

Classification of Cutting Tool Wear Evolution in CNC Turning Using Convolutional Neural Networks

Dimitrios Apollon Gkartzonikas^{1,a}, Nikolaos Karkalos^{1,b},
Angelos Markopoulos^{1,c} and Panorios Benardos^{1,d*}

¹National Technical University of Athens, School of Mechanical Engineering, Manufacturing Technology Laboratory, Heron Polytechniou 9, GR15772, Athens, Greece

^adimitrisgkartzonikas@gmail.com, ^bnkark@mail.ntua.gr, ^camark@mail.ntua.gr,
^dpbenard@mail.ntua.gr
(*corresponding author)

Keywords: Machine learning, machine vision, tool wear classification, flank wear

Abstract. This study proposes an automated framework for online cutting tool wear classification in CNC turning using low-cost optical equipment and Convolutional Neural Networks (CNNs). Longitudinal turning experiments were performed on CK45 medium carbon steel using a HAAS TL1 lathe under dry machining conditions. Tool wear evolution was monitored via a lathe-mounted digital microscope, with images classified into three distinct stages: Low ($V_b < 160 \mu\text{m}$), Medium ($160 \leq V_b \leq 200 \mu\text{m}$), and Critical ($V_b > 200 \mu\text{m}$). A shallow CNN architecture, consisting of three convolutional blocks and a Softmax output layer, was developed to balance model complexity with computational efficiency for potential edge deployment. To enhance robustness against positional changes, data augmentation techniques including random translations and rotations were applied. The results demonstrate good performance, with the model achieving 94.7% accuracy and a weighted F1-score of 95.4% on the testing subset. While the model showed exceptional performance in identifying Low and Medium wear, data scarcity in the Critical wear class remained a limiting factor for recall. Overall, the study confirms that shallow CNNs can accurately capture spatial hierarchies for image-based wear assessment.

Introduction

In the era of Industry 4.0, the development of automated monitoring frameworks reduces reliance on manual tasks and enables data-driven decision making [1]. In this context, cutting tool wear monitoring still remains a topic that gathers a lot of interest due to the fact that continuous, accurate tool wear classification can directly improve productivity, quality and cost. Productivity is improved since adaptive control strategies can be applied, where feeds and speeds can be optimized [2]. Additionally, the number of unplanned downtimes for replacing worn out or broken tools can be reduced [3]. Quality is improved by achieving surface finish and dimensional accuracy values with less variability [4]. Cost is reduced as tool life can be more accurately estimated and therefore predictive maintenance strategies can be applied [5]. Finally, deploying automated monitoring frameworks at the process level enables 100% inspection strategies. This eliminates the sampling risk associated with Type I and Type II errors, thereby reducing uncertainty in quality assurance.

Typically, automated cutting tool wear classification is based on indirect methods where various signal types such as cutting forces, vibrations, acoustic emissions, motor loads etc., provide corresponding data modalities that are used by machine learning (ML) models that try to infer the actual wear state of the cutting tool [6]. However, when it comes to transferring these research approaches from the lab to real shop-floors, four main challenges appear. Firstly, instrumenting the machine tools requires significant cost for sensors, data acquisition and storage setups [7]. Secondly, human expertise is also a requirement since the collected signals need to be preprocessed and suitable features need to be extracted to develop the ML predictive models. Thirdly, a lot of relevant data need to be collected so that the ML models can achieve satisfactory performance [8]. Finally, particularly

for the case of vision systems, environmental variability in terms of inconsistent lighting, coolant spray and equipment vibrations plays a significant role for their practical deployment [9].

