INTERNATIONAL JOURNAL OF AUTOMOTIVE AND MECHANICAL ENGINEERING
ISSN: 2229-8649 e-ISSN: 2180-1606
VOLUME 22, ISSUE 1, 2025, 12174 - 12185
DOI: https://doi.org/10.15282/ijame.22.1.2025.17.0934

RESEARCH ARTICLE

Tool Wear Prediction Based on Adaptive Feature-Temporal Weighted Method and Long Short-term Memory Model

Wanzhen Wang[1], Sze Song Ngu[1]*, Miaomiao Xin[1,2], Xiaomei Ni[1,2], Yuan Liu[2], Man Qiu[2], Qian Wang[2], Hongyan Zang[2], Lin Wang[2], Na Li[2]

[1]Faculty of Engineering, Universiti Malaysia Sarawak,94300 Kota Samarahan, Sarawak, Malaysia
[2]School of Intelligent Manufacturing and Control Engineering, Qilu Institute of Technology, 250200 Jinan, China

ABSTRACT – Tool wear prediction constitutes a critical enabler for intelligent manufacturing systems, providing scientific decision support for proactive maintenance scheduling. Tool wear prediction is a critical challenge in manufacturing, as accurate predictions can optimize production processes, reduce downtime, and lower maintenance costs. However, existing methods face significant limitations: traditional machine learning approaches rely heavily on complex feature engineering, requiring extensive domain expertise, which limits their generalizability and scalability. Additionally, conventional Long Short-term Memory (LSTM) models, while effective for structured data on tool machining, struggle to adequately assign importance to features, leading to suboptimal prediction accuracy. These gaps highlight the need for a more robust and efficient approach to tool wear prediction. This study develops an adaptive LSTM framework with dual attention mechanisms to automate tool wear prediction in smart manufacturing. The method processes vibration signals through four stages−signal division, multi-domain feature extraction, adaptive feature-temporal weighting, and wear quantification. The proposed method is experimentally validated, and the results demonstrate superior performance with lower errors than conventional LSTM models, achieving a reduction of 12.27%, 34.31%, and 40.21% in RMSE on C1, C4, and C6, respectively, and 20.48%, 37.71%, and 39.12% in MAE. The developed model helps producers determine the precise timing of tool changes, reducing the risk of downtime due to tool failure and thus increasing productivity.

ARTICLE HISTORY
Received : 03rd Apr. 2024
Revised : 06th June 2024
Accepted : 05th Feb. 2025
Published : 19th Mar. 2025

KEYWORDS
Tool wear prediction
Feature-weighted
Temporal-weighted
Multi-domain feature

1. INTRODUCTION

Prognostics and Health Management (PHM) is a hot research topic within the manufacturing sector. In intelligent manufacturing, intelligent predictive maintenance for the next generation of Computer numerical control (CNC) machine tools with data-driven methods such as artificial intelligence algorithms has become a popular direction. Tools gradually wear out during the machining process, and this is detrimental to the quality of the workpiece. When it can not meet the requirements of machining quality, it will bring about sudden downtime losses and disrupt the production plan. Accurate prediction of tool conditions becomes critical.

Monitoring tool wear conditions have undergone expert knowledge-based inspection methods, statistical methods, and data-driven methods. In general, tool condition information such as tool wear value or tool wear status can be obtained by direct measurement or indirect mapping methods. The measurement is generally done through non-stop inspection, including image acquisition or measurement. Directly obtaining tool wear information works more accurately, but downtime inspection has brought about a decline in productivity. Non-stop inspection in the processing of tool wear image acquisition will also interfere with the process. Many researchers have built devices that use offline or online processing intervals to make measurements. Taking the milling tools as an example, the main methods are the measurement of the end teeth and the measurement of the main cutting-edge or off-line measurement, as shown in Figure 1. However, the efficiency and effectiveness of inspections can not be guaranteed. The indirect method achieves the purpose of tool status classification or tool wear value prediction by mining patterns of tool wear degradation information in historical data. The data-driven method was used with the indirectly acquired signals. Indirect methods offer advantages in online monitoring of machine tool machining status. Many studies have used indirect signals such as vibration [1], [2], force [3], [4], Acoustic Emission (AE) signals [5], [6], [7], etc. to predict the value of tool wear.

***CORRESPONDING AUTHOR | S. S. Ngu** | ✉ ssngu@unimas.my
© 2025 The Author(s). Published by Universiti Malaysia Pahang Al-Sultan Abdullah Publishing. This is an open access article under the CC BY-NC 4.0 International license

