Proceedings of the 8th International Conference on Communication and Electronics Systems (ICCES 2023)
IEEE Xplore Part Number: CFP23AWO-ART; ISBN: 979-8-3503-9663-8

# A Deep Learning-based Innovative Approach for Enhancing the Precision of Tool Wear Prediction

[1]Dr.D. vinod
*Department of Computing Technologies,*
*School of Computing,*
*College of Engineering and Technology,*
*SRM Institute of Science and Technology,*
*SRM Nagar, Kattankulathur, Chennai.*
dvinopaul@gmail.com

[2]Karimulla Syed
*Assistant Professor*
*Mechanical Engineering*
*Koneru Lakshmaiah Education Foundation-522501*
syedkarimulla@kluniversity.in

[3]Sampath Kumar S
*Assistant Professor*
*Computer Science and Engineering*
*Sri Eshwar College of Engineering, Coimbatore.*
Sampath.shanmugam@gmail.com

[4]Dr. B. Sankara Babu
*Professor, Department of CSE,*
*Gokaraju Rangaraju Institute of Engineering and Technology (Autonomous),*
*Bachupally, Telangana*
sankarababu.b@griet.ac.in

[5]Shatrudhan Pandey
*Research Scholar,*
*Department of Production and Industrial Engineering*
*Birla Institute of Technology, Mesra, Ranchi, 835215.*
er.shatrudhanp@gmail.com

[6]Dr. Raja Velur Loganathan
*Professor/ Department of mechanical engineering,*
*Loyola institute of technology, Chennai*
vlraja73@gmail.com

2023 8th International Conference on Communication and Electronics Systems (ICCES) | 979-8-3503-9663-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCES57224.2023.10192664

**Abstract- Throughout the cutting operation, among the most crucial steps is that has a direct impact on machining accuracy and product quality is tool wear. If you can accurately forecast how quickly tools will wear out, you can make the necessary adjustments early on, minimizing downtime and maximizing product quality. The evaluation of cutting tool wears during the production stage is crucial. The main objective of this study is to track tool degradation over time to guarantee regular tool changes and prevent workpiece waste and potential machine damage from excessive tool wear or unexpected tool breakage. For this reason, it is necessary to develop a system that is both intelligent and capable of providing accurate solutions to these issues. The criteria of intelligent production are great, though, and traditional methods simply cannot match them. Because of this, a deep learning-based innovative approach is presented to advance the precision of tool wear prediction. Parallel convolutional neural networks were constructed and used to achieve multi-scale feature fusion. To improve the model's efficiency, the channel attention mechanism was integrated with the residual connection to take into account the relative importance of each feature map. To prove that the created approach is superior, several experiments were carried out to anticipate tool wear, and the findings are more reliable and accurate than those obtained using existing methods. To guarantee the excellence of the engine cylinder and the steadiness of the machining procedure, a tool wear monitoring system was created. Using the PHM 2010 milling cutter wear dataset, tests were performed to validate the model effect. Based on the experimental data, this model has an average RMSE of 2.67 and an average MAE of 2.5. The trial outcomes proved the effectiveness of the suggested approach in evaluating the tool's wear.**

***Keywords— Tool Wear, Cutting Process, Parallel CNN, Attention Mechanism, Intelligent Production.***

## I. INTRODUCTION

More complex manufacturing is an indicator of a country's economic health [1]. In today's industrial sectors, milling is a critical operation, and optimizing the milling process is essential for realizing significant financial advantages and enhancing customer satisfaction. Factors including cutting tool, power, machining productivity, and machined surface finish have a significant impact on manufacturing costs [2]. The state of the cutting tools' wear has a significant impact on both the reliability of the machinery and the reliability of the finished goods throughout production. In addition, it is important to keep track of the tool's wear state so that you may change it at the appropriate time; replacing a tool too soon increases costs while doing it too late can degrade the eminence of the workpiece and potentially harm the machine. Cutting tools and replacements account for 3–12% of the entire processing cost in machining [3], whereas cutting-tool-related downtime can account for 20% [4]. Additionally, the tools experience a wide range of failure modes, including adhesion, attrition, cracking, dissemination, and ductile materials. It is hard to avoid tool wear due to the intricate mechanics involved in the phenomenon. Hence, in order to enhance and preserve the quality of the finished product, tool wear management is crucial in modern production. For these reasons, several tool wear monitoring research projects have focused on creating an automated and precise tool wear surveillance system capable of producing alerts for tool wear.

