Current Science
CS 5 (5), 4813-4827 (2025)
Received 19 July 2025 | **Revised 16 August** 2025 | **Accepted 12 September** 2025 | **Online Available 15 October** 2025
https://doi.org0/10.52845/CS/2025-5-5-35

Current Science

Current Science

OPEN ACCESS

ISSN(O)2795-8639

Research Article

Tool Wear Prediction Based on LSTM-MSCNN Fusion Model and Multi-source Time-frequency Characteristics

Zhang Yanping[1*, 2], Kho Lee Chin[1], Feng Xiansong[3], Zhang Mingqiang[4], Yuan Dongfeng[5]

[1]**Faculty of Engineering, University Malaysia Sarawak, Sarawak, Malaysia**

[2]**Shandong Provincial Key Laboratory of Industrial Big Data and Intelligent Manufacturing, Qilu Institute of Technology, Shandong, China**

[3]**School of Information and Electrical Engineering, Shandong Jianzhu University, Shandong, China**

[4]**School of Cyber Science and Engineering, Qufu Normal University, Shandong, China**

[5]**School of QILU Transportation, Shandong University, Shandong, China**

[*]**Corresponding Author: Zhang Yanping**

Abstract:
Accurate tool wear prediction is pivotal for ensuring machining quality, extending tool lifespan, and reducing operational and maintenance costs. However, modelling the complex spatio-temporal dependencies in multi-source heterogeneous sensor data remains a major challenge. This study proposes *LSTM-MSCNN*, a hybrid deep learning framework that synergistically combines Long Short-Term Memory (*LSTM*) networks for modelling long-range temporal dependencies with Multi-Scale Convolutional Neural Networks (*MSCNN*) for extracting hierarchical spatial features across multiple receptive fields. The model incorporates five heterogeneous sensor modalities--cutting force, torque, image, vibration, and acoustic emission--to build a unified time-frequency feature representation, consisting of 13 time-domain and 8 frequency-domain descriptors, enabling the joint exploitation of temporal dynamics and spatial correlations. The proposed architecture is validated on the Qilu Institute of Technology Coated End Milling Cutter (*QIT-CEMC*) dataset, obtained from real-world coated end milling operations under varying cutting parameters and material conditions. Comparative experiments against strong baselines, including standalone LSTM, ResNet, and MSCNN models, demonstrate that *LSTM-MSCNN* achieves superior predictive performance, with a mean squared error (*MSE*) of 0.406 and a mean absolute error (*MAE*) of 0.286. These results confirm the model's high accuracy, enhanced generalization capability, and robustness in complex machining environments. The proposed method provides a practical and scalable solution for intelligent manufacturing systems, offering valuable guidance for predictive maintenance, process optimization, and decision-making in advanced machining applications.

Keywords: Tool wear prediction; LSTM-MSCNN; Multi-source fusion; Time-frequency features; Deep learning

1. Introduction

1.1 Research Background

Milling is a widely used material removal process in discrete manufacturing, extensively applied in aerospace, precision molding, and automotive industries. However, under prolonged high-speed and high-load operations, tool wear becomes inevitable, adversely affecting surface quality, geometric accuracy, and production efficiency. Consequently, Tool Condition Monitoring (*TCM*) has emerged as a critical research focus in

intelligent manufacturing and industrial automation [1-2].

Modern TCM systems collect multi-modal signals from various sensors, such as spindle vibration [3], cutting force [4], motor current and power [5], acoustic emission [6], and visual data [7], to assess tool health status. These multi-source signals enable modeling of wear evolution through data-driven methods. However, single-modal signals suffer from interference sensitivity, low temporal resolution, and limited feature representation, which restricts high-accuracy wear prediction under complex machining conditions [8]. For example, vibration signals are easily disturbed by environmental and process variations, while image-based data, despite their intuitive nature, often face difficulties in acquiring high-resolution labels for real-time applications. In addition, traditional handcrafted feature extraction relies heavily on expert experience, lacking adaptability and the ability to fully characterize the nonlinear wear mechanisms.

1.2 Literature Review

1.3 In recent years, researchers have increasingly explored the integration of deep learning and multimodal information fusion into TCM systems, achieving notable progress in data augmentation, feature learning, and model optimization.

For data augmentation and feature learning, Quan *et al*. [9] proposed a hybrid data augmentation mechanism (HDAM), significantly improving recognition accuracy under small-sample conditions. Chauhan *et al.*[10] combined stationary wavelet transform (*SWT*) with local linear embedding (*LLE*) to develop a mixed extreme learning machine (*MLELM*), enhancing prediction accuracy and generalization capability. Liu et al. [11] constructed a regularized modeling framework (*TCMo*R), incorporating physical characteristics into model training to strengthen feature representation.

In the field of visual inspection, Lin *et al.* [12] employed a U-Net with a pseudo-label mechanism to improve tool image segmentation accuracy. Zhao *et al.* [13] introduced lightweight deep network architectures to achieve efficient end-to-end segmentation. Li *et al.* [14,15] applied transfer learning to small-sample image recognition tasks and achieved promising results.

In integrating deep learning with transfer modeling, Fang *et al.* [16] proposed a TWM hybrid model with dual knowledge embedding, combining data augmentation and physical constraints to improve prediction accuracy. Wang *et al.* [17] developed a denoising Transformer auto encoder (*DTAE*), enhancing multisensor feature extraction and noise robustness. Huang *et al.* [18] adopted a domain-adversarial domain confusion network (*DADCN*) to strengthen cross-domain modeling capability. Wang *et al.* [19] introduced a pretraining strategy integrating MP-CNN and SCINet, achieving superior performance under cross-data set conditions.

Despite these advances, several limitations remain. First, limited practical adaptability: most approaches rely on public datasets with fixed cutting parameters and materials (e.g., PHM2010 [20-23]), and only a few studies (e.g., [24]) have conducted multi-condition validation. Even so, these studies are largely restricted to mechanical sensors and fail to capture the complexity of coated end milling in industrial environments, including material variability and dynamic cutting speed. Second, over reliance on single wear indicators: existing studies typically focus either on state classification [20,22,25-27] or on remaining useful life (*RUL*) prediction [23,24], lacking a unified framework. Third, insufficient exploration of joint time-frequency features: many studies rely solely on statistical features [21] or single models [10], overlooking the synergy of time-frequency information, which results in incomplete feature representation.

Quantitative comparisons further illustrate performance differences. Zhao *et al.* [28] proposed a BWO-LSTM-AdaBoost model for tool wear prediction, reporting an RMSE of 6.471. Xiao *et al.* [29] developed an ISCSO-LSTM model that achieved lower errors, with an MAE of 2.847 and RMSE of 3.6799, demonstrating higher accuracy and robustness. Guo *et al.* [30] applied a TCN-BiLSTM model for bearing wear prediction, yielding an *MAE* of *8.1* and *RMSE* of *9.6*, highlighting variability across conditions. In classification tasks, Wang *et al.* [31] reported an average accuracy of *92.6%* using ResNet-18 for tool wear recognition, whereas Tong *et al.* [32] achieved *99.10%* accuracy on bearing recognition with a multiscale network model.

