euspen's 24th International Conference & Exhibition, Dublin, IE, June 2024
www.euspen.eu

A review and benchmark study of tool state recognition in the CNC milling process

Chen Yin[1], Jeong Hoon Ko[2*]

[1]Hong Kong Institute for Data Science, School of Data Science, City University of Hong Kong, Kowloon, Hong Kong
[2]Taizhou Institute of Zhejiang University, 618, West Section of Shifu Avenue, Taizhou City, Zhejiang Province

E-mail address: jhkolioneagle@gmail.com

Abstract

CNC milling, one of the most essential and popular machining processes in the manufacturing industry, shows highly time-varying and complicated dynamical characteristics. The importance and complexity of the milling processes have made tool condition monitoring (TCM) a hot issue over the past decades. Recently, the rapid development of machine learning has set off new waves in various fields of industry. Numerous TCM methods utilizing machine learning methods have been explored, and most of them focus on tool wear monitoring, including wear state identification and remaining useful life prediction. However, there is still a lack of capability to predict tool breakage, a more severe and unexpected cutting tool failure mode, concurrently with tool wear status using machine learning. Therefore, the article provides a state-of-the-art review of tool state recognition (TSR), indicating the identification of the holistic tool states from health, wear, and breakage. Specifically, the main sections outline traditional machine learning methods that require signal processing and feature extraction and advanced neural network models that can detect tool states across different working conditions. Three primary methodologies are selected to present a more reliable analysis and intuitive comparison, including typical traditional methods, advanced machine learning, and transfer learning. Benchmark studies are carried out for a tool vibration dataset collected by milling experiments under different working conditions to compare the recognition accuracy and computational efficiency quantitatively. The comparison results address the primary strengths and weaknesses of current methods for TSR. Finally, potential research directions are concluded to enhance TSR's accuracy, efficiency, and reliability.

Milling, Neural networks, Pattern recognition, Tool

1. Introduction

CNC milling, which utilizes rotational cutting tools to intermittently cut workpieces into desired geometric surfaces, is one of the most popular and efficient machining processes in the manufacturing industry. Due to the complex mixed physical and chemical effects caused by forces, shocks, and heat, cutting tools, which are the most active cutting element during the milling process, have high failure risks [1]. Recently, due to the fast development of sensing and information technology, various monitoring signals have been collected during the milling process, and machine learning has become the foremost tool in TCM. However, most TCM studies focus on progressive tool wear, such as wear state identification, wear volume estimation, and remaining useful life prediction. Much less attention has been paid to tool breakage, which is a more severe and unexpected tool failure mode during the milling process.

In order to illustrate the difference between tool breakage and progressive tool wear, a schematic diagram of degradation curves for both situations is given in Figure 1. The average flank wear width (VB) is a widely accepted metric to evaluate the tool life. As shown in Figure 1(a), there are three distinct stages for a progressively worn tool: initial wear, normal wear, and severe wear. The VB values of the tool increase rapidly in the initial wear stage and severe wear stage and vary slowly in the normal wear stage. However, as shown in Figure 1(b), the VB values suddenly jump to a high level close to failure once tool breakage occurs. Usually, tool breakage is prone to occur in the initial wear stage due to improper setting of cutting parameters at the beginning and in the severe wear stage due to the rapidly growing forces acting on the tool with accumulation wear. Tool breakage may also occur when the machining workpiece has a high hardness and the tool is relatively brittle. Additionally, milling chatter is another main cause of tool breakage. Although analytical models were developed to guide parameter selections [2], the changing geometrical status of cutting tools and unexpected fluctuations in cutting depth could turn stable milling into unstable conditions, and further lead to tool breakage [3,4]. Compared to monitoring progressive tool wear, the detection of tool breakage is more difficult because the tool breakage occurs randomly and instantaneously without warning. Therefore, the quality of the monitoring signals and the detection algorithms need to be further improved to identify the tool breakage status [5].