With the proliferation across industries and the rise in computer capacity, deep learning has gained traction as a powerful tool for machine learning and information algorithms. Because of its robust feature abstraction, feature merging, and concept abilities, it is increasingly being utilized for predicting tool wear [5]. This is because it does not rely on human experience but rather can directly extract characteristics from the unique data. He used a 1D convolutional neural network (CNN) to concatenate and pool the original data along the time series to produce the compressed timing characteristics, which he then sent into a bidirectional long short-term memory network (LSTM) to forecast tool wear. In order to anticipate tool wear, Chan et al. [6] presented Holistic-Local -LSTM, which employs CNN to extract characteristics from original data and employs the derived findings as input of LSTM. When raw data is fed into a network, however, the resulting model is often overfitted with an excessive number of parameters and may not perform optimally due to ambiguous features.

The deep learning (DL) approach, on the other hand, has an extraordinary feature learning capacity and can automatically extract characteristics from unprocessed information without the need for domain expertise. But the main point is that present designs like RNN, LSTM, and their variations have been shown to be difficult to train and reexamine, and they do not take spatial information into account. CNN also excels in feature extraction, but only in high-dimensional settings; it cannot handle low-dimensional data. These issues have an effect on the methods currently used to anticipate tool wear. The tapping technique is also one of the most often used in today's industrial sector. Tapping is the last step in the production process. The machined thread toughness and surface consistency are connected to the final assembly quality compared to the tapped thread on the object. Good morphological and dimensional accuracy and surface finishing with zero clearance are required for optimal assembly performance. Tap wear has a major effect on the qualities of a hole drilled, especially its surface smoothness and thread toughness. The performance of the assembly will suffer, and there's a risk of fracture failure on the thread if that happens. To the extent feasible, however, there is less scholarly attention paid to the problem of predicting tap wear.

Consequently, in order to excerpt more complex patterns from physical features [7], some researchers syndicate feature engineering with deep learning. There have been some successful outcomes. By the GAF imaging method, Martinez et al. [8] were able to transform the time series into pictures, which they then fed into a convolutional neural network (CNN) to use in training predictions and tool wear categorization. The wear on tool franks was forecasted by Marani et al. just use an LSTM-based model. [9], and the findings suggest that the best LSTM model confirms its potential to capture tool frank wear. In order to forecast tool wear, Zhao et al. [10] introduced local feature-based gated recurrent unit systems, which merged manually designed features with automated feature learning. In order to realize the adaptive excavation of tool wear functionalities and the categorization of wear extent, the initial motions were transformed into corresponding power spectra utilizing wavelet packet transmogrify, and a tool wear able to monitor model was constructed to use a DCNN. For accurate feature extraction of wavelet decomposition in several frequency bands, Duan et al.[11] presented a multi-frequency-band extracting features model based on a DCNN framework.

Automated tool wears identification in face milling using a CNN that can accurately determine wear rates has been created [12]. Hyperparameter modification was studied by Kothuru et al. [13] to enhance CNN's tool predictive maintenance in face milling. Dzulfkri et al. [14] have suggested a deep metric learning technique for diagnosing the state of stamping tools. Several DML methods were compared to find the most effective for evaluating stamping equipment. Overall, the scientists found that the triplet network produced the best outcomes. To effectively monitor tool status and anticipate tool wear, researchers have found that there is a shortage of sufficient experimental data to construct machine learning models. When deep learning models need to be constructed, the work becomes more difficult since the models require extensive empirical training information.