In summary, current research has advanced

prediction accuracy and feature representation through deep learning and multimodal approaches. Nevertheless, challenges remain in adapting models to complex working conditions, developing unified frameworks that integrate multiple wear indicators, and fully exploiting joint time–frequency features. These gaps point to key directions for future research.

1.3 The Main Contributions of This Paper

To address the above challenges, this paper proposes a deep fusion tool wear prediction method for multi-source heterogeneous signals, named LSTM-MSCNN. The main contributions are as follows:

Multi-source heterogeneous signal integration: Five types of sensor signals, including cutting force, torque, vibration, acoustic emission, and visual data, are integrated to comprehensively capture tool wear information, overcoming the limitations of single-modal or single-index approaches.

Advanced feature modeling: LSTM and MSCNN are combined to capture long-term temporal dependencies and extract multi-scale spatial features. A unified time–frequency representation containing 13 time-domain descriptors and 8 frequency-domain descriptors is constructed, enhancing nonlinear modeling capacity and feature richness.

Industrial-relevant validation: Experiments are conducted on the QIT-CEMC coated end mill data set, covering diverse cutting parameters to better reflect industrial applications. Compared with baseline methods such as LSTM, ResNet, MSCNN, and existing literature [23,33]($MAE = 6.15$ [23], $RMSE = 9.369$ [33]), the proposed LSTM-MSCNN demonstrates superior accuracy and generalization ability, highlighting its industrial applicability.

In summary, this work addresses the deficiencies of single-modal and single-index modeling, validates the effectiveness of multi-source deep fusion under complex real-world conditions, and provides a new solution for tool condition monitoring and prediction in intelligent manufacturing.

2. System Model Design

2.1 Data Acquisition and Data Set Analysis

2.1.1 Construction of data acquisition platform

To obtain tool wear data reflecting multi-source and heterogeneous signal characteristics, such as cutting force, torque, visual imagery, vibration, and acoustic emission, *QIT* constructed a multi-modal data acquisition system under controlled milling conditions. The data set is called *QIT-CEMC* and is publicly available through the international authoritative journal Scientific Data. It can be accessed at https://www.qlit.edu.cn/datasets2/.

The experimental setup employed a vertical machining center (VDF-850) manufactured by Dalian Machine Tool Group as the primary processing equipment. This machine provides a maximum spindle speed of 8000 rpm and a 1700 mm/min peak feed rate. The machining tool employed is a coated four-edge end mill, and the workpiece material is titanium alloy, which conforms to the machining characteristics of typical difficult-to-machine metals.

Cutting force signals were captured using a Kistler 9170B251 rotary dynamometer mounted on the tool holder. The signals were converted by a charge amplifier and transmitted to a host computer for real-time force waveform monitoring.

Vibration signal and acoustic emission data were obtained through the DH5922D triaxial accelerometer and the acoustic emission sensor installed on the machine spindle. These were integrated into the HS922D dynamic signal acquisition system (Donghua Hi-Tech), which interfaced with the host system via USB for synchronized data display and logging.

Tools were removed post-machining for image acquisition, and the bottom and peripheral cutting edges were inspected using a 19Jc universal tool microscope. Wear features were manually annotated, and image clarity was enhanced by adjusting the focal plane through QMS3D-M (v2.3.3) software to ensure clear visualization of the wear profiles. The measurement procedure is illustrated in Figure 1.

To guarantee data consistency and experimental repeatability, all milling operations were executed automatically using pre-defined tool paths and machining programs. Detailed machining parameters and sample statistics are summarized in Table 1.

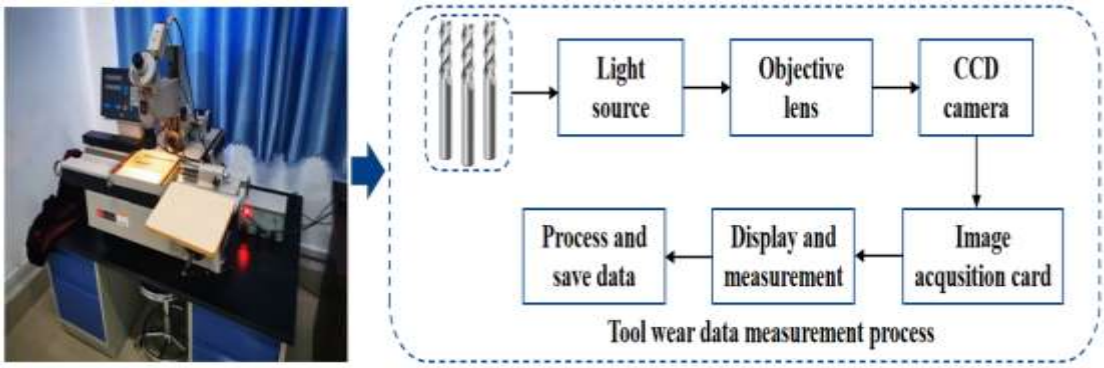


Figure 1 The description of a tool wear measurement procedure

Table 1 Tool wear experiment platform processing parameters

Parameter	Specification	Paramder	Value
Machine Tool	VDF-850 Vertical Machining Center	Spindle Speed (r/min)	1430
Tool Type	VSM-4E-D10.0	Feed Rate (mm/min)	130
Vibration and Sound Sensor	Donghua HS922D	Sampling Resolution/mm	0.2
Rotating Dynamometer	Kistler 9170B251	Cutting Width/mm	8
Data Acquisition System	WebDAQ-504	Sampling Frequency/kHz	10
Wear Measurement Instrument	19JC Digital Tool Microscope	Number of Machining Instances	68

2.1.2 Data Introduction and Analysis

The *QIT-CEMC* multi-modal tool wear data set contains approximately 40 GB of data, covering five types of machining signals: cutting force, spindle torque, image, vibration, and acoustic signals. These signals comprehensively reflect the multi-physical characteristics and evolution of tool wear during milling processes. The tool wear health indicators CSV file includes 12 initial columns, each corresponding to the maximum flank wear width (*VBmax*), half-depth wear width (*VB in 1/2* a_p), and wear area (*S*) of four cutting edges. Subsequent columns contain maximum flank wear widths (*VBmax*) and wear areas (*S*) at the tooth tips of these cutting edges, as shown in Figure 2. A total of 68 experiments were conducted in this study, with each experiment recording corresponding multi-source signals and tool wear labels. To ensure data integrity and modeling quality, the original data was defined based on acquisition process records and data quality checks. After data cleaning, 59 experimental datasets were retained and organized into training and test sets. Each data set underwent microscope inspection to match with tool wear labels, following standard microscopic imaging and manual labeling protocols for accuracy. Figure 3 illustrates the distribution of wear labels in the cleaned data set, which can be used for analyzing sample distributions and subsequent model training and evaluation.

