Tool Wear Prediction Combining Global Feature Attention and Long Short-Term Memory Network

Wanzhen Wang[1], Sze Song Ngu[1,*], Miaomiao Xin[2], Xiaomei Ni[2], Beibei Kong[2], Kaiyuan Wu[2], Ruyue Han[2]

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

[2]School of Intelligent Manufacturing and Control Engineering, Qilu Institute of Technology, Jinan, China

Received 28 August 2024; received in revised form 04 September 2024; accepted 05 September 2024

DOI: https://doi.org/10.46604/peti.2024.14201

Abstract

This study aims to accurately predict tool flank wear in milling and simplify the complexity of feature selection. A hybrid approach is proposed to eclectically integrate the advantages between the long short-term memory (LSTM) network and the global feature attention (GFA) module. First, the feature matrix is calculated using the multi-domain feature extraction method. Subsequently, a parallel network is employed to achieve feature fusion. The stacked LSTM network extracts the temporal dependencies between features and the GFA module is used to adaptively complement key features representing global information of samples. Finally, the output features are concatenated, and tool wear prediction is achieved through a fully connected layer. Experiments demonstrate the optimal performance in predicting tool flank wear. Specifically, using the proposed GFA-LSTM framework reduces the mean absolute error (MAE) by 36.9%, 17.7%, and 25.2% in three experiments compared to the simple LSTM, validating the effectiveness of the proposed method.

Keywords: tool wear, LSTM, global feature attention, feature selection

1. Introduction

Intelligent technology has empowered the traditional manufacturing industry, and the intelligence of the manufacturing process has become a trend, especially the machining center tools instrumental in improving productivity and reducing losses. This section focuses on intelligent prediction of tool wear and describes the research background, related literature survey, research gaps, research purpose, and the structure of this study.

1.1. Background

Prognostics and health management (PHM) of equipment conduces to significant productivity gains and loss reductions [1]. Developments such as artificial intelligence (AI) are considered the fourth industrial revolution, and these new-generation information technologies in PHM have significantly contributed to economic development. Smart manufacturing has evolved in this context. Metal workpieces are machined into high-quality formed parts through various machining processes. Since milling is one of the more common machining processes, in which the tool processes a qualified workpiece by removing the material from the machined workpiece, the milling tool will gradually wear to reach the failure point and be changed in time, whether in roughing or fine machining. Tool changeover time becomes a key factor in productivity.

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

English language proofreader: Chih-Wei Chang

Determining the tool change time has been gradually valued by monitoring the tool wear status (TWS) to avoid delaying the delivery of parts or products. Traditional methods rely on tremendous costs and experienced workers to listen to the sound of the cutting process, look at the color of the chips, feel the vibration of the machine, etc., which have gradually been abandoned by managers due to the inability to quantify and automate online monitoring. In addition to some physical-driven and statistical methods, data-driven approaches are gradually becoming a widely researched direction for researchers to monitor the TWS due to the end-to-end prediction mechanisms. TWS can be divided into the prediction of tool wear indications and the classification of tool usage status.

1.2. Literature survey

In data-driven approaches, history data including data acquired from machine processing is called process data, and, on the other hand, tool wear indicators measured offline are called health indicators (HI) and are used to train some AI models. Researchers expect the trained AI model to perform well on new data. From erstwhile studies, TWS monitoring is broadly categorized into three types: studies on machine data, studies on tool degradation features, and studies on TWS monitoring models.

Data is the key to a data-driven approach in monitoring the TWS. Since the acquisition of high-quality processing data is time-consuming and costly, several researchers have released open-source datasets to support the research. The PHM 2010 dataset [2] and the NASA milling dataset [3] are pervasively deployed datasets. Datasets in this field show a trend towards developing in the direction of multimodality. For example, the MATWI dataset [4] provided structured data including force, vibration, acoustic emission (AE) signals, and image data. The Mudestreda dataset [5] provided two types of images including the wear area images of the tools and force spectrograms converted by time series data of force signals. However, acquiring images online requires additional equipment, such as industrial cameras, and faces difficulties in achieving high-resolution images from complex machining environments.

In addition to the machining process data, the health index (HI) of the tool needs to be measured to implement supervised AI model training. Numerous researchers have carried out studies on different HIs and achieved optimistic prediction performance. Moreover, several researchers use some HIs including tool flank wear values in wear area B (VB), wear area, surface quality of workpieces, etc. Tool wear was predicted by establishing a mapping relationship between vibration signals and wear areas [6]. On the other hand, the surface morphology of the workpiece, which provided a new idea for tool wear, was investigated [7]. Statistically, VB is the most used HI to train AI models so far.

Regarding the tool degradation information, extracting effective features from the processing data is crucial to training AI models. Since raw machine data tend to have high sampling rates, signal channels interfere with each other. High-quality feature extraction requires domain experience, and its selection and validity analysis are manual and complex. Feature selection and downscaling studies have been conducted. For instance, different feature selection methods were compared including the Pearson correlation coefficient (PCC), random forest (RF), principal component analysis, and hybrid methods [8]. Different variants of the long short-term memory (LSTM) model were used to validate the performance of these feature selection methods in predicting tool wear. Furthermore, extensive feature collection was deployed in the time domain and compared three feature selection methods with decision tree regressor and RF regressor [9]. In this context, the collection selected the key features and performed well in tool wear prediction. Together with the use of cross-validation, the feature selection process was complex and time-consuming.

Some studies designed different models including both machine learning and deep learning models, thereby achieving decent performance in tool wear prediction or TWS classification. For example, the Bayesian-optimized support vector regression (SVR) performed well in achieving tool wear prediction [10]. The RF method was used to achieve classification

[11] or prediction [12] under different working conditions in the tool wear prediction field and other fields. Compared with the extraction of shallow features by traditional machine learning methods, deep learning methods mine deeper tool wear degradation information and achieve ideal prediction results. For example, a model-based approach for tool wear monitoring based on an adaptive neuro-fuzzy inference system for a turning process was presented, and an effective prediction performance of online monitoring was obtained [13]. To improve the convergence speed, particle swarm optimization (PSO) was used to optimize the model and enhance its predictive performance [14-15].

Since tool wear has strong time series characteristics, some models derived from the recurrent neural network (RNN), such as LSTM, have strong advantages in predicting tool wear. In addition, some studies have developed hybrid models combining convolutional neural networks (CNNs) and LSTMs to compensate for the shortcomings of LSTMs, which only mine the temporal dependencies of data and lack spatial relationship mining. For example, better results were achieved using a combination of CNN and LSTM than using a single model [16]. The LSTM model was tested using spindle motor current signals collected during experiments performed on a turning machine and demonstrated its capability to capture tool flank wear during the machining process [17]. Besides, as one of the variants of RNN and an improved model of LSTM, bidirectional gated recurrent unit (BiGRU) was also used to predict tool wear [18]. To focus on the impact of key features on model predictions, many studies have used various forms of attention mechanisms such as self-attention [19], convolutional block attention [20], etc.

