International Journal of Engineering and Technology Innovation, vol. 14, no. 3, 2024, pp. 271-284

Tool Wear Prediction Based on Adaptive Feature and Temporal Attention with Long Short-Term Memory Model

Wanzhen Wang[1,2,*], Sze Song Ngu[2], Miaomiao Xin[1,2], Rong Liu[1,2], Qian Wang[2], Man Qiu[2], Shengqun Zhang[1,2]

[1]Department of Intelligent Manufacturing and Control Engineering, Qilu Institute of Technology, Jinan, China

[2]Faculty of Engineering, Universiti Malaysia Sarawak, Kota Samarahan, Sarawak, Malaysia

Received 16 February 2024; received in revised form 18 March 2024; accepted 19 March 2024

DOI: https://doi.org/10.46604/ijeti.2024.13387

Abstract

Effective monitoring of tool wear status can improve productivity and reduce losses. In previous studies, extensive feature selection was required when using the traditional machine learning method. The gating mechanism in the traditional long short-term memory (LSTM) model may incur information loss and a weaker representation of global sequential dependencies in handling long sequences. This paper aims to enhance the performance of the LSTM model in tool wear prediction by combining feature and temporal attention. Firstly, the original vibration signal is divided into sub-sequences and related features extracted. Secondly, the ability to capture global sequential dependencies using the LSTM model is improved by feature and temporal attention. Finally, a fully connected layer is used to predict tool wear values. Compared to traditional LSTM, the proposed method performs best in three evaluation metrics, RMSE, MAE, and the coefficient of determination.

Keywords: feature attention, temporal attention, tool wear, LSTM

1. Introduction

In recent years, predictive maintenance, especially predictive maintenance of cutting tools, has become an increasingly important technical aspect of intelligent manufacturing of computer numerical control (CNC) machine tools. During the machining process, the tool may gradually wear out until it fails. Moreover, chipping, chattering, and other faults may emerge subsequently. All these tool states affect the machining quality and productivity. Statistically, approximately 20% of downtime and economic losses are attributed to the cutting tool failure [1]. Monitoring tool wear state (TWS) enables tool change decisions and optimization of machining parameters, and managers can utilize it to rationalize production, improve tool efficiency, and reduce losses.

In the machining process, the machining environment is often more intricate due to the difficulty of directly measuring the TWS. In the past, relying on expert judgment unable to monitor online and experienced operators to stop and inspect consumed lots of time. Therefore, practically, developing automated, non-stop monitoring, iterative monitoring systems has become a research hotspot. Several research paths have yielded good results in tool wear monitoring. Specifically, these models can be summarized as physical-based models, data-driven models, and hybrid methods.

First, the physical-based model is based on an understanding and modeling of the physical processes of a system. In the modeling process, expert experience and domain knowledge of the tool wear are both required. For example, in Zhang et al. [2], a physical model was established to analyze milling force considering the effects of cutter runout and tool wear, and the

* Corresponding author. E-mail address: 22010329@siswa.unimas.my

English language proofreader: Chih-Wei Chang

theoretical relationship among spindle box vibration, cutting torque, and milling force was elucidated. Furthermore, the force homogenization effect of multi-tooth with tool wear is found. The revelation of wear mechanisms related to tool wear during these machining processes significantly enhances the predictive performance of the model. However, with the increasing complexity of the machining system, the physical-based methods hinder it in analyzing the tool wear accurately [3].

Many researchers utilize data-driven approaches because of their capability to handle complex and nonlinear relationships in the data without relying on explicit models. The indirect signals such as images, vibration, cutting force, acoustic emission, spindle motor power, and current collected during online machining are commonly used to explore the mapping relationship between these signals and tool degradation information. Concerning these signals, the cutting force signal and vibration are more intuitive, and some of the other signals require expensive sensors or have poor signal quality. From the perspective of an application, the trade-off between economics and monitoring performance is required.

Regarding the number of signals used, single signal, effective multi-sensor signal fusion, and multi-modal data fusion have become one of the research priorities. For example, in Twardowski et al. [4], vibration acceleration signals were collected as the input data of the machine learning model to predict cutting-edge wear, compared to the regression models, it outperforms. Online cutting force data was processed by using fractal analysis to generate signal features mainly used for assessing tool wear and effectively identifying different wear stages [5]. Audio signals were converted to images, and an ensemble convolutional neural network (CNN) was used to predict tool wear values, which demonstrates the effectiveness and accuracy across different cutting conditions [6]. The one-dimensional force signals were converted to two-dimensional image signals by the Gramian angular summation field (GASF) method, and then a modified AlexNet network was used to properly classify TWS [7].

By combining machine vision and acoustic emission (AE) signals, researchers in Chen et al. [8] constructed the mapping between the wear mount extracted by the machine vision method and the AE feature vector which is acknowledged to be practical and versatile in tool condition monitoring. Feed-motor current was measured as the input data to estimate the feed-cutting force by Li et al. [9], and such a method correlated the change in feed-cutting force with the tool wear rate, substantiating its effectiveness and industrial applicability. To attain the high precision requirement in micromachining, a single-image super-resolution approach for direct tool wear estimation in micro-milling was developed, as mentioned in Zhu et al. [10]. The reconstructed high-resolution image can be conveniently applied to wear monitoring.

Using historical data, a data-driven approach including both traditional machine learning methods and deep learning methods facilitates the mapping of data to targets and obtain. Feature engineering including feature extraction and selection should be reduced concerning the dimensionality of the data and obtain more sensitive features. In traditional machine learning methods, feature engineering can significantly determine the performance and generalization of the model. For example, in Li et al. [11], based on the clustering algorithm and support vector regression (SVR) model, a steps-ahead tool wear prediction method was proposed which showed high accuracy.

Furthermore, a method based on kernel principal component analysis and Gaussian process regression (GPR) to accurately estimate the flank wear width with good robustness was proposed. Meanwhile, a random forest model for classifying TWS in milling under varied cutting conditions was presented in Li and Lin [12], beyond such application, the random forest model with data normalization was found to be more robust to varied cutting conditions, and spindle speed had a more significant effect than feeds on classification accuracy.

In Xue et al. [13], a TWS recognition algorithm based on wavelet threshold de-noising (WTD), variational mode decomposition (VMD), manifold learning, and weighted random forest based on K-nearest neighbor (KWRF) was proposed and showed high accuracy. Regarding multi-domain feature engineering and combinatorial prediction techniques, the complexity of machine learning methods has increased. Many studies focused on the classification of the TWS and prediction

of the tool wear value using features extracted from the single signal or hybrid signals called multi-sensory signals. Feature selection methods were used to acquire more sensitive features such as the Pearson correlation coefficient [14], metrics of Mon and Rob [15], Spearman's coefficient [16], recursive feature addition [17], mutual information [18], etc.