Side teeth

Cutting edge1			Cutting edge2			Cutting edge3			Cutting edge4		
VBmax (mm)	1/2ap (mm)	S(mm²)	VBmax (mm)	1/2ap (mm)	S(mm²)	VBmax(mm)	1/2ap (mm)	S(mm²)	VBmax (mm)	1/2ap(mm)	S(mm²)

End teeth

Cutting edge1		Cutting edge2		Cutting edge3		Cutting edge4	
VBmax (mm)	S(mm²)	VBmax (mm)	S(mm²)	VBmax (mm)	S(mm²)	VBmax (mm)	S(mm²)

Figure 2 Data set structure

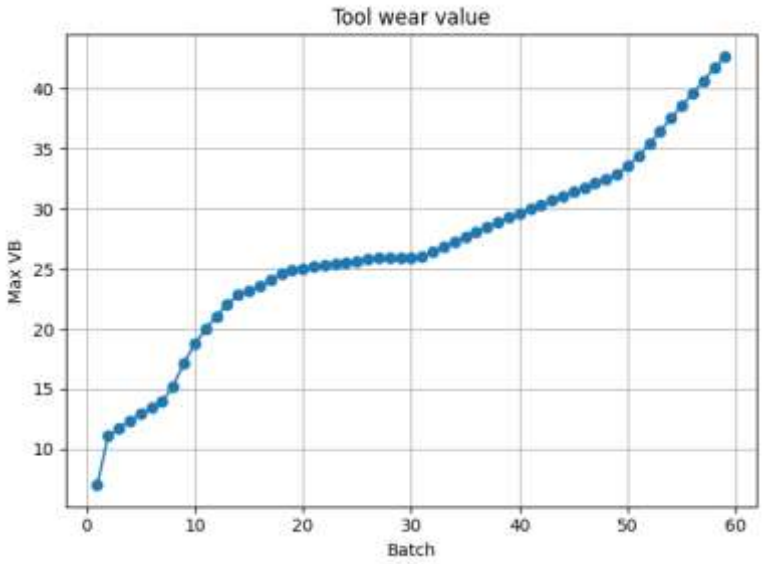
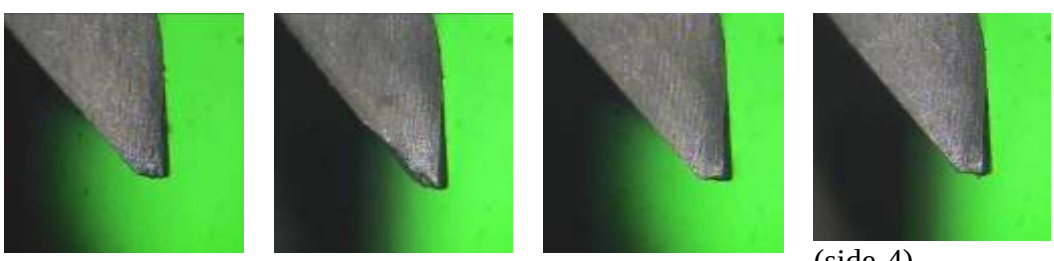


Figure 3: Visualization of 59 groups of labels

In order to standardize the image dimensions and streamline the model processing, the original image with a resolution of pixels was resized to pixels. Besides, the RGB images were converted to a single-channel gray-scale representation to reduce the computational complexity. For each experimental trial, a se-quence of eight images was compiled and stacked to create an 8-channel input tensor. These tensors were saved in Tensor format to ensure compatibility with the multimodal model. For illustrative purposes, the 25th image in the wear sequence was randomly selected and presented in Figure 4.



(side-1) (side-2) (side-3) (side-4)

Figure 4 8-channel tensor constructed from microscopic images

2.2 Multi-modal time-frequency feature extraction

To fully explore the temporal and spectral characteristics of the multimodal signals, this paper conducts joint feature extraction from force, torque, vibration, and acoustic data across both time and frequency domains. The resulting features are integrated to construct a robust and informative vector that effectively represents the tool wear condition, serving as the foundational input for the deep learning model.

2.2.1 Time domain feature extraction

Time-domain characteristics directly reflect the statistical behavior of signals over time, providing strong interpretability while maintaining low computational complexity. In this paper, 13 typical time-domain features were extracted from each sampled signal $x_n = \{x_1, x_2, \cdots, x_i, \cdots, x_n\}$, where n denotes the signal length. The corresponding mathematical formulations are as detailed below.

$$Max: x_{max} = max\{x_1, x_2, \cdots, x_n\} \quad (1)$$

$$Min: x_{min} = min\{x_1, x_2, \cdots, x_n\} \quad (2)$$

$$Mean: \mu = \frac{1}{n}\sum_{i=1}^{n} x_i \quad (3)$$

$$Std: \sigma = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)} \quad (4)$$

$$RMS: x_{RMS} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \quad (5)$$

$$Peak: x_{Peak} = max|x_i - \mu| \quad (6)$$

$$Peak-to-Peak: x_{P2P} = x_{max} - x_{min} \quad (7)$$

$$Variance: Var = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \quad (8)$$

$$Skewness: Skew = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^3 \quad (9)$$

$$Kurtosis: Kurt = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - \mu}{\sigma}\right)^4 \quad (10)$$

$$ShannonEntropy: H(x) = -\sum_{i=1}^{n} p(x_i)\log p(x_i) \quad (11)$$

$$Std(Sin): Std_{(Sin)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(sin(x_i) - \overline{sin(x)})^2} \quad (12)$$

$$Std(Arctan): Std_{Arctan} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(arctan(x_i) - \overline{arctan(x)})^2} \quad (13)$$

2.2.2 Frequency domain feature extraction

Frequency-domain characteristics contain information on signal periodicity, resonance patterns, and spectral distribution, which are critical for identifying the corresponding frequency characteristic changes across different stages of tool wear. This paper transforms the

original time-domain signals into their spectral representation using the Fast Fourier Transform (*FFT*). Then, eight frequency-domain characteristics are extracted to determine the signal's frequency behaviour. The corresponding computational formulas are presented below:

$$MeanFrequency: f_{mean} = \frac{\sum_{i=1}^{n} f_i P(f_i)}{\sum_{i=1}^{n} P(f_i)} \quad (14)$$

$$MedianFrequency: \sum_{i=1}^{k} P(f_i) = \frac{1}{2}\sum_{i=1}^{n} P(f_i) \quad (15)$$

$$TotalPower: P_{total} = \sum_{i=1}^{n} P(f_i) \quad (16)$$

$$OccpiedBandwidth: \alpha = 90\% \quad (17)$$

$$PowerBandwidth: B_p = f_{high} - f_{low} \quad (18)$$

$$MaxPSD: PSD_{max} = max(P(f_i)) \quad (19)$$

$$MaxAmplitude: A_{max} = max(|x(f_i)|) \quad (20)$$

$$FreMaxAmplitude: f_{A_{max}} = \arg max(|x(f_i)|) \quad (21)$$

f_i is the median frequency; $P(f_i) = |x(f_i)|^2$ represents the power spectral density.