1.3. Research gap

In the aforementioned methods, there are some limitations:

(1) Traditional machine learning methods require complex feature engineering. Complex manual feature selection is still used in some deep learning models. The use of manual cross-validation and selection to screen for valid features is overly complex.

(2) LSTM has been verified to be effective in tool wear prediction, whereas some of the overall sample information is lost in the division of the sample. Many researchers used CNN and other automatic extractions of features from the original processing data to achieve a certain prediction effect. This is still localized spatial information mining and lacks global information, which incurs reduced prediction performance.

1.4. The purpose of the study

Based on the gap, the purpose of the study is to accurately predict tool wear in milling and simplify the feature selection process. This paper proposes an improved LSTM model with the global feature attention (GFA) method named GFA-LSTM and enhances the performance of tool wear prediction. Specifically, the GFA method is used to adaptively enhance the key features and simplify the feature selection process. Combined with the ability of LSTM to extract time dependencies between features, the model gets high performance in tool wear prediction compared with simple LSTM and other machine learning methods. The contribution of this paper is:

(1) A GFA layer will be used to learn the importance of features, which are related to tool degradation information adaptively enhanced and suppressed invalid features, thereby achieving a simplified representation of the global information of the samples.

(2) GFA uses the max pooling, average pooling, and sum operation to represent the overall information of sample multi-step. It can enhance the effective features and at the same time compensate for the disadvantage of insufficient characterization of global information when LSTM mines the time dependence of local features.

The subsequent sections include the introduction of related methods, the setup of the experimental environment, the model training situation, the results and analyses, and finally, the conclusion.

2. Methodology

In this section, the LSTM model is first introduced, and the reason for using the model is described in detail. Thereafter, the core GFA method is shown in detail including the input data, the calculated algorithm, and the output data used in this method. Finally, the overall structure of the proposed model is shown at the end of this section.

2.1. LSTM model

The LSTM [21] network was designed to extract time dependencies in features, especially long-term dependencies, and had a greater advantage over the gradient explosion and other problems of RNN. The data collected during tool machining features strong time series characteristics, and the historical machining data features a greater impact at the next moment. With this advantage, LSTM can deeply mine the features related to tool wear and achieve higher accuracy in tool wear prediction. The structure of the LSTM cell is shown in Fig. 1. The formulas for its gating mechanism are given below.

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

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

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

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

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

In each formula, i, f, and o denote the input, forget, and output gate. h and C denote the hidden state and the cell state value. t refers to the current moment, and the corresponding t–1 refers to the previous moment. Thus i_t, f_t, o_t, and h_t refer to the values of the input, forgetting, output, and hidden states at the moment t, respectively. W and b refer to the weights and biases corresponding to the subscripts. σ and tanh denote the activation function.

The memory output of the previous moment is achieved by the input of the previous moment and the current moment through the calculation of the three gating mechanisms. The stack of multiple LSTM units and layers enables deeper mining and extracting more effective tool degradation information in tool wear machining data.

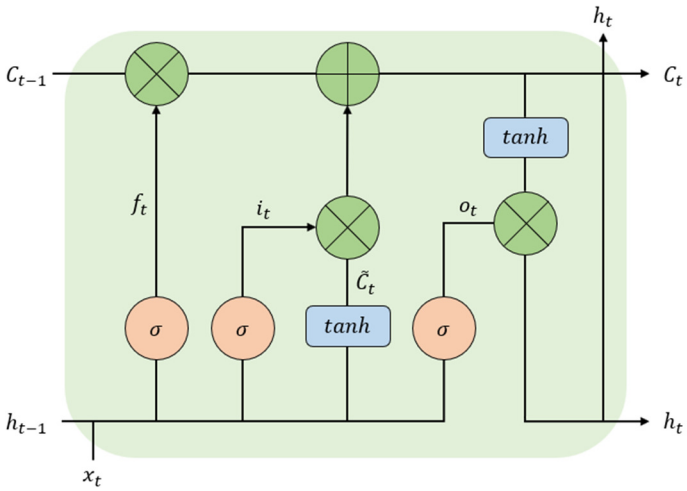

Fig. 1 The structure of the LSTM cell

2.2. Global feature attention method

LSTM can extract temporal dependencies between features and is a local representation of the sample. However, it lacks a representation of the global information of the samples. Therefore, this paper proposes a global feature representation method.

When expressing the global features of the sample, feature importance signifies tool degradation information mining. Considering the complexity of feature extraction and selection, this paper is inspired by the Squeeze-and-Excitation (SE) network [22] and proposes a GFA method. The inter-feature attention module is used to achieve adaptive enhancement of important features and inhibit invalid features.

The sample data input to the LSTM enters the parallel GFA module, postulating that the input features of the sample are expressed by:

$$X_t = \{x_1, x_2, \ldots, x_t, \ldots, x_T\} \in R^{T \times N} \tag{6}$$

where x_t is the feature vector at moment t. N and T refer to the number of features and time steps of the input data.

Three different ways of squeezing are performed according to the time-step dimension, where each feature at multiple time steps is represented as the maximum F_{max}, mean F_{avg}, and sum values F_{sum}, denoted as the formula below, respectively.

$$F_{max} = \left\{\mathrm{MAX}\left(x_1^{(1)}, x_2^{(1)}, \ldots, x_T^{(1)}\right), \ldots, \mathrm{MAX}\left(x_1^{(N)}, x_2^{(N)}, \ldots, x_T^{(N)}\right)\right\} \tag{7}$$

$$F_{avg} = \left\{\mathrm{AVG}\left(x_1^{(1)}, x_2^{(1)}, \ldots, x_T^{(1)}\right), \ldots, \mathrm{AVG}\left(x_1^{(N)}, x_2^{(N)}, \ldots, x_T^{(N)}\right)\right\} \tag{8}$$

$$F_{sum} = \left\{\mathrm{SUM}\left(x_1^{(1)}, x_2^{(1)}, \ldots, x_T^{(1)}\right), \ldots, \mathrm{SUM}\left(x_1^{(N)}, x_2^{(N)}, \ldots, x_T^{(N)}\right)\right\} \tag{9}$$

where MAX, AVG, and SUM represent operations to find the maximum value, the average value, and the sum, respectively. In $x_a^{(b)}$, The subscript a denotes the feature vector of a time step, and the superscript b denotes the b-th element of the vector.