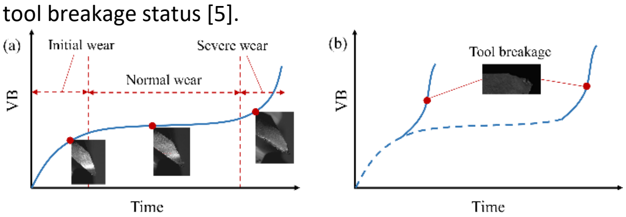

Figure 1. Schematic diagram of progressive wear and breakage degradation curves. (a) Progressive tool wear. (b) Tool breakage.

2. TSR review

2.1. Feature extraction for TSR

Generally, feature extraction methods could be categorized as time-domain methods, frequency-domain methods, and time-frequency domain methods. Time-domain analysis has the advantage of inexpensive computation. Cutting force signals can directly reflect the dynamic variation between tools and workpieces, so multiple studies extracted time-domain features from force signals for TSR [6]. Altintas et al. [7] utilized the difference of force signals to detect the tool breakage in the milling process. Since the current/power varies rapidly with the cutting force and its measurement does not require additional sensors, various time-domain statistical features like the maximum, average, and standard deviation can be extracted from current/power signals for TSR [8]. Another efficient signal for time-domain analysis is acoustic emission, the primary advantage of which is that the significant frequency range relevant to tool status is much higher than that of the environmental noise and machine tool vibration [9].

The frequency-domain analysis utilizes the fast Fourier transform (FFT) to convert time-domain signals to the frequency domain, and tool failure features are then extracted from the frequency spectra [10]. Compared to pure frequency-domain analysis, the time-frequency analysis is more powerful and more appropriate for the nonstationary and nonlinear monitoring signals [11]. Short-time Fourier transform (STFT), empirical mode decomposition (EMD), and wavelet transform (WT) are mainstream time-frequency methods. STFT is an extension of FFT and can simultaneously analyze signals in both time and frequency domains [12]. WT and its variants, such as discrete wavelet transform (DWT), continuous wavelet transform (CWT), and wavelet packet decomposition (WPT) could be the most popular time-frequency methods. The ability to use high-frequency resolution makes them powerful in the feature extraction of TSR [13]. Compared to WT and its variants, EMD can adaptively decompose signals into a series of intrinsic mode functions (IMFs), which has the advantage of not requiring any predetermined parameters and functions [14]. Hilbert transform [15] and energy-based analysis [16] are widely combined with EMD for TSR.

2.2. Machine learning-based TSR

Machine learning is widely used in TSR to predict the tool states from extracted features, including support vector machine (SVM), hidden Markov model (HMM), random forest (RF), clustering, and artificial neural network (ANN). SVM, developed by statistical learning theory and structural risk minimization principle, is a popular machine learning algorithm [17]. HMM consists of a Markov process that describes transition sequences of hidden states and a random process that establishes observation sequences of hidden states [18]. RF is a typical ensemble learning method that combines the output of multiple decision trees to give a comprehensive prediction [19]. Dahe et al. [20] extracted statistical features from vibration signals and utilized RF to recognize tool conditions. Jogdeo et al. [21] utilized a statistical analysis method to tune hyperparameters of the random forest and achieved robust recognition of tool states.

ANN has become the most popular decision-making method in various domains, which shows excellent nonlinear learning ability to recognize tool breakage and tool wear from signal features [22]. Huang et al. [23] proposed a probabilistic neural network for the decision-making analysis of a tool breakage detection system. Different from other machine learning algorithms, clustering is an unsupervised learning method that can be used for anomaly detection purposes [24]. Torabi et al. [25] extracted wavelet features of force and vibration signals for the clustering analysis, and the results showed that clustering methods are repeatable and noise-robust in TSR. Gui et al. [26] utilized the clustering method to analyze the time-domain features for real-time tool breakage detection.