2.3 LSTM-MSCNN Multimodal Fusion Network Model

To effectively capture both temporal patterns and spatial features within sequential data, this paper proposes a heterogeneous deep prediction model--LSTM-MSCNN--which integrates LSTM with an MSCNN. As illustrated in Figure 5, the model adopts a dual-branch architecture to process time-series and image data independently. It enables complementary feature extraction and deep multimodal integration, the robustness and predictive accuracy.

Time-series signals such as vibration, acoustic emission, and cutting force in the input stages are initially transformed into the frequency domain using FFT to extract spectral features. These features are then input into a two-layer stacked *LSTM* network to capture long-range temporal dependencies and complex nonlinear behaviours inherent in the signal dynamics. In parallel, the collected image data (e.g., images of worn tool surfaces) undergo pre-processing steps such as normalisation and grayscale conversion. Then, the processed images are input into the MSCNN branch, where convolution kernels of varying scales extract spatial features across multiple receptive fields, enabling the model to learn from localised texture variation and structural patterns.

The outputs from both the *LSTM* and the *MSCNN* branches are transmitted to the feature fusion layer. Here, cross-modal information is integrated through feature concatenation and aligned in dimension to ensure coherence. The fused features are then linearly transformed and reduced in dimensionality to eliminate redundancy, followed by flattening and input into a fully connected regression layer to generate the predicted tool wear value.

The architecture of the *LSTM-MSCNN* model is specially designed to address the dynamic modelling of temporal features and the multi-scale representation of spatial features. This enables effective capture of spatiotemporal coupling patterns in complex sequences, making it particularly suitable for predictive tasks involving strong temporal dependencies and spatial variability.

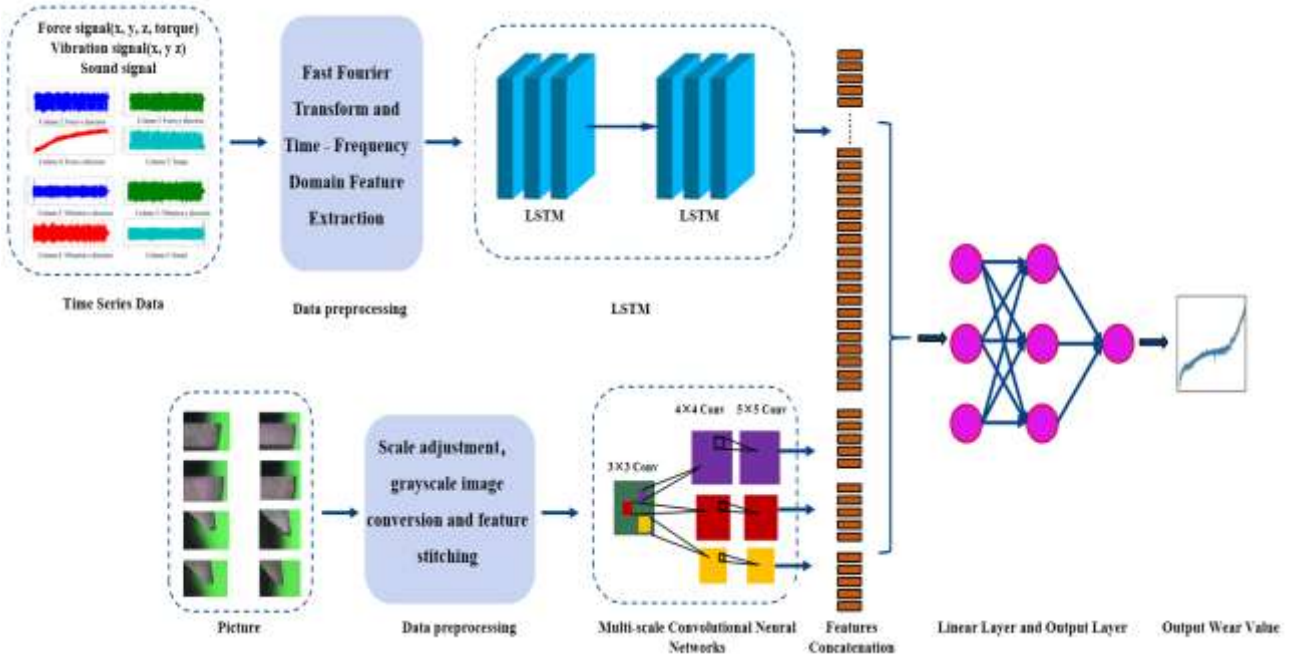


Figure 5 LSTM-MSCNN multimodal fusion network model

3. Experimental Design

3.1 Experimental Setup

In the model training process, the mean square error loss function (*MSE_Loss*) is selected as the regression objective function to enhance the prediction accuracy. The training adopts a sample-by-sample updating strategy with a batch size of 1, the initial learning rate is set to 0.00003, and the fixed learning rate scheduling approach. The model is trained over 150 epochs. The data set is divided into a training set and a test set according to the ratio of 7:3 to ensure the model evaluation's reliable performance and generalisation ability.

3.2 Evaluation index

This paper adopts the *MAE* and *MSE* as the primary evaluation metrics to assess the regression performance of the proposed and comparative models. *MAE* measures the average absolute difference between the predicted and actual values of the model. Its mathematical definition is expressed as follows:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left| f_i - f_i \right|^2 \quad (22)$$

Where N denotes the total number of samples, f_i is the actual value, and f_i is the predicted value of the model. *MAE* reflects the actual average amplitude of the error; the smaller the *MSE* value, the higher the accuracy of the model.

In contrast, *MSE* emphasises larger prediction errors by squaring the residuals, thereby penalising outliers more heavily. The metric is formally defined as:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(f_i - f_i)^2 \quad (23)$$

3.3 Network Parameters of the LSTM-MSCNN Model and Comparison Model

MSE is more sensitive to the samples with large deviations in prediction errors and reflects the model's stability in extreme or outlier values.

Through the *MSE* and *MAE* indicators, this paper offers a comprehensive evaluation of the model's accuracy and consistency. Together, these metrics support a fair and transparent comparison of model performance, highlighting the relative strengths and weaknesses of the approaches under investigation.

This paper proposes the *LSTM-MSCNN*, a hybrid deep learning model that integrates temporal sequence modelling with multi-scale spatial feature extraction. Five widely recognised models, *LSTM*, *BiLSTM*, *ResNet*, *CNN*, and *MSCNN*, are selected as baselines for comparative analysis to evaluate the proposed model's effectiveness. These models are also chosen because they are commonly used in industrial applications. The details of network architecture and parameter configurations for all models are outlined below.