This paper aims to address these gaps by proposing a low-cost, image-based approach for the online classification of the cutting tool's flank wear focusing on turning processes. More specifically, the proposed approach involves the use of convolutional neural networks (CNNs) in order to associate image features with each of the three typical cutting tool wear stages (initial, steady-state and accelerated wear) [10]. Using images as inputs to the CNN addresses two of the previously identified challenges. Images are easier and less costly to acquire compared to other signal types. The CNN's ability to automatically extract and evaluate spatial features from raw pixel data makes them exceptional for more complex visual pattern recognition and reduces data preprocessing efforts. Finally, the dataset size limitation can be partially addressed by adopting data augmentation techniques.

In order to obtain the required experimental dataset, dry machining experiments were performed with constant cutting parameters. After a pre-determined number of passes were completed, an image of the cutting tool was captured using a low-cost digital microscope that was mounted on the lathe and then the cutting tool was removed to measure its actual flank wear using a stereoscopic microscope. The following sections provide the details regarding both the experimental as well as the CNN model development procedures and the achieved results.

Materials and Methods

Materials. The selected workpiece material for performing the experimental investigation was CK45 medium carbon steel due to its very good machinability and wide range of applications across various industrial sectors such as automotive, heavy machinery and construction. Table 1 presents the typical chemical composition of the selected material.

Table 1. CK45 steel chemical composition.

	C (wt%)	Si (wt%)	Mn (wt%)	P (wt%)	S (wt%)
CK45	0.42 – 0.50	≤ 0.40	0.50 – 0.80	≤ 0.035	≤ 0.03

Turning experiments. Longitudinal turning experiments were performed using two CK45 cylindrical workpieces (76mm diameter, 610mm length) on a HAAS TL1 CNC lathe (Fig. 1) using a TNMG160408-UG YG3020 carbide insert that was mounted on a MTJNR 2525 M16 tool holder. The cutting parameters were selected to accelerate flank wear development, while remaining within the supplier's recommended ranges. Thus, cutting speed (V_c), feed (f) and depth of cut (a_p) were set at 280 m/min, 0.35 mm/rev and 0.8 mm, respectively. The turning experiments were performed under dry machining conditions, i.e. no cutting fluid was used.



Fig. 1. HAAS TL1 CNC lathe.

Wear evaluation. The cutting tool wear was evaluated in terms of flank wear width using the V_b parameter and according to the ISO 3685 standard. A Leica MZ6 stereo-microscope was used along

with the ImageJ image analysis software to capture and analyze images of the cutting insert, resulting in the precise measurement of the maximum flank wear width (Fig. 2). In order to ensure consistent monitoring of the actual wear evolution, the cutting tool was removed from the lathe and inspected in the stereo-microscope each time that a cutting length of 1000 mm, corresponding to two cutting passes, had been completed. Similarly, after each cutting interval and before the tool was removed, images of the flank face were captured at a 640x480 pixel resolution with the help of a low-cost digital microscope, which was directly mounted on the lathe (Fig. 2). The cutting tool wear evolution is shown in Fig. 3 as a function of machining time. Three distinct stages can be observed, corresponding to initial, steady-state and accelerated wear patterns, respectively.

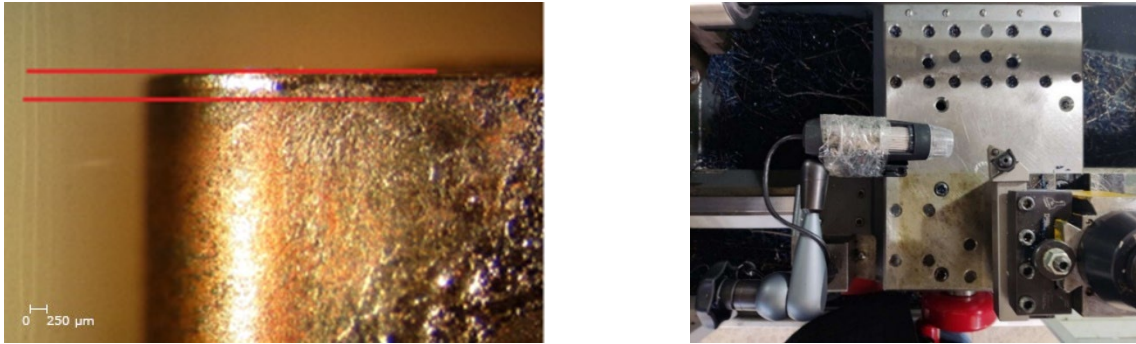


Fig. 2. Labelled image of cutting tool captured by Leica MZ6 (left) and low-cost digital microscope mounted on the CNC lathe (right).

