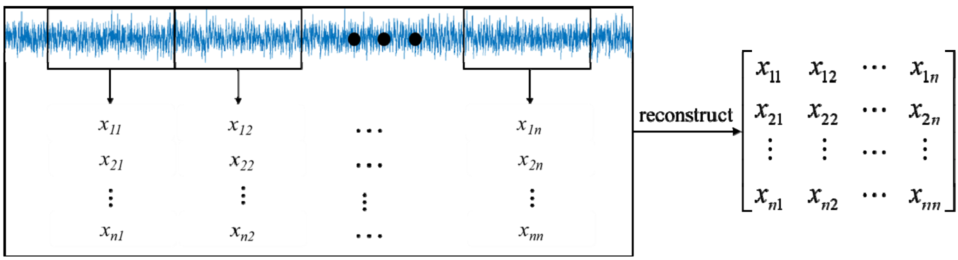
Domain	*Features*	*Formula representation*
Time domain	Root mean square (Rms)	$x_1 = \{E(x_i^2)\}^{1/2}$
	Peak value (Peak)	$x_2 = \max(\lvert x_i \rvert)$
	Crest factor	$x_3 = \dfrac{x_2}{x_1}$
	Kurtosis (Kur)	$x_4 = E\{[(\lvert x_i \rvert - \mu) / \sigma]^4\}$
	Skewness (Ske)	$x_5 = E\{[(\lvert x_i \rvert - \mu) / \sigma]^3\}$
	Square root amplitude	$x_6 = \left(\dfrac{1}{N}\sum_{i=1}^{N}\sqrt{\lvert x_i \rvert}\right)^2$
	Margin factor (L)	$x_7 = \dfrac{x_2}{x_6}$
	Average rectified value	$x_8 = \dfrac{1}{N}\sum_{i=1}^{N}\lvert x_i \rvert$
	Impulse factor	$x_9 = \dfrac{x_2}{x_8}$
	Form factor	$x_{10} = \dfrac{x_1}{x_8}$
Frequency domain	Centre of gravity frequency	$x_{11} = \dfrac{\int_0^\infty fS(f)df}{\int_0^\infty S(f)df}$
	Mean square frequency	$x_{12} = \dfrac{\int_0^\infty f^2S(f)df}{\int_0^\infty S(f)df}$
	Root mean square frequency	$x_{13} = \sqrt{x_{12}}$
	Frequency variance	$x_{14} = \dfrac{\int_0^\infty (f - x_{11})^2S(f)df}{\int_0^\infty S(f)df}$
	Frequency standard deviation	$x_{15} = \sqrt{x_{14}}$

Notes: Where x represent the input sequence signal and μ represent the mean and σ represent the standard deviation, f is the frequency and $S(f)$ is the frequency amplitude.

3.2 Feature sequence extraction

In order to improve the efficiency of feature extraction, 15 dimensional and dimensionless feature parameters in the time and frequency domains (Chauhan et al., 2021; Vashishtha and Kumar, 2022b; Zhu et al., 2021) are extracted to reduce the impact of different working conditions, shown in Table 2.

Figure 4 Features sequence reconstructed (see online version for colours)



Features are extracted from every 200 points in the original signal to construct a feature sequence matrix, shown in Figure 4. The feature sequence matrix could retain the feature trend but also reduce noise and dimension of the original signal.

4 Experimental observation

4.1 Experimental setup

The TCM experiments uses a vertical machining centre (Dalian machine tool DMTG VDL850A) to finish milling process. The material of the workpiece is 45 steels (shown in Table 3), and the experimental tools are uncoated tungsten steel end mill cutters (10 mm diameter, three-edged), which both are commonly used in industrial manufacturing. The experimental device of the monitoring system is shown in Figure 5(a).