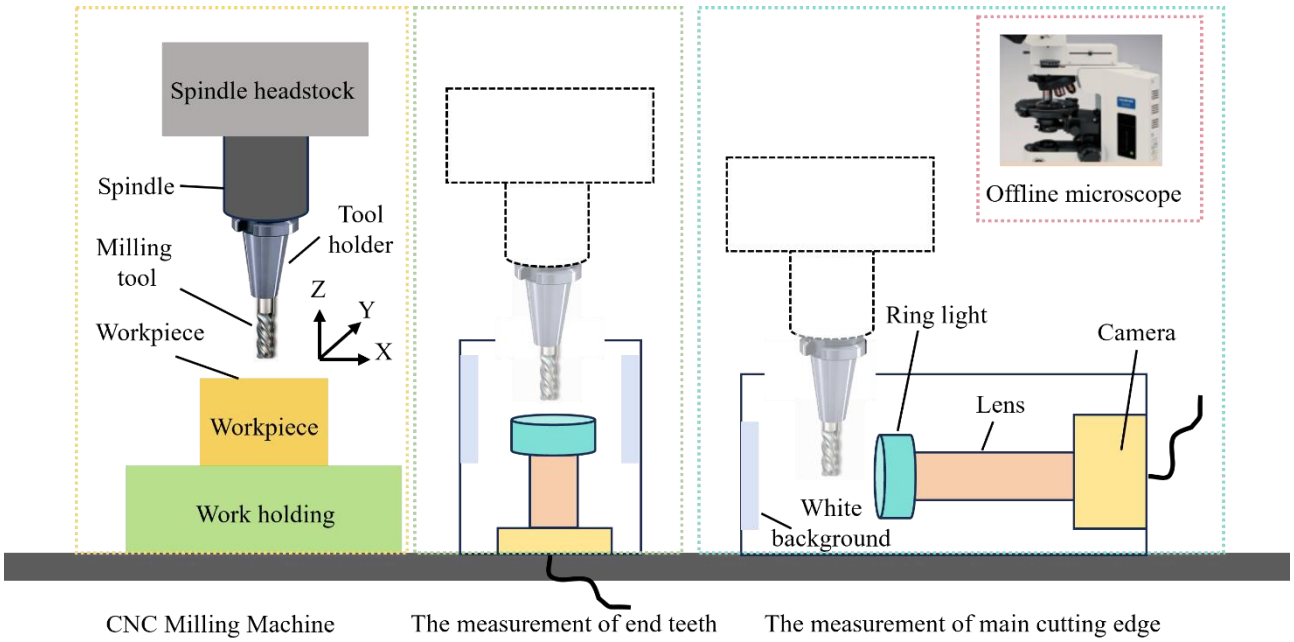


Figure 1. The main methods of direct health indicator measurement

Data-driven methods such as machine learning are more common in predicting tool wear, with faster training speed and less resource consumption. However, these methods have higher requirements regarding the quantity and quality of samples and features. Machine learning methods mining shallow tool degradation information often require complex feature engineering. Many studies have done a lot of research in feature extraction and screening with Support Vector Regression (SVR) [8], [9], Extreme learning machine (ELM) [10], [11], Hidden Markov Model (HMM) [12], Decision Tree (DT) [13], etc. Ensemble learning has become one of the research points, but further improvement of prediction accuracy also requires mining more depth features.

In recent years, deep learning has maintained its advantage in processing complex nonlinear data and can mine deeper and more complex features. Some research [3], [6] used Convolutional Neural Networks (CNN) to mine spatial relationships with the original signals. In [14], Recurrent Neural Networks (RNN) models were used to mine temporal dependencies between tool wear features. In research[15], [16] LSTM was used to mine long-term dependencies between features when RNN encounters a gradient problem. The Gated Recurrent Unit (GRU) model has been introduced in tool state discrimination due to its computational cost-effectiveness. In terms of the number of signal sources, tool wear prediction has experienced single-signal prediction, multi-signal fusion prediction and prediction with mixed processing information. Some researchers have proposed some mechanisms such as residual module [17], attention mechanism [18], new loss function [19], multi-scale convolution [20], and parallel mechanism [14] to improve prediction performance. Among these, the attention mechanism has been refined and applied to the field at different levels to enhance model performance by focusing on the most relevant features. Specifically, the attention mechanisms adaptively learn importance in multiple dimensions to prioritize critical information. In 2D or 3D CNN, different variants are used to capture spatial and channel-wise dependencies, improving feature representation in various scenarios. However, the incorporation of attention mechanisms often increases model complexity, requiring careful design to balance performance and cost.

Image-based methods can perform well but require a lot of computational resources. The LSTM model and GRU model, which are based on the original timing signal for prediction, become the primary choice. Effective feature extraction and selection can improve the prediction performance of LSTM. There are many studies applying traditional feature selection techniques and dimension reduction methods. LSTM can also increase the length of the sequences, which can be used to improve prediction performance.

Although these works have provided valuable insights, they are limited by several factors:

i) Existing methods lie in the feature selection process, which often requires complex and labor-intensive manual intervention, heavily relying on specialized domain knowledge. This not only increases the time and cost of implementation but also limits the generalizability and accessibility of these methods for non-expert users.

ii) For models like LSTM, which process time-step-based data, the feature importance across different time steps is not explicitly characterized. This lack of representation undermines the effectiveness of feature selection and negatively impacts the overall performance of the model.

iii) The importance of time-step outputs in LSTM models is not adequately characterized, which diminishes their ability to effectively extract temporal dependencies. This limitation reduces the model's interpretability and performance in capturing long-term patterns.

In this paper, the adaptive sequence attention mechanism is introduced to address these challenges effectively. Sequences most relevant to the current prediction task are selected rather than treating the entire sequence uniformly. By dynamically adjusting the importance of different time steps, critical temporal dependencies are captured. First, considering the input format of the proposed model, the vibration time series data is converted to sub-sequences. Then, 24 features are extracted from every sub-sequence across multiple domains. After that, the feature-weighted mechanism named FW was used to process the features. Max pooling and average pooling methods with Softmax function enhance the important features and suppress some lower relevance features. Then, multiple LSTM layers use these features and leverage the advantage of LSTM in time-dependency feature extraction. A temporal-weight layer named TW is designed to learn the weights between sequences. Output features are flattened and used to predict final wear values with a fully connected layer. The main work is summarised as follows:

i) A method for assigning importance to features named FW using max pooling and average pooling between features of subsequences is designed. The attention weights are modified through back-propagation and do not need manual feature engineering.

ii) A temporal-weighted mechanism named TW based on max pooling and average pooling between subsequences is proposed, where temporal importance is trained and then assigned to each set of features at the time step. This further exploits the temporal dependence between the subsequences and improves the prediction performance.

Specifically, the next sections elaborate on several sections. Section 2 describes the traditional structure of methods used. The data used, the comparison setup, and the main results are described in the third part. Finally, relevant conclusions and research perspectives are presented.

2. METHODS AND MATERIAL

In this section, the basic structure of LSTM and the proposed FW and TW mechanisms are introduced in detail. Finally, the architectural diagram of the whole model and its analysis are elaborated.

2.1 Structure of Traditional LSTM

LSTM shows advantages in sequence modeling tasks, making it one of the preferred models for processing sequence data. The individual indirect signals acquired during tool machining have strong temporal characteristics, where the signal at the next moment is influenced by the previous moment. To improve the gradient problem in the process of long sequences data fitting using the RNN model, LSTM with gating mechanisms can allow information to propagate over long distances in a sequence. The structure of the LSTM unit is illustrated in Figure 2. The formulae associated with the LSTM cells are shown in Eq. (1) to Eq. (5).

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

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

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

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

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

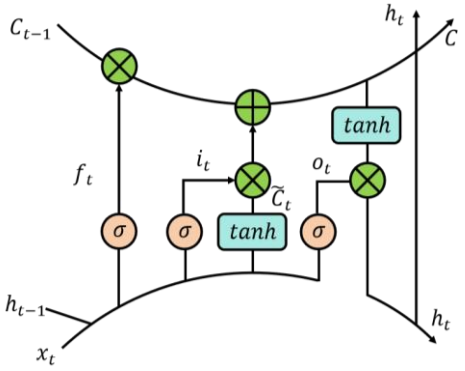


Figure 2. The basic structure of LSTM

Where x_t, $\tilde{C}_t$ and h_{t-1} are the current features, the cell state, and the hidden state at the previous time, respectively. W and b represent different weights and biases. σ is the activate function named Sigmoid.