2.3. Deep learning-based TSR

Deep learning models with powerful nonlinear fitting abilities have the advantage of handling large and complex datasets [27]. Typical deep learning models for TSR include auto-encoder (AE), recurrent neural network (RNN), and convolutional neural network (CNN). AE is a powerful unsupervised learning algorithm for the extraction of tool failure features [28]. Kin et al. [29] proposed a stacked AE-based CNC machine tool diagnosis system. Popular RNNs include long short-term memory (LSTM) [30] and gated recurrent units (GRU) [31], which are ideal options for the process of time-series tool monitoring signals [32]. Nam and Kwon [33] proposed a tool breakage monitoring system with LSTM-based autoencoders. Due to the outstanding ability of nonlinear mapping, CNN has become the actual standard in deep learning communities and is widely used in TSR [34]. Yin et al. [35] combined the one-dimensional CNN (1D-CNN) and deep generalized canonical correlation analysis for tool failure diagnosis based on multiple sensor signals.

The application of deep learning models in TSR requires a large amount of training data. However, in the practical milling process, the cutting tool is only allowed to work in health conditions. Once the tool wear/breakage occurs, the CNC machine tool will shut down immediately. Namely, limited failure samples could be collected for TSR in the practical milling process. In this case, transfer learning methods are studied to solve the data imbalanced problem [36]. Li et al. [37] proposed a Wasserstein generative adversarial network to monitor tool breakage under data-imbalanced conditions.

3. Benchmark study

3.1. Experimental setup

To perform the benchmark study of typical TSR techniques, milling experiments were carried out on a five-axis machining center. As shown in Figure 2, a three-axis accelerometer was mounted on the spindle box to collect the cutting vibration signals, and the machined workpiece is a brick with a material of #45 steel. The four-edge end milling cutter with a diameter of 12 mm was studied in the experiments, and those in health, wear, and breakage status, as shown in Figure 3, were used to machine the workpiece. The cutting depth was 1 mm, and the feed rate was set to 0.1 mm/rev. Moreover, rotation speeds of 2000 RPM, 2600 RPM, and 3200 RPM were used to collect cutting vibration signals with a sampling frequency of 12 kHz.

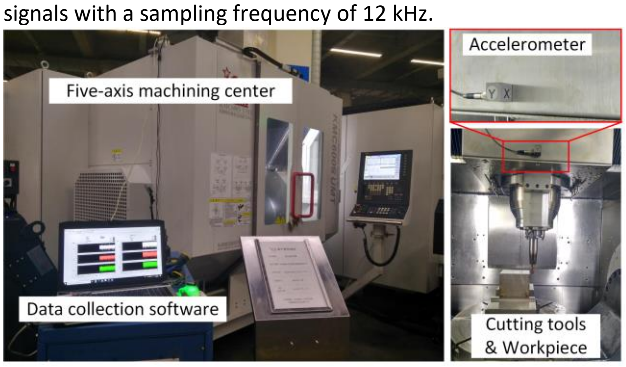

Figure 2. Milling experimental setup

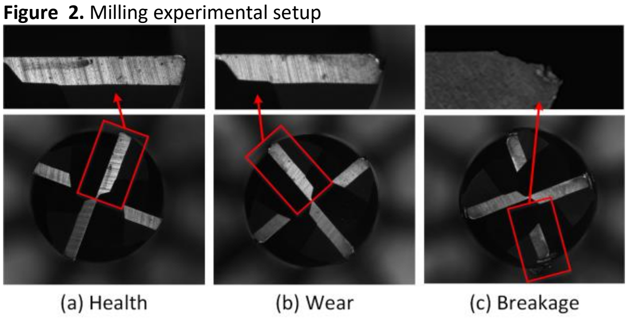

Figure 3. Three different cutting tools status

After the milling experiments, vibration signals of cutting tools under different health conditions (health, wear, and breakage) and various working conditions (2000 RPM, 2600 RPM, and 3200 RPM) were obtained. Since these vibration signals were collected under actual machining processes, signal pre-processing techniques were performed to remove the signal segments collected during air cutting. Finally, the X-directional vibration signals collected under different working conditions are utilized to organize data samples, and a total of 1500 data samples were obtained for each condition. The constructed three datasets are shown in Table 1. The raw vibration signals as well as the frequency spectra of data samples in dataset C are visualized in Figure 4.