Deep learning methods with improved models were also used to predict tool wear value or classify TWS such as the dissociation artificial neural network (Dis-ANN) [19], siamese neural network (SNN) [20], cross-domain adaptation network based on attention mechanism [21], CNN, recurrent neural networks (RNN) [22], long short-term memory (LSTM) [23], etc. Feature extraction based on original data and adaptive feature importance weight assignment provides a research path for fast acquisition of sensitive features to improve prediction performance.

In addition, to enhance the generalization capability, tool wear prediction under variable machining conditions and variable data distribution has been investigated. For example, to overcome the adaptation to physical-based models to variable processing conditions with high complexity, adaptation to variable data distributions, and over-reliance on data in data-driven models, a multi-domain mixture density network was proposed in Kim et al. [24], in which domain-invariant representations and domain-invariant features are obtained by a Bayesian learning-based feature extractor and an adversarial learning approach, respectively.

Given the performance improvement of traditional machine learning models requires complex feature filtering for dimensionality reduction, data-driven approaches are adopted. The data collected in the machining process has strong time-series characteristics, and existing LSTM models can effectively handle long-term dependencies. However, as the length of the sequence increases, the gating mechanism leads to a decrease in the efficiency of information transfer and fails to capture the correlation between the sequences globally. Consequently, a hybrid feature attention long short-term memory temporal attention (FA-LSTM-TA) model combining input feature self-attention and temporal attention is proposed to predict tool wear values, in which the mechanisms of attention improve the performance of the model prediction without the complex feature screening.

Firstly, the vibration single-channel signal is divided into sub-sequences according to certain time steps and some features are extracted in the time, frequency, and time-frequency domains of each time step. Subsequently, the input features are weighted using the input feature self-attention layer which enhances the features that are related to the tool degradation information and suppresses the irrelevant ones. The processed features are input into the two LSTM layers according to the time steps to learn the dependence between the time steps. The output features of the LSTMs are weighted by another temporal attention to adaptively learn the in-between importance. The weighted features are used to predict by a fully connected prediction layer.

The proposed model can deeply explore the complex relationship between the original signal features and the time steps to optimize the prediction performance. The main contributions of this paper are:

(1) A feature-weighted mechanism based on self-attention of input features is proposed, where the input features will adaptively enhance the features related to the tool degradation information and suppress the irrelevant features through attention, and the attention scores are learned through backpropagation, eliminating the need for complex feature engineering.

(2) A temporal attention method is proposed based on the attention of the time steps to determine the influence of the relationship between time steps on tool wear prediction. This method can mine deeper temporal dependence between multiple time steps of the signal.

Specifically, the next sections are organized as follows: Section 2 describes the basis model and proposed model. The experimental setup, the evaluation of the models, the analysis, and the discussion of the result are described in Section 3. Section 4 concludes the findings.

2. Methodology

In this section, the multi-domain feature extractions from the time domain, frequency domain, and time-frequency domain are introduced. After detailed elaboration of the feature and temporal attention, the main architecture of the proposed model is presented.

2.1. Multi-domain feature extraction

This paper uses one channel of the vibration signal from the raw data for tool wear prediction. Assumptions are a data point of the vibration signal. To reduce the data dimensionality and computational complexity, multi-domain feature extractions in the time domain, frequency domain, and time-frequency domain are performed to extract important representations of the original signal. Among them, 12 features extracted in the time domain and their computational formulas are shown in Table 1.

Table 1 Equations for features in the time domain

Feature name	Expression	Feature name	Expression
Absolute mean value (AMV)	$\frac{1}{T}\sum_{i=1}^{T}\lvert p_i\rvert$	Shape factor (Shape)	$\frac{\sqrt{\frac{1}{T}\sum_{i=1}^{T}p_i^2}}{AMV}$
Peak value (PV)	$\max(p)$	Pulse factor (PF)	$\frac{\max(p)}{AMV}$
Root mean square (RMS)	$\frac{1}{T}\sqrt{\sum_{i=1}^{T}p_i^2}$	Skewness factor (SF)	$\frac{SK}{\left(\frac{1}{T}\sum_{i=1}^{T}p_i^2\right)^3}$
Root amplitude (RA)	$\left(\frac{1}{T}\sum_{i=1}^{T}\sqrt{\lvert p_i\rvert}\right)^2$	Crest factor (CF)	$\frac{max(p)}{\sqrt{\frac{1}{T}\sum_{i=1}^{T}p_i^2}}$
Skewness (SK)	$\frac{1}{T}\sum_{i=1}^{T}\left(\lvert p_i\rvert-\frac{1}{T}\sum_{i=1}^{T}\lvert p_i\rvert\right)^3$	Clearance factor (Clf)	$\frac{max(p)}{RA}$
Kurtosis (Kur)	$\frac{1}{T}\sum_{i=1}^{T}p_i^4$	Kurtosis factor (KurF)	$\frac{T\sum_{i=1}^{T}p_i^4}{(\sum_{i=1}^{T}p_i^2)^2}$

Fast Fourier transform (FFT) is utilized to transform the time-domain data into frequency-domain data. A total of 4 features including centroid frequency, mean square frequency, root mean square frequency, and frequency variance are extracted. Using the discrete wavelet transform (DWT) with the Daubechies wavelet (db3) and a symmetric mode, the input data is decomposed into 8 frequency sub-bands up to a maximum level of 3. Regarding each sub-band, the Euclidean norm, representing the energy or magnitude of each sub-band is calculated. After the feature extraction using vibration signals, 24 features in total are obtained.

2.2. LSTM

The signals collected during machining are time series signals, and the extracted features have obvious time-dependent characteristics. For example, cutting and the vibration induced by the continuous wear of the tool are also gradually changing in the process. The machining state in the previous moment has a significant effect on the current. Technically, LSTM is a variant of RNN designed to overcome the problem of gradient vanishing and gradient explosion that RNN is prone to in processing long sequences. LSTM models are widely used in several application scenarios, e.g., natural language processing, time series prediction, etc. LSTM achieves a high degree of modularity and flexibility using memory cells and gating mechanisms, which improves the ability to capture important information in long sequences. LSTM improves the RNN by proposing the input gate, forgetting gate, and output gate mechanisms. The structure is shown in Fig. 1. The computational steps for updating an LSTM cell are shown as the following sequences.

$$\tilde{C}_t = \tanh\left(W_c \bullet \left[h_{t-1}, x_t\right] + b_i\right) \tag{1}$$

$$i_t = \sigma\left(W_t \bullet \left[h_{t-1}, x_t\right] + b_i\right) \tag{2}$$

$$f_t = \sigma\left(W_f \bullet \left[h_{t-1}, x_t\right] + b_f\right) \tag{3}$$

$$o_t = \sigma\left(W_o \bullet \left[h_{t-1}, x_t\right] + b_o\right) \tag{4}$$

$$h_t = o_t \otimes \tanh\left(C_t\right) \tag{5}$$

where x_t, h_{t-1}, and C_t denote the input data, the hidden state, and the cell state of the LSTM cell at time step t, respectively. W and b denote the weights and biases, respectively. σ denotes the sigmoid function.