After normalizing to the same dimension, the above three feature vectors are fed into the multilayer perceptron (MLP) network with shared weights for feature importance learning. The outputs are summed element by element and then processed using softmax activation to obtain their importance weights. The weights are obtained by learning through the backpropagation algorithm. The mathematical expression for this operation is shown below:

$$\alpha_{\mathrm{GFA}} = \sigma\left\{\mathrm{SUM}\left[\mathrm{MLP}\left(F_{max}\right), \mathrm{MLP}\left(F_{avg}\right), \mathrm{MLP}\left(F_{sum}\right)\right]\right\} \tag{10}$$

where MLP is the multi-layer perceptron network with shared parameter values. The number of neural units in the hidden layer is opted to be half of the input layer. σ is the softmax activation function, and α_{GFA} is the feature importance weights learned after sample learning.

After the importance weights are obtained, their dimensionality is expanded and then multiplied element by element with the original input features using the Hadamard Product. The detailed formula is given by:

$$F_{\mathrm{GFA}} = O_{\mathrm{expand}}\left(\alpha_{\mathrm{GFA}}\right) \otimes X_t \tag{11}$$

where O_{expand} is the dimension expansion operation to match the dimensions of the input data X_t. The symbol $\otimes$ is the operation of Hadamard Product. F_{GFA} denotes the feature that represents the global information of the sample after the GFA method.

The output of the final GFA is spliced with the output of the LSTM as the input for the next stage. A fully connected layer with an L1 regularization term is used to predict the tool wear. To prevent overfitting, the dropout and batch normalization (BN) layer are also used in the model to enhance the generalization ability.

2.3. Overall GFA-LSTM model structure

The temporal dependencies between the features learned by LSTM are utilized as local feature information, which is spliced with the global feature information learned by GFA. The spliced features are used as the input of the fully connected layer, and the tool wear prediction results are finally obtained. The structure of the whole model is shown in Fig. 2.

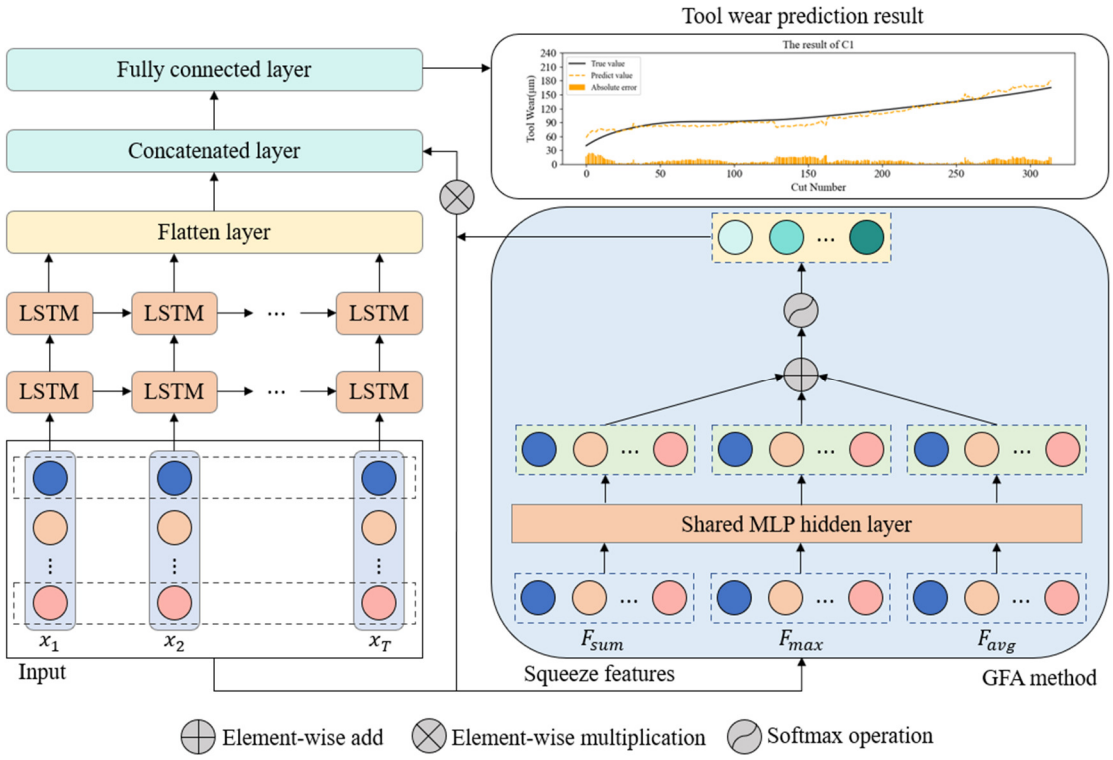

Fig. 2 The structure of the proposed GFA-LSTM model

3. Experiment Works

In this section, the data is described first, including the experiment setup, machine parameters, and sample number. Subsequently, the processing of all the data is shown meticulously. After elaborating on the feature extraction, the training progress including all the settings of model parameters, and some compared models is introduced.

3.1. Data description

In this paper, the PHM 2010 dataset [2] is used. Milling tools with three cutting edges were employed to mill a square carbon steel workpiece with constant machine conditions. Vibration signals, AE signals, and force signals were acquired online when machining. After every cycle of the machine operation, it was stopped, and 3 tool flank wear values named VB were measured offline. The experiment setup and the machine conditions are shown in Fig. 3 and Table 1, respectively.

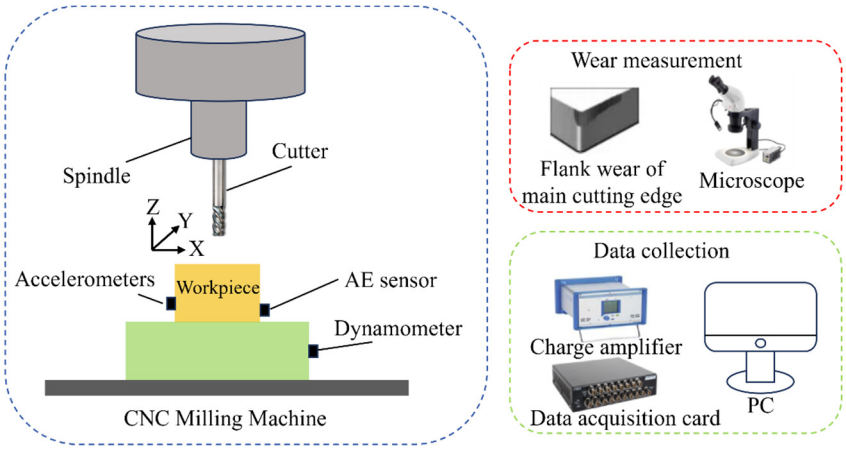

Fig. 3 Details of the experiment setup

As shown in Table 1, the dataset acquired data from 6 tools but only measured the VB of 3 tools named C1, C4, and C6. Every tool has 315 samples with about 200,000 rows and 7 columns representing the *x*, *y*, and *z* direction of force and vibration, and the AE signal, respectively.