Table 1 Dataset definition

Dataset	Spindle speed	Tool condition	Sample Number
Dataset A	2000 RPM	Health/wear/breakage	500/500/500
Dataset B	2600 RPM	Health/wear/breakage	500/500/500
Dataset C	3200 RPM	Health/wear/breakage	500/500/500

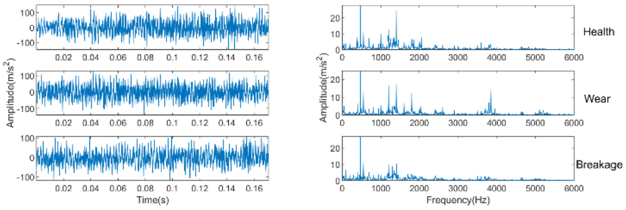

Figure 4. X-directional vibration signals collected under 3200 RPM

3.2. TSR methods selection

Three typical TSR methods are selected to present the comparative studies, which include typical machine learning methods, deep learning neural networks, and transfer learning techniques. A brief introduction to these approaches is given as follows,

1) Feature extraction with SVM (FE-SVM). As one of the most popular shallow learning approaches, SVM is widely used in TSR. Fault features extracted from the time domain, frequency domain, and time-frequency domain are utilized as the model inputs [38]. FE-SVM is introduced in comparative studies to demonstrate the performance of traditional shallow models for TSR.

2) 1-D CNN. The 1-D CNN is a standard of deep learning approaches and is also widely used in TSR. Thus, a 1-D CNN with a typical structure of two convolutional layers and three fully connected layers is studied [34], which can give a comparison of deep models to shallow ones. The frequency spectra of vibration signals are utilized as the model inputs.

3) Cross-domain adaptation networks with attention mechanism (CDATT). CDATT is an advanced transfer learning model that utilizes the attention mechanism to capture the significant fault features, and a joint distribution adaptation regularization term is constructed to solve the performance degradation under variable working conditions [39].

3.3. Results and analysis

The comparative studies are performed under nine scenarios, and the identification results are given in Table 2. It is worth noting that the ratio of training data to test data in all scenarios is 7:3, and the training and testing data are the same for all methods. For scenarios 1, 5, and 9 where the testing samples are from the same dataset as the training samples, the identification results show that the identification accuracies of FE-SVM are close to 1-D CNN and CDATT, indicating that shallow learning models can achieve competitive performance with deep learning models in the scenario of same working conditions. However, for the other scenarios where the testing samples are from different datasets, the identification results show that FE-SVM suffers from a significant performance degradation compared to 1-D CNN and CDATT in TSR. On the other hand, although the identification results of 1-D CNN are much better than FE-SVM, there is a distinct performance gap between 1-D CNN and CDATT. Namely, the results indicate that the tool failure features extracted by 1-D CNN are more robust than hand-crafted features utilized in FE-SVM, but these features cannot be well adapted to a fresh situation in TSR. In such different situations, transfer learning models like CDATT can be a good option for TSR. The ability to learn cross-domain tool failure features can solve the domain discrepancy caused by different working conditions.

Table 2 Identification accuracy (%) for different models under different scenarios.