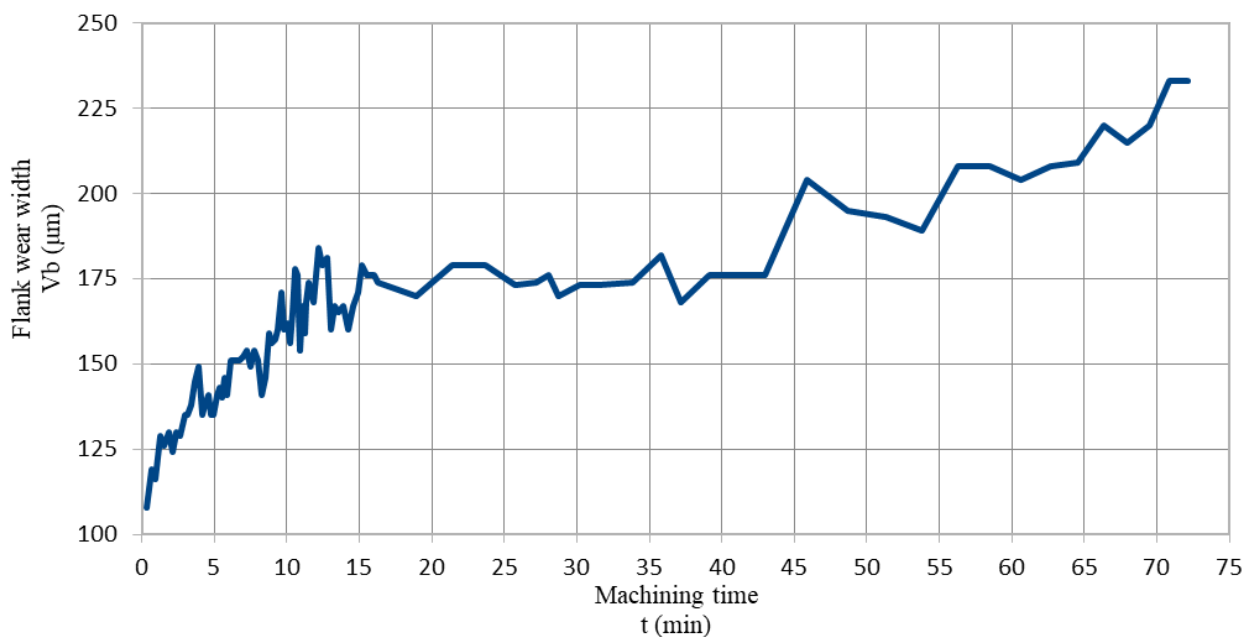


Fig. 3. Cutting tool wear evolution.

It has to be emphasized that the cutting tool's removal from the lathe and subsequent wear measurement is only necessary so that the ground truth regarding wear evolution is established. This is a necessary step in any approach that involves the development of deep learning models and is performed as part of the model's training phase. After the model has been trained, it is expected to be able to classify the tool wear using only the images captured directly from the low-cost microscope and therefore no tool removal from the lathe is required. This corresponds to the model's inference phase.