(1) LSTM-MSCNN (Ours)

The proposed model integrated the temporal modelling capability of *LSTM* and the spatial perception strengths of M*SCNN* to construct a heterogeneous feature learning structure. The *LSTM* branch employs a two-layer stacked architecture, with an input and hidden layer dimension of 100. The output is flattened to generate a compact time series feature vector. Meanwhile, the *MSCNN* branch extracts multi-scale spatial information through convolutional kernels with varying sizes $(3\times3, 4\times4, 5\times5)$. The receptive field is further enhanced by combining different stride sizes and pooling operations. After feature concatenation and convolution dimensionality reduction, the outputs are merged with those from the time series branch. The final prediction is generated through a fully connected layer.

(2) LSTM

A standard two-layer stacked *LSTM* network is constructed to model temporal dependencies in the input sequence. The hidden dimension of each layer is 100, and a fully connected layer generates the predicted values following a flattening process. However, this approach lacks the ability to capture spatial dependencies in the data.

(3) BiLSTM

The bidirectional *LSTM* architecture enhances the modelling of forward and backward dependencies within the temporal features. Integrating the outputs from both directions improves the ability to expressiveness of the time-series feature. However, this structure does not incorporate spatial structure information.

(4) ResNet

Based on residual networks, deep convolution layers and residual connections are used to extract high-level spatial features from image-based data. Feature extraction is achieved through four sets of residual blocks, and final prediction results are produced through global average pooling followed by a fully connection layer.

(5) CNN

A single-scale convolutional network is constructed using different convolution kernel sizes to extract localized spatial features. Leaky *ReLU* is used as the activation function, and stride convolution is employed to achieve down-sampling. However, this model does not take into consideration the temporal correlation in the data.

(6) MSCNN

The multi-scale convolution structure is similar to the spatial branch of the proposed architecture. Different convolution kernels and stride combinations are integrated to extract spatial features at multiple scales. However, like ResNet-18 and traditional *CNN*, it does not incorporate temporal sequence modelling. Details of each model's core components-such as the spatio-temporal feature processing mechanism, convolution kernel design, and parameter settings for the full connected layer -are provided in Table 2.

Table 2 Blocks and parameter configuration of different models

Model	Blocks	Time-Spatial feature processing	FC layer dimension
LSTM-MSCNN (Ours)	LSTM+ MSCNN	T: Two-layer LSTM(100-dimensional) 3/4/5×5 S: MSCNN+stitching+ dimension reduction	18169→100→1
LSTM	Simple LSTM network	T: Two-layer LSTM (100-dimensional) S: --	16800→100→1

BiLSTM	BiLSTM network	T: Double BiLSTM (two-way splicing) S: --	33600→100→1
ResNet-18	Residual convolution block	T: -- S: 3×3 residual convolution (4 basic blocks)+pooling	16→1
CNN	Single scale convolution	T: -- S: Multi-branch convolution (3/4/5/6×5) + Phased downsampling	10952→ 50→1
MSCNN	MSCNN+fusion	T: -- S: 3/4/5×5 MSCNN+ stitching +dimension reduction	1369→ 50→1

Explanation: T: Time feature; S: Spatial feature.

4. Experimental Results and Discussion

4.1 Performance Comparison of Tool Wear Prediction

In order to evaluate the prediction performance of the proposed *LSTM-MSCNN* model, two widely recognized indicators, *MSE* and *MAE*, were employed to assess the error magnitude and robustness of the model on the *QIT-CEMC* data set. Table 3 shows the results of the compared models, including *LSTM*, *BiLSTM*, *ResNet*, *CNN*, *MSCNN*, and the proposed model in this paper. As shown in Table 3, the proposed *LSTM-MSCNN* model achieved an *MSE* of 0.406 and an *MAE* of 0.286, outperforming all other benchmark models. These results indicate that the proposed model has higher accuracy and better generalisation in tool wear prediction tasks. A detailed comparison of the predictive error across different models on the *QIT-CEMC* data set is summarised in Table 3.

Table 3 The prediction errors of the compared models on the QIT-CEMC data set

No.	Model_name	Average_MSE	Average_MAE
1	LSTM	10.501	1.645
2	BiLSTM	15.177	2.013
3	ResNet	77.741	6.664
4	CNN	22.532	3.742
5	MSCNN	15.333	2.700
6	LSTM-MSCNN（Ours）	0.406	0.286

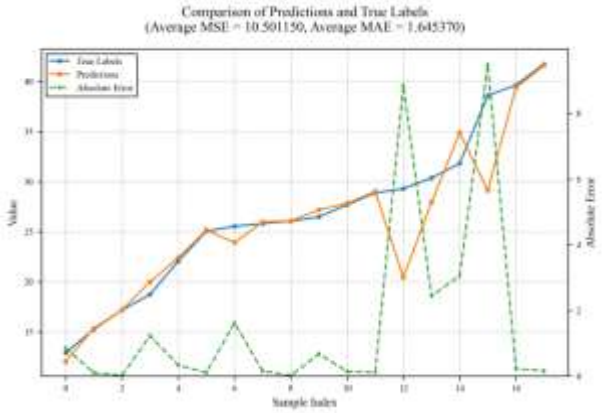
The ablation experiments are as follows:The ablation study results presented in Table 4 demonstrate that the *LSTM-MSCNN* model exhibits marginally lower performance than LSTM when using time-domain features, whereas it achieves the highest performance among the three models when using frequency-domain features.

Table 4 Error comparison of Time/Frequency domain features

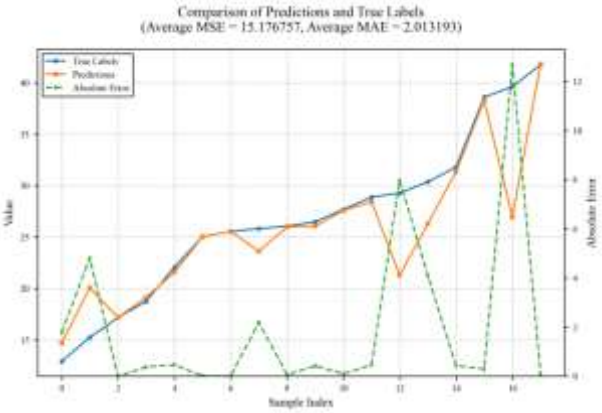
Model_name	Only Time-domain Features		Only Frequency-domain Features	
	Average_MSE	Average_MAE	Average_MSE	Average_MAE
LSTM	2.026	1.184	58.146	6.198
BiLSTM	34.370	3.929	55.695	5.887
LSTM-MSCNN（Ours）	12.333	1.560	34.984	4.730

Figure 6 (a)-(f) shows the prediction performance of different models on the test set. LSTM can capture the overall upward trend of wear, but there is a systematic overestimation, with local absolute error peaks reaching 10-15 (*MSE =10.501*, *MAE=1.645*); the errors of BiLSTM and