No.	Scenario	FE-SVM	1-D CNN	CDATT
1	Dataset A to A	99.1	100	100
2	Dataset A to B	23.87	61.93	99.29
3	Dataset A to C	25.46	57.54	86.59
4	Dataset B to A	27.32	63.1	99.09
5	Dataset B to B	88.9	98.57	99.87
6	Dataset B to C	32.2	61.47	90.32
7	Dataset C to A	32.4	66.31	91.48
8	Dataset C to B	37.75	73.94	99.5
9	Dataset C to C	93.16	99.42	100

The computation burden and efficiency of the identification algorithm are other evaluation metrics in TSR. The training and testing time of different algorithms in scenario 9 are presented in Table 3. It can be seen that the training time of FE-SVM is much lower than deep learning models, and the computation burden increases with the complexity of deep learning models. Nevertheless, well-trained 1-D CNN and CDATT are more efficient in the testing stage than SVM. The results indicate that although deep learning models require more computation burden in the model training stage, they could be more efficient than shallow learning models.

Table 3 Computation time for different models in scenario 9.

Model	Training time (s)	Testing time (ms)
FF-SVM	4.31	377.65
1-D CNN	72.56	8.92
CDATT	167.05	16.14

4. Conclusion

This paper provides a comprehensive review of TSR, while feature extraction-based, machine learning-based, and deep learning-based methods are detailed and summarized. Moreover, milling experiments under different working conditions are carried out, and benchmark studies among three popular TSR approaches are presented through the collected tool data. Based on the review and benchmark studies, conclusions and suggestions for TSR, especially potential challenges for the practical application of deep learning models in TSR are summarized as follows,

1) Literature review indicates that machine learning and deep learning methods have become state-of-the-art techniques in TSR. The results of benchmark studies demonstrate that deep learning models show better identification accuracy than typical feature extraction-based shallow learning methods. Therefore, exploring more accurate and robust deep learning models in TSR can be a good research direction.

2) Regarding computation burden and efficiency, the results of benchmark studies reveal that deep learning models take more computation time but operate more efficiently in the testing stage. However, some occasions, such as online monitoring, require the TSR model iterative upgrades with increasing milling data. So, the requirements on heavy computation burden may still be a nonnegligible drawback that restricts the application of deep learning models. Therefore, research on simplifying models without degrading model performance is still a necessary and promising topic.

3) Since collecting sufficient data samples with specific milling conditions to train a deep learning model from scratch is always costly and time-consuming, the development of transfer learning TSR models can be a good solution to tackle this problem. The results of benchmark studies exhibit that the transfer learning-based model can perform well in unseen work conditions. Therefore, exploring robust and accurate TSR models on limited or even no data conditions needs more attention in future work.

4) Although advanced deep learning models show superior performance than traditional feature extraction-based models in TSR, the black-box nature and complex information mapping process make them difficult for users to understand. However, in industrial scenarios, especially high-value milling processes, the explainability and reliability of the identification algorithm are of great importance. Therefore, integrating various knowledge like physics, simulation, or theory in deep learning models and improving their interpretability is an urgent and essential topic.

5) Current TSR research concentrates on developing models with higher detection accuracy and lower prediction error but ignores the inevitable effects of uncertainties on prediction results. Typical uncertainties include milling environment fluctuation, data collection device degradation, and noise interference. Therefore, considering the uncertainty and transforming the point prediction framework into an interval prediction framework to improve model practicality are also challenging and valuable topics.

References

[1] Zhou Y. and Xue W. 2018 Review of tool condition monitoring methods in milling processes. *Int J Adv Manuf Technol.* **96** 2509–2523

[2] Ko J.H. 2015 Time domain prediction of milling stability according to cross edge radiuses and flank edge profiles. *Int J Mach Tools Manuf.* **89** 74–85

[3] Yin C., Wang Y., Ko J.H., Lee H.P. and Sun Y. 2023 Attention-driven transfer learning framework for dynamic model guided time domain chatter detection. *J Intell Manuf.* 1-19

[4] Ko J.H. 2022 Machining Stability Categorization and Prediction Using Process Model Guided Machine Learning. *Metals* **12**, 298

[5] Nath C. 2020 Integrated Tool Condition Monitoring Systems and Their Applications: A Comprehensive Review. *Procedia Manuf* **48** 852–863

[6] Zhu K., Mei T. and Ye D. 2015 Online Condition Monitoring in Micromilling: A Force Waveform Shape Analysis Approach. *IEEE Trans Ind Electron* **62** 3806–3813

[7] Altintas Y., Yellowley I. and Tlusty J. 1988 The Detection of Tool Breakage in Milling Operations. *J. Eng. Ind.* **110** 271–277

[8] Bhuiyan M.S.H. and Choudhury I.A. 2014 Review of Sensor Applications in Tool Condition Monitoring in Machining. *Comprehensive Materials Processing*. **13** 539–569.