In this study, the input to the LSTM has a dimensionality of (BS, T, N), which represents the batch size, time steps, and feature numbers, respectively. The output of the LSTM retains the same number of time steps as the input, producing a sequence of hidden states that encode temporal information. Multiple LSTM layers can be stacked to enhance their ability. A key focus of this work is investigating the importance of the time step and feature dimensions. Specifically, this work proposed the feature-weighted and temporal-weighted mechanisms to explore how these dimensions influence the ability to capture the underlying dynamics of tool degradation, thereby improving prediction accuracy and robustness.

2.2 Feature-Weighted Mechanism

Many existing static methods of feature importance selection use a single condition and face many shortcomings. For example, PCA is a linear transformation that can only discover linear structures in data and cannot handle non-linear relationships. KPCA can discover non-linear relationships and can find more complex structures in the data, but KPCA may be time-consuming and lead to overfitting or performance degradation if not chosen properly. Adaptive feature attention can dynamically learn the importance of each feature and adaptively adjust the weights of the features according to different situations of the input data. The sub-sequences are generated according to the sliding window method to utilize the adaptive feature attention mechanism. Features can be extracted manually or used in other models. The raw data S_r can be represented as Eq. (6) and Eq. (7).

$$S_r = \left(s^1, s^2, ...s^t, ...s^T\right)^T \in R^{T \times N} \tag{6}$$

$$S^t = \left(s_1{}^t, s_2{}^t, ...s_i{}^t, ...s_N{}^t\right)^T \in R^{N \times 1} \tag{7}$$

where s^t denotes the t-TH subsequences.

The features have different impacts on the predicted values, sensitive features need to be given larger weights and insensitive features will be given lower weights. To determine the importance of features inspired by [21], a feature-weighted mechanism named FW is proposed. It features average pooling and max pooling, which are calculated in FW. Max pooling extracts the most prominent features by focusing on the highest values, which helps identify critical local patterns. Average pooling provides a global perspective by smoothing out variations and capturing the overall trend of the data. Average and max pooling operations are performed sequentially on N features. The two results are sent into a dense layer of weight sharing. After the two-output data are summed, a Softmax function is used to learn the importance between features by balancing the contributions of both pooling strategies. After extending the weights obtained in the previous step to the feature dimensions, the weighted feature matrix S_o is calculated. The mathematical expressions are summarized in Eq. (8).

$$S_o = S_r \square f\left(A\left(f_{avg}\left(S_r\right)\right) + A\left(f_{\max}\left(S_r\right)\right)\right) \tag{8}$$

where f_{avg}, $f_{\max}$, A and f denote the average and max pooling operation, the ReLU and Softmax activation function, respectively.

Ultimately, Figure 3 clearly depicts the entire calculation process.

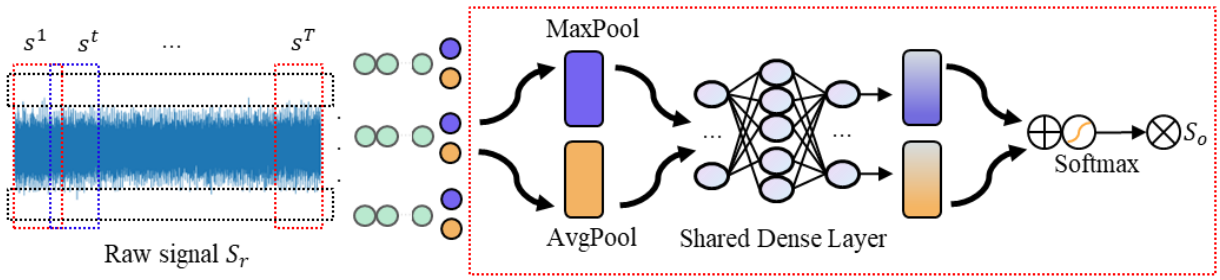


Figure 3. The structure of the FW mechanism

2.3 Temporal-Weighted Mechanism

Because the signals obtained from a single machining process contain both the unstable wear during the tool's initial entry and exit, as well as the stable wear in the intermediate period, the individual sub-sequences contribute to tool degradation to significantly different degrees. The temporal-weighted mechanism is proposed in that at the temporal dimension, it not only makes predictions based on the current input but also assigns different weights based on the information at each time step in the input sequence. The operation makes it possible to represent the effect of different sequences in the sample on tool degradation.

The input of the TW mechanism can be expressed as Eq. (9) and Eq. (10).

$$X_r = \left(x^1, x^2, ...x^t, ...x^T\right)^T \in R^{T\times N} \tag{9}$$

$$x^t = \left(x_1^t, x_2^t, ...x_i^t, ...x_N^t\right)^T \in R^{N\times 1} \tag{10}$$

where X_r denotes the input of the TW mechanism. x^t and x_i^t denote the subsequences and the features in one subsequence.

It can be seen from Figure 4 that the feature dimensions are squeezed using the maximum and average values, respectively. Then, the result of the two pooling layers is used as the input of two parallel weight-shared dense layers with a hidden layer. The outputs are summed and calculated by a Softmax activation to learn the importance between time steps. The mathematical expressions are summarized in Eq. (11).

$$X_o = X_r \square \sigma\left(A\left(f_{avg}\left(X_r\right) + A\left(f_{\max}\left(X_r\right)\right)\right)\right) \tag{11}$$

where X_o, f_{avg}, $f_{\max}$, and σ denote the weighted features, squeezed average, max values, and the Softmax function, respectively.

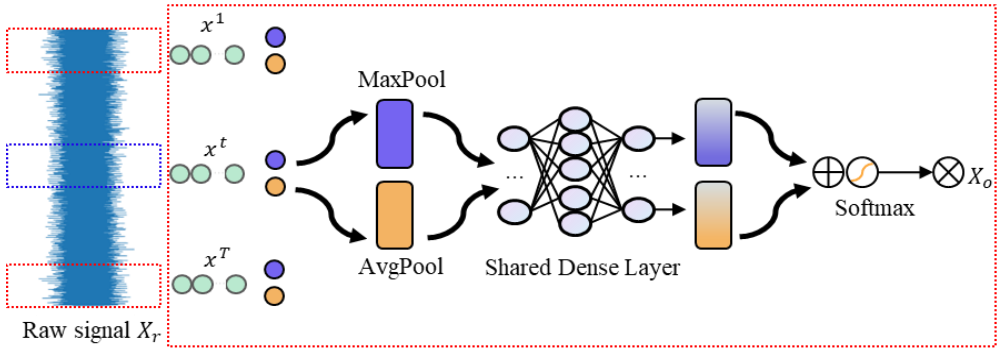


Figure 4. The structure of the TW mechanism

2.4 The Structure of the Proposed Model

It can be seen from Figure 5 that the output data of FW are processed by the stacked LSTM layers, and features are adjusted by FW to represent their importance. Using the gating mechanism, the stacked LSTM is implemented to further fuse features. After learning the importance between sequences through the FW module, the flank wear is predicted by the fully connected layer.

