Deep Reinforcement Learning-based Parameter Optimization in Milling: A Novel Approach for Enhancing Tool Life

Stefania Ferrisi[1,*], Mohamadreza Afrasiabi[2], Rosita Guido[1], Domenico Umbrello[1], Giuseppina Ambrogio[1] and Markus Bambach[2]

[1]Department of Mechanical, Energetic and Management Engineering, University of Calabria, Ponte P. Bucci, Rende, 87036, CS, Italy

[2] Advanced Manufacturing Lab, ETH Zurich, Zurich, Switzerland

Abstract

Sustainable manufacturing is recognized as one of the core values of Industry 5.0. The adoption of sustainable practices has become a pivotal factor in enhancing resource efficiency and reducing waste, thereby ensuring the competitiveness of industries. Among the critical factors influencing sustainability in machining processes is tool wear progression, which affects surface quality, dimensional accuracy, energy efficiency, and material utilization. The machining parameters have been shown to have a significant impact on the progression of tool wear. The setting of these parameters is usually performed manually by operators, who exploit their experience.

This study aims to develop an optimization model for setting cutting parameters based on deep reinforcement learning (DRL). The objective is to enhance resource usage and production efficiency, reduce waste, extend tool life, and, consequently, improve the sustainability of the entire machining process. The integration of DRL techniques facilitates the development of autonomous systems capable of formulating an adaptive strategy for the selection of cutting parameters, thereby enabling the realization of specific objectives. To this end, a custom simulation environment was developed to capture the dynamics of a milling process, incorporating two competing objectives: production efficiency and tool efficiency. The experimental results demonstrate the efficacy of the proposed methodology in optimizing production efficiency and tool efficiency through the application of DRL algorithms. These findings underscore the potential of DRL in driving intelligent and sustainable machining processes, thereby aligning with the overarching objectives of Industry 5.0 by reducing human dependency, improving system adaptability, and enhancing sustainability goals.

Keywords

Deep reinforcement learning, Tool wear, Machining parameter optimization, Milling process

1. Introduction

In the era of Industry 5.0, integrating sustainable practices into production processes has become increasingly important for industries to remain competitive in the marketplace and meet existing environmental and social regulations. Manufacturing processes are the main drivers of global warming [1]. Tool wear progression has been demonstrated to have a significant impact on the sustainability of production processes. The rapid deterioration of cutting tools can lead to poor surface quality and short tool life. It results in increased material waste and energy consumption, ultimately compromising production efficiency and sustainability. The monitoring of tool wear progression has been a subject of interest for researchers since the introduction of the Industry 4.0 paradigm.

The advent of digital technology in manufacturing has facilitated the monitoring of machinery conditions, the diagnosis of root causes of failure, and the prediction of the remaining useful life (RUL) of mechanical systems or components [2]. In this context, the construction of predictive models based on Artificial Intelligence (AI) techniques is considered a powerful solution for the predictive maintenance of machining operations [3]. Developing predictive models for tool wear enables the identification of

2nd Workshop on Green-Aware Artificial Intelligence, 28th European Conference on Artificial Intelligence (ECAI 2025), October 25–30, 2025, Bologna, Italy

*Corresponding author.

stefania.ferrisi@unical.it (S. Ferrisi)

0009-0002-7783-2756 (S. Ferrisi)

the optimal time to replace the cutting tool, thereby preventing tool breakage and surface defects on the workpiece.

An accurate predictive model for tool wear was developed by Lin et al. [4], combining XGBoost feature selection from vibration and cutting force signals with a PSO-BP network. Rao [5] developed a CNN-LSTM prediction model to obtain a precise prediction of tool wear during the machining of AISI D_2 steel with different combination of cutting parameters. High prediction performance were demonstrated by adopting a GRU-LSTM prediction model for the RUL of the cutting tool [6]. A new framework for tool wear prognostics across various machining scenarios was proposed by Han et al. [7], demonstrating lower prediction error in forecasting future tool wear. Hao et al. [8] adopted a multimodal large language model architecture to forecast future wear time series.

In the newer vision of Industry 5.0, the traditional emphasis on developing purely predictive models to address sustainability challenges in machining has become increasingly insufficient. Recent perspectives have emphasized the necessity for autonomous and adaptive systems that are capable of dynamically determining optimal machining parameters. The objective of these systems is twofold: firstly, to enhance operational efficiency; secondly, to improve key sustainability indicators throughout the machining process. In this view, Reinforcement learning (RL) has emerged as a promising technology adopted to reach this objective. It is a machine learning technique in which an agent learns optimal actions through trial and error by interacting with an environment [9], with the main goal of enhancing performance by maximizing the cumulative reward [10]. It enables intelligent decision-making in uncertain and dynamic environments. Usually, the environment is modeled by states while the agent can take certain actions as a function of the current state. Following the selection of an action at each time step, the agent is given a reward and transitions to a new state, acquiring the capacity to learn a policy strategy. The primary objective is to exploit the interactions between the agent and its immediate environment to derive the most advantageous policy that can maximize rewards received over time. Typically, RL is combined with deep neural networks, defining a deep reinforcement learning (DRL) technique that facilitates more scalable learning in high dimensional spaces. In the field of machining, the use of DRL offers a promising solution for improving the automation, efficiency, and adaptability of processes in response to different production conditions [11], thereby facilitating a more cognitive and personalized manufacturing paradigm [12].