[9] Hassan M., Sadek A., Damir A., Attia M.H. and Thomson V. 2016 Tool Pre-Failure Monitoring in Intermittent Cutting Operations. *ASME Int Mech Eng Congress Expo*. **50527** V002T02A049

[10] Hamidieh Y. A., Seth B. B. and Sridhara R. 2001. Enhancement for Frequency Domain Based Tool Breakage Detection System. In *ASME Int Mech Eng Congress Expo*. **35487** 309-316.

[11] Yin C., Wang Y., Lee H.P., He J., He Y. and Sun Y. 2021 Robust wheel wear detection for solid carbide grinding under strong noise interference: A new approach based on improved ensemble noise-reconstructed empirical mode decomposition. *Wear.* **486** 204112**.**

[12] Zamudio-Ramirez I., Osornio-Ríos R.A., Diaz-Saldaña G., Trejo-Hernández M. and Antonino-Daviu J.A. 2020 STFT-based induction motor stray flux analysis for the monitoring of cutting tool wearing in CNC machines. *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*. 2511–2516.

[13] Sevilla-Camacho P.Y., Herrera-Ruiz G., Robles-Ocampo J.B. and Jáuregui-Correa J.C. 2011 Tool breakage detection in CNC high-speed milling based in feed-motor current signals. *Int J Adv Manuf Technol.* **53** 1141–1148

[14] Yin C., Wang Y., Ma G., Wang Y., Sun Y. and He Y. 2022 Weak fault feature extraction of rolling bearings based on improved ensemble noise-reconstructed EMD and adaptive threshold denoising. *Mech Syst Signal Process.* **171** 108834

[15] Peng Y. 2004 Empirical Model Decomposition Based Time-Frequency Analysis for the Effective Detection of Tool Breakage. *J. Manuf. Sci. Eng.* **128** 154–166

[16] Chen Q., Xie M., Zou Z., and Yang Z. 2023 Multi-signal comparison and analysis of tool wear based on EMD energy decomposition method. *2023 4th International Conference on Mechatronics Technology and Intelligent Manufacturing* 517-521.

[17] Sun S., Hu X., Cai W. and Zhong J. 2019 Tool Breakage Detection of Milling Cutter Insert Based on SVM. *IFAC-PapersOnLine.* **52** 1549–1554

[18] Lu M.-C. and Wan B.-S. 2013 Study of high-frequency sound signals for tool wear monitoring in micromilling. *Int J Adv Manuf Technol.* **66** 1785–1792

[19] Wu D., Jennings C., Terpenny J., Gao R.X. and Kumara S. 2017 A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests. *J. Manuf. Sci. Eng.* **139**

[20] Virendra Dahe S., Sai Manikandan, G. Jegadeeshwaran, R. Sakthivel, G. and Lakshmipathi J. 2021 Tool condition monitoring using Random forest and FURIA through statistical learning. *Mater Today Proc.* **46** 1161–1166

[21] Jogdeo A.A., Patange A.D., Atnurkar A.M. and Sonar P.R. 2023 Robustification of the Random Forest: A Multitude of Decision Trees for Fault Diagnosis of Face Milling Cutter Through Measurement of Spindle Vibrations. *J. Vib. Eng. Technol.* 1-19

[22] Corne R., Nath C., El Mansori M. and Kurfess T. 2017 Study of spindle power data with neural network for predicting real-time tool wear/breakage during inconel drilling. *J Manuf Syst.* **43** 287–295