## II. METHODOLOGY

Convergence issues of deep networks may be circumvented and network deterioration brought on by layering can be avoided by using a deep CNN network with residual blocks. Fig.1 depicts the learning process as a series of nonlinear computation layers in a constant stack trying to fit the residual here between input data and the output data.

$$G(x) = f(x) + X \tag{1}$$

Residual learning builds on top of the standard linear network architecture by fusing a shortcut into the main path using an additive technique.

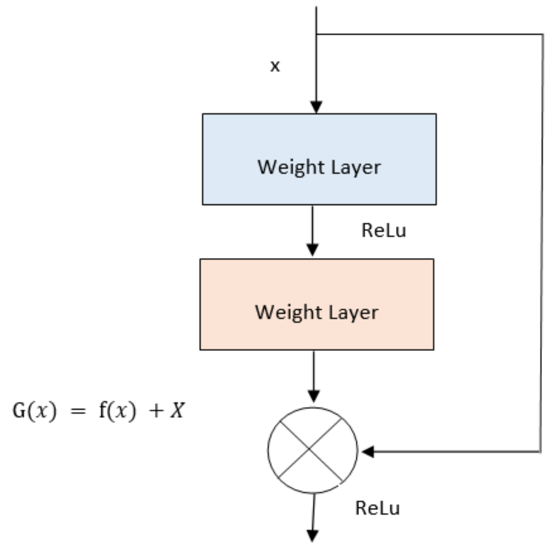

Fig.1. Residual Module

A total of 16 residual blocks, each with its own set of 38 convolution layers and 18 maximum pool layers, are fed the original input from the previous layers. Three 1D convolutional layers, a batch normalisation layer, a ReLU activation layer, a dropout layer, and a maxpooling layer make up one of the two types of residual modules. There are $32\times2^k$ convolution kernels in each convolutional layer. The second through sixteenth residual blocks differ from the first in that they contain additional batch normalisation layers, ReLU activation levels, and dropout layers. When the fast connection and the second convolutional layer have completed their work, the residual module merges the two sets of results by summing them. The length of the feature map is halved when it travels through a max-pooling layer with a pool size of 2. With a pool size of 1, the feature map is unaffected, hence just eight layers are used in this section of ResNet. In the process of local edge detection, the input is down sampled by a frequency of 28.

The primary function of the convolutional layers is feature extraction from input data. Specifically, it has a functional form of

$$I_i^k = \sum_{l \in N_i} I_l^{k-1} \cdot w_{li}^k + b_i^k \tag{2}$$

where $I_i^k$ is the lth layer's $i$th feature map, and $I_i^{l-1}$ is the $k$th layer's $l$th feature map as an output. The convolution operation seen among $l$ th feature space of the for mer layer and the $i$th feature map is denoted by $I_i^k$. The effort feature size, $N_i$, is represented by $b_i^k$, while $b_i^k$ is the matching bias. That's why we say a 1D CNN is

$$I_i^k = \sum_{l \in N_i}^{N_{i-1}} Conv1D(I_l^{k-1} \cdot w_{li}^k + b_i^k) + w_l^k \qquad (3)$$

By reducing the number of parameters while keeping translational invariance, the pooling layer reduces the dimensionality of the feature map. Maximum pooling and average pooling are two examples of the broader pooling techniques. Maximum pooling was used to process each feature map and pick the strongest features for this paper. It has a mathematical expression of:

$$Q_i = \max_{I_i^k \in p} I_i^k \qquad (4)$$

The activation function, which is not continuous As the patterns learned from the data input are typically asymmetric, we employed ReLU as defined by.

$$ReLu(z) = \begin{cases} 0, & z < 0 \\ z, & z > 0 \end{cases} \qquad (5)$$

### *A. Attention Mechanism*

The attention model has been presented to examine the intrinsic properties of data and increase the competence of data dispensation. By assigning different weights to transcripts and encoding the information more effectively, neural translator machines are one illustration of a technology that uses an attention mechanism. It offers the framework for versions of the following attention methods. It is well-known that the traffic-flow fascinating discover at different times may not be similarly appropriate for forecast..