1.1. Related work

In recent years, several studies have adopted DRL techniques to solve a cutting-parameters optimization problems for machining processes [13– 14]. In traditional production, processing parameters are typically determined by technicians based on their experience and expertise. However, these parameters are rarely adapted or optimized in accordance with the actual processing conditions, which may result in suboptimal outcomes [15]. The adoption of experienced learning such as DRL allows the development of strategies during the data generation process [16]. In order to ensure quality, efficiency, and sustainability in a machining process, it is necessary to make dynamic process adjustments [13]. These studies formulate the cutting parameter optimization problem by defining a set of objectives. The most common objectives considered are the maximization of material removal rate and the minimization of machining time to enhance production efficiency, the reduction of energy consumption [17] and emissions of carbon footprint [18] to improve environmental sustainability, the maintenance of surface quality [19], the reduction of production costs for economic sustainability [14]. All these objectives are either considered individually or combined in a multi-objective optimization model, allowing the agent to learn trade-offs between conflicting requirements.

Despite the significant impact of tool wear progression on all the aspects of production efficiency, workpiece quality, and the economic and environmental sustainability of manufacturing processes, few studies consider its influence on parameter optimization. These recent studies have begun to model tool wear progression within sequential decision-making frameworks by integrating the measured tool wear into the state that describes the environment.

A significant reduction in energy consumption was demonstrated by an optimization model that takes tool wear progression into account [20]. Xie et al. [21] defined an optimization model with the objectives of improving energy consumption, energy efficiency and surface quality, by considering the progression of tool wear for a turning process. Li et al. [18] reported improvements of 6.72% and 8.60% in energy consumption and production time respectively, compared with a model that did not consider tool wear. The findings of these studies pave the way for a new framework of cutting-parameter adaptive optimization models, by taking into account tool wear.

1.2. Our contribution

Despite the impact of tool wear progression on the sustainability goals for a machining operation and the consolidated demonstration that take into account tool wear on optimization problems provide significant improvement in terms of energy efficiency and surface quality, no study has yet considered the introduction of an objective that seeks to extend the lifespan of cutting tools. This represents a gap in the literature, as tool wear not only influences surface quality and production efficiency, but also plays a crucial role in the overall sustainability of machining operations by influencing energy consumption and material waste. Including a measure of tool efficiency in the objective function of an optimization model transforms it from a passive state variable into an active driver of decision-making, enabling the agent to proactively select actions that preserve tool life and promote long-term process sustainability.

Motivated by this gap in the literature, we propose a novel decision-making model that explicitly incorporates tool life extension as a core optimization objective. The reward function developed in this work is designed to balance two competing goals: maximizing production efficiency by considering the material removal rate, and enhancing tool efficiency by promoting strategies that extend tool lifespan. In contrast to previous methodologies, which considered tool wear exclusively as a component of the state, our model integrates it directly into the optimization objective. This approach facilitates proactive decision-making, enabling active management of the cutting tool in response to tool wear, thereby ensuring more sustainable and cost-effective machining operations. To support this approach, we conducted an experimental campaign to collect data followed by the development of a machine learning model to simulate tool wear progression. Based on this, a custom environment was designed to represent the milling operation of a 3-axis CNC machine. Finally, an RL agent was trained to dynamically adjust three cutting parameters by testing and evaluating the performance of four different DRL algorithms.

The remainder of the paper is organized as follows. Sec. 2, describes the materials and methods adopted in this study, including the experimental setup, problem statement, and RL-based optimization framework. Sec. 3 discusses the results obtained from different DRL algorithms. Finally, Sec. 4 provides the conclusion and outlines possible directions for future research.

2. Materials and Methods

This section describes the experimental setup adopted to collect data, with details on the workpiece and the cutting tool considered. Then, the formulation of the optimization model aimed at maximizing the tool and production efficiency is presented. Finally, the RL framework adopted to address the problem is described.

2.1. Experimental setup

Four experimental campaigns were designated for the collection of tool wear progression data during the milling of ASTM A516 Gr. 70 steel plates. They are a specific grade of carbon steel plate typically employed in refineries, chemical and power plants, and other moderate to low temperature applications. The primary factors contributing to the significant utilization of this material are its mechanical performance efficiency under stress and its comparatively low cost [22]. The plates adopted during the

experimentation had a dimension of 20x100x400 mm. All the tests were conducted in the 2003 Mazak Nexus Model 410A CNC Vertical Machining Center. The CoroMill 245 cutter, with five cutting edges, was used. Fig. 1 shows the plate and the cutting tool mounted in the CNC machine.