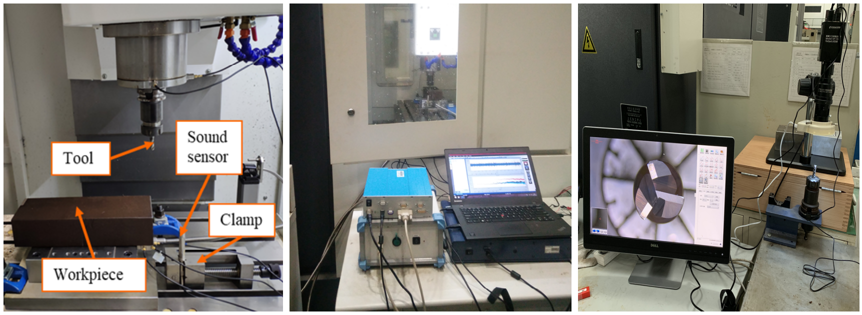
Table 3 Chemical properties of workpiece material

Carbon (C)%	Silicon (Si)%	Manganese (Mn)%	Nickel (Ni)%	Chromium (Cr)%	Copper (Cu)%
0.42-0.50	0.17-0.37	0.50-0.80	< 0.30	< 0.25	< 0.25

The sound sensor is located on the machine table and acquired by a signal acquisition device [ECON Dynamic Signal Analyzer 16 channels, shown in Figure 5(b)]. The signal obtained from the data acquisition were amplified, and then transmitted to the PC. The sampling frequency is 12 kHz. A high-definition electron microscope (GP-300C) is adopted to measure the tool wear value after finishing each milling processing [Figure 5(c)].

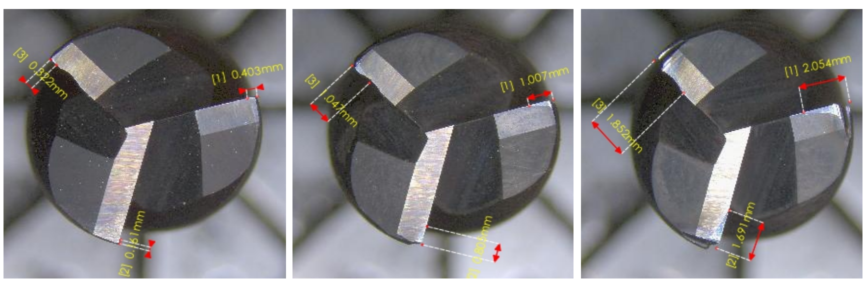
The max tool wear value of the three inserts is used as the tool wear value to evaluate the tool wear condition. Figure 6 shows the tool wear progress after 1, 5, and 10 milling on a single workpiece surface. Figure 7 shows the change of the tool wear value of one tool.

Figure 5 The experimental setup, (a) experimental platform (b) data acquisition system (c) tool microscope (see online version for colours)



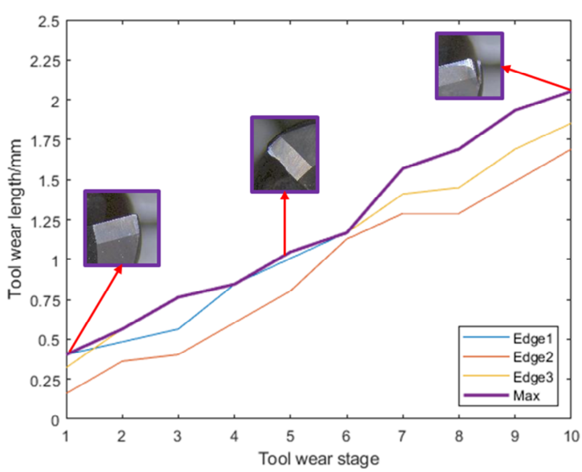
(a) (b) (c)

Figure 6 Tool wear images in different phase, (a) first milling phase (b) fifth milling phase (c) tenth milling phase (see online version for colours)



(a) (b) (c)

Figure 7 Tool wear length value of the first milling tool (see online version for colours)



According to the variable level of cutting parameters, we use the orthogonal table L9(34) to select the parameter combinations for experiments. In addition, considering the cost of the experiment, five parameter combinations were randomly selected from the remaining parameter combinations to join the experiment. Therefore, a total of 14 parameter combinations (shown in Table 4) were employed in our experiments, and each parameter combination correspond to a new tool for milling experiment.