Data preprocessing. Based on Fig.3, the wear measurements were divided in distinct classes: a) Low Wear ($V_b < 160 \mu\text{m}$), b) Medium Wear ($160 \leq V_b \leq 200 \mu\text{m}$) and c) Critical Wear ($V_b > 200 \mu\text{m}$), thereby formulating the investigated problem as a three-class classification task. Therefore, following this procedure, a dataset was created that consisted of 93 images, captured by the lathe-

mounted microscope, and associated flank wear width measurements, each belonging in the above classes. In order to accommodate CNN model development, additional preprocessing techniques were applied. Data augmentation was performed to increase dataset size and variability by applying random translations (± 5 pixels) and rotations ($\pm 15^\circ$) to the captured images. These geometrical transformations aim to increase model robustness against positional and orientation changes. Furthermore, the dataset was randomly split into training, validation and testing subsets following a 80/10/10 ratio of total available data. The validation subset was used for early stopping, while the testing subset was used to evaluate the model's generalization performance when presented with unknown data.

Convolutional Neural Networks. CNNs are a specialized class of deep learning models, which is designed to process data with a known grid-like topology. Unlike traditional multilayer Artificial Neural Networks (ANNs) that flatten input data into a single vector, thereby discarding crucial structural information, CNNs preserve spatial and temporal correlations through the application of learnable kernels (filters) [11]. A typical CNN is constructed by stacking different layers that transform the input into an output class score through a differentiable function. Three main layer types are involved; convolutional, pooling and fully connected. The convolutional layer consists of learnable kernels that convolve across the dimensions of the input (typically width and height of an image). This process produces 2D activation maps (feature maps) that represent the response of that filter at every spatial position, effectively automatically extracting hierarchical spatial features such as edges, textures etc. The pooling layer is inserted between successive convolutional layers so as to perform non-linear downsampling. Its primary function is to progressively reduce the spatial size of the representation, essentially reducing the number of parameters. Max pooling, which outputs the maximum value within a defined rectangular window, is the most common approach. The fully connected layers mirror the structure of a feedforward neural network. The final fully connected layer typically uses a Softmax activation function to produce probability distributions over the target classes. For the investigated problem, the input to the CNN is the image captured by the low-cost microscope mounted on the CNC lathe and the output is the level of the cutting tool wear as it is described by one of the three defined classes. The expectation is that through training the model will be able to associate image features with the wear evolution mechanisms.

Results and Discussion

Shallow CNN model. The architecture of the developed CNN model consisted of three Convolutional Blocks, each containing a Convolutional Layer, a Batch Normalization and a ReLU activation function. It also included two Max-Pooling Layers (2x2) after its first two blocks for dimensional reduction, a dropout layer (rate = 0.5) before the final layers that helps to prevent overfitting, a fully connected layer and a Softmax layer for the final classification into the three wear classes (Fig. 4). The model's hyperparameters corresponding to the convolutional layers are presented in Table 2. The choice of a shallow architecture was made as a compromise between model complexity and computational efficiency, which are important considering the limited dataset size.