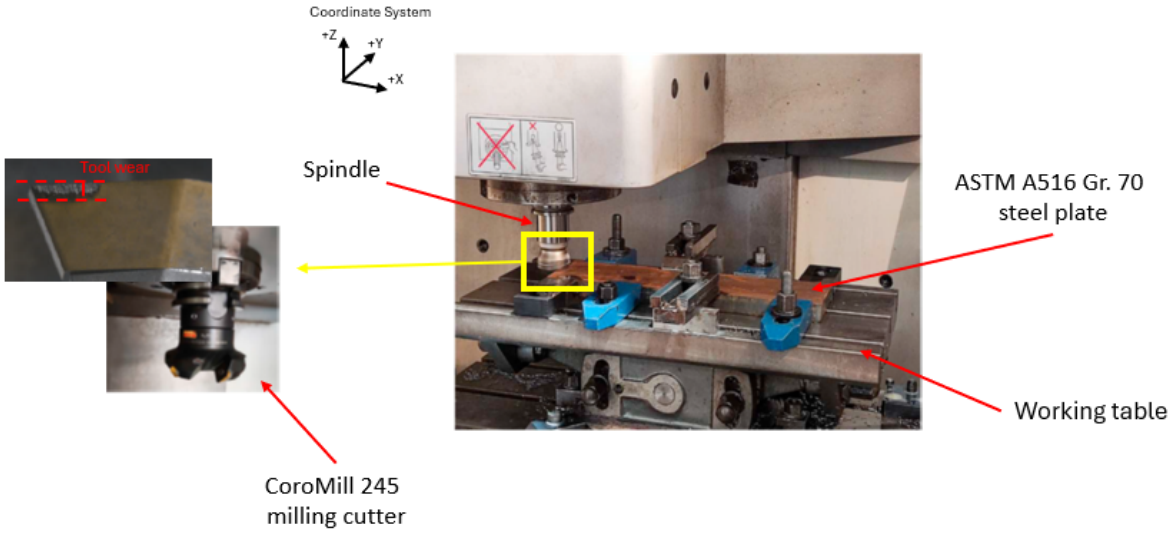
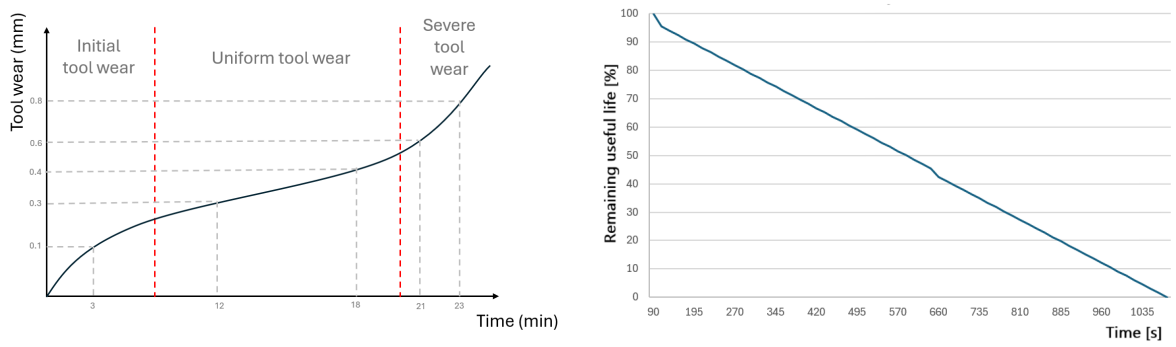


Figure 1: CNC machine configuration for experimental campaigns.

Tool wear measurements were taken after each milling operation, employing a Leica DM4000M metallographic optical microscope. Tool wear progression follows a typical pattern, as shown in Fig 2a. This pattern can be divided into three regions: (i) an initial tool wear zone characterized by rapid tool wear; (ii) a uniform tool wear zone; and (iii) a severe tool wear zone, which leads to tool failure. Measuring tool wear during the experiments makes it possible to collect data and identify the exact moment when the cutting tool enters the third zone and, consequently, when replacement is required. This also allows recording the tool's life and evaluating its remaining useful life (RUL) after each milling operation. The RUL can be described as a percentage value: a value of 100% indicates that the tool has never been used; the percentage value drops when the tool is used, and, as a consequence, the tool wear rate increases. The corresponding RUL trend derived from the measured tool wear is shown in Fig. 2b.

Table 1 lists the ranges of the cutting parameter values adopted during the experimental campaign.



(a) (b)

Figure 2: Tool degradation trends: (a) typical tool wear progression (b) corresponding RUL derived from wear measurements.

Table 1
Range values for the cutting parameters.