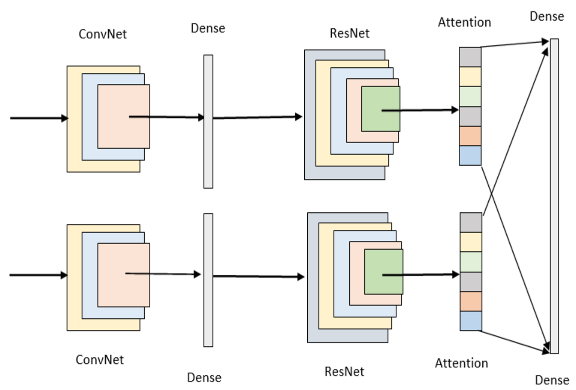


Fig.2. Parallel Deep learning model with attention mechanism

One of the main ideas behind attention mechanisms is to model the way in which human attention works in order to boost deep learning's efficiency. In order to construct an output sequence from an input one, we may use the posterior distribution of attention to adjust the weighting parameters of the items in the input pattern. The essential of the attention function may be defined as a mapping from a search to a set, as seen in fig.2. Multiplication in Equation 1, concatenation in Equation 2, and the perceptron in Equation 3 all realize the standard similarity functions.

$$h(P, M_j) = P^t V_x M_j \qquad (6)$$

$$h(P, M_j) = V_x[P : M_j] \qquad (7)$$

$$h(P, M_j) = V^t{}_x tanh[V_x P + M_j] \qquad (8)$$

Before each convolution layer, the network is batch normalised to guarantee uniform convergence. Meanwhile, applying the ReLU activation function can drastically decrease the required iterations required for a deep learning network to merge and enhance the training network's effectiveness. The Adam optimizer's training process was initially set at $10^{-2}$, and the dropout chance was initially set to 0.3. To measure how much of a gap there was between the production and the reference labels, we turned to the cross-entropy function. The smaller the amount of cross entropy, the closer the observed dispersal of data is to the density function. The cross entropy method is used to decide when to stop training a model. If there is no change in the cross-entropy value after eight iterations, the training of the model will be terminated.

$$L(Z, k) = -log \frac{exp(Q(Z, k))}{\sum_{j=0}^{M} exp(Q(Z, k))} \qquad (9)$$

### *B. Parameter Optimization*

To ensure the suggested method is generalizable, the cross-validation technique is used. If examples C1 and C4 are used for training, then C6 will be used for testing. In terms of the main operation, we have 1D CNN, Max pooling, BN, dropout, channel technique with residual link, and the FCL layer. In particular, characteristics from various signals are extracted making use of the parallel CNN. CNNs typically use a 3x3 kernel size. Next, when each layer of the CNN is completed, the residual connectivity and channel attention method are employed. The final step is for the completely linked layer to produce an estimate of tool wear.

## III. EXPERIMENTAL SETUP

Using the tool wear big dataset provided by PHM 2010, the effectiveness of the suggested tool wear led to intents was verified. There are six separate entries for tools c1–c6 in the dataset. Seven-dimensional sensor data from 315 cuts is included in files C1, C4, and C6, together with the matching three-dimensional tool wear value. More than 200,000 measurements are taken each time data is gathered from the sensors, at a rate of 50KHz. Down sample c1, c4, and c6 using the previously described technique for doing so.

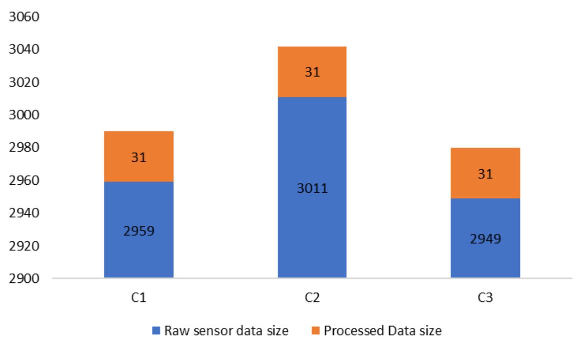


Fig.3. Data comparison from the dataset

Transport data is down-sampled by 99%, from 3GB to 31MB for each tool, with a reconstruction sampling interval of 1000 and complexity of 7. In this study, we utilise the deformation in the X direction to explain the process of data preparation by addressing the multimodal input, which consists of three channels. In order to retrieve the legitimate data segment of each cut, the signal power is terminated in the centre, where 10,240 values were captured as shown in fig.3.