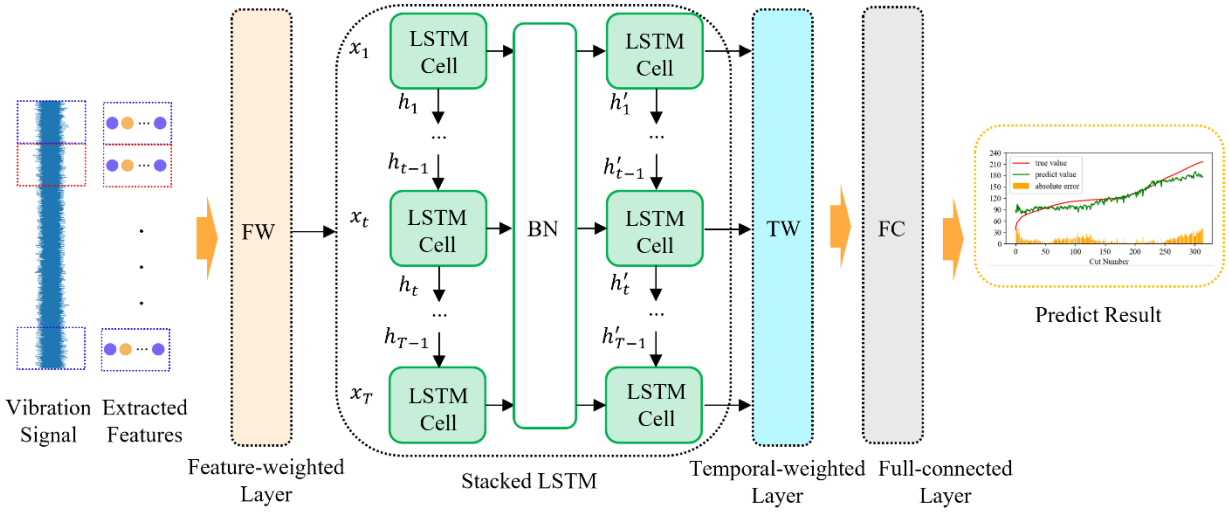


Figure 5. Flow chart of FW-LSTM-TW

3. EXPERIMENTAL RESULTS

Firstly, the method of processing and splitting the datasets is described. A basic feature extraction method is introduced. Contrasting models and the proposed FW-LSTM-TW method are estimated, respectively. Then, the results and discussion are detailed and elaborated at the end of the section.

3.1 Data introduction

The PHM2010 competition datasets [22] are used in this paper. Six sets of milling experiments using six tools named C1 to C6 were carried out according to the same machining conditions, and vibration caused by machine process, force and Acoustic Emission (AE) signals were collected 315 times during the machining process. In every tool named C1, C4 and C6, the flank wear from the three main cutting edges was measured after each machining. The other tools were used as competition data with unpublished wear labels. The location of sensors in the experiment and related details of data acquisition equipment are introduced in Figure 6.

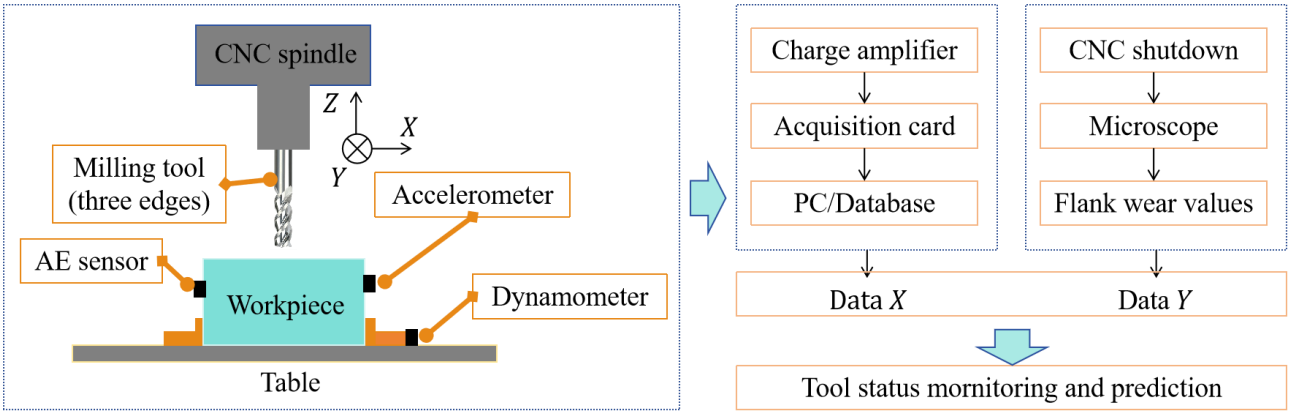


Figure 6. Experimental setup of the PHM2010 datasets

When estimating the performance of different models, the vibration signals collected in the x-axis direction are used as data input. The data from the three tools with the true value label are split into the train and test datasets, and three experiments were carried out, reserving a complete data sample of one tool for each test set. To train the best models, the training set combining two complete tools continues to be divided according to a verification set ratio of 0.2. Details of the total data division are shown in Table 1.

Table 1. Details of data organization for training

Combination of data	Sample size of the training data	Sample size of the validation data	Test data from one separate tool	Sample size of the test data
C1+C4	504	126	C6	315
C1+C6	504	126	C4	315
C4+C6	504	126	C1	315

3.2 Data Processing

Invalid data points and outliers are processed first. Because SVR is a traditional machine learning method and up to 200,000 rows in every sample can be obtained from a single machine process, feature extraction is needed. For traditional LSTM and the proposed model, every sample is split into fixed-time steps. The same feature extraction method is used for convenience. The multi-domain feature extraction method is used, including 12 time-domain features representing the waveform, amplitude, and phase of the signal change with time, 4 frequency domain features describing the amplitude and phase of the various frequency components in the signal and 8 time-frequency domain features considering the two dimensions of time and frequency. Centroid frequency, mean square frequency, root mean square frequency, and frequency variance are calculated to represent the characteristic in the frequency domain. In the time-frequency domain, the discrete wavelet transform (DWT) method is applied to extract the 8-wavelet energy as the features. Table 2 shows some of the formulas in detail.