Cutting parameter	Unit of measure	Range values
Feed rate (f)	mm/min	800 – 1600
Spindle speed (n)	rpm	700 – 1400
Depth of cut (DOC)	mm	3.0 – 4.5

2.2. Problem Statement and Optimization Model Formulation

The present study aims to formulate a bi-objective optimization model based on three cutting parameters in a milling process—feed rate (f), spindle speed (n), and depth of cut (DOC)—to improve production efficiency and tool efficiency. The two optimization objectives are the material removal rate (MRR) and the relative number of cutting passes (R_{ct}).

MRR is considered to maximize the production efficiency. It can be computed as follows:

$$MRR = DOC \times f \times v \tag{1}$$

where v is the cutting speed that can be derived by the spindle speed and the diameter of the tool D as $\pi \times D \times n$.

R_{ct} is the factor that considers the tool efficiency. It is defined as the ratio between the number of passes executed by the tool up to a given moment and the maximum number of passes the tool can perform before replacement is required. The maximum number of cutting passes for a single tool was determined experimentally by recording the point at which tool wear reached the threshold requiring replacement. Consequently, R_{ct} is defined as follows:

$$R_{ct} = \frac{\#pass}{max(\#pass)} \tag{2}$$

where the maximum number of cutting passes is dynamically updated if a new strategy adopted during the milling operation exceeds the previous value. A value for R_{ct} close to zero indicates that the tool experienced premature wear, which can be indicative of lower efficiency, process stability issues, or more severe operating conditions.

The two objectives are in the following Eq. (3):

$$OF = \begin{matrix} \max MRR \\ \max R_{ct} \end{matrix} \tag{3}$$

Finally, the following set of constraints are defined:

$$f_{min} \leq f \leq f_{max} \tag{4}$$
$$n_{min} \leq n \leq n_{max} \tag{5}$$
$$DOC_{min} \leq DOC \leq DOC_{max} \tag{6}$$
$$RUL \geq RUL_{tlv} \tag{7}$$

Constraints (4)–(6) ensure that the cutting parameters remain within their minimum and maximum safety limits, denoted by the subscripts $\min$ and $\max$, respectively. Constraint (7) guarantees that the cutting tool is replaced when it reaches the end of its lifespan. Indeed, to maintain safe operating conditions, the threshold (RUL_{tlv}) is defined to avoid the breakage of the cutting tool and the damage to the workpiece surface. From a visualization of the RUL distribution in the collected dataset, a threshold of 15% is assigned as the critical point at which the severe tool wear stage begins for the cutting tool, as shown in Fig. 3.

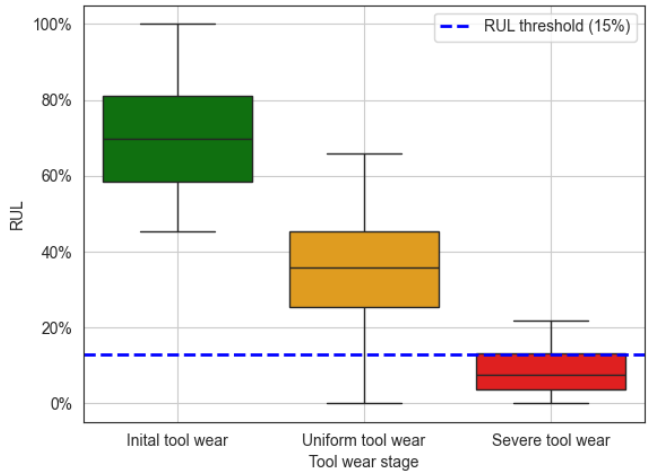


Figure 3: RUL distribution across tool wear stages.

2.3. RL-based optimization framework

The proposed RL-based cutting parameter optimization framework, summarised in Fig. 4, comprises three main phases, described as follows.

1. Let L be the length of material worked. A state s is defined by five parameters: $s = (f, n, DOC, L, RUL)$. A predictive model for RUL was developed to estimate the expected RUL of the next state given a starting state. For example, given the initial state $s_0 = (f_0, n_0, DOC_0, L_0, RUL_0)$, where RUL_0 is set to 100%, the model predicts the expected RUL of the next state $s_1 = (f_1, n_1, DOC_1, L_1, RUL_1)$. We trained a Linear Regression model in Python using the scikit-learn package [23]. The model achieved high prediction performance on the test set (R^2 = 0.99, RMSE= 0.008, and MAE= 0.007).
2. A custom environment that represents our *Milling Environment* was defined in Python by using the open-source library Gymnasium [24]. Continuous action and observation spaces were defined. The reward function was defined as the sum of the two objective functions in Eq. (3) as follows:

$$r = \omega_1 * MRR + \omega_2 * R_{ct} \tag{8}$$

   with MRR and R_{ct} normalized to [0, 1]. With the main goal of obtaining a policy that balance the two objectives, same weights for the two terms in Eq. (8) were considered. The two main methods of *step* and *reset* were constructed. The *step* method receives the action of the agent and returns the next state and the reward associated. The *reset* method enables the re-initialization of the environment at the end of each episode.
3. To solve the problem formulated in Sec. 2.2 with the single objective function (8), we applied four DRL algorithms – Proximal Policy Optimization (PPO) [25], Soft Actor-Critic (SAC) [26], Twin Delayed Deep Deterministic Policy Gradient (TD3), and Deep Deterministic Policy Gradient (DDPG) [27] [28]– in Python, by using the package stable-baseline3 [29].