Table 4 Experimental milling parameter setting

Test no.	*Spindle speed (rpm)*	*Cutting depth (mm)*	*Feed speed (mm/min)*	*Type*
1	2,300	0.4	400	Train
2	2,300	0.5	450	Train
3	2,300	0.6	500	Test
4	2,400	0.4	450	Train
5	2,400	0.5	500	Test
6	2,400	0.6	400	Train
7	2,500	0.4	500	Train
8	2,500	0.5	400	Test
9	2,500	0.6	450	Train
10	2,300	0.4	500	Test
11	2,300	0.6	400	Train
12	2,500	0.6	500	Train
13	2,500	0.6	400	Train
14	2,500	0.4	400	Train

In a tool wear experiment, the surface of the workpiece was cut ten times. Limited by the conditions of the experimental equipment and the cost of the experiment, 14 sets of experiments are carried out for the tool, ten sets of samples under different working conditions are selected for training, and four sets of samples for testing.

4.2 Sample division

Sound sensor signals are used in the network. Sound signals can be obtained by placing the sensor in the vicinity of the processed workpiece and the tool. It has the advantages of convenient installation, wide application range and does not damage the workpiece and the tool. There are 392 training sets and 160 test sets used in the network. There is no identical tool wear data in the training and test sets. All results of the training set are recorded in a way that records the root mean square error (RMSE).

For all architectures, the complete error gradient is calculated, and the weights are trained by using gradient descent with momentum. In all experiments, keep the same training parameters: randomly assign initial weights, and keep the training algorithm and parameters unchanged, allowing us to focus on the impact of changing the architecture. The experimental research was conducted on a computer equipped with a 2.90 GHz Intel Xeon E3 1240 v3 processor.

4.3 Model evaluation metrics

The three commonly used indicators of machine learning network regression prediction are mean absolute error (MAE), RMSE and correlation coefficient R^2. RMSE and R^2 are suitable for tool wear prediction. These metrics are calculated in equations (10), (11) and (12):

$$MAE = \frac{1}{k}\sum_{i=1}^{k}\left|y_i - \hat{y}_i\right| \tag{10}$$

$$RMSE = \sqrt{\frac{1}{k}\sum_{i=1}^{k}\left(y_i - \hat{y}_i\right)^2} \tag{11}$$

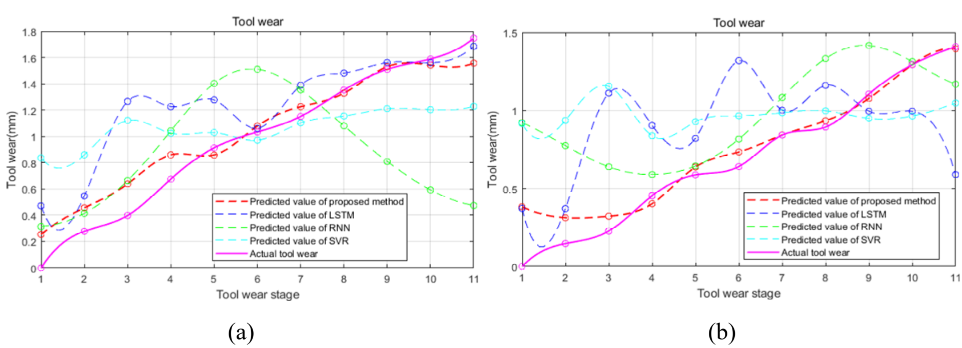
$$R^2 = 1 - \frac{\frac{1}{k}\sum_{i=1}^{k}\left(y_i - \hat{y}_i\right)^2}{\frac{1}{k}\sum_{i=1}^{k}\left(y_i - \overline{y}_i\right)^2} \tag{12}$$

where y_i is the measured value of tool wear experiment and $\hat{y}_i$ is the output value in the network, m refers to the number of samples. MAE and RMSE are related to the actual measured value, and R^2 is scale irrelevant as a dimensionless metric. When the calculated values of the MAE and RMSE metrics are smaller, the performance of the prediction model is better. For the R^2 metric, when the calculated value approaches 1, the predicted result is closer to the actual measured value.