Considering the different magnitudes between features, the training data are normalized using a z-score to improve model training stability and speed up model convergence, as shown in Eq. (12).

$$x_{norm} = \frac{x_i - \mu}{\sigma} \tag{12}$$

Where x_i and x_{norm} denote the input and standardized data using the average value μ and standard deviation value σ of x_i. The same parameters are used to normalize other data to prevent information leakage.

Table 2. Formulas of the 12 features

Name	Formula	Name	Formula
Absolute Mean Value	$\frac{1}{T}\sum_{i=1}^{T}\lvert p_i\rvert$	Shape Factor	$\frac{\sqrt{\frac{1}{T}\sum_{i=1}^{T}p_i^{2}}}{AMV}$
Peak Value	$\max(p)$	Pulse Factor	$\frac{\max(p)}{AMV}$
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	$\frac{\max(p)}{RA}$
Kurtosis	$\frac{1}{T}\sum_{i=1}^{T}p_i^{4}$	Kurtosis Factor	$\frac{T\sum_{i=1}^{T}p_i^{4}}{\left(\sum_{i=1}^{T}p_i^{2}\right)^{2}}$

According to ISO8688-2:1989 [23], flank wear *VB* is recommended as one of the effective forms of tool failure. As mentioned earlier, three flank wear values from one tool were recorded. For the purpose of reducing measurement errors, average wear values are calculated as the final label to train models. The four values from the three tools are shown in Figure 7.

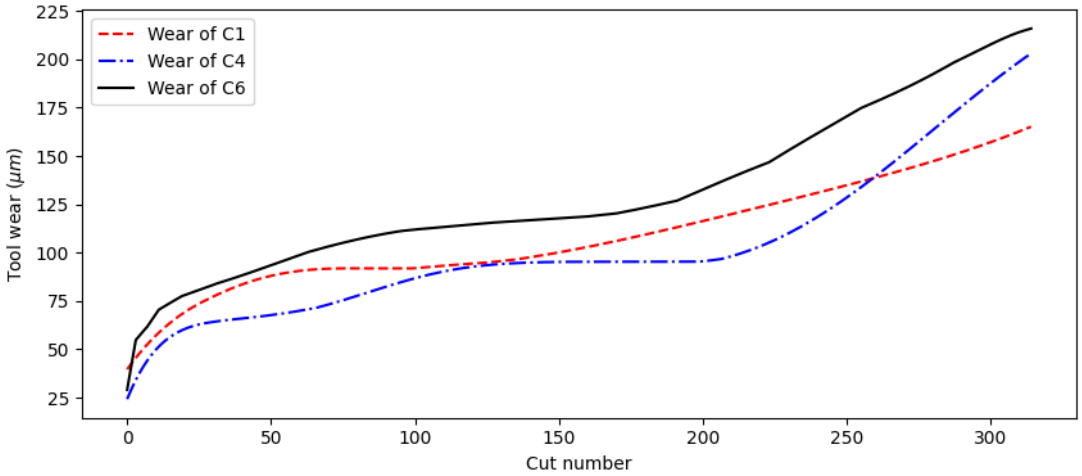


Figure 7. The final average wear values from the three datasets

As machining proceeds, the wear value keeps on growing to max wear $MAX(y_i)$, based on this, in order to train the model efficiently, average wear values y_i are treated as Equation (13) to get normalized y_{norm}. All the wear values are scaled to within 0-1. The predicted values of the final model are back-normalized according to this formula to get the final predicted values.

$$y_{norm}=\frac{y_i}{MAX(y_i)} \tag{13}$$

3.3 Model Training and Testing

For all the LSTM training, a small batch size of 16 Nadam optimizers was selected, considering the small sample size. *RMSE* , *MAE*, and R^2 are used to evaluate the models quantitatively. The three evaluated metrics are shown in Equation (14), Equation (15) and Equation (16), respectively.

$$RMSE = \sqrt{\frac{1}{s}\sum_{i=1}^{s}\left(\hat{y}_i - y_i\right)^2} \tag{14}$$

$$MAE = \frac{1}{s}\sqrt{\sum_{i=1}^{s}\left|\left(\hat{y}_i - y_i\right)\right|} \tag{15}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{s}\left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{s}\left(y_i - \bar{y}\right)^2} \tag{16}$$

Where $\hat{y}_i$ and y_i refer to the i TH real and prediction values. $\bar{y}$ is the average value that expresses the size of the machine cycles.

To verify the validity of the FW and TW mechanism, some models, such as SVR and the traditional staked-LSTM model, which have two layers, are used to train as the comparison models.

In the experiment of SVR, different kernels, including the radio bias function, linear and poly are tested using the grid search method. The features are extracted in the same way, except that original signals are not divided into sub-sequences. Each sample extracts 24 features as input of the SVR. Feature filtering was also used using Pearson's coefficient method, and a total of the top 10 sensitive features were selected for training and testing. When training LSTM, a batch normalization operation is used to make the input more stable, and a dropout layer before the last full connect layer is used to prevent overfitting. The rate of dropout after the dense layer is 0.3. The number of neurons in the LSTM layers was tuned using a grid search method, with values ranging from 16 to 32 in steps 2, as well as 48, 64, and 96. The dropout rate was selected from {0.1, 0.2, 0.3}. To optimize resource utilization, an early stopping mechanism was employed, which terminates training if the validation loss fails to decrease for 8 consecutive epochs. The final optimized parameters are listed in Table 3.

Table 3. Tuned parameters of the proposed model

Name of parameters	Detailed Value
Numbers of subsequences	30
Leaning rate	0.0001
Number of LSTM layers	2
L2 regularizer	0.01
Unites number in 1st LSTM	20
Unites number in 2nd LSTM	18
Dropout rate in Dense layer	0.3
Batch size	15
Epoch and early stopping epoch	100/8
Optimizer	Nadam

3.4 Analysis of the Results

The results of traditional LSTM without proposed FW and TW are shown in Figure 8. The details of RMSE, MAE and R^2 are shown in Figure. 9. Although the predicted performance of SVR is further improved after feature screening, prediction performance is less effective than deep learning methods. On C1, the LSTM model shows a 34.32% and 31.64% improvement over SVR in RMSE and MAE. Similar comparisons are made for C4 and C6. The improvements are 28.13%, 22.20% in RMSE 37.71%, and 24.06% in MAE, respectively. Because the temporal dependencies between the sub-sequences are characterized, LSTM can mine the temporal dependencies between different subsequences, further enriching the feature information related to tool degradation. SVR needs to manually select import features and is only able to mine the shallow information, and the tool degradation information can not be completely characterized.