3. Results and Discussion

The four DRL algorithms were trained in the custom *Milling Environment*, described in Sec. 2.3. The training was performed on a machine equipped with Windows 11 operating system, 16 GB of system memory, an NVIDIA GeForce RTX 4060 GPU (8 GB VRAM), and an Intel Core i7 processor.

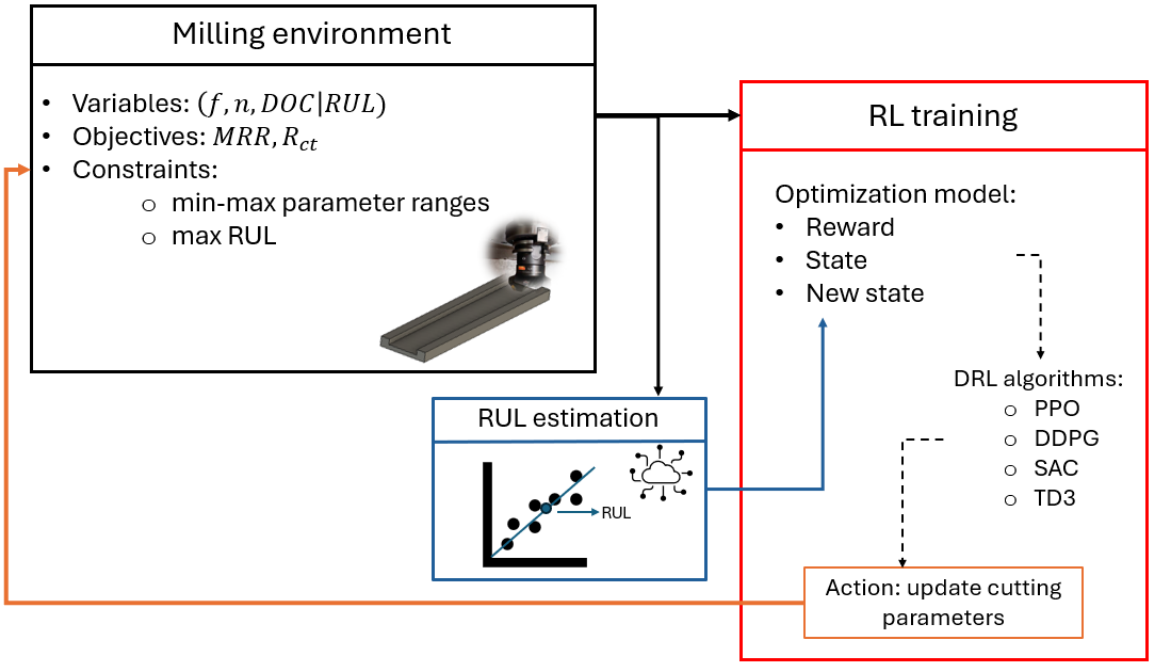


Figure 4: RL-based optimization framework.

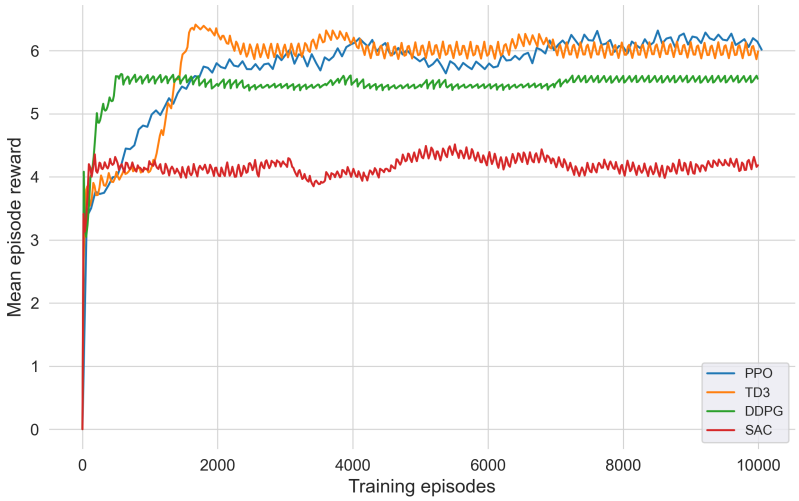
The four algorithms implemented (PPO, SAC, TD3, and DDPG) were applied using the *MLPPolicy* architecture. This type of policy defines multi-layer perceptrons with an architecture of two hidden layers with 64 neurons, used as the fuction approximator for both actor and critic networks of the DRL algorithms. The training progress of the four algorithms were monitored using Tensorboard [30]. The training was conducted over 10000 episodes.