Table 1 Machine conditions of the dataset

Tool No.	Conditions and details				
	Spindle speed (rpm)	Feed speed (mm/min)	Y depth of cut (radial) (mm)	Z depth of cut (axial) (mm)	The sample size of the dataset
C1	10400	1555	0.125	0.2	Every tool has 315 files and every file has about 200,000 data points and 7 columns (signal channels)
C2 (no label)					
C3 (no label)					
C4					
C5 (no label)					
C6					

3.2. Data processing

Because the data processing center control and the acquisition software control belong to two sets of control systems, the start and end of the acquired signals incur the unnecessary presence of invalid signals before and after the signals. In addition, there are some outliers in the signals, which are detrimental to model training and need to be removed. In this paper, these two aspects of data processing are carried out. To simplify the calculations, feature extraction is performed on the first 6 columns approximately 200,000 rows of data for each sample, and AE signals are removed.

Finally, 144 features are extracted from 10 subsequences of every sample according to the method of multi-domain feature extraction proposed by Wang et al. [19] including 12 features from the time domain, 4 features from the frequency domain, and 8 features from time-frequency domains. In the time and frequency domain, some statistical features are extracted using the formula, which is shown in Table 2. In the frequency domain and the time-frequency domain, 8 energies from the 8 sub-bands decomposed by the discrete wavelet transform method are extracted.

Table 2 Features extracted from the time domain and frequency domain

Feature name	Expression	Feature name	Expression
Absolute mean value	$\frac{1}{T}\sum_{i=1}^{T}\lvert p_i\rvert$	Shape factor	$\frac{\sqrt{\frac{1}{T}\sum_{i=1}^{T}p_i^2}}{\frac{1}{T}\sum_{i=1}^{T}\lvert p_i\rvert}$
Peak value	$\max(p)$	Pulse factor	$\frac{\max(p)}{\frac{1}{T}\sum_{i=1}^{T}\lvert p_i\rvert}$
Root mean square	$\frac{1}{T}\sqrt{\sum_{i=1}^{T}p_i^2}$	Skewness factor	$\frac{SK}{\left(\sqrt{\frac{1}{T}\sum_{i=1}^{T}p_i^2}\right)^3}$
Root amplitude	$\left(\frac{1}{T}\sum_{i=1}^{T}\sqrt{\lvert p_i\rvert}\right)^2$	Crest factor	$\frac{\max(p)}{\sqrt{\frac{1}{T}\sum_{i=1}^{T}p_i^2}}$
Skewness	$\frac{1}{T}\sum_{i=1}^{T}\left(\lvert p_i\rvert-\frac{1}{T}\sum_{i=1}^{T}\lvert p_i\rvert\right)^3$	Clearance factor	$\frac{\max(p)}{\left(\frac{1}{T}\sum_{i=1}^{T}\sqrt{\lvert p_i\rvert}\right)^2}$
Kurtosis	$\frac{1}{T}\sum_{i=1}^{T}p_i^4$	Kurtosis factor	$\frac{T\sum_{i=1}^{T}p_i^4}{(\sum_{i=1}^{T}p_i^2)^2}$
Centroid frequency	$\frac{\sum_{i=1}^{N}f_i\cdot P(f_i)}{\sum_{i=1}^{N}P(f_i)}$	Mean square frequency	$\frac{\sum_{i=1}^{N}f_i^2\cdot P(f_i)}{\sum_{i=1}^{N}P(f_i)}$
Root mean square frequency	$\sqrt{\frac{\sum_{i=1}^{N}f_i^2}{N}}$	Frequency variance	$\frac{\sum_{i=1}^{N}(f_i-f_c)^2\cdot P(f_i)}{\sum_{i=1}^{N}P(f_i)}$

In Table 2, p and T represent the input data sequence and the number of data points, respectively. Through the fast Fourier transform, four features including centroid frequency, root mean square frequency, mean square frequency, and frequency variance are extracted, where f_i is the frequency, $P(f_i)$ is the power or amplitude at that frequency, and N is the number of frequencies. To mitigate the errors, which are presented in human measurements, the average of the three cutting edges on the tool is taken as the final wear label. The measured 3 flank wear values on the three tools and the calculated average VB are shown in Figs. 4-6