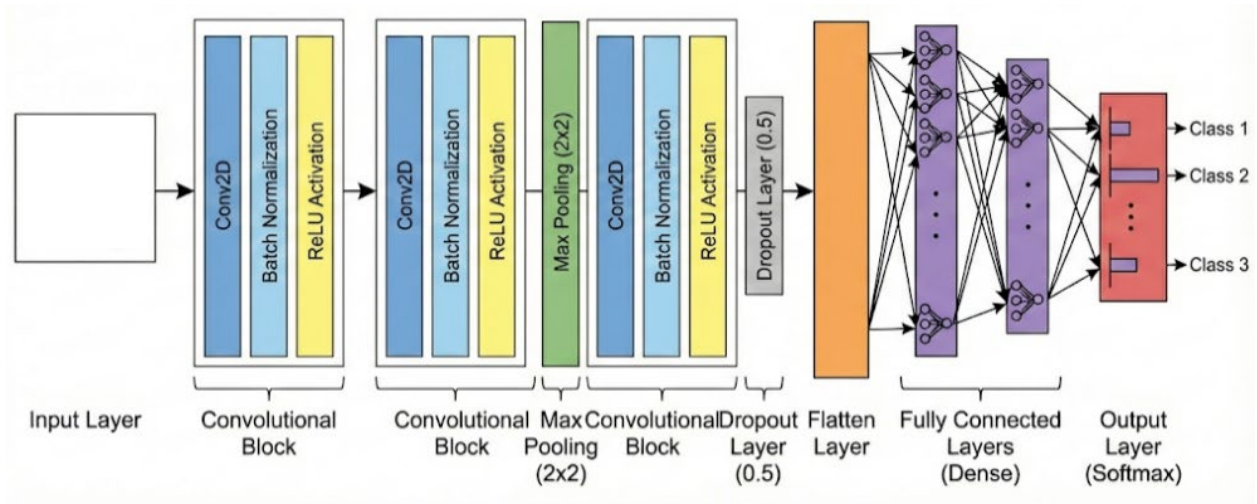


Fig. 4. CNN model architecture.

Table 2. CNN model’s convolutional layer hyperparameters.

Convolutional Layer	Type	Kernel Size	Stride	Padding	Filters
1	Conv2D	5x5	1	Same	16
2	Conv2D	3x3	1	Same	32
3	Conv2D	3x3	1	Same	64

The network was trained using the Stochastic Gradient Descent with Momentum (SGDM) optimizer with an initial learning rate of 0.001, a mini-batch size of 32, over 25 epochs. The training performance, in terms of achieved accuracy and cross-entropy loss over each iteration, is shown in Fig. 5. As can be observed both training and validation curves exhibit smooth convergence behavior, finally reaching 100% and 94.7% accuracy, respectively. The complete results across training, validation and testing subsets are visualized through the respective confusion matrices in Fig. 6a-c.