Fig. 5a compares the four trained algorithms, showing the mean reward values across each episode for the four trained algorithms. The algorithm TD3 demonstrated the most stable learning behavior, converging to an average final reward of approximately 6 after 2000 episodes. PPO achieved a similar final reward but exhibited a less rapid initial learning curve, requiring 4000 episodes to achieve equivalent performance to TD3. Finally, DDPG outperformed SAC in terms of final reward, although both algorithms showed lower performance than the other two trained algorithms. Fig 6 provides some statistics about the rewards obtained during the training phase of each DRL algorithm: the average reward with its standard deviation, and the maximum and minimum reward values.

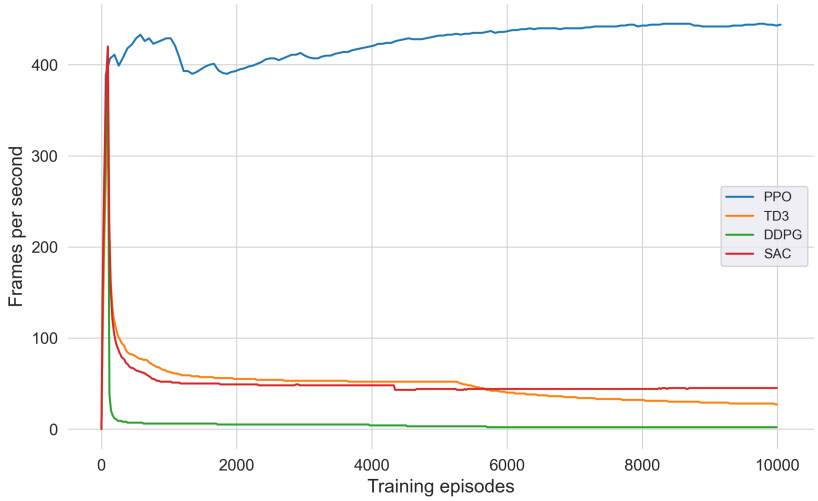
With regard to the computational efficiency of the four DRL algorithms, Fig. 5b presents the results in terms of frames per second, i.e. the number of iterations of the *Milling environment* that are completed per second. PPO emerged as the most efficient algorithm for the problem examined. As a result, although the TD3 algorithm demonstrated rapid convergence in terms of average reward, it demonstrated inferior performance with respect to computational efficiency in comparison to PPO. This finding suggests that PPO emerged as the most stable algorithm in terms of both convergence and computational efficiency, thereby achieving higher overall performance than the other three algorithms employed in these studies.

Finally, a trade-off analysis was performed for the two terms of the multi-objective reward formulation defined in Eq. (8), using the weight combinations (ω_1, ω_2) listed in Table 2.

Fig. 7 shows the results obtained by training the four DRL algorithms with each configuration of weights. The results highlight a clear trade-off between the two objectives: as the weight of one term increases, its corresponding performance improves at the expense of the other. This inverse relationship is particularly evident in the case of DDPG and TD3, which show large fluctuations depending on the weight configuration. Conversely, PPO and SAC exhibit more stable behavior across varying weight combinations. Moreover, the combination of balanced weights provides the best overall compromise



(a)



(b)

Figure 5: Performance comparison of the DRL algorithms: (a) mean reward per episode; (b) computational efficiency measured in frames per second.

across all algorithms, as it yields high values for both terms of the reward. This finding indicates that the algorithms are capable of satisfying both objectives without significantly compromising one in preference to the other. Additionally, the inclusion in the reward formulation of a term related to the tool efficiency clearly enhances the performance in that dimension. This indicates that the learning agent adapts its policy to reduce tool degradation, by defining strategies that are more sustainable in the long term compared to approaches that optimize only the MRR.

4. Conclusions

The present study was conducted with the objective of extending the tool lifespan for a milling process. The identification of a strategy that facilitates the adaptation of cutting parameters has the potential to model the process with respect to the tool wear progression – a key factor affecting the sustainability of a machining process. This strategy can contribute to the reduction of machine downtime, energy