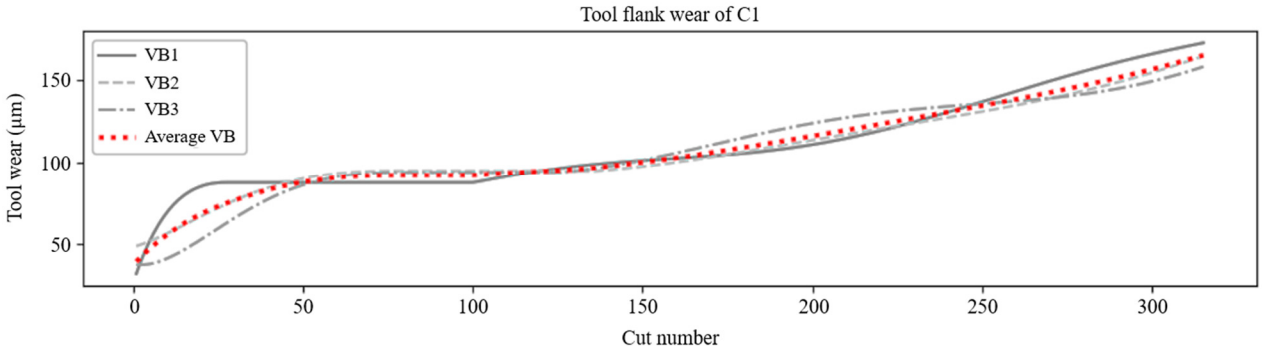

Fig. 4 Tool flank wear of the three cutting edges and the average value on C1

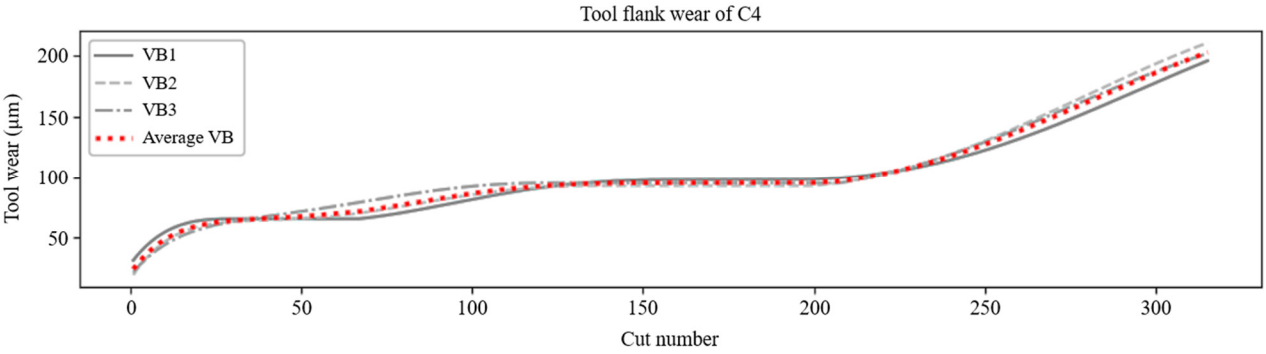

Fig. 5 Tool flank wear of the three cutting edges and the average value on C4

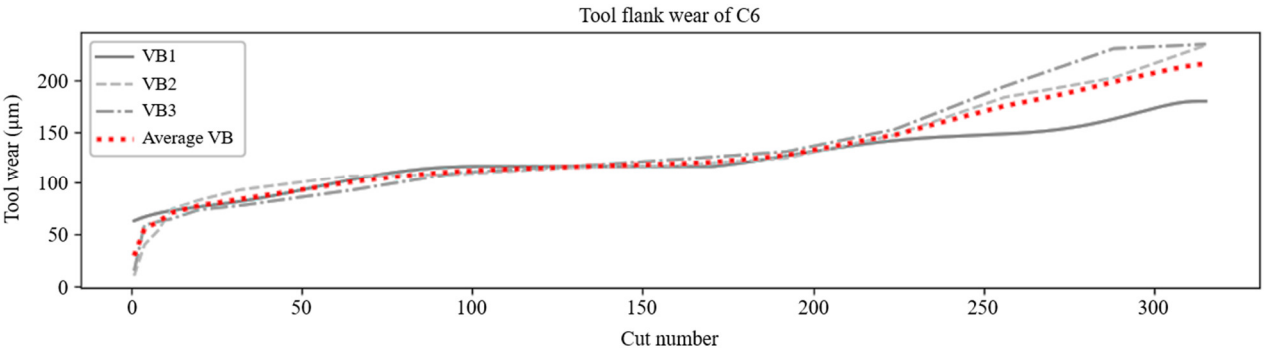

Fig. 6 Tool flank wear of the three cutting edges and the average value on C6

3.3. Model training

To compare the performance of the proposed GFA-LSTM, simultaneous experiments with SVR, MLP, and Simple LSTM are conducted. Features are normalized before being input into the model. Its mathematical expression is as follows:

$$x_{norm} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{12}$$

where x and x_{norm} denote the input features and the normalized data. The min and max represent the min and the max values in the training features. The same operation is used in the validation and test data to prevent data leakage.

All experiments are conducted three times, i.e., three tools are trained in two-by-two combinations and the remaining one is tested to verify the validity of the model. The ratio of training samples, validation samples, and test samples is 8:1:1. The sample split result is shown in Table 3.

Table 3 Sample split details

Case	Training set size (rate = 0.8)	Validation set size (rate = 0.2)	Testing set size (data from the new tool)
1	C1+C4: 504	C1+C4: 126	C6: 315
2	C4+C6: 504	C4+C6: 126	C1: 315
3	C1+C6: 504	C1+C6: 126	C4: 315

For the SVR and MLP models, the input feature data are selected using the PCC method. Given the use of merging the data of the 2 tools for the training set, instead of simply calculating the correlation, this paper filters out the features with strong correlation by calculating the correlation individually. After filtering features from the two parts of the training data, the final filtered features are obtained by finding the intersection. Eventually, the results and details of feature selection are shown in Table 4.

Table 4 The feature selection details

Case	Number of the training set (Part 1)	Number of the training set (Part 2)	Number of the final selected features
1	86	82	75
2	82	87	78
3	86	87	78

Five-fold cross-validation and grid search are used in SVR to tune the parameters. The PCC between two data series, represented by p, is calculated using the following formula:

$$p = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{13}$$

where x and y represent two vectors or sequences of the same dimension. The correlation coefficient between all input feature sequences and the tool wear label sequences is calculated. Feature sequences with strong correlations indicate that these features contain more information concerning tool degradation. Conversely, feature sequences with weak correlations should be removed. The threshold for the correlation coefficient is set to 0.85.

For the simple LSTM model, the BN layer, dropout, and L1 operations are used to prevent overfitting. Additionally, model checkpoints are set up to save the model with the lowest loss on the validation set during training iterations. The architecture and related parameter settings for the simple LSTM are shown in Table 5.

Table 5 Architecture and parameters of the simple LSTM

Layer name	Value	Parameters	Value
Input layer	(10,144)	Optimizer	adam
LSTM layer	128	Learning rate	0.001
Dropout	0.2	Batch size	32
BN	-	Epochs	150
Flatten	1280	Checkpoint	True
Dropout	0.2	Time step	10
Fully connected layer	1		

The proposed model uses the GFA module for adaptive feature importance selection of the input raw features. To verify the superiority of the proposed GFA module, ablation experiments are conducted to compare the effectiveness including the

normal MLP and LSTM without GFA, GFA-MLP, and the proposed GFA-LSTM model. The root mean square error (RMSE), mean absolute error (MAE), and R square (R^2) are selected as the evaluation metrics to evaluate the performance. The calculation formulas of these 3 metrics are given by:

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{true} - y_{pre}\right)^2} \tag{14}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_{true} - y_{pre}\right| \tag{15}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_{ture} - y_{pre}\right)^2}{\sum_{i=1}^{n}\left(y_{ture} - \overline{y}_{true}\right)^2} \tag{16}$$

where y_{true}, y_{pre}, and $\overline{y}_{true}$ denote the real value, prediction value, and the average real value.