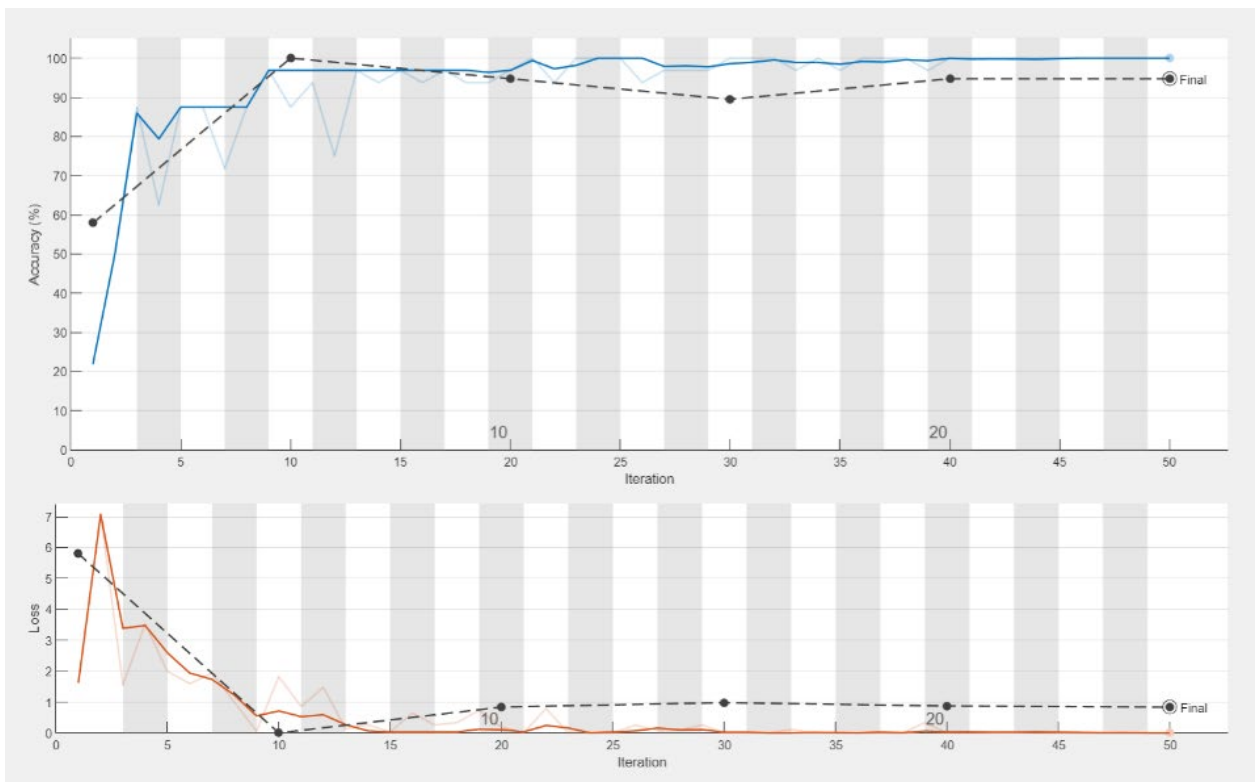


Fig. 5. Shallow CNN model training history.

True Class	Low Wear	44		
	Medium Wear		23	
	Critical Wear			7
		Low Wear	Medium Wear	Critical Wear
		Predicted Class		

Fig. 6a. Confusion matrix for training subset.

True Class	Low Wear	11		
	Medium Wear	1	5	
	Critical Wear			2
		Low Wear	Medium Wear	Critical Wear
		Predicted Class		

Fig. 6b. Confusion matrix for validation subset.

True Class	Low Wear	11		
	Medium Wear		6	
	Critical Wear		1	1
		Low Wear	Medium Wear	Critical Wear
		Predicted Class		

Fig. 6c. Confusion matrix for testing subset.

Additional performance metrics such as precision, recall and F1-score are also calculated in order to fully evaluate the model's predictive performance. Table 3 presents the relevant values for all the above metrics, focusing on the testing subset in order to assess the model's generalization capability. Out of the 19 cases of the testing subset, only one was misclassified by the shallow CNN model (a case of Critical Wear was misclassified as Medium Wear), resulting in overall accuracy equal to 94.74%.

Table 3. CNN's performance metrics for the testing subset.

Class	Precision	Recall	F1-score	Accuracy	Weighted F1-score
Low Wear	1.000	1.000	1.000	0.947	0.954
Medium Wear	0.857	1.000	0.923		
Critical Wear	1.000	0.500	0.667		

The model demonstrates exceptional performance in identifying Low Wear, with perfect precision and recall. For Medium Wear, the model achieves perfect recall, meaning it successfully captured all true medium wear instances; however, its precision is slightly lower (0.857) because it misclassified one "Critical Wear" instance as "Medium". For Critical Wear, while the precision is perfect, i.e. the model never falsely identifies a lower wear state as critical, the recall is only 0.500. This is again due to the fact that out of the two "Critical Wear" instances, one was misclassified as "Medium Wear". Since there are only two instances of the "Critical Wear" class, it is not straightforward to draw meaningful conclusions regarding the model's reduced performance regarding this class.

In order to deal with this issue, the weighted average F1-score metric, which accounts for class imbalance by weighting each class score by its number of instances, was also calculated and found equal to 0.954. The weighted F1-score reflects how the model would perform on a random sample from this specific, imbalanced population, indicating that for nearly 95% of the cases encountered in this specific environment, the model will be correct and reliable. In any case, to avoid any errors related to the Critical Wear class further experimental investigation and/or additional data augmentation methods should be investigated to account for the class imbalance problem.

Conclusions

This work presented an automated framework for online cutting tool wear monitoring utilizing low-cost optical equipment and CNNs. The achieved results show that shallow CNN architectures are highly effective at capturing the spatial hierarchies necessary for accurate image-based tool wear assessment. The main conclusions can be summarized as follows:

- The proposed shallow CNN achieved an accuracy of 94.7% and a weighted F1-score of 95.4% in the testing subset.
- The class imbalance associated with the Critical Wear class is the limiting factor for even better model performance, especially regarding recall.
- Computationally efficient models provide a viable solution suitable for edge deployments in real manufacturing environments.