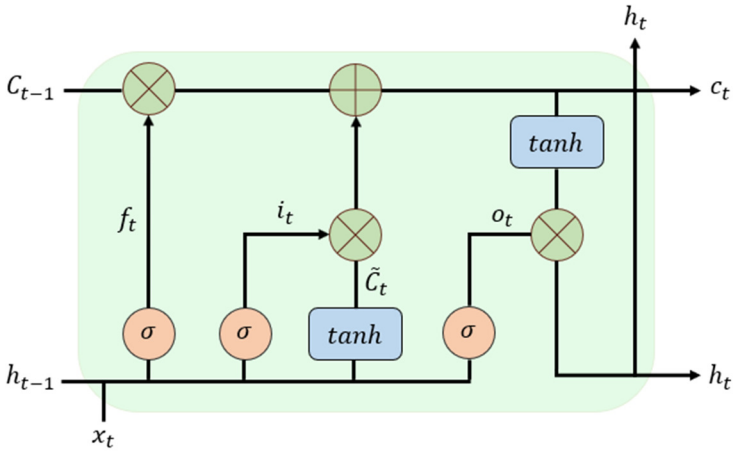


Fig. 1 The gating mechanisms of the LSTM cell

2.3. *Feature attention mechanism*

Despite the advantages of processing long sequence dependencies, the computational efficiency and the forgetting of important historical information or the disappearance of the gradient can incur the deterioration of global dependency capture. In Vaswani et al. [25], the self-attention mechanism was introduced, in which sequences are processed by the self-attention mechanism. Each sequence element adaptively calculates the correlation with the rest of the sequences, which achieves the goal of considering the information at all positions sequentially to achieve global capture of important information. Self-attention dynamically regulates the relationship of each position sequentially, and LSTM can dynamically control the flow of information through the gating mechanism. Consequently, the combination can achieve more flexible information interaction and flow in the model, thus better adapting to different sequence data characteristics.

For an input sequence, to capture the dependencies between the elements of the sequence, the degree of association between those elements is expressed by assigning a weight to each element of the input sequence. Postulate the kth input feature sequence is given as $x^k = (x_1^k, x_2^k, \cdots, x_T^k)^T \in R^T$, Q, K, and V are obtained through the learnable W_q, W_k, as presented in,

$$Q_n = W_q \bullet x \tag{6}$$

$$K_n = W_k \bullet x \tag{7}$$

$$V_n = W_v \bullet x \tag{8}$$

where Q denotes the set of features to be queried for calculating the similarity sequentially scores features with other locations. K denotes the set of references for comparing the query with the features of other locations. V contains the relevant information to be queried.

The similarity matrix f within the features is obtained by the dot product operation shown in:

$$f = K_n^T \bullet Q_n \in R^{n \times n} \tag{9}$$

The attention matrix α is obtained by normalizing each column, as shown in:

$$\alpha = ColumnsSoftmax(f) \tag{10}$$

Attention scores $\tilde{x}^t$ are obtained by weighted summation of values through the attention matrix below the formula and then are used as the input of LSTM layers.

$$\tilde{x}^t = V_n \alpha \tag{11}$$

Overall, the entire mechanism of feature self-attention can be represented as the architecture in Fig. 2.

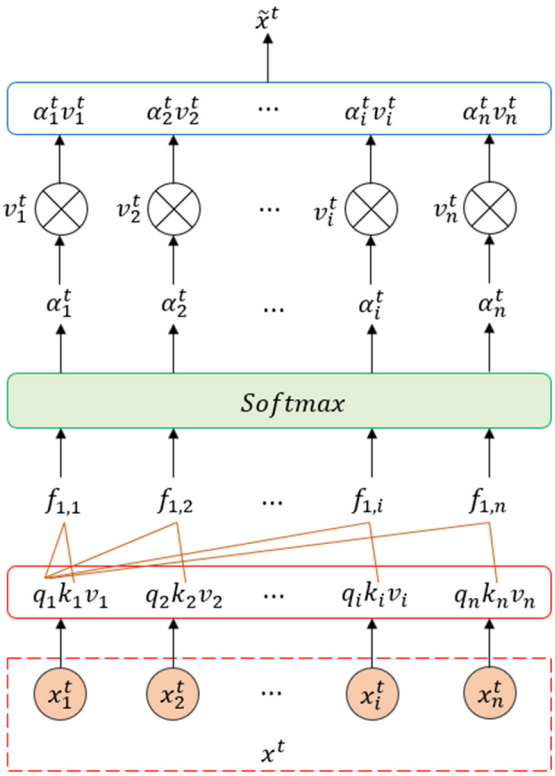


Fig. 2 Feature attention mechanism

2.4. *Temporal attention mechanism*

The addition of the self-attention layer to LSTM can further enhance the ability to characterize features at each time step sequentially. By combining the sequence modeling capability of LSTM with the global interaction capability of the self-attention mechanism, more representational and semantically rich feature representations can be generated, which improves the performance of the model in predicting tool wear using sequence input.

Like feature attention, for the kth sample of the input data, each sample has T time steps, and each time step has n features, The output $h' = (h'_1, h'_2, \cdots, h'_T)$ after LSTM training is subjected to self-attention calculation as depicted below,

$$Q_t = W'_q \bullet h' \tag{12}$$

$$K_t = W_k \bullet h' \tag{13}$$

$$V_t = W_v \bullet h \tag{14}$$

$$l = W_t^T \bullet Q_t \tag{15}$$

$$\beta = ColumnsSoftmax(l) \tag{16}$$

$$output = V_t \beta \tag{17}$$

The output data is used for tool wear prediction using the fully connected layer. Overall, the architecture of the temporal self-attention layer is shown in Fig. 3. The architecture of the whole model is shown in Fig. 4, where BN, dropout layers, and the regularisation term terms are applied to prevent overfitting.