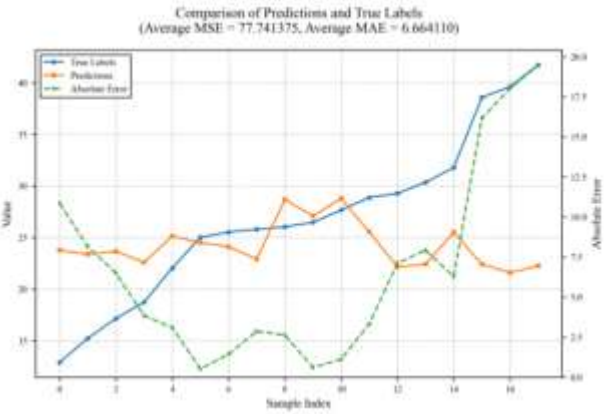
CNN are even larger, reaching *MSE = 15.176*, *MAE =2.013* and *MSE = 22.531*, *MAE=3.742* respectively, making it difficult to accurately depict nonlinear mutations and long-range dependencies. MSCNN has improved in multi-scale feature extraction, but there is still a persistent deviation in the high wear range (*MSE =15.333*, *MAE=2.700*). ResNet performs the worst, with the prediction curve being disconnected from the true value (*MSE =77.741*, *MAE= 6.664*). In contrast, the prediction curve of LSTM-MSCNN almost coincides with the true curve, with only minor deviations in a few intervals. Its MSE and MAE are 0.406 and 0.286 respectively, significantly outperforming other models, verifying the effectiveness of integrating temporal dependencies and multi-scale features.



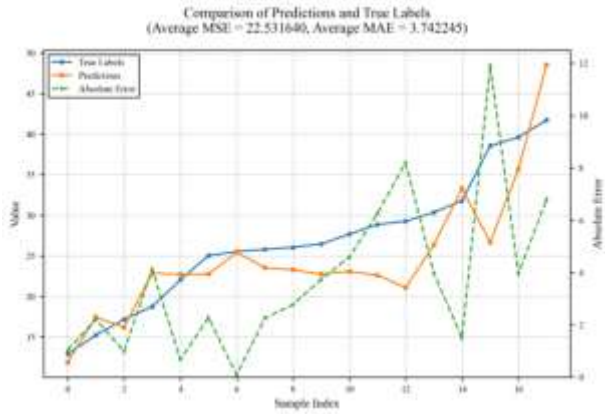
(a) LSTM



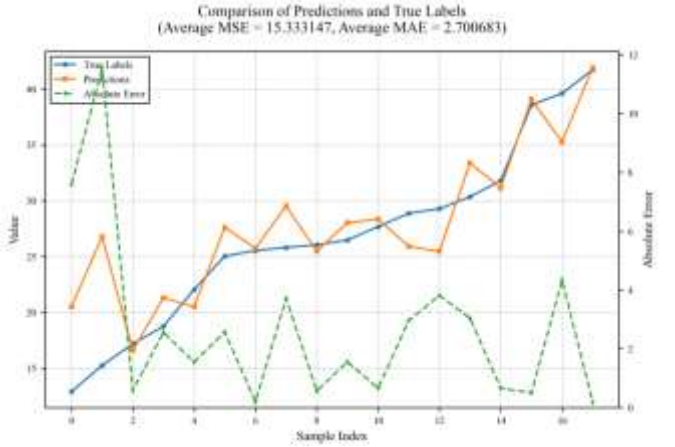
(b) BiLSTM



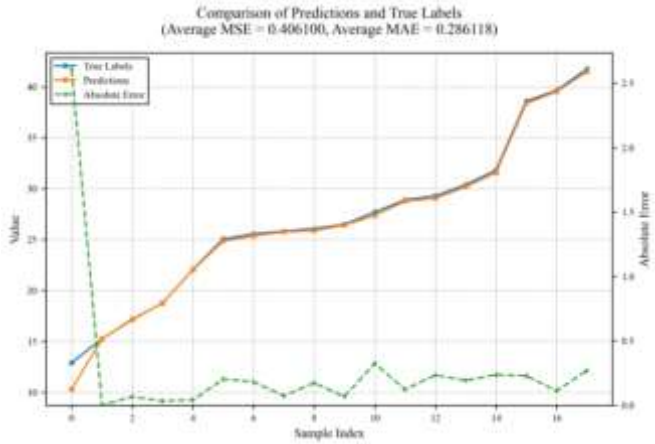
(c) RESNet



(d) CNN



(e) MSCNN

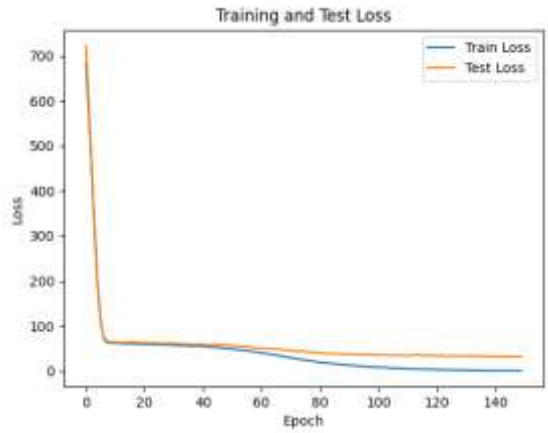


(f) LSTM-MSCNN(ours)

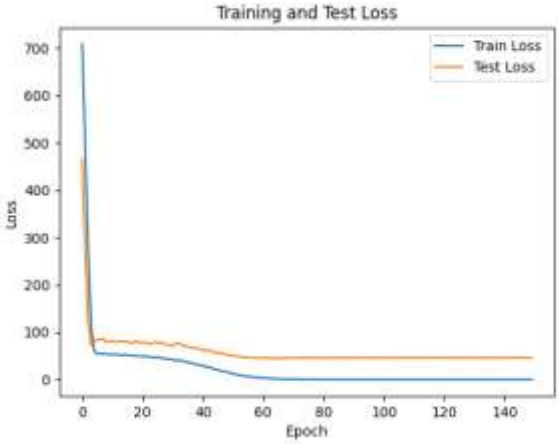
Figure 6 Comparison curve between the predicted output of different models and real labels

4.2 Analysis of Model Training and Testing Loss

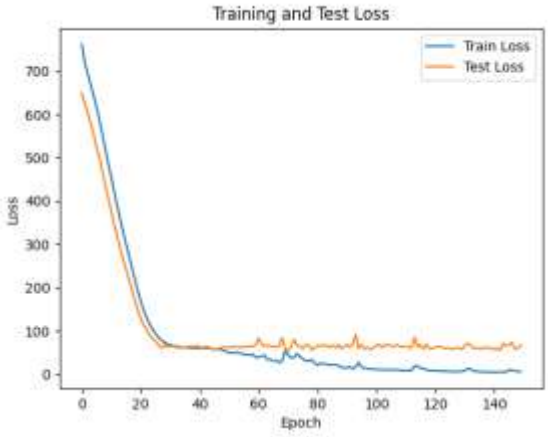
To further analyse the models' training process and generalization capability, Figure 7 shows the training and testing loss trend for each model throughout the training process. As observed, the proposed *LSTM-MSCNN* model exhibits fast convergence and a minimal gap between the training and testing curves. This indicates mild over fitting and strong generalization ability. In contrast, the test loss curves of the *ResNet* and *CNN* models exhibit significant fluctuations, reflecting lower stability and suboptiml convergence behaviour compared to the proposed model.