4.4 Results and analysis

This section compares the effects of traditional machine learning models and DL models in tool wear prediction. The AM combined LSTM model established by ten cutters training sampling data is applied to predict the wear amount of another four cutter. When network train stop, 4 mean value of tool wear are calculated. To determine the run time of our program, we use a workstation with the following characteristics: 128 GB, 64-bit architecture. It executes instructions sequentially. The program runs about five minutes.

Figure 8 The prediction of tool wear, (a) no. 3 tool (b) no. 10 tool (see online version for colours)



From Figure 8, the relationship between the predicted tool value and the actual wear under the new test data and different working conditions. It can be easily got information that the trend of the overall predicted value is the same as the actual wear value in no. 3 tool and no. 10 tool, the intermediate error is smaller than the initial error and the error at some stages is less than 0.1 or even close to the wear value. To further show the advantages of AM combined LSTM model method, the traditional methods such as RNN and SVR and LSTM are also used to predict tool wear. The same training dataset and test dataset are used for training and testing to evaluate the predictive performance of all models in tool wear condition. The predictive performance of these models is evaluated using MSE and RMSE and R^2 in Tables 5–7.

Table 5 MAE result of prediction for test set

Testing tool	*MAE*			
	SVR	*RNN*	*LSTM*	*Proposed method*
3	0.3746	0.4993	0.2792	*0.1194*
5	0.4157	0.2993	0.1648	*0.1219*
8	0.3733	0.2488	0.273	*0.1095*
10	0.4343	0.3255	0.4033	*0.0838*

Table 6 RMSE result of prediction for test set

Testing tool	*RMSE*			
	SVR	*RNN*	*LSTM*	*Proposed method*
3	0.4527	0.604	0.3771	*0.1472*
5	0.474	0.4194	0.2089	*0.1377*
8	0.4877	0.3182	0.3062	*0.1303*
10	0.5211	0.4124	0.473	*0.1348*

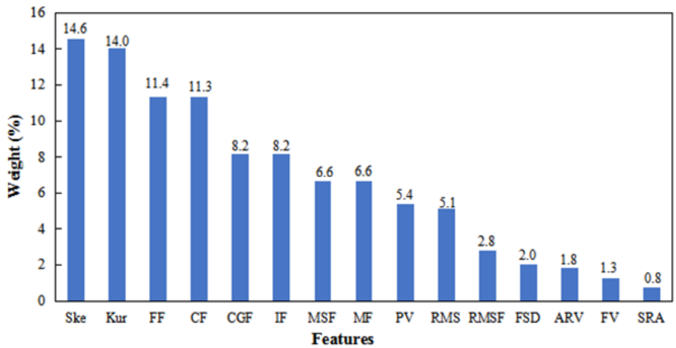
Table 7 R^2 result of prediction for test set

Testing tool	R^2			
	SVR	*RNN*	*LSTM*	*Proposed method*
3	0.3792	–0.1053	0.5692	*0.9344*
5	0.3259	0.4722	0.869	*0.9431*
8	–0.1129	0.5262	0.5612	*0.9205*
10	–0.2511	0.2163	–0.031	*0.9163*

As shown in Tables 5–7, the AM-combined LSTM (AMLSTM) model performed well on the three evaluation metrics, for the four testing tools with different working conditions, the value of MAE and RMSE are less than 0.15, and the R^2 value are above 0.9, significantly better than that of the other three methods. Therefore, the performances of the proposed AMLSTM method are wonderful and reliable under different working conditions.