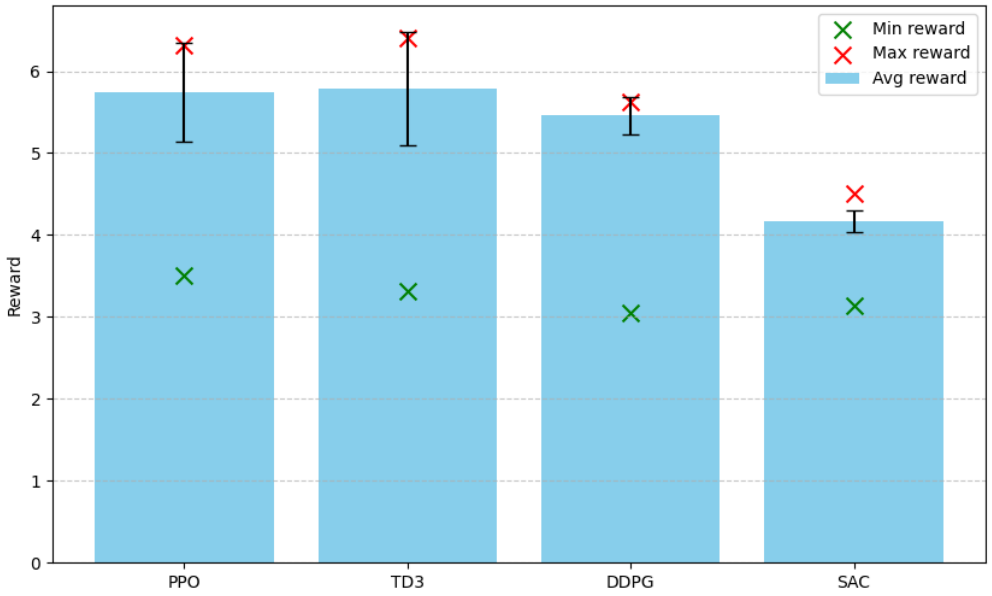


Figure 6: Statistics of the performance of the DRL algorithms.

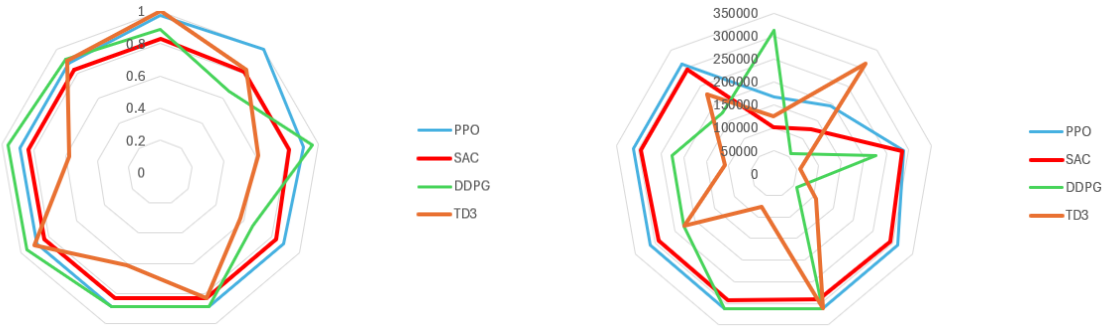
Table 2
Weight combinations used for the trade-off analysis.

ω_1	ω_2
0	1
0.1	0.9
0.25	0.75
0.4	0.6
0.5	0.5
0.6	0.4
0.75	0.25
0.9	0.1
1	0

consumption, and resource waste, thereby promoting a more sustainable approach to milling operations.

In this work the goal of extending tool lifespan is addressed alongside the objective of increasing the material removal rate for each milling operation. The aim of the study is twofold: firstly, to enhance the sustainability of the milling process, and secondly, to maintain a high level of efficiency. DRL techniques were applied to derive an adaptive strategy for selecting cutting parameters in a milling process. A custom-designed *Milling Environment* was utilised to evaluate the four DRL algorithms (PPO, TD3, SAC, and DDPG). Among them, PPO demonstrated superior performance, achieving a high mean reward and excellent computational efficiency during training. Specifically, PPO maintained a stable tool lifespan above 88% of the maximum across most weight combinations while simultaneously reaching the highest attainable MRR of approximately 312,000 mm^3/min. By contrast, SAC preserved a tool lifespan of ∼83% but with a lower MRR (approximately 295,000 mm^3/min), while DDPG achieved a competitive tool lifespan (∼96%) at the expense of reduced MRR (approximately 226,000 mm^3/min). Finally, TD3 generally resulted in significantly lower MRR and tool lifespan values. These results demonstrate that PPO achieves the most balanced trade-off, ensuring both machining efficiency and tool sustainability.

Future research could focus on enhancing the reliability of the proposed approach by incorporating real-time sensor data for the online monitoring of tool wear and process conditions. This would enable the introduction of additional constraints related to process stability, further aligning the strategy with



(a) (b)

Figure 7: Trade-offs in the multi-objective reward function obtained by varying the weights assigned to the R_{ct} and MRR terms during DRL algorithm training. The radar plots show: a) results related to the R_{ct} term; b) results related to the MRR term.

industrial needs. Moreover, the integration of multi-agent reinforcement learning could be explored to simultaneously optimize conflicting objectives, thereby supporting more advanced and balanced decision-making frameworks for smart manufacturing.

Acknowledgments

This work contributed to the basic research activities of the WP9.6: "AI for Green" supported by the PNRR project FAIR—Future AI Research (PE00000013), Spoke 9—Green-aware AI, under the NRRP MUR program funded by the NextGenerationEU.