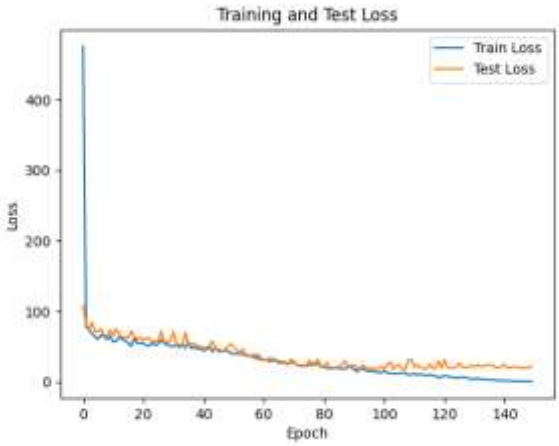
(a) LSTM



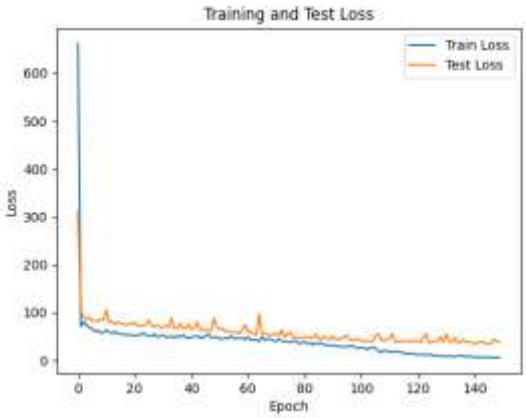
(b) BiLSTM



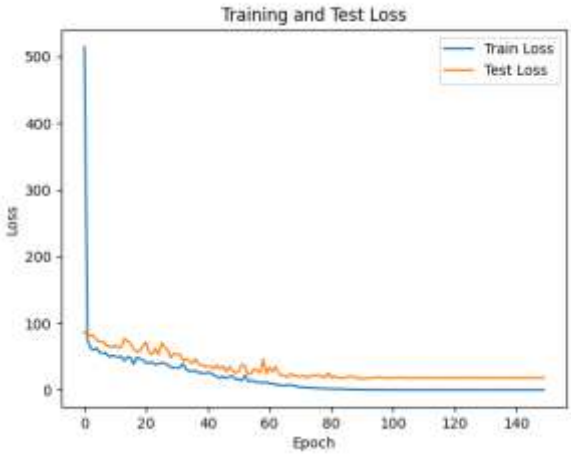
(c) RESNet



(d) CNN



(e) MSCNN



(f) LSTM-MSCNN(ours)

Figure 7 Comparison of training loss and test loss curves of different models

5. Conclusions

This study proposes a hybrid *LSTM-MSCNN* model for tool wear prediction in complex machining environments, combining *LSTM's* temporal modelling with *MSCNN's* multi-scale spatial feature extraction. The model outperforms benchmark approaches, achieving high accuracy, stable convergence, and strong agreement between predicted and actual wear curves. These results demonstrate its potential for real-time, multi-modal monitoring in intelligent manufacturing. Future work will focus on integrating wear mechanism knowledge to enhance interpretability and incorporating uncertainty modelling and attention mechanisms to improve adaptability.

Funding: The work was supported in part by the Key Research and Development (Major Scientific and Technological Innovation) Project of Shandong Province under Grant No. 2019TSLH0202 and No. 2019JZZY01011, and in part by the Guangdong Basic and Applied Basic Research Foundation under Grant No. 2021B1515120066.

Conflict of Interest Statement

The authors declare no conflict of interest.

References

1. LI Kai, LI Zhou-long, JIA Xian-shi, et al. (2024). A domain adversarial graph convolutional network for intelligent monitoring of tool wear in machine tools, Computers & Industrial Engineering, 187, 109795. https://doi.org/10.1016/j.cie.2023.109795
2. REN Ze-yu. (2019). Research status and development trend of tool wear monitoring technology,China Plant Eng, 2,76-77. https ://kns. cnki.net.
3. LI Xin, ZHANG Yu, ZHU Kun-peng. (2022). Tool wear monitoring method based on s-transform time-frequency characteristics, Modular Machine Tool & Automatic Manufacturing Technique, 10, 88-91+96. http://doi.org/10.13462/j.cnki.mmtamt.2022.10.019
4. TIAN Ying, WANG Wen-hao. (2021). A method for tool condition monitoring based on spindle power, J. Tianjin Univ.(Sci. Technol.), 54(11), 1179-1186. https://kns.cnki.net.
5. LI L, YAN Q S, LI K. (2023). Tool wear online monitoring during shearing process strip based on acoustic emission signal, Journal of Mechanical & Electrical Engineering, 40(7), 1102-1111. https://kns.cnki.net.
6. LI Shan-shan, LIU Li-bing, LI Li, et al. (2017). A method of CNC tool wear condition monitoring based on region growing arithmetic, Manufacturing Technology & Machine Tool, 66(2), 132-136. http://doi.org/10.19 287/j.cnki.1005-2402. 2017.02.025
7. ZHAO Dong-xu, YUAN Zhi-xiang, YI Si-guang, et al. (2024). Tool wear identification based on dual-channel convolutional information fusion, Modern Manufacturing Engineering, 45(1), 124-129. http://doi.org/10.16731/j.cnki.1671-3133.2024. 01.018
8. DI Zijun, YUAN Dongfeng, LI Dongyang, et al. (2024). Tool Fault Diagnosis Method Based on Multiscale-Efficient Channel Attention Network, Journal of Mechanical Engineering, 60(6), 82-90. https://kns.cnki.net.
9. Y. Quan, C. Liu, Z. Yuan, and B. Yan. (2024). Hybrid Data Augmentation Combining Screening-Based MCGAN and Manual Transformation for Few-Shot Tool Wear State Recognition, IEEE Sensors Journal, 24(8), 12186-12196. http://doi.org/10.1109/JSEN.2024.3372438.
10. S. Chauhan, R., Trehan, R. P. Singh, and V. S. Sharma. (2024). Assessment of Machining Performance for Intelligent Tool Wear Prediction Using Hybrid Extreme Learning Machine, IEEE Sensors Journal, 24(22),37915-37922. https://doi.org/10.1109/JSEN.2024.3458394
11. Z. Liu, Z. Q. Lang, Y. Gui, Y. P. Zhu, H. Laalej and D. Curtis. (2024). Vibration Signal-Based Tool Condition Monitoring Using Regularized Sensor Data Modeling and Model Frequency Analysis, IEEE Transactions on Instrumentation and Measurement, 73, 1-13. https://doi.org/10.1109/TIM.2023.3343825
12. Y. S. Lin and M. S. Tsai.(2025). Development of SAM-augmented U-Net model with Transfer Learning for Multiple Tool Wear Detection, IEEE Transactions on Instrumentation and Measurement, 74,1-8. https://doi.org/10.1109/TIM.2025.3527601
13. Zhao P, Li Z, You Z et al. (2024). SE-U-Lite: Milling Tool Wear Segmentation Based on