4. Results and Discussion

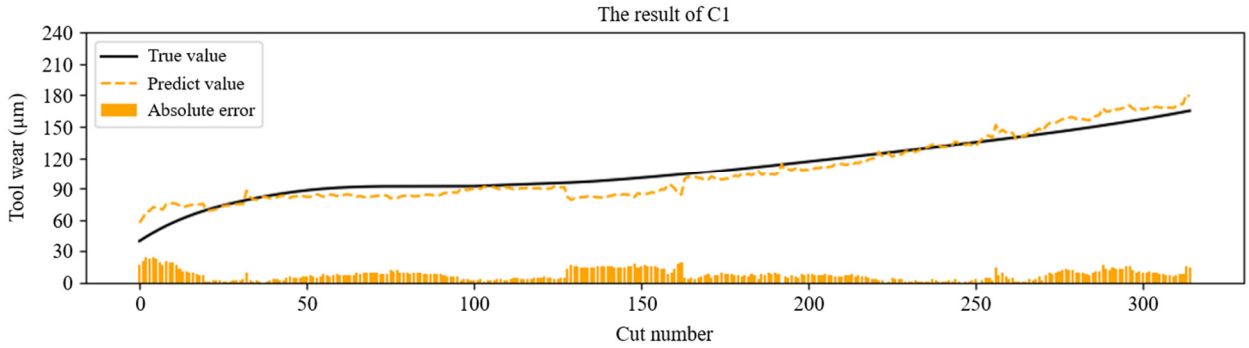

Fig. 7 The result of the GFA-LSTM model on C1

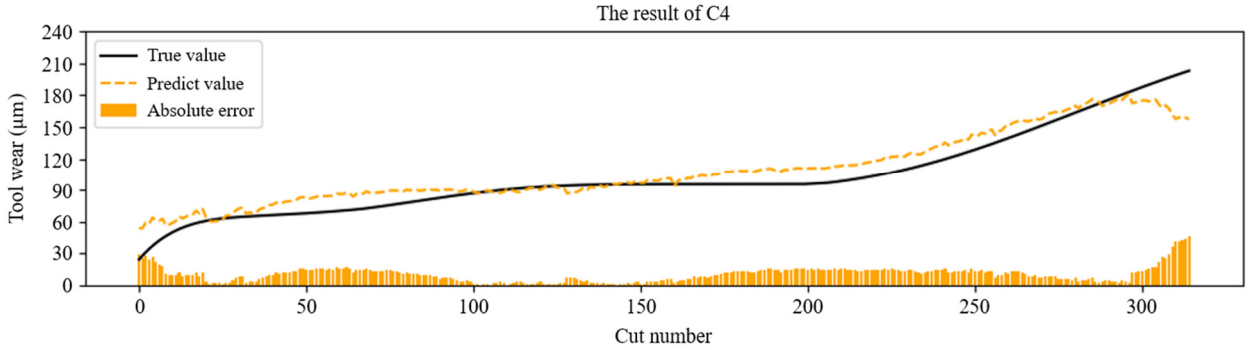

Fig. 8 The result of the GFA-LSTM model on C4

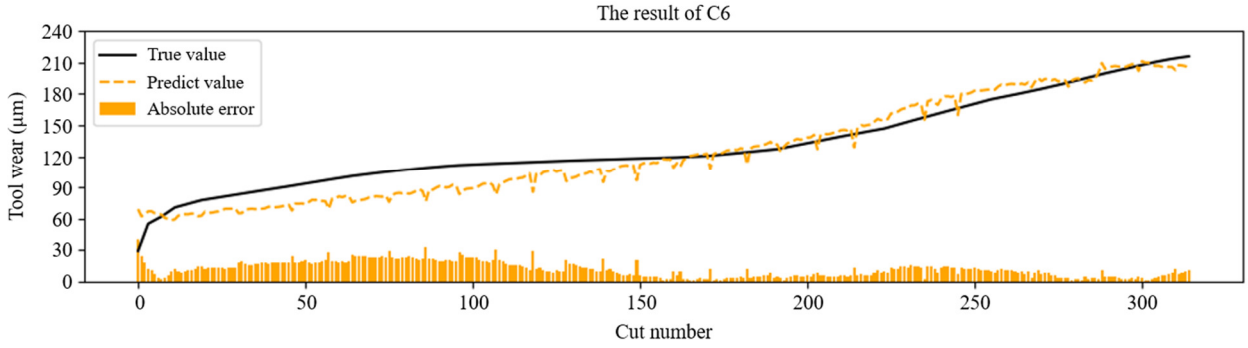

Fig. 9 The result of the GFA-LSTM model on C6

Figs. 7-9 show the results of the proposed GFA-LSTM model on tools C1, C4, and C6, respectively. The proposed model can capture the trend of tool wear with less overall error. However, the predicted results fluctuate more during the initial wear and severe wear stages, which may be attributed to these two rapid phases of wear and fluctuating machining data. Increased uncertainty in machining data during rapid wear poses a challenge for accurate prediction of tool wear.

Table 6 and Figs. 10-12 show the results of the comparison and the proposed model. After complex operations such as feature selection and cross-validation, SVR can attain certain prediction results. MLP has poor prediction results using the original unscreened features. The MLP, with the addition of the GFA module, has a greater improvement in prediction performance despite the use of the original unscreened features. The MAE is reduced by 7.8%, 21.9%, and 10.9% in the C1, C4, and C6 experiments. The LSTM model evinces significant improvement in RMSE, MAE, and R^2 compared to the SVR and MLP. With the addition of the GFA module, its prediction results exhibit significant improvement. Compared with the simple LSTM, the MAE is reduced by 36.9%, 17.7%, and 25.2% in the three sets of experiments, C1, C4, and C6, respectively. The proposed GFA-LSTM model gets the best RMSE, MAE, and R^2 among all the compared models, which verifies the effectiveness of the proposed GFA module.

Table 6 Results of the compared models and the proposed model

Case	Metrics	SVR	MLP	LSTM	GFA-MLP	GFA-LSTM
C1	RMSE	16.76	19.57	12.94	11.05	9.06
	MAE	12.71	8.95	11.85	8.25	7.48
	R^2	0.6233	0.4863	0.7754	0.8364	0.8900
C4	RMSE	18.15	18.28	16.17	14.83	12.52
	MAE	13.96	15.18	12.36	11.86	10.17
	R^2	0.7710	0.7677	0.8182	0.8472	0.8909
C6	RMSE	16.91	19.05	16.59	16.70	13.60
	MAE	13.57	14.93	15.38	13.30	11.50
	R^2	0.8221	0.7741	0.8288	0.8264	0.8850