## IV. PERFORMANCE EVALUATION

Tool wear prediction measures include root mean squared error (RMSE), mean absolute error (MAE), and the determination coefficient (R2).

$$MAE = \frac{1}{M}\sum_{j=1}^{m}|y - y_j| \qquad (10)$$

$$RMSE = \sqrt{\frac{1}{M}\sum_{j=1}^{m}(y - y_j)^2} \qquad (11)$$

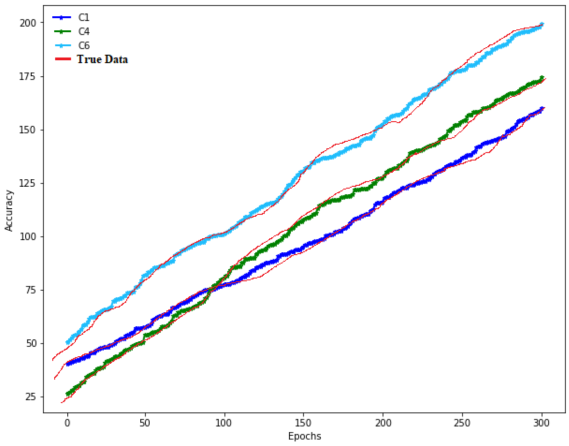


Fig.4. The results of predicted tool wear

$$R^2 = 1 - \frac{\Sigma(y - y_j)}{\Sigma(y - y\widehat{A_\tau})} \qquad (12)$$

Fig.4 depicts the projected tool wear outcomes of the created model across the numerous testing datasets. Measured manually using a microscope, the real data reflect actual tool wear, while the anticipated tool wear is derived using the suggested technique. When the data from three different cutters are examined, it is evident that the projected tool wear is quite similar to the measured tool wear.

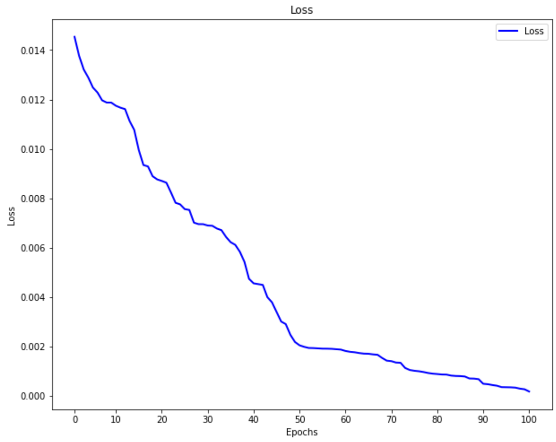


Fig.5. Loss of the proposed system

Average error, root-mean-squared error, and other metrics of judgment, correlation coefficient, and training loss were calculated and plotted in Fig. 4,5,6, and fig.7 to better illustrate the efficacy of the proposed method.

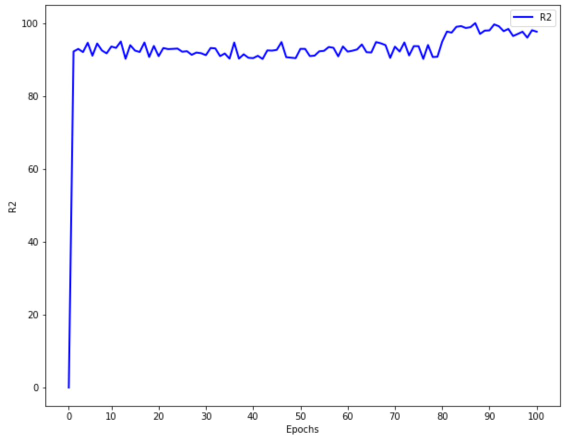


Fig.6. Evaluation of $R^2$ of proposed model

Figure 5 shows that the training loss eventually converges to zero at extremely short epochs. Accuracy was very close to 1 as shown in Fig. 6. The MAE and RMSE both converged, as shown in Fig. 7 and 8. This allows for a swift and simple convergence of all criteria used for evaluation. That's why the developed model is so useful for predicting tool wear.