Additionally, feature weights can be calculated with the AM using the training set in the network. Figure 9 shows the final normalised weights of features (in Table 2) after training. It can be found that skewness (Ske) and kurtosis (Kur) are the two most important characteristics in the extracted features for this tool wear experiment. Skewness and kurtosis can be applied to assess the asymmetry and steepness of the probability distribution of the sound time series signal respectively. In addition, the weight of square root amplitude (SRA) is the smallest, which reflects the poor performance of the tool wear.

Figure 9 Feature weights (see online version for colours)



Hence, the AMLSTM model can be used in many similar problems and the feature weights can be calculated automatically in the network. From the above analysis, it can easily be determined that the AMLSTM model results are better than those of the LSTM, SVR, and RNN in the test samples. Furthermore, it can be seen that skewness and kurtosis are two important features in the extracted features.

4.5 NASA data test

The NASA milling dataset employed for validation testing of the proposed TCM method was obtained from the Matsuura Machining Center (MC-510V) during dry rough milling processes of cast iron or stainless steel J45 workpieces using a six-tooth face milling cutter with KC710 carbide inserts under different cutting parameters (García et al., 2016).

The experimental data is collected at multiple locations by three types of sensors (AE sensors, vibration sensors, and current sensors), but no sound sensor. To verify the method proposed, the vibration signals of the spindle direction are selected. The depth of cut and feed rate of tools 4 and 12 are different from that of other tools, these two tools are selected as testing samples, and the remaining 14 tools are used as training samples.

The prediction results of three methods are shown in Table 8. By comparing the RNN and SVM, the RMSE and R^2 of the proposed method are better than that of RNN and SVM. For example, the correlation coefficient R^2 of the proposed method are above 0.85, which is at least 6% and 39% higher than that of RNN and SVM respectively.

Table 8 Prediction results of NASA dataset

Tool	*MAE*			*RMSE*			R^2		
	SVR	*RNN*	*Proposed*	*SVR*	*RNN*	*Proposed*	*SVR*	*RNN*	*Proposed*
4	0.2268	*0.0562*	0.0920	0.2678	0.0691	*0.0595*	0.4976	0.8308	*0.8920*
12	0.3693	*0.0877*	0.1042	0.4304	0.0931	*0.0613*	0.3887	0.7490	*0.8727*

5 Conclusions

In this paper, an AMLSTM model is proposed to improve the prediction performance of LSTM-based TCM method. Firstly, sound time series signal obtained from a machining process is converted into several feature sequences, and these feature sequences are input to the proposed AMLSTM model to train the weight of the feature sequences. Finally, the trained AMLSTM model with the optimal weight of the feature can be used to estimate the tool wear value. The application of the proposed method in milling TCM experiments shows that the AMLSTM-based method is significantly better than SVR-based, RNN-based, and LSTM-based methods under different working conditions. Moreover, skewness and kurtosis are two important features for TCM. The proposed method is suitable for time series sequence, but in the field of image, it is necessary to redesign or change the structure of the network to be suitable for image analysis.

Acknowledgements

This work was supported in part by the National Natural Science Foundation of China under Grant U1909217, in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LQ21E050003, and in part by the Wenzhou Science Technology Project of China under Grant ZG2021027.

References

Bhattacharyya, P., Sengupta, D. and Mukhopadhyay, S. (2007) 'Cutting force-based real-time estimation of tool wear in face milling using a combination of signal processing techniques', *Mechanical Systems and Signal Processing*, Vol. 21, No. 6, pp.2665–2683.

Bhuiyan, M.S.H., Choudhury, I.A., Dahari, M., Nukman, Y. and Dawal, S.Z. (2016) 'Application of acoustic emission sensor to investigate the frequency of tool wear and plastic deformation in tool condition monitoring', *Measurement*, Vol. 92, No. 10, pp.208–217.