[23] Huang P.B., Ma C.-C. and Kuo C.-H. 2015 A PNN self-learning tool breakage detection system in end milling operations. *Appl Soft Comput.* **37** 114–124

[24] Fu P., Li W. and Guo L. 2011 Fuzzy Clustering and Visualization Analysis of Tool Wear Status Recognition. *Procedia Eng.* **23** 479–486

[25] Torabi A.J., Er M.J., Li X., Lim B.S. and Peen G.O. 2016 Application of Clustering Methods for Online Tool Condition Monitoring and Fault Diagnosis in High-Speed Milling Processes. *IEEE Syst J.* **10** 721–732

[26] Gui Y., Lang Z.-Q., Liu Z., Zhu Y., Laalej H. and Curtis D. 2023 Unsupervised Detection of Tool Breakage: A Novel Approach Based on Time and Sensor Domain Data Analysis. *IEEE Trans Instrum Meas.* **72** 1–13

[27] Xiao W., Huang J., Wang B. and Ji H. 2022 A systematic review of artificial intelligence in the detection of cutting tool breakage in machining operations. *Measurement.* **190** 110748

[28] Song G., Zhang J., Zhu K., Ge Y., Yu L. and Fu Z. 2023 Tool wear monitoring based on multi-kernel Gaussian process regression and Stacked Multilayer Denoising AutoEncoders. *Mech Syst Signal Process.* **186** 109851

[29] Kim J., Lee H., Jeon J.W., Kim J.M., Lee H.U. and Kim S. 2020 Stacked Auto-Encoder Based CNC Tool Diagnosis Using Discrete Wavelet Transform Feature Extraction. *Processes.* **8** 456

[30] Zhang J., Zeng Y. and Starly B. 2021 Recurrent neural networks with long term temporal dependencies in machine tool wear diagnosis and prognosis. *SN Appl. Sci.* **3** 442

[31] Wang J., Li Y., Zhao R. and Gao R.X. 2020 Physics guided neural network for machining tool wear prediction. *J Manuf Syst.* **57** 298–310

[32] Zegarra F.C., Vargas-Machuca, J. and Coronado A.M. 2023 A Comparative Study of CNN, LSTM, BiLSTM, and GRU Architectures for Tool Wear Prediction in Milling Processes. *J of Mach Eng.* **23** 122-136

[33] Nam J.S. and Kwon W.T. 2022 A Study on Tool Breakage Detection During Milling Process Using LSTM-Autoencoder and Gaussian Mixture Model. *Int. J. Precis. Eng. Manuf.* **23** 667–675

[34] Yin C., Wang Y., He Y., Liu L., Wang Y. and Yue G. 2021 Early fault diagnosis of ball screws based on 1-D convolution neural network and orthogonal design. *P I MECH ENG O-J RIS*. **235** 783-797

[35] Yin Y., Wang S. and Zhou J. 2023 Multisensor-based tool wear diagnosis using 1D-CNN and DGCCA. *Appl Intell.* **53** 4448–4461

[36] Deng C., Tang J., Miao J., Zhao Y., Chen X. and Lu S. 2023 Efficient stability prediction of milling process with arbitrary tool-holder combinations based on transfer learning. *J Intell Manuf.* **34** 2263–2279

[37] Li X., Yue C., Liu X., Zhou J. and Wang L. 2024 ACWGAN-GP for milling tool breakage monitoring with imbalanced data. *Robot Comput Integr Manuf.* **85** 102624

[38] Liao X., Zhou G., Zhang Z., Lu J. and Ma J. 2019 Tool wear state recognition based on GWO–SVM with feature selection of genetic algorithm. *Int J Adv Manuf Technol.* **104** 1051–1063

[39] He J., Sun Y., Yin C., He Y. and Wang Y. 2023 Cross-domain adaptation network based on attention mechanism for tool wear prediction. *J Intell Manuf.* **34** 3365–3387