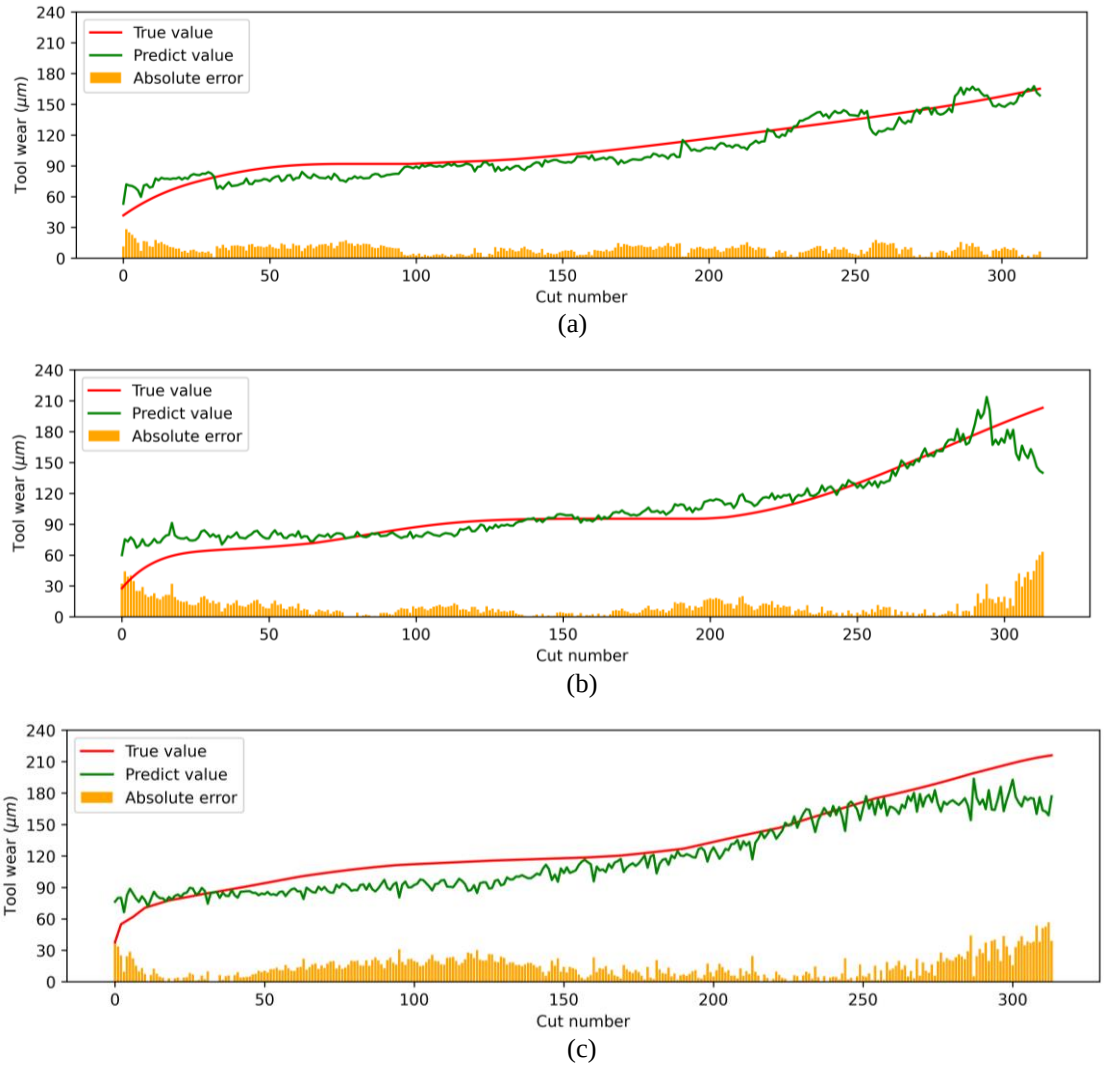


Figure 8. Comparative trends of the traditional LSTM (a) C1, (b) C4, (c) C6

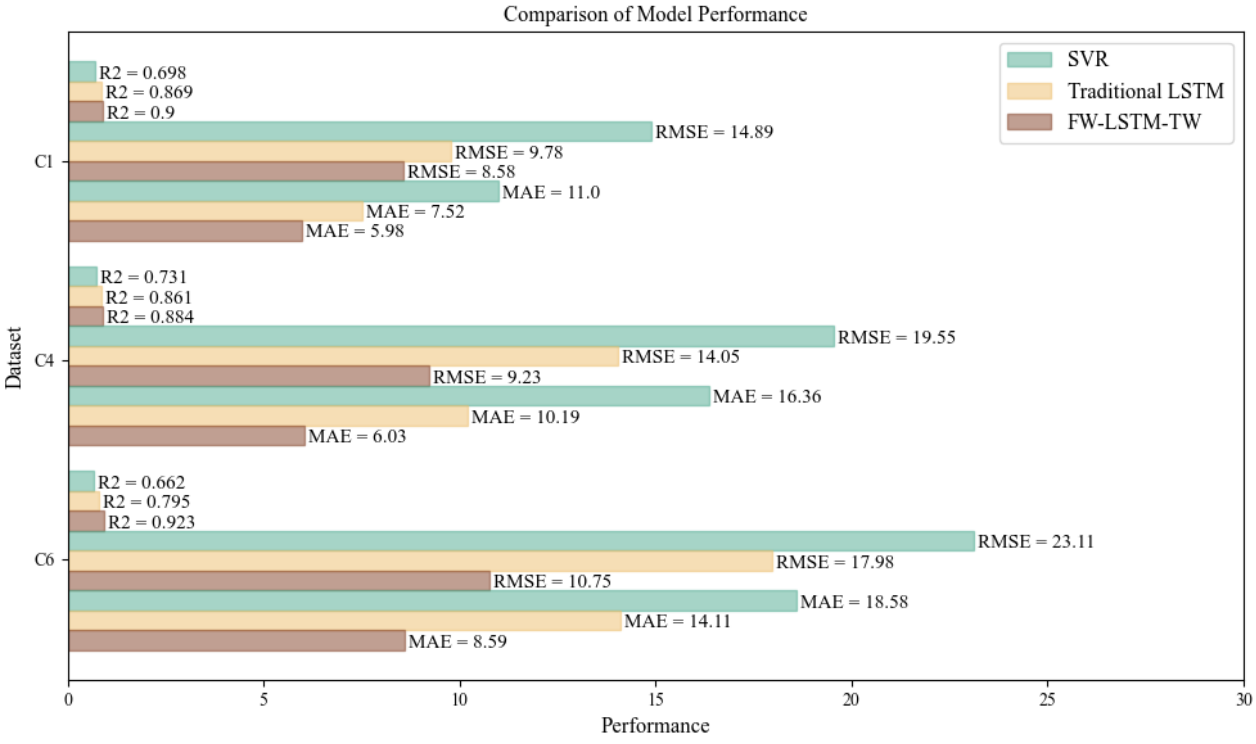
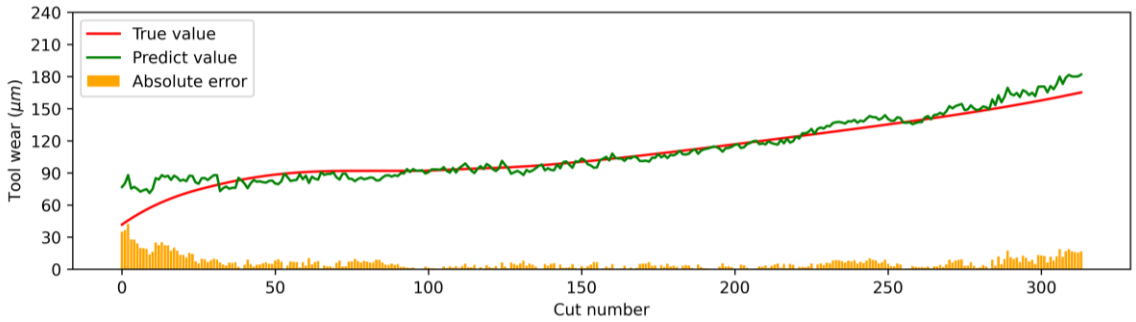