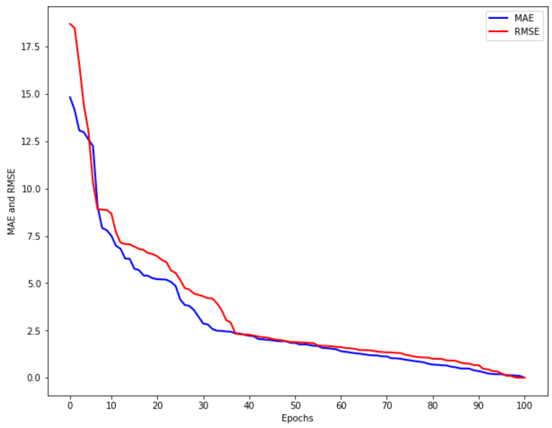


Fig.7. MAE and RMSE of proposed model

The proposed method utilizes a number of reiterated and applied to learn the status of the selected features via the channel attention mechanism and the residue link. This helps to further boost the prediction performance of tool wear.

TABLE I

PERFORMANCE MEASURES OF MODELS

| Model | MAE | | | RMSE | | |
|---|---|---|---|---|---|---|
| | C1 | C4 | C6 | C1 | C4 | C6 |
| SVM | 24.2 | 23.3 | 23.1 | 22.2 | 22.5 | 23.4 |
| CNN | 18.1 | 17.3 | 18.7 | 17.5 | 17.3 | 18.3 |
| LSTM | 20.3 | 20.2 | 20.0 | 19.8 | 19.4 | 19.6 |
| CNN+LSTM | 21.3 | 21.5 | 21.7 | 21.9 | 20.6 | 20.7 |
| RNN | 19.6 | 19.5 | 19.7 | 20.4 | 19.8 | 20.4 |
| ResNet | 8.6 | 8.7 | 8.2 | 8.4 | 8.0 | 8.1 |
| Proposed model | 3.1 | 2.8 | 2.4 | 3.0 | 2.9 | 2.7 |

Table 1 is a listing of the MAE and RSME findings obtained from three different testing cutters using all of the aforementioned procedures. The findings of the RMSE and MAE of the 3 cutters using the suggested approach are, as can be seen, the least significant of those obtained using any of the other methods.

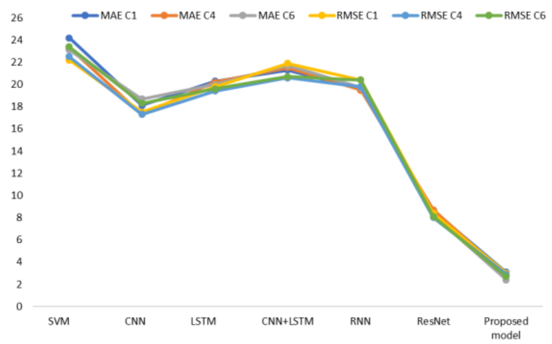


Fig.8. MAE and RMSE of different models

From the Fig.7, It is clear that the suggested technique with the multichannel attention mechanism produces tool wear prediction results that are more closed and stable than the real tool wear. It can be shown that the most common approaches used now, which include channel mechanisms, have attained a higher level of prediction. In conclusion, the extensive comparative findings that have been shown above demonstrate that the suggested design is more talented than previous architectures in terms of tool wear forecasting and is talented to reliably and accurately anticipate tool wear in a variety of scenarios.