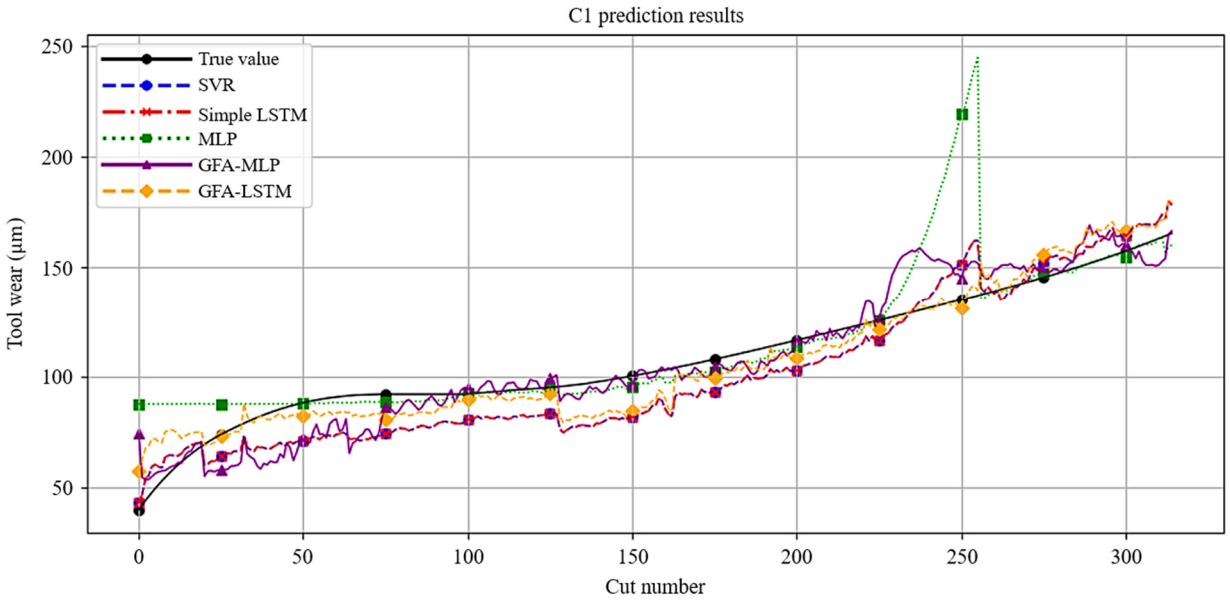

Fig. 10 Result of compared model and GFA-LSTM in C1

From Figs. 10-12, the SVR and MLP models can capture certain wear trends, whereas the error is undeniably large. Especially in the prediction of C1, the MLP presents poor prediction results in the severe wear stage, which indicates that it does not mine the potential patterns in the data well. Simple LSTM has larger errors and more drastic fluctuations in the prediction of wear in the initial and severe stages, and its prediction structure is affected by redundant features. GFA-MLP has some improvement over MLP, but its performance enhancement is constrained by the shallow feature mining of MLP. The

proposed GFA-LSTM model performs best in the three sets of experiments with relatively less error. However, the prediction performance at both ends is poor compared to the middle stable cutting stage. The number of samples and data imbalance may have affected the performance of the proposed model.

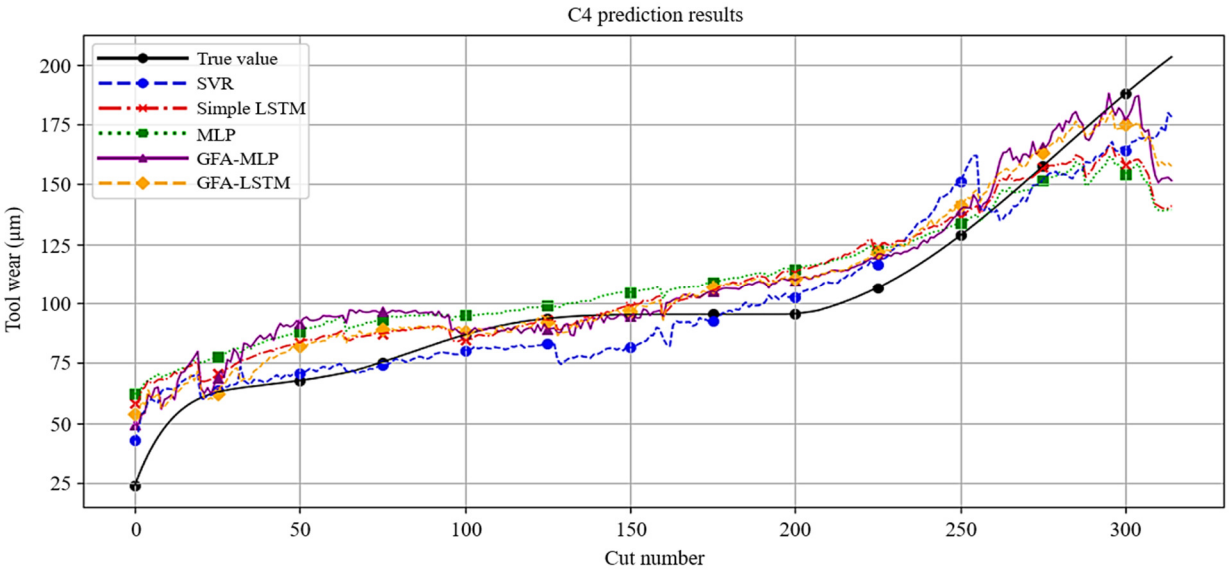

Fig. 11 Result of compared model and GFA-LSTM in C4

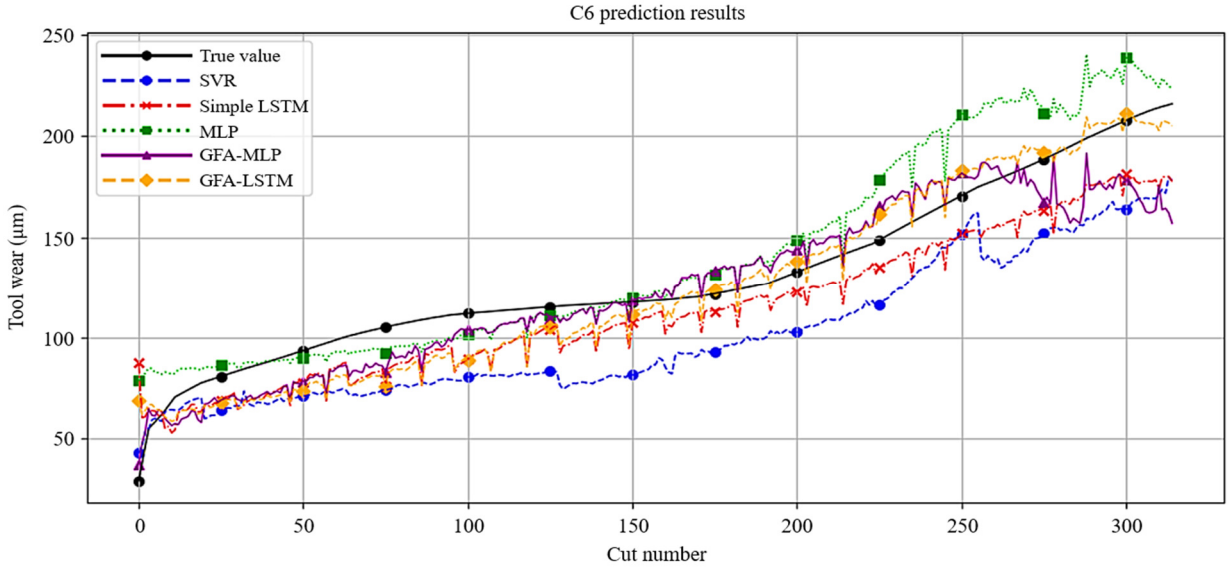

Fig. 12 Result of compared model and GFA-LSTM in C6

Compared with some models in the literature [23], the GFA-LSTM yield the apposite performance in RMSE according to Table 7. Using the shallow LSTM to predict tool wear results in a relatively high RMSE. The deep-LSTM method outperformed shallow LSTM. Regarding the CNN-LSTM model, after adding CNN for spatial feature extraction, there was only a slight improvement on C6, whereas the overall enhancement was not significant. Although the proposed model renders limited improvement compared with the deep-LSTM and the CNN-LSTM model on C4, using shallow LSTM overall achieves comparable or higher tool wear prediction performance. This partially validates the advantages of the proposed model.