Cai, W., Zhang, W., Hu, X. and Liu, Y. (2020) 'A hybrid information model based on long short-term memory network for tool condition monitoring', *Journal of Intelligent Manufacturing*, Vol. 31, No. 6, pp.1497–1510.

Cao, H., Shao, H., Zhong, X., Deng, Q., Yang, X. and Xuan, J. (2022) 'Unsupervised domain-share CNN for machine fault transfer diagnosis from steady speeds to time-varying speeds', *Journal of Manufacturing Systems*, Vol. 62, No. 1, pp.186–198.

Cao, X.C., Chen, B.Q. and Zeng, N.Y. (2020) 'A deep domain adaption model with multi-task networks for planetary gearbox fault diagnosis', *Neurocomputing*, Vol. 409, No. 10, pp.173–190.

Chauhan, S., Singh, M. and Aggarwal, A.K. (2021) 'Bearing defect identification via evolutionary algorithm with adaptive wavelet mutation strategy', *Measurement*, Vol. 179, No. 7, p.109445.

Chu, Q., Ouyang, W., Li, H., Wang, X., Liu, B. and Yu, N. (2017) 'Online multi-object tracking using CNN-based single object tracker with spatial-temporal attention mechanism', *2017 IEEE International Conference on Computer Vision (ICCV)*, pp.4846–4855.

Gao, C., Xue, W., Ren, Y. and Zhou, Y. (2017) 'Numerical control machine tool fault diagnosis using hybrid stationary subspace analysis and least squares support vector machine with a single sensor', *Applied Sciences-Basel*, Vol. 7, No. 3, pp.1–12.

García, N., Paulino, G., Esperanza, O.G. et al. (2016) 'Hybrid ABC optimized MARS-based modeling of the milling tool wear from milling run experimental data', *Materials*, Vol. 9, No. 2, pp.1–17.

He, D., Li, R. and Bechhoefer, E. (2012) 'Stochastic modeling of damage physics for mechanical component prognostics using condition indicators', *Journal of Intelligent Manufacturing*, Vol. 23, No. 2, pp.221–226.

Javed, K., Gouriveau, R., Li, X. and Zerhouni, N. (2018) 'Tool wear monitoring and prognostics challenges: a comparison of connectionist methods toward an adaptive ensemble model', *Journal of Intelligent Manufacturing*, Vol. 29, No. 8, pp.1873–1890.

Karim, F., Majumdar, S., Darabi, H. and Chen, S. (2018) 'LSTM fully convolutional networks for time series classification', *IEEE Access*, Vol. 6, No. 12, pp.1662–1669.

Kumar, A. and Kumar, R. (2020) 'Development of LDA based indicator for the detection of unbalance and misalignment at different shaft speeds', *Experimental Techniques*, Vol. 44, No. 2, pp.21–229.

Kumar, A., Vashishtha, G., Gandhi, C.P., Tang, H.S. and Xiang, J.W. (2021) 'Sparse transfer learning for identifying rotor and gear defects in the mechanical machinery', *Measurement*, Vol. 179, No. 7, p.109494.

Lee, H., Jeong, H., Koo, G., Ban, J. and Kim, S.W. (2021) 'Attention recurrent neural network-based severity estimation method for interturn short-circuit fault in permanent magnet synchronous machines', *IEEE Transactions on Industrial Electronics*, Vol. 68, No. 4, pp.3445–3453.

Lei, Z., Zhou, Y.Q., Sun, B.T. and Sun, W.F. (2020) 'An intrinsic timescale decomposition-based kernel extreme learning machine method to detect tool wear conditions in the milling process', *Int. J. Adv. Manuf. Technol.*, Vol. 106, Nos. 3–4, pp.1203–1212.

Liu, C., Wang, G.F. and Li, Z.M. (2015) 'Incremental learning for online tool condition monitoring using ellipsoid ARTMAP network model', *Applied Soft Computing*, Vol. 35, No. 10, pp.186–198.