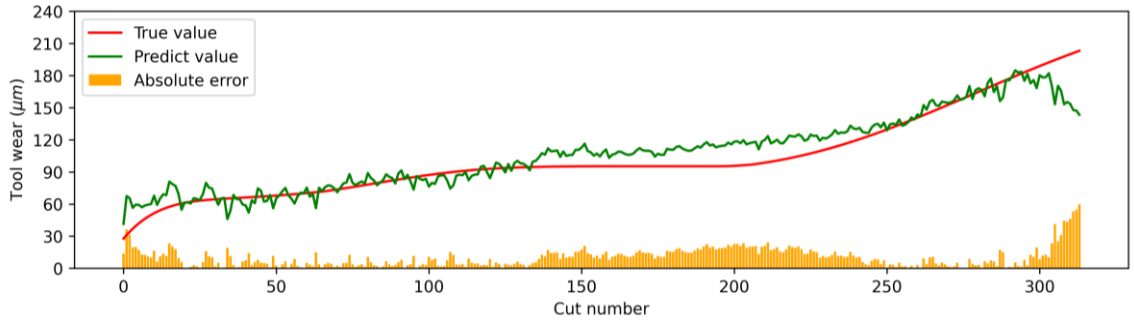
Figure 9. Performance comparison between SVR, traditional LSTM and proposed FW-LSTM-TW model

As shown in Figure 9 and Figure 10, the proposed model performs well in C1, C4, and C6. The RMSE of the proposed model is 8.58, 9.23, 10.75 with a 95% confidence interval of [7.45, 9.68], [7.96, 10.40], [9.98, 11.45] obtained through bootstrapping in C1, C4, and C6, respectively. The MAE of the proposed model is 5.98(confidence interval: [5.33, 6.64]), 6.03(confidence interval: [5.34, 6.79]), and 8.60(confidence interval: [7.91, 9.28]) in C1, C4, and C6, respectively.

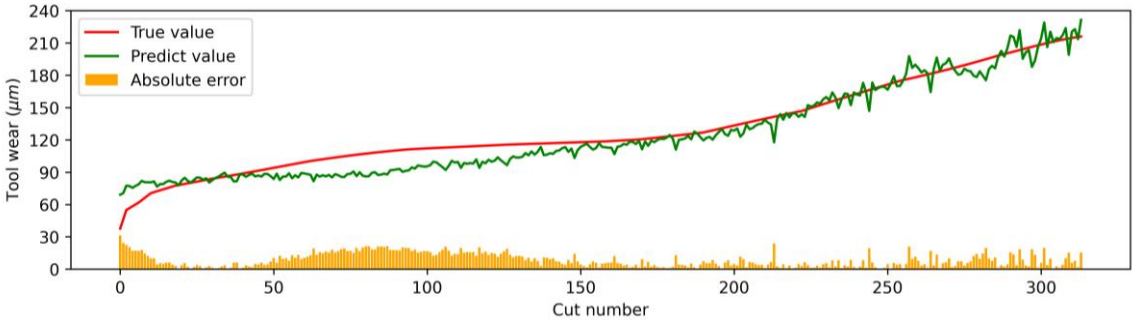
Traditional LSTM with the FW-TW mechanism performed better with a reduction of 12.27%, 34.31% and 40.21% in RMSE on C1, C4 and C6, respectively. The model with the addition of the proposed mechanism is better than the traditional LSTM model with 20.48%, 37.71% and 39.12% reduction in MAE. This validates the effectiveness of the designed TW and FW mechanisms. The FW mechanism reassigns weights to the features of each subsequence, reinforcing the features. The effect of the relationships between the sequences on the prediction results is characterized by learning the weights in the TW mechanism, which further improves the feature-to-target mapping.



(a)



(b)



(c)

Figure 10. Comparative trend of the proposed FW-LSTM-TW model (a) C1, (b) C4, (c) C6

During the experiments, the early and late wear phases were poorly predicted. This can be attributed to several factors:

i) Invalid and unstable signals:

In both early and late phases, the data contain invalid and unstable signals. Although we manually removed these invalid values during feature extraction, some errors may have been introduced, affecting the model's performance. These unstable signals may have influenced the training process, but it remains an open question whether they contain valuable information about tool degradation.

ii) Rapid wear fluctuations and limited sample size:

A rapid rise in wear values occurred during these phases, causing the signal to fluctuate more significantly. This increased variability made it challenging for the model to capture consistent patterns. The early phase and the rapid wear phase at the end are characterized by fast wear progression and short durations, resulting in a limited number of samples. This small sample size, combined with the single signal's limitations, posed challenges for effective model training and reduced prediction accuracy.