Table 7 Comparison of existing literature and the method in this paper based on RMSE

Compared model	C1	C4	C6
Shallow LSTM	24.30	19.50	28.90
Deep-LSTM	12.10	10.20	18.90
CNN-LSTM	13.77	11.85	14.33
GFA-LSTM	9.06	12.52	13.60

5. Conclusions

To simplify the feature selection and accurately predict tool wear in milling, this paper proposed a GFA-LSTM model. Firstly, extensive initial features were extracted from the force and vibration signals using a multi-domain feature extraction method. Subsequently, a parallel strategy was used to extract the temporal dependencies among the features using stacked LSTMs, while the feature selection was adapted using the proposed GFA module. Afterwards, the outputs of the LSTMs, which represent the local information, and the outputs of the GFA, which represent the global information, were concatenated together. Finally, a fully connected layer is used to predict tool wear. The conclusions of the study are presented below:

(1) The feature selection is instrumental to the data-driven approach. SVR performs appropriately after manual feature selection and simple MLP with original features did not perform as expected, shallow feature mining and the negative impact of redundant features on model training make its predictive performance limited.
(2) The proposed GFA module can adaptively select key features related to the tool wear degradation information and significantly improve the MLP and simple LSTM network.

In the field of tool wear prediction, the paucity of prior knowledge in data-driven approaches has led to a bottleneck in improving predictive performance. In the long run, research on model interpretability, such as digital twin-driven tool wear prediction and physics-informed neural networks (PINN) that combine physical mechanisms with data-driven models for tool wear prediction, will potentially become a research hotspot.

Acknowledgment

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

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Y. Hu, X. Miao, Y. Si, E. Pan, and E. Zio, "Prognostics and Health Management: A Review From the Perspectives of Design, Development and Decision," Reliability Engineering & System Safety, vol. 217, article no. 108063, January 2022.
[2] X. Li, B. S. Lim, J. H. Zhou, and S. Huang, "Fuzzy Neural Network Modelling for Tool Wear Estimation in Dry Milling Operation," Annual Conference of the Prognostics and Health Management Society, vol. 1, no. 1, pp. 1-11, 2009.
[3] K. Goebel and A. Agogino, "Milling Data Set," NASA Prognostics Data Repository, NASA Ames Research Center, Moffett Field, CA, 2007.
[4] L. De Pauw, T. Jacobs, and T. Goedemé, "Matwi: A Multimodal Automatic Tool Wear Inspection Dataset and Baseline Algorithms," Computer Vision Systems – 14th International Conference, pp. 255-269, September 2023.
[5] H. Truchan, E. Naumov, R. Abedin, G. Palmer, and Z. Ahmadi, "Multimodal Isotropic Neural Architecture With Patch Embedding," Neural Information Processing – 30th International Conference, pp. 173-187, November 2023.
[6] B. Cardoz, H. N. E. A. Shaikh, S. M. Mulani, A. Kumar, and S. G. Rajasekharan, "Random Forests Based Classification of Tool Wear Using Vibration Signals and Wear Area Estimation From Tool Image Data," The International Journal of Advanced Manufacturing Technology, vol. 126, no. 7-8, pp. 3069-3081, June 2023.
[7] L. Cao, T. Huang, X. M. Zhang, and H. Ding, "Generative Adversarial Network for Prediction of Workpiece Surface Topography in Machining Stage," IEEE/ASME Transactions on Mechatronics, vol. 26, no. 1, pp. 480-490, February 2021.
[8] S. Sayyad, S. Kumar, A. Bongale, K. Kotecha, G. Selvachandran, and P. N. Suganthan, "Tool Wear Prediction Using Long Short-Term Memory Variants and Hybrid Feature Selection Techniques," The International Journal of Advanced Manufacturing Technology, vol. 121, no. 9-10, pp. 6611-6633, August 2022.

[9] 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.

[10] M. S. Babu and T. B. Rao, "An In-Process Tool Wear Assessment Using Bayesian Optimized Machine Learning Algorithm," International Journal on Interactive Design and Manufacturing, vol. 17, no. 4, pp. 1823-1845, August 2023.

[11] 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.

[12] D. K. Dhaked, P. Kumar, and S. Ganguly, "Development of Data Driven Model for Proton Exchange Membrane Fuel Cell Using Machine Learning Approaches," IEEE 3rd International Conference on Control, Instrumentation, Energy & Communication, pp. 67-72, January 2024.

[13] M. Marani, M. Zeinali, J. Kouam, V. Songmene, and C. K. Mechefske, "Prediction of Cutting Tool Wear During a Turning Process Using Artificial Intelligence Techniques," The International Journal of Advanced Manufacturing Technology, vol. 111, no. 1-2, pp. 505-515, November 2020.

[14] S. Wang, Z. Yu, G. Xu, and F. Zhao, "Research on Tool Remaining Life Prediction Method Based on CNN-LSTM-PSO," IEEE Access, vol. 11, pp. 80448-80464, 2023.

[15] V. Lakshmi Narayanan, D. Kumar Dhaked, and R. Sitharthan, "Improved Machine Learning-Based Pitch Controller for Rated Power Generation in Large-Scale Wind Turbine," Renewable Energy Focus, vol. 50, article no. 100603, September 2024.

[16] B. Li, Z. Lu, X. Jin, and L. Zhao, "Tool Wear Prediction in Milling CFRP With Different Fiber Orientations Based on Multi-Channel 1DCNN-LSTM," Journal of Intelligent Manufacturing, vol. 35, no. 6, pp. 2547-2566, August 2024.

[17] M. Marani, M. Zeinali, V. Songmene, and C. K. Mechefske, "Tool Wear Prediction in High-Speed Turning of a Steel Alloy Using Long Short-Term Memory Modelling," Measurement, vol. 177, article no. 109329, June 2021.

[18] R. Wang, Q. Song, Y. Peng, P. Jin, Z. Liu, and Z. Liu, "A Milling Tool Wear Monitoring Method With Sensing Generalization Capability," Journal of Manufacturing Systems, vol. 68, pp. 25-41, June 2023.

[19] W. Wang, S. S. Ngu, M. Xin, R. Liu, Q. Wang, M. Qiu, et al., "Tool Wear Prediction Based on Adaptive Feature and Temporal Attention With Long Short-Term Memory Model," International Journal of Engineering and Technology Innovation, vol. 14, no. 3, pp. 271-284, July 2024.

[20] 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.

[21] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735-1780, November 1997.

[22] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 7132-7141, June 2018.

[23] F. C. Zegarra, J. Vargas-Machuca, and A. M. Coronado, "Comparison of CNN and CNN-LSTM Architectures for Tool Wear Estimation," IEEE Engineering International Research Conference, pp. 1-4, October 2021.