Declaration on Generative AI

During the preparation of this work, the authors used DeepL and Grammarly for: Grammar and spelling check. After using these tools, the authors reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] G. M. Minquiz, M. A. Meraz-Melo, J. Flores Méndez, et al., Sustainable assessment of a milling manufacturing process based on economic tool life and energy modeling, Journal of the Brazilian Society of Mechanical Sciences and Engineering 45 (2023) 365. doi:10.1007/s40430-023-04189-8.

[2] D. Wu, C. Jennings, J. Terpenny, R. X. Gao, S. Kumara, A comparative study on machine learning algorithms for smart manufacturing: Tool wear prediction using random forests, Journal of Manufacturing Science and Engineering 139 (2017) 071018. doi:10.1115/1.4036350.

[3] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, R. X. Gao, Deep learning and its applications to machine health monitoring, Mechanical Systems and Signal Processing 115 (2019) 213–237. URL: https://www.sciencedirect.com/science/article/pii/S0888327018303108. doi:https://doi.org/10.1016/j.ymssp.2018.05.050.

[4] Z. Lin, Y. Fan, J. Tan, et al., Tool wear prediction based on xgboost feature selection combined with pso-bp network, Scientific Reports 15 (2025) 3096. doi:10.1038/s41598-025-85694-9.

[5] K. V. Rao, Assessment of tool condition and surface quality using hybrid deep neural network: Cnn-lstm-based segmentation and statistical analysis, Journal of Tribology 147 (2025) 084201. doi:10.1115/1.4067496.

[6] C. Liu, Y. Quan, Y. Zhou, et al., Intelligent rul prediction method of cutting tools based on gru-lstm, Journal of the Brazilian Society of Mechanical Sciences and Engineering 47 (2025) 278. doi:10.1007/s40430-025-05553-6.

[7] S. Han, U. Awasthi, G. M. Bollas, Physics-informed symbolic regression for tool wear and remaining useful life predictions in manufacturing, Journal of Manufacturing Systems 80 (2025) 734–748. doi:10.1016/j.jmsy.2025.03.023.

[8] C. Hao, Z. Wang, X. Mao, S. He, B. Li, H. Liu, F. Peng, W. Li, A novel and scalable multimodal large language model architecture tool-mmgpt for future tool wear prediction in titanium alloy high-speed milling processes, Computers in Industry 169 (2025) 104302. doi:10.1016/j.compind.2025.104302.

[9] R. S. Sutton, A. G. Barto, et al., Reinforcement learning: An introduction, volume 1, MIT press Cambridge, 1998.

[10] B. Kommey, O. J. Isaac, E. Tamakloe, D. Opoku4, A reinforcement learning review: Past acts, present facts and future prospects, IT Journal Research and Development 8 (2024) 120–142. URL: https://journal.uir.ac.id/index.php/ITJRD/article/view/13474. doi:10.25299/itjrd.2023.13474.

[11] H. Zhang, W. Wang, Y. Wang, Y. Zhang, J. Zhou, B. Huang, S. Zhang, Employing deep reinforcement learning for machining process planning: An improved framework, Journal of Manufacturing Systems 78 (2025) 370–393. doi:https://doi.org/10.1016/j.jmsy.2024.12.010.

[12] C. Li, P. Zheng, Y. Yin, B. Wang, L. Wang, Deep reinforcement learning in smart manufacturing: A review and prospects, CIRP Journal of Manufacturing Science and Technology 40 (2023) 75–101. doi:https://doi.org/10.1016/j.cirpj.2022.11.003.

[13] P. Wang, Y. Cui, H. Tao, X. Xu, S. Yang, Machining parameter optimization for a batch milling system using multi-task deep reinforcement learning, Journal of Manufacturing Systems 78 (2025) 124–152. doi:10.1016/j.jmsy.2024.11.013.

[14] W. Li, B. Li, S. He, X. Mao, C. Qiu, Y. Qiu, X. Tan, A novel milling parameter optimization method based on improved deep reinforcement learning considering machining cost, Journal of Manufacturing Processes 84 (2022) 1362–1375. doi:10.1016/j.jmapro.2022.11.015.

[15] W. Li, C. Hao, S. He, C. Qiu, H. Liu, Y. Xu, B. Li, X. Tan, F. Peng, Multi-agent reinforcement learning method for cutting parameters optimization based on simulation and experiment dual drive environment, Mechanical Systems and Signal Processing 216 (2024) 111473. doi:https://doi.org/10.1016/j.ymssp.2024.111473.

[16] S. Dharmadhikari, N. Menon, A. Basak, A reinforcement learning approach for process parameter optimization in additive manufacturing, Additive Manufacturing 71 (2023) 103556. doi:https://doi.org/10.1016/j.addma.2023.103556.

[17] F. Lu, G. Zhou, C. Zhang, Y. Liu, F. Chang, Z. Xiao, Energy-efficient multi-pass cutting parameters optimisation for aviation parts in flank milling with deep reinforcement learning, Robotics and Computer-Integrated Manufacturing 81 (2023) 102488. doi:10.1016/j