Pal, S., Heyns, P.S., Freyer, B.H., Theron, N.J. and Pal, S.K. (2011) 'Tool wear monitoring and selection of optimum cutting conditions with progressive tool wear effect and input uncertainties', *Journal of Intelligent Manufacturing*, Vol. 22, No. 4, pp.491–504.

Szydlowski, M., Powalka, B., Matuszak, M. and Kochmanski, P. (2016) 'Machine vision micro-milling tool wear inspection by image reconstruction and light reflectance', *Precision Engineering – Journal of the International Societies for Precision Engineering and Nanotechnology*, Vol. 44, No. 4, pp.236–244.

Vashishtha, G. and Kumar, R. (2021) 'Pelton wheel bucket fault diagnosis using improved Shannon entropy and expectation maximization principal component analysis', *Journal of Vibration Engineering & Technologies*, DOI: 10.1007/s42417-021-00379-7.

Vashishtha, G. and Kumar, R. (2022a) 'Autocorrelation energy and aquila optimizer for MED filtering of sound signal to detect bearing defect in Francis turbine', *Measurement Science and Technology*, Vol. 33, No. 1, p.15006.

Vashishtha, G. and Kumar, R. (2022b) 'An amended grey wolf optimization with mutation strategy to diagnose bucket defects in Pelton wheel', *Measurement*, Vol. 187, No. 6, p.110272.

Xue, H.T., Ding, D.Y., Zhang, Z.M., Wu, M. and Wang, H.Q. (2021) 'A fuzzy system of operation safety assessment using multi-model linkage and multi-stage collaboration for in-wheel motor', *IEEE T. Fuzzy Syst.*, DOI: 10.1109/TFUZZ.2021.3052092.

Zeng, J.F., Ma, X. and Zhou, K. (2019) 'Enhancing attention-based LSTM with position context for aspect-level sentiment classification', *IEEE Access*, Vol. 7, No. 1, pp.20462–20471.

Zhang, C. and Zhang, J. (2013) 'On-line tool wear measurement for ball-end milling cutter based on machine vision', *Computers in Industry*, Vol. 64, No. 6, pp.708–719.

Zhao, R., Wang, D.Z., Yan, R.Q., Mao, K.Z., Shen, F. and Wang, J.J. (2018) 'Machine health monitoring using local feature-based gated recurrent unit networks', *IEEE Transactions on Industrial Electronics*, Vol. 65, No. 2, pp.1539–1548.

Zhao, R., Yan, R., Wang, J. and Mao, K. (2017) 'Learning to monitor machine health with convolutional bi-directional LSTM networks', *Sensors*, Vol. 17, No. 2, p.273.

Zheng, G., Sun, W., Zhang, H., Zhou, Y. and Gao, C. (2021) 'Tool wear condition monitoring in milling process based on data fusion enhanced long short-term memory network under different cutting conditions', *Eksploatacja i Niezawodnosc – Maintenance and Reliability*, Vol. 23, No. 4, pp.612–618.

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

Zhou, Y., Zhi, G., Chen, W., Qian, Q., He, D., Sun, B. and Sun, W. (2022a) 'A new tool wear condition monitoring method based on deep learning under small samples', *Measurement*, Vol. 189, No. 2, p.110622.

Zhou, Y., Sun, B., Sun, W. and Lei, Z. (2022b) 'Tool wear condition monitoring based on a two-layer angle kernel extreme learning machine using sound sensor for milling process', *Journal of Intelligent Manufacturing*, Vol. 33, No. 1, pp.247–258.

Zhu, K. and Liu, T. (2018) 'Online tool wear monitoring via hidden semi-Markov model with dependent durations', *IEEE Trans. Ind. Inf.*, Vol. 14, No. 1, pp.69–78.

Zhu, Q., Sun, B., Zhou, Y., Sun, W. and Xiang, J. (2021) 'Sample augmentation for intelligent milling tool wear condition monitoring using numerical simulation and generative adversarial network', *IEEE Transactions on Instrumentation and Measurement*, Vol. 70, No. 3, p.3516610.