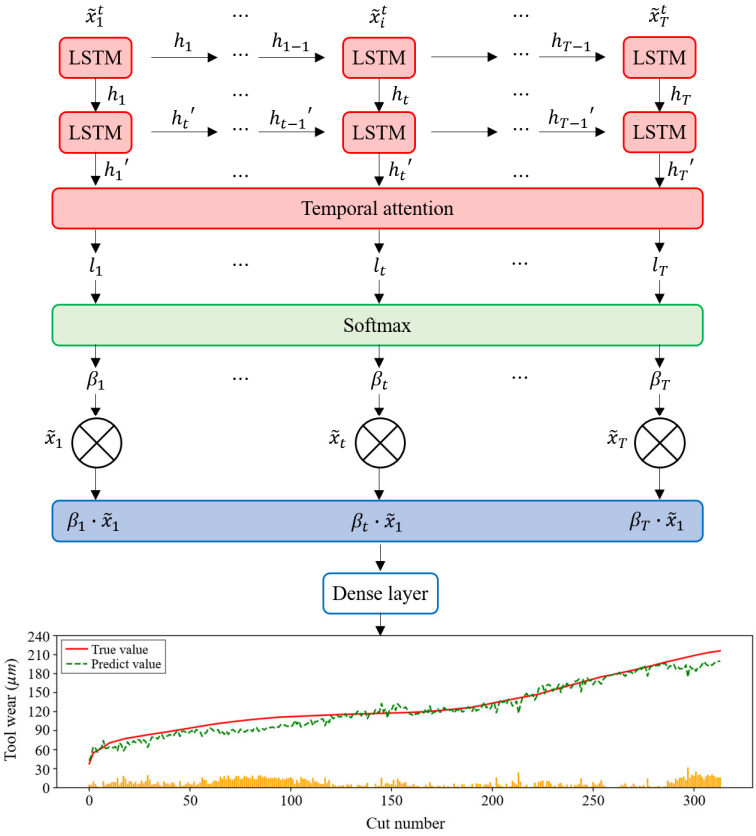


Fig. 3 Temporal attention mechanism and the dense layer

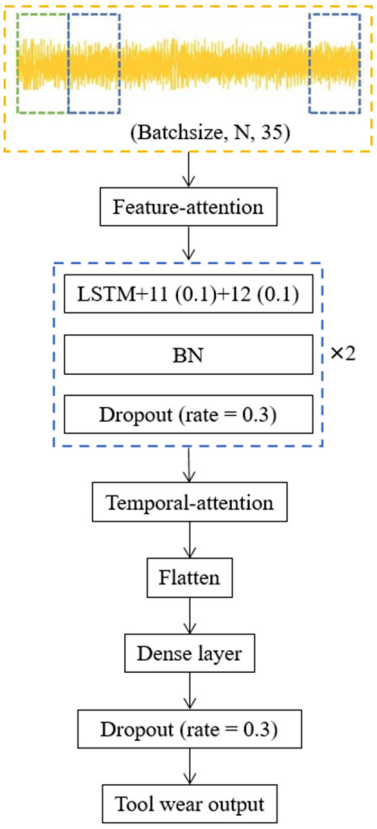


Fig. 4 The architecture of the proposed model

3. Experimental Results

In this section, details of the data collection are described by introducing the machining environment and conditions. The data was processed according to the different models compared, and the dataset was divided into training, validation, and testing sets. Subsequently, the SVR, the traditional LSTM, and the proposed model are trained. Finally, the results of the model evaluation are compared and analyzed.

3.1. Data description

The dataset used for this experiment is the PHM2010 competition dataset [26]. Apart from the availability of validating classification tasks in TWS monitoring, the dataset is more suitable for regression tasks in tool wear values prediction due to the relatively larger number of samples. Multi-signals were collected using sensors when milling with the CNC machine center. The experimental platform and main hardware components are shown in Fig. 5. In the dataset, milling data from six tools were collected with the same machining conditions. Among the tools, tools named C1, C4, and C6 were measured with 3 flank wear values after each machining process, while C2, C3, and C5 have no measured wear values. Each tool was machined 315 times and contained force signals, vibration signals with 3 channels named x, y, z, and AE signals with only 1 channel.

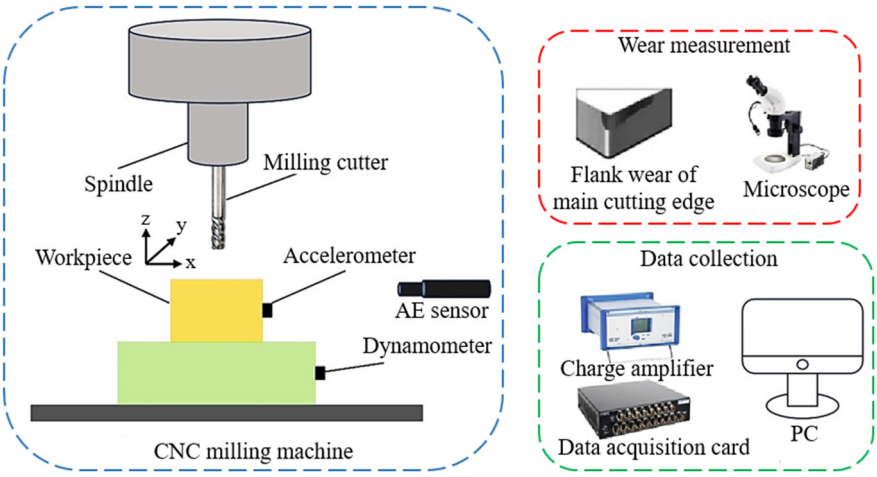


Fig. 5 Experimental platform and main hardware components

3.2. Data processing

A single channel of data in the *x*-direction of the vibration signal was used for the experiment. Two-by-two combinations of 3 tools with labeled values were used as the training set and the rest as the test set, i.e., 630 sets of samples were used as the training set, and 315 sets of data were used as the test set. The training set is again used as the validation set according to the separation rate of 0.2. The detailed division of the dataset used is listed in Table 2.

Table 2 Training, validation, and testing datasets

Training set	Training data size	Validation data size	Test set	Test data size
C1+C4	504	126	C6	315
C1+C6	504	126	C4	315
C4+C6	504	126	C1	315

To accommodate different model inputs like SVR, a traditional machine learning method, feature extraction is performed on the full vibration signal of the original sample without dividing the subsequence. In the LSTM and the proposed method, the vibration signals will be divided into sub-sequences according to a non-overlapping sliding window, after which the features are extracted on the sub-sequences separately as shown in Fig. 6.

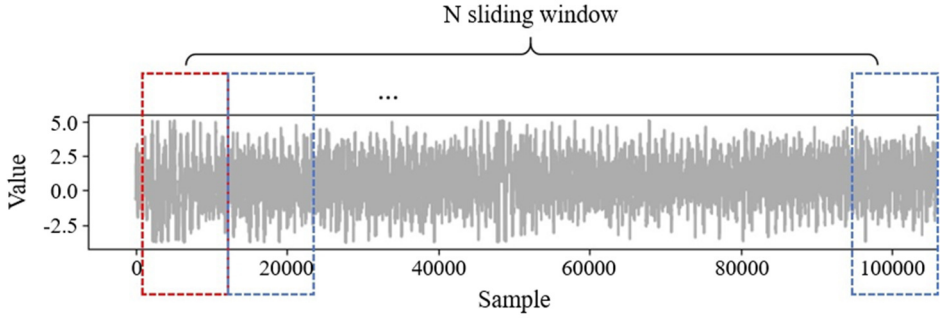


Fig. 6 The sliding window method

Assuming the number of sub-sequences is *N*, for each *N* according to the multi-domain feature extraction method, feature extraction is performed from three domains including time, frequency, and time-frequency domain. A total of 24 features are extracted in every sub-sequence. The dimension of each data is (315, N, 24) and the features are normalized as:

$$x_{norm} = \frac{x_i - \mu}{\sigma} \tag{18}$$

where x_i and x_{norm} denote the original feature value and the normalized feature value, μ and σ denote the average and standard deviation of the trained data. For validate and test data the same parameters are used for normalization.