To improve this work several future research directions can be identified. The main priority will be to address data scarcity of the critical wear class. Several alternatives will be considered ranging from additional experimental investigations with pre-worn tools, to adopting class-weighted loss functions and even examining synthetic data generation through Generative Adversarial Networks (GANs). Transfer learning approaches through publicly available datasets will also be adopted to enhance model generalization. Furthermore, formulating the wear evolution as a regression rather than a classification problem will allow the direct prediction of the flank wear width and therefore offer a more refined and useful monitoring framework. Finally, additional validation by considering different cutting parameters, materials and tool geometries will be performed.

References

- [1] B.T. Saatci, M. Ulas, T. Gurgenc, Hybrid AI Systems for Tool Wear Monitoring in Manufacturing: A Systematic Review, *Appl. Sci.* 16(1) (2026) 208. <https://doi.org/10.3390/app16010208>
- [2] Y. Jiang, J. Huang, K. Liu, D. Zhao, H. Liu, Y. Wang, Intelligent machining technology for online monitoring and control of tool condition: A review, *Front. Mech. Eng.* 20 (2025) 50. <https://doi.org/10.1007/s11465-025-0866-y>
- [3] S. Li, M. Li, Y. Gao, Deep Learning Tool Wear State Identification Method Based on Cutting Force Signal, *Sensors.* 25(3) (2025) 662. <https://doi.org/10.3390/s25030662>
- [4] K. Wang, A. Wang, L. Wu, G. Xie, Machine Tool Wear Prediction Technology Based on Multi-Sensor Information Fusion, *Sensors.* 24(8) (2024) 2652. <https://doi.org/10.3390/s24082652>
- [5] A.D. Patange, R. Jegadeeshwaran, N.C. Dhobale, Milling cutter condition monitoring using machine learning approach, *IOP Conf. Ser.: Mater. Sci. Eng.* 624 (2019) 012030. <https://doi.org/10.1088/1757-899X/624/1/012030>
- [6] P. Unal, B.U. Deveci, A.M. Ozbayoglu, A Review: Sensors Used in Tool Wear Monitoring and Prediction, In: Awan, I., Younas, M., Poniszewska-Marańda, A. (eds) *Mobile Web and Intelligent Information Systems. MobiWIS 2022. LNCS 13475* (2022). Springer, Cham. https://doi.org/10.1007/978-3-031-14391-5_15
- [7] K. Jemielniak, Contemporary challenges in tool condition monitoring, *Journal of Machine Engineering*, 19(1) (2019) 48–61.
- [8] F. Ramzan, D. Reforgiato Recupero, A Literature Review on Enhancing Predictive Maintenance in Smart Manufacturing Industries: Fostering Human-Technology Collaboration and Overcoming Data Scarcity Limitations with Advanced AI Models, *Oper. Res. Forum* 6 (2025) 181. <https://doi.org/10.1007/s43069-025-00584-0>
- [9] A.S. Kumar, A. Agarwal, V.G. Jansari, K.A. Desai, C. Chattopadhyay, L. Mears, Realizing on-machine tool wear monitoring through integration of vision-based system with CNC milling machine, *J. Manuf. Syst.* 78 (2025) 283-293. <https://doi.org/10.1016/j.jmsy.2024.12.004>
- [10] B.J. Sujay, S. A. Singh, A. Agarwal, K.A. Desai, L. Mears, Identifying tool wear stages in turning process through machined surface image analysis using convolutional neural network, *Manuf. Lett.* 44 (2025) 678–686. <https://doi.org/10.1016/j.mfglet.2025.06.079>
- [11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016. <https://doi.org/10.1007/s10710-017-9314-z>