Lightweight U-Net Model With Squeeze-and-Excitation Module, IEEE Transactions on Instrumentation and Measurement, 73, 1-8. https://doi.org/10.1109/TIM.2024.3385833
14. Li R, Wei P, Liu X, et al. (2023). Cutting tool wear state recognition based on a channel-space attention mechanism, Journal of Manufacturing Systems, 69, 135-149. https://doi.org/10.1016/j.jmsy.2023.06.010
15. Aitha Sudheer Kumar, Ankit Agarwal, Vinita Gangaram Jansari, K A Desai, Chiranjoy Chattopadhyay, Laine Mears. (2025). Realizing on-machine tool wear monitoring through integration of vision-based system with CNC milling machine, Journal of Manufacturing Systems, 78, 283-293. https://doi.org/10.1016/j.jmsy.2024.12.004
16. Xiaohui Fang, Qinghua Song, Jing Qin, Zhenyang Li, Haifeng Ma, Zhanqiang Liu. (2025). A dual knowledge embedded hybrid model based on augmented data and improved loss function for tool wear monitoring, Robotics and Computer-Integrated Manufacturing, 92, 102901. https://doi.org/10.1016/j.rcim.2024.102901
17. Wang, H., Wang, S., Sun, W., & Xiang, J. (2024). Multi-sensor signal fusion for tool wear condition monitoring using denoising transformer auto-encoder Resnet, Journal of Manufacturing Processes, 124, 1054-1064. https://doi.org/10.1016/j.jmapro.2024.07.002
18. Huang, Z., Shao, J., Zhu, J., Zhang, W., & Li, X. (2024). Tool wear condition monitoring across machining processes based on feature transfer by deep adversarial domain confusion network, Journal of Intelligent Manufacturing, 35(3), 1079-1105. https://doi.org/10.1007/s10845-023-02088-2
19. Wang, Y., Gao, J., Wang, W., Du, J., & Yang, X. (2024). A novel method based on deep transfer learning for tool wear state prediction under cross-dataset, International Journal of Advanced Manufacturing Technology, 131(1), 171-182.https://doi.org/10.1007/s00 170-024-13055-3
20. Y. Huai, C. Chen, J. Shi, D. Yue and C. Bo. (2025). A Tool Wear State Recognition Model Based on Multi-Scale Convolutional Neural Network and Transformer Encoder, 2025 Joint International Conference on Automation-Intelligence-Safety (ICAIS) & International Symposium on Autonomous Systems (ISAS), Xi'an, China, 1-6. https://doi:10.1109/ICAISISAS 64483.2025. 11051483
21. T. Liu and D. Niu. (2025). Tool Wear Condition Monitoring Based on Multi-Sensor Data Fusion, 2025 37th Chinese Control and Decision Conference (CCDC), Xiamen, China, 155-160. https://doi:10.1109/CCDC 65474.2025.11090922.
22. L. Feng， S. Zhang， Z. Li and X. Feng，(2025). Research on Classification and Recognition of Micro Milling Tool Wear based on Improved DenseNet，IEEE Access, 13,65659-65671. https://doi10.1109/ACCESS .2025.3557057.
23. C. Dong and J. Zhao, (2025). An Augmented AutoEncoder With Multi-Head Attention for Tool Wear Prediction in Smart Manufacturing, IEEE Access, 12,79128-79137. https://doi: 10.1109/ACCESS.2024.3406568.
24. Y. Xie, J. Zhu, Y. Dai and C. Zhang, (2025). Multicondition Tool Wear Assessment for Cutting Tools Based on Kernel Principal Component Analysis and Integrated Transfer Learning, IEEE Transactions on Instrumentation and Measurement, 74, 1-13. https://doi: 10.1109/TIM.2025.3552442.
25. 4. A. Ur Rehman, T. Salwa Rabbi Nishat, M. Uddin Ahmed, S. Begum and A. Ranjan, (2025). Chip Analysis for Tool Wear Monitoring in Machining: A Deep Learning Approach, IEEE Access, vol. 12, pp. 112672-112689. https://doi:10.1109/ACCESS.2024. 34435
26. Y. -S. Lin and M. -S. Tsai, (2025). Development of SAM-Augmented U-Net Model With Transfer Learning for Multiple Tool Wear Detection, IEEE Transactions on Instrumentation and Measurement, 74, 1-8. https://doi: 10.1109/TIM.2025.3527601.
27. M. Bilal, R. Podishetti, L. Koval, M. A. Gaafar, D. Grossmann and M. Bregulla, (2024). Automatized End Mill Wear Inspection Using a Novel Illumination Unit and Convolutional Neural Network, IEEE Access, 12,124282-124297. https://doi:10. 1109/ACCESS.2024.3454692.
28. Zhao, X., Yang, W., Hu, S., Xun, K., & Tan, Q. (2024). Milling tool wear prediction based on improved LSTM-AdaBoost. Machine Tool & Hydraulics, 52(10), 14–20. https://doi.org/ CNKI:SUN:JCYY.0.2024-10-003.

29. Xiao, B., Li, Y., Duan, Z., & Chen, L. (2023). Tool wear prediction based on ISCSO-LSTM model. Modular Machine Tool & Automatic Manufacturing Technique, (06), 102–105, 110. https://doi.org/10.13462/j.cnki.mmtamt. 2023. 06.023
30. Y. Guo, J. Zhou, Z. Dong, H. She and W. Xu.(2024). "Rolling Bearing RUL Prediction Based on Fusion of Multi-Head Attention and Improved TCN-BiLSTM," in IEEE Access, vol. 12, pp. 95641-95658. https://doi:10. 1109/ACCESS.2024.3424521.
31. Wang, H., Wang, S., Sun, W., & Xiang, J..(2024).Multi-sensor signal fusion for tool wear condition monitoring using denoising transformer auto-encoder Resnet.Journal of Manufacturing Processes,124,1054-1064. https://doi.org/10.1016/J.JMAPRO.2024.07. 002.
32. J. Tong, C. Liu, J. Bao, H. Pan and J. Zheng.(2023). "A Novel Ensemble Learning-Based Multisensor Information Fusion Method for Rolling Bearing Fault Diagnosis," in IEEE Transactions on Instrumentation and Measurement, vol. 72, pp. 1-12, no. 9501712. https://doi: 10.1109/TIM.2022.3225910.
33. J. Yang, X. Yu, C. Hu, J. Liang, J. Duan and T. Shi, (2025). Hybrid Global and Local Neural Network With Lightweight Multiscale Fusion for Tool Wear Prediction, IEEE Transactions on Instrumentation and Measurement, 74,1-15. https://doi:10.1109 /TIM. 2025. 3579834.