According to ISO8688-2:1989 [27], VB characterizes the tool wear. For the aforementioned tool wear values, three wear values from C1, C4, and C6 were collected. To reduce the errors in the acquisition process, the average value is used as the final tool wear value label. The four tool wear values from C1, C4, and C6 are shown in Fig. 7, Fig. 8, and Fig. 9, respectively.

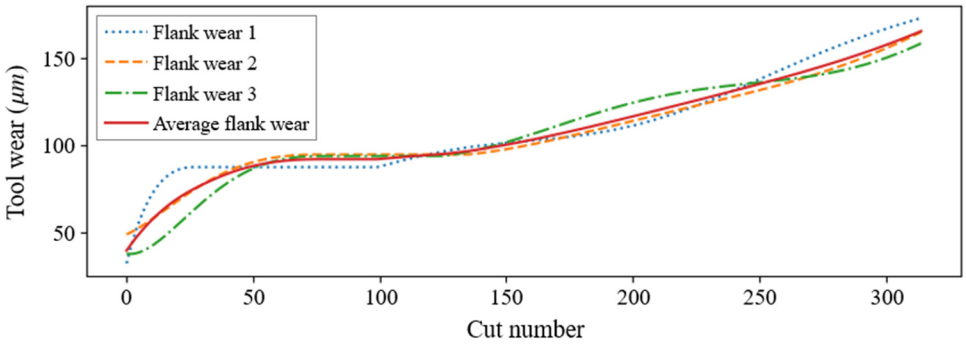


Fig. 7 Three tool wear and average value from C1

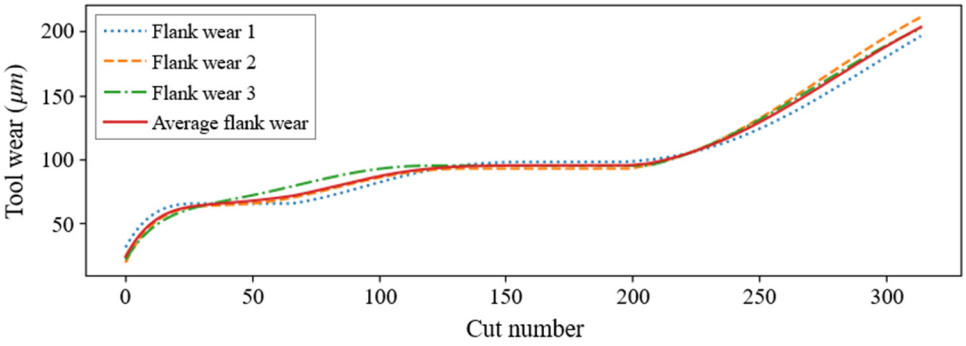


Fig. 8 Three tool wear and average value from C4

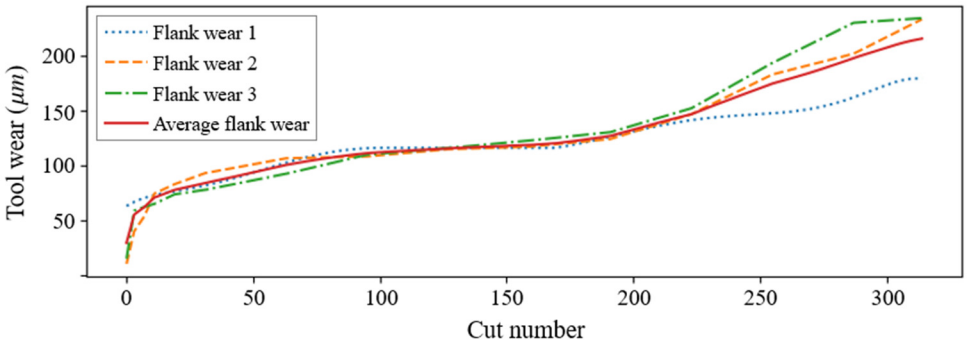


Fig. 9 Three tool wear and average value from C6

To improve the training efficiency, the tool wear value labels are processed accordingly, and the labels of the training data, validation data, and test data are scaled to within 0-1 using the formula depicted in:

$$y_{norm} = \frac{y_i}{MAX\left(y_i\right)} \tag{19}$$

where y_i and y_{norm} denote the original tool wear values of trained data. *MAX* denotes the maximum values of tool wear in trained data. When test results are obtained, the predicted value needs to be multiplied by the maximum value of the training labels to obtain the final predicted value.

3.3. Model training and testing

A small batch size is used to train the model, while the batch size is 16. Nadam is used as the optimizer of the model. Root mean square error (RMSE), mean absolute error (MAE), and R-square (R^2) are used to quantitatively estimate the performance of models. The calculation formulas of the criteria are expressed in,

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}_i - y_i\right)^2} \tag{20}$$

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|\hat{y}_i - y_i\right| \tag{21}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m}\left(\hat{y}_i - y_i\right)}{\sum_{i=1}^{m}\left(\overline{y} - y_i\right)} \tag{22}$$

where $\hat{y}_i$ and y_i denote the ith real and prediction value. $\bar{y}$ stands for the average value of the real tool wear values. m expresses the total number of test labels.

To verify the effectiveness of the proposed tool wear prediction method and find the best number of sub-sequences, comparative experiments were designed. It is noteworthy that selecting the number of sub-sequences is crucial. As the original data has more than 200,000 points in every sample, more sub-sequences yield more time steps which hinders the training process, i.e., fewer sub-sequences will lose some information related to tool wear. Therefore, to choose the optimal sub-sequence number, different sub-sequences divided by the moved window method with no overlap are used to train the model.

3.4. Analysis of the result

A genetic algorithm (GA) is used to select optimal hyper-parameters for the unit number of the two LSTM layers, the unit number of the connected layer, batch size, and epochs. For instance, the final optimal hyper-parameters selected when estimating tool C1 are shown in Table 3.