4. CONCLUSIONS

This paper proposed an FW-LSTM-TW method to estimate the tool flank wear. To comprehensively extract effective features from machining data, a feature mixing and adaptive selection method based on feature-weighted and temporal-weighted is proposed, which more effectively mines the features related to wear degradation, suppresses related features, and deeply explores the time-dependent information between features. Experiment validation with the PHM competition dataset verifies better performance than traditional LSTM with the following conclusions.

i) FW and TW can enhance the effect of model prediction by adaptively enhancing the sensitive features to suppress the insensitive features, and the relationship between time steps has a certain influence on the prediction results, which can enhance the prediction effect.

ii) Despite complex feature engineering, machine learning methods such as SVR have shown limited improvement in effectiveness. Deep learning methods showed better mapping ability in the experiments performed in this paper. The combination of the two mechanisms makes the traditional LSTM perform better.

In conclusion, the proposed attention-enhanced LSTM model significantly improves tool wear prediction accuracy, offering a practical solution for monitoring in intelligent manufacturing fields. This advancement contributes to the optimization of machining processes, reduction of downtime, and enhancement of overall productivity in smart factories, aligning with the goals of Industry 4.0. Its performance in early and late wear phases remains suboptimal due to limited sample sizes and unstable signals. Future work will focus on data augmentation to address sample imbalance, investigate the potential degradation information in unstable signals, and explore enhanced methods for feature extraction. Comparative experiments with variable machining conditions and multi-sensory signals will further improve the model's robustness and generalization capability.

ACKNOWLEDGEMENTS

This work was supported by funding sources, including the Research Program of Qilu Institute of Technology [No.: QIT23NN043].

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

AUTHORS CONTRIBUTION

W. Wang (Conceptualisation; Methodology; Software; Writing - original draft)
S. S. Ngu (Supervision)
M. Xin (Data curation; Visualisation; Writing - review & editing)
X. Ni (Resources; Validation)
Y. Liu (Methodology) Supervision
M. Qiu (Writing - review & editing)
Q. Wang (Writing - review & editing)
H. Zang (Resources; Validation)
L. Wang (Data curation; Resources)
N. Li (Data curation; Supervision)

REFERENCES

[1] Z. Liu, Z.-Q. Lang, Y. Gui, Y.-P. Zhu, H. Laalej, and D. Curtis, "Vibration signal-based tool condition monitoring using regularized sensor data modeling and model frequency analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–13, 2024.

[2] B. Yang, M. Wang, Z. Liu, C. Che, T. Zan, X. Gao et al., "Tool wear process monitoring by damping behavior of cutting vibration for milling process," *Journal of Manufacturing Processes*, vol. 102, pp. 1069–1084, 2023.

[3] L. Wu, K. Sha, Y. Tao, B. Ju, and Y. Chen, "A hybrid deep learning model as the digital twin of ultra-precision diamond cutting for in-process prediction of cutting-tool wear," *Applied Sciences*, vol. 13, no. 11, p. 6675, 2023.

[4] 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, 2020.

[5] M. Ma, C. Sun, X. Chen, X. Zhang, and R. Yan, "A deep coupled network for health state assessment of cutting tools based on fusion of multisensory signals," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 12, pp. 6415–6424, 2019.

[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, 2022.

[7] D. T. Hoang, N. V. Thien, T. T. T. Pham, and T. D. Nguyen, "Combined analysis of acoustic emission and vibration signals in monitoring tool wear, surface quality and chip formation when turning Scm440 steel using MQL," *Eureka: PE*, no. 1, pp. 86–101, 2023.

[8] J. Wang, S. Zhang, C. Li, L. Wu, and Y. Wang, "A data-driven method with mode decomposition mechanism for remaining useful life prediction of lithium-ion batteries," *IEEE Transactions on Power Electronics*, vol. 37, no. 11, pp. 13684–13695, 2022.

[9] C. Chen, N. Lu, B. Jiang, and C. Wang, "A risk-averse remaining useful life estimation for predictive maintenance," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 2, pp. 412–422, 2021.

[10] Z. Gao, Q. Hu, and X. Xu, "Condition monitoring and life prediction of the turning tool based on extreme learning machine and transfer learning," *Neural Computing and Applications*, vol. 34, no. 5, pp. 3399–3410, 2022.

[11] H. Li and R. Deng, "Cutting tool wear detection method based on fusion of ELM and wavelet packet decomposition," in *International Conference on Advanced Algorithms and Neural Networks (AANN 2022)*, R. Tiwari, Ed., Zhuhai, China: SPIE, 2022, p. 61.

[12] M. Lin, S. Wanqing, D. Chen, and E. Zio, "Evolving connectionist system and hidden semi-markov model for learning-based tool wear monitoring and remaining useful life prediction," *IEEE Access*, vol. 10, pp. 82469–82482, 2022.

[13] 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, 2023.

[14] 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, p. 118548, 2023.

[15] H. Guo, X. Lin, and K. Zhu, "Pyramid LSTM network for tool condition monitoring," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–11, 2022.

[16] F. C. Zegarra, J. Vargas-Machuca, and A. M. Coronado, "Comparison of CNN and CNN-LSTM architectures for tool wear estimation," in *2021 IEEE Engineering International Research Conference (EIRCON)*, Lima, Peru: IEEE, 2021, pp. 1–4.

[17] J. Ou, H. Li, B. Liu, and D. Peng, "Deep transfer residual variational autoencoder with multi-sensors fusion for tool condition monitoring in impeller machining," *Measurement*, vol. 204, p. 112028, 2022.

[18] N. Hu, Z. Liu, S. Jiang, Q. Li, S. Zhong, and B. Chen, "Remaining useful life prediction of milling tool based on pyramid CNN," *Shock and Vibration*, vol. 2023, p. 1830694, 2023.

[19] 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, pp. 1–11, 2021.

[20] B. Wang, Y. Lei, N. Li, and W. Wang, "Multiscale convolutional attention network for predicting remaining useful life of machinery," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 8, pp. 7496–7504, 2021.

[21] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Computer Vision – ECCV 2018*, vol. 11211, V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds., Cham: Springer International Publishing, 2018, pp. 3–19.

[22] "2010 PHM Society Conference Data Challenge," PHM Society. Accessed: Feb. 16, 2024. [Online]. Available: https://phmsociety.org/phm_competition/2010-phm-society-conference-data-challenge/

[23] "ISO 8688-2:1989," ISO. Accessed: Feb. 16, 2024. [Online]. Available: https://www.iso.org/standard/16092.html