## V. CONCLUSION

Using a CNN model and a channel learning algorithm, a deep learning-based approach is suggested to advance the accuracy of tool wear estimate. The suggested technique takes into account the relative importance of each feature. For accurate tool wear prediction, the parallel architecture of CNN and ResNet50 was designed to achieve multi sensor feature fusion, bringing together information from several sensors to use their own strengths. Selecting the relevant features to enhance tool wear prediction outcomes using the channel attention method with the residual connection. Because of this comparison with state-of-the-art methodologies, proposed method confidently attest to the efficiency and generalizability, which has proven to be more accurate in its predictions. Thus, this study presents a novel approach to predicting tool wear, and the proposed approach has great promise for the future of today's industrial sector. The model's performance is meaningfully improved than that of competing approaches. The model's initial wear phase and rapid wear live performance, as well as the model's resilience, might be enhanced in future study. The tool wear monitoring system has to be further developed and the suggested approach needs to be validated in a variety of settings before it can be used widely.

Nevertheless, a significant quantity of training data is required for a deep learning model, such as the one presented for machined surface-based tool wear condition monitoring, in order to attain remarkable monitoring performance. So, the future direction of the research will be on developing appropriate unsupervised models to address this difficulty. Furthermore, the infiltration of chips and crude oils that cause machined surface pollution presents a significant analytical problem for an intelligent monitoring platform wear system throughout the picture gathering process. To increase the generalizability of the projected method a wide range of variables, including different cutting instruments and characteristics are investigated.

## REFERENCES

[1] Haraguchi, N.; Cheng, C.F.C.; Smeets, E. The Importance of Manufacturing in Economic Development: Has This Changed World Dev. 2017, 93, 293–315.
[2] Luan, X.; Zhang, S.; Li, J.; Mendis, G.; Zhao, F.; Sutherland, J.W. Trade-off Analysis of Tool Wear, Machining Quality and Energy Efficiency of Alloy Cast Iron Milling Process. Procedia Manuf. 2018, 26, 383–393.
[3] Liu, C.; Wang, G.F.; Li, Z.M. Incremental Learning for Online Tool Condition Monitoring Using Ellipsoid ARTMAP Network Model. Appl. Soft Comput. 2015, 35, 186–198.
[4] Zhou, Y.; Sun, B.; Sun, W. A Tool Condition Monitoring Method Based on Two-Layer Angle Kernel Extreme Learning Machine and Binary Differential Evolution for Milling. Measurement 2020, 166, 108186.
[5] Wang, M., Zhou, J., Gao, J., Li, Z. & Li, E. Milling tool wear prediction method based on deep learning under variable working conditions. IEEE Access. 8, 140726–140735 (2020)
[6] Chan, Y. et al. Tool wear prediction using convolutional bidirectional LSTM networks. J. Supercomput. 78, 810–832 (2022)
[7] Zheng, H. & Lin, J. A deep learning approach for high speed machining tool wear monitoring. In 2019 3rd International Conference on Robotics and Automation Sciences (ICRAS), 63–68 (2019).
[8] Martínez-Arellano, G., Terrazas, G. & Ratchev, S. Tool wear classifcation using time series imaging and deep learning. Int. J. Adv. Manuf. Technol. 104, 3647–3662 (2019).
[9] Marani, M., Zeinali, M., Songmene, V. & Mechefske, C. Tool wear prediction in high-speed turning of a steel alloy using long short-term memory modelling. Measurement 12, 109329 (2021)
[10] Zhao, R. et al. Machine health monitoring using local feature-based gated recurrent unit networks. IEEE Trans. Ind. Electron. 65, 1539–1548 (2018)
[11] Duan, J., Duan, J., Zhou, H., Zhan, X. & Shi, T. Multi-frequency-band deep CNN model for tool wear prediction. Meas. Sci. Technol. 32, 065009 (2020).

[12] 24. Wu X, Liu Y, Zhou X, Mou A (2019) Automatic identification of tool wear based on convolutional neural network in face milling process. Sensors 19(18):3817. https://doi.org/10.3390/s19183817
[13] Kothuru A, Nooka SP, Liu R (2019) Application of deep visualization in CNN-based tool condition monitoring for end milling. Procedia Manuf 34:995–1004. https://doi.org/10.1016/j.promfg. 2019.06.096
[14] Dzulfkri Z, Su P-W, Huang C-Y (2021) Stamping tool conditions diagnosis: a deep metric learning approach. Appl Sci 11(15):6959. https://doi.org/10.3390/app11156959