Table 3 Initial and optimal parameters of GA

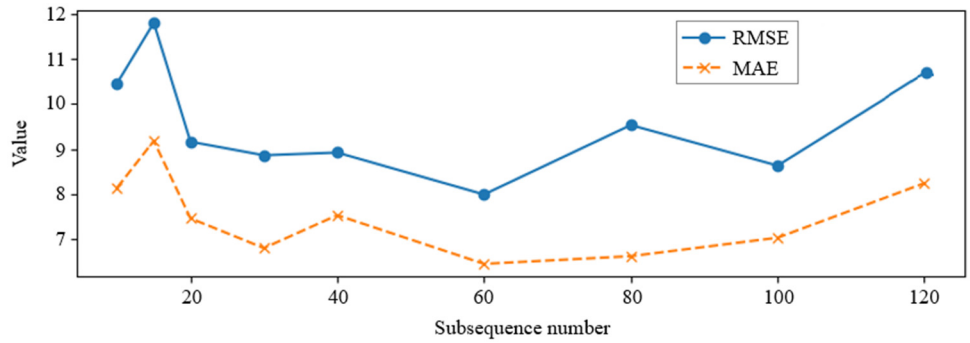
Parameter	Initial range	Optimal value
Unites number of LSTM1	32-128	53
Unites number of LSTM2	32-128	49
Unite the number of the dense layer	5-128	110
Batch size	8, 16, 32, 64, 128	16
Epoch	10, 20, 30, 50,100	20

Table 4 The result of C1, C4, and C6 with different sub-sequences

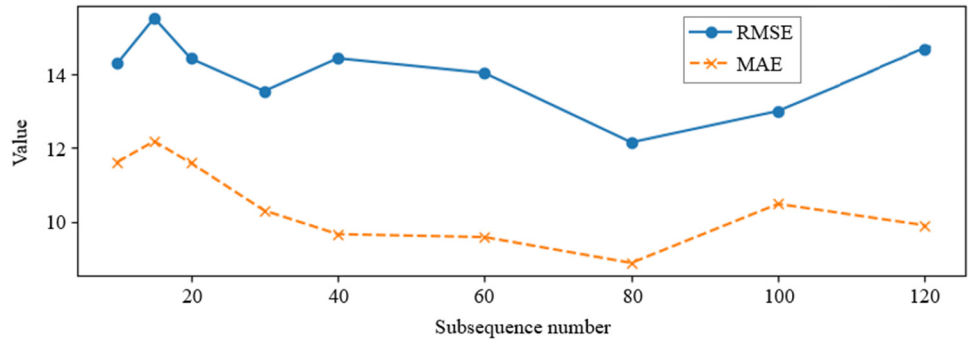
Number of C1	RMSE	MAE	R^2	Number of C4	RMSE	MAE	R^2	Number of C6	RMSE	MAE	R^2
C1-10	10.46	8.14	0.851	C4-10	14.30	11.62	0.856	C6-10	16.52	13.63	0.827
C1-15	11.80	9.18	0.810	C4-15	15.52	12.18	0.831	C6-15	13.27	11.27	0.888
C1-20	9.16	7.46	0.885	C4-20	14.42	11.59	0.854	C6-20	13.90	10.40	0.878
C1-30	8.86	6.80	0.893	C4-30	13.54	10.30	0.871	C6-30	**9.44**	**7.39**	**0.944**
C1-40	8.92	7.53	0.891	C4-40	14.43	9.66	0.854	C6-40	16.89	13.35	0.819
C1-60	**7.99**	**6.45**	**0.913**	C4-60	14.03	9.58	0.862	C6-60	16.95	11.37	0.818
C1-80	9.53	6.62	0.876	C4-80	**12.15**	**8.88**	**0.896**	C6-80	18.84	13.03	0.775
C1-100	8.63	7.03	0.898	C4-100	13.00	10.48	0.881	C6-100	16.82	13.06	0.821
C1-120	10.66	8.24	0.845	C4-120	14.67	9.90	0.849	C6-120	21.60	17.83	0.705

A total of 9 from 10, 15, 20, 30, 40, 60, 80, 100 to 120 sub-sequences are tested. RMSE, MAE, and R-square of the model with different numbers of sub-sequences are shown in Table 4, and the trends are illustrated in Fig. 10. It can be interpreted if the number of sub-sequences is designed to be excessively small or large, the model fails to obtain high performance. Training LSTM with a low number of sub-sequences indicates, in other words, fewer time steps, and the time-dependent information of features between individual sub-sequences is insufficiently mined, leading to insufficient mapping results between features and

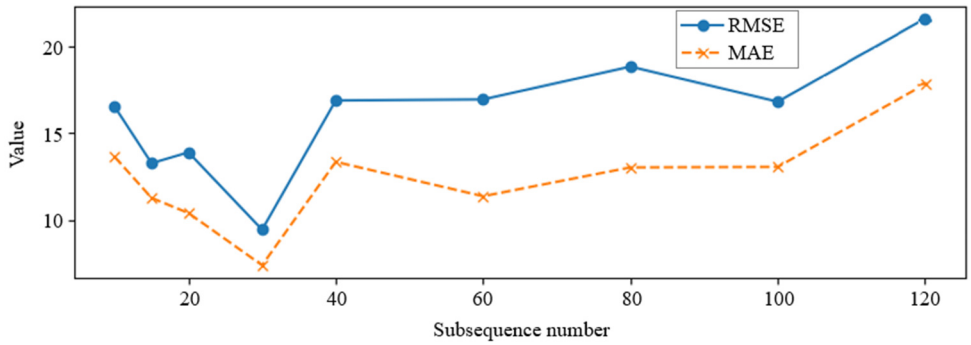
tool degradation information. When training with a higher number of sub-sequences, LSTM models can fully utilize the dependencies between time-step features to achieve better prediction performance. However, excessive time steps can incur more time consumption for prediction. The optimal number of sub-sequences for C1, C4, and C6 is 60, 80, and 30, respectively.



(a) Test result of C1 with different sub-sequences



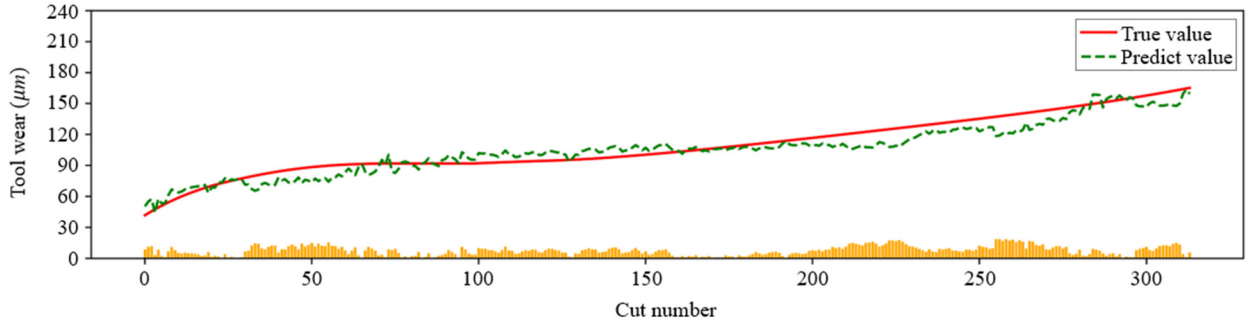
(b) Test result of C4 with different sub-sequences



(c) Test result of C6 with different sub-sequences

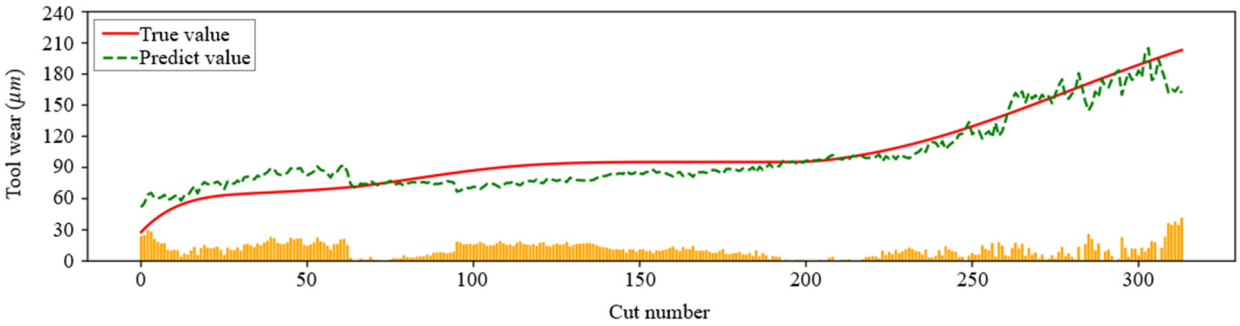
Fig. 10 Test result from C1, C4, and C6 with different sub-sequences

All the comparison results from different models including SVR, traditional LSTM, and the proposed method are shown in Table 5 and Fig. 11 including true, predicted values, and absolute errors. The prediction results of tool wear show that the model can accurately capture the trend of tool wear, and meanwhile the overall error is small.

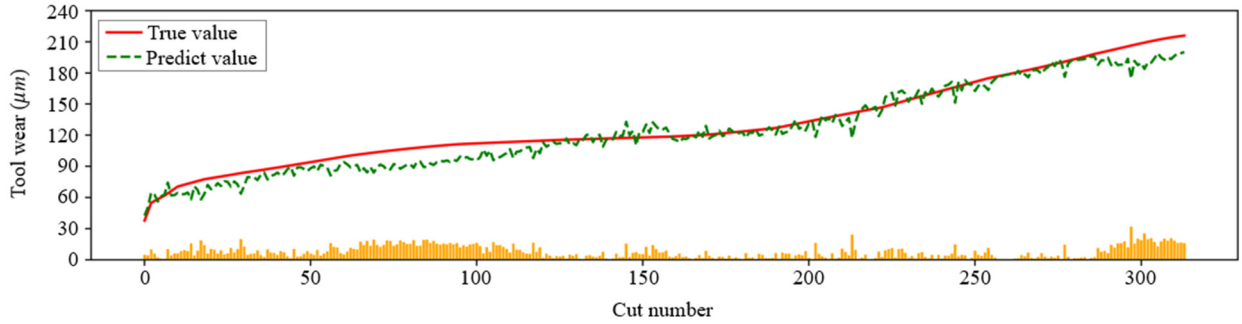


(a) Test result of C1

Fig. 11 Test result of the proposed model for C1, C4, and C6



(b) Test result of C4



(c) Test result of C6

Fig. 11 Test result of the proposed model for C1, C4, and C6 (continued)

Table 5 The result of C1, C4, and C6 with different models

Test tool and metrics		SVR	LSTM	FA-LSTM-TA
C1	RMSE	14.89	9.78	**7.99**
	MAE	11.00	7.52	**6.45**
	R^2	0.698	0.869	**0.913**
C4	RMSE	19.55	14.05	**12.15**
	MAE	16.36	10.19	**8.88**
	R^2	0.731	0.861	**0.896**
C6	RMSE	23.43	14.53	**9.44**
	MAE	18.53	12.67	**7.39**
	R^2	0.658	0.862	**0.944**

Given the values depicted in Table 5, it is noteworthy that SVR outperformed the highest RMSE and MAE among the deep learning models herein. However, the R-square of SVR stands lower. Specifically, the SVR performs well on linearly divisible data or with fewer features presented, and the data in the machining process has characteristics such as non-smoothness and non-linearity, which leads to a less effective prediction than LSTM. Compared to the SVR, the traditional LSTM model reduces the RMSE in C1, C4, and C6 by 34.32%, 28.13, and 37.99%, respectively. MAE was reduced by 31.64%, 37.71%, and 31.62%, respectively.

LSTM is a type of RNN designed for modeling sequential data, capable of capturing long-term dependencies within the data. LSTM possesses powerful nonlinear modeling capabilities, enabling it to learn complex data patterns and relationships compared with SVR. The proposed method obtains the best performance with RMSE and MAE metrics on C1, C4, and C6 improved by 18.30%, 13.52%, 35.03% and 14.23%, 12.86%, 41.67% over the conventional LSTM, respectively. The attention mechanism herein enables the model to accentuate different elements within an input sequence. Given the use of feature attention and temporal attention in LSTM, the proposed method can further enhance global sequential dependencies which improves the efficiency of information transfer in the LSTM gating mechanism.

According to the result of the proposed method in Fig. 11, the prediction of the initial and late phases of wear is less satisfactory and is particularly poorly predicted at the boundary. In these two phases, moreover, tools wear faster, and the indirect signals become complex and unstable, hindering mining complex tool degradation information using only a single signal.

Compared to the larger number of smoother samples in the intermediate stable wear phase, the number of samples in both initial and late phases is smaller, and signals show large fluctuations. Given such aspects, further investigation is needed concerning the solution of unbalanced samples.

4. Conclusions

This paper proposes a new method combining feature and temporal attention to enhance the capabilities of LSTM in extracting global sequential dependencies and predicting flank wear. The proposed model is validated using the prognostics and health management (PHM) competition dataset with the following conclusions.

(1) Single-channel subsequence division helps to improve the LSTM model prediction performance, a lower number of sub-sequences cannot deeply dig into the temporal dependence between the features of each time step incurring poor prediction. Instead, a higher number of sub-sequences can improve the performance of the model prediction.

(2) Traditional LSTM models outperform SVR because of the strong ability to capture long-term dependencies between sequences using data with time series characteristics, e.g., non-smoothness and non-linearity. After increasing feature attention and temporal attention, the proposed method obtains the best performance with RMSE and MAE metrics on C1, C4, and C6 improved by 18.30%, 13.52%, 35.03% and 14.23%, 12.86%, 41.67% over the traditional LSTM, respectively. The proposed method performs well due to the enhancement of the global sequential correlation by self-attention, which improves the delivery efficiency of the LSTM gating mechanism and the capture of temporal dependencies in long sequences.

However, a limitation is delineated herein, i.e., the real machining environment is variable, and the distribution between the training data and the actual predicted data obtained may change. The data from the PHM dataset was obtained using the same processing method under one processing condition, i.e., the ability of the model to generalize might be constrained under the variation of different machining conditions. Hence, transfer learning, domain adaption, uncertainty quantification, digital twins, and methods that combine physical-based and data-driven are perceived to be investigated in the future.

Acknowledgment

This work is supported by the research program of Qilu Institute of Technology (No.: QIT23NN043 and No.: QIT22NN007). This work is also supervised by professors from Universiti Malaysia Sarawak.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] D. Kong, Y. Chen, and N. Li, "Gaussian Process Regression for Tool Wear Prediction," Mechanical Systems and Signal Processing, vol. 104, pp. 556-574, May 2018.

[2] X. Zhang, Y. Gao, Z. Guo, W. Zhang, J. Yin, and W. Zhao, "Physical Model-Based Tool Wear and Breakage Monitoring in Milling Process," Mechanical Systems and Signal Processing, vol. 184, article no. 109641, February 2023.

[3] G. Wang and F. Zhang, "A Sequence-to-Sequence Model with Attention and Monotonicity Loss for Tool Wear Monitoring and Prediction," IEEE Transactions on Instrumentation and Measurement, vol. 70, article no. 3525611, 2021.

[4] P. Twardowski, J. Czyżycki, A. Felusiak-Czyryca, M. Tabaszewski, and M. Wiciak-Pikuła, "Monitoring and Forecasting of Tool Wear Based on Measurements of Vibration Accelerations During Cast Iron Milling," Journal of Manufacturing Processes, vol. 95, pp. 342-350, June 2023.

[5] M. Jamshidi, X. Rimpault, M. Balazinski, and J. F. Chatelain, "Fractal Analysis Implementation for Tool Wear Monitoring Based on Cutting Force Signals During CFRP/Titanium Stack Machining," The International Journal of Advanced Manufacturing Technology, vol. 106, no. 9-10, pp. 3859-3868, February 2020.
[6] Z. Li, X. Liu, A. Incecik, M. K. Gupta, G. M. Królczyk, and P. Gardoni, "A Novel Ensemble Deep Learning Model for Cutting Tool Wear Monitoring Using Audio Sensors," Journal of Manufacturing Processes, vol. 79, pp. 233-249, July 2022.
[7] X. Zhou, T. Yu, G. Wang, R. Guo, Y. Fu, Y. Sun, et al., "Tool Wear Classification Based on Convolutional Neural Network and Time Series Images During High Precision Turning of Copper," Wear, vol. 522, article no. 204692, June 2023.
[8] M. Chen, M. Li, L. Zhao, and J. Liu, "Tool Wear Monitoring Based on the Combination of Machine Vision and Acoustic Emission," The International Journal of Advanced Manufacturing Technology, vol. 125, no. 7-8, pp. 3881-3897, April 2023.
[9] X. Li, A. Djordjevich, and P. K. Venuvinod, "Current-Sensor-Based Feed Cutting Force Intelligent Estimation and Tool Wear Condition Monitoring," IEEE Transactions on Industrial Electronics, vol. 47, no. 3, pp. 697-702, June 2000.
[10] K. Zhu, H. Guo, S. Li, and X. Lin, "Online Tool Wear Monitoring by Super-Resolution Based Machine Vision," Computers in Industry, vol. 144, article no. 103782, January 2023.
[11] Y. Li, X. Huang, J. Tang, S. Li, and P. Ding, "A Steps-Ahead Tool Wear Prediction Method Based on Support Vector Regression and Particle Filtering," Measurement, vol. 218, article no. 113237, August 2023.
[12] K. M. Li and Y. Y. Lin, "Tool Wear Classification in Milling for Varied Cutting Conditions: With Emphasis on Data Pre-Processing," The International Journal of Advanced Manufacturing Technology, vol. 125, no. 1-2, pp. 341-355, March 2023.
[13] Z. Xue, L. Li, Y. Wu, Y. Yang, W. Wu, Y. Zou, et al., "Study on Tool Wear State Recognition Algorithm Based on Spindle Vibration Signals Collected by Homemade Tool Condition Monitoring Ring," Measurement, vol. 223, article no. 113787, December 2023.
[14] A. Colpani, A. Fiorentino, E. Ceretti, and A. Attanasio, "Tool Wear Analysis in Micromilling of Titanium Alloy," Precision Engineering, vol. 57, pp. 83-94, May 2019.
[15] B. Zhang, L. Zhang, and J. Xu, "Degradation Feature Selection for Remaining Useful Life Prediction of Rolling Element Bearings," Quality and Reliability Engineering International, vol. 32, no. 2, pp. 547-554, March 2016.
[16] L. Colantonio, L. Equeter, P. Dehombreux, and F. Ducobu, "Comparison of Cutting Tool Wear Classification Performance with Artificial Intelligence Techniques," Materials Research Proceedings, vol. 28, pp. 1265-1274, 2023.
[17] V. Warke, S. Kumar, A. Bongale, and K. Kotecha, "Robust Tool Wear Prediction Using Multi-Sensor Fusion and Time-Domain Features for the Milling Process Using Instance-Based Domain Adaptation," Knowledge-Based Systems, vol. 288, article no. 111454, March 2024.
[18] M. Hu, W. Ming, Q. An, and M. Chen, "Tool Wear Monitoring in Milling of Titanium Alloy Ti–6Al–4 V Under MQL Conditions Based on a New Tool Wear Categorization Method," The International Journal of Advanced Manufacturing Technology, vol. 104, no. 9-12, pp. 4117-4128, October 2019.
[19] S. Y. Wong, J. H. Chuah, H. J. Yap, and C. F. Tan, "Dissociation Artificial Neural Network for Tool Wear Estimation in CNC Milling," The International Journal of Advanced Manufacturing Technology, vol. 125, no. 1-2, pp. 887-901, March 2023.
[20] J. Duan, J. Liang, X. Yu, Y. Si, X. Zhan, and T. Shi, "Toward Practical Tool Wear Prediction Paradigm with Optimized Regressive Siamese Neural Network," Advanced Engineering Informatics, vol. 58, article no. 102200, October 2023.
[21] J. He, Y. Sun, C. Yin, Y. He, and Y. Wang, "Cross-Domain Adaptation Network Based on Attention Mechanism for Tool Wear Prediction," Journal of Intelligent Manufacturing, vol. 34, no. 8, pp. 3365-3387, December 2023.
[22] J. Duan, X. Zhang, and T. Shi, "A Hybrid Attention-Based Paralleled Deep Learning Model for Tool Wear Prediction," Expert Systems with Applications, vol. 211, article no. 118548, January 2023.
[23] W. Yu, H. Huang, R. Guo, and P. Yang, "Tool Wear Prediction Based on Attention Long Short-Term Memory Network with Small Samples," Sensors and Materials, vol. 35, no. 7(2), pp. 2321-2335, 2023.
[24] G. Kim, S. M. Yang, S. Kim, D. Y. Kim, J. G. Choi, H. W. Park, et al., "A Multi-Domain Mixture Density Network for Tool Wear Prediction Under Multiple Machining Conditions," International Journal of Production Research, in press. https://doi.org/10.1080/00207543.2023.2289076.
[25] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., "Attention Is All You Need," https://doi.org/10.48550/arXiv.1706.03762, June 12, 2017.
[26] "2010 PHM Society Conference Data Challenge," https://phmsociety.org/phm_competition/2010-phm-society-conference-data-challenge, February 16, 2024.
[27] Tool Life Testing in Milling— Part 2: End Milling, ISO 8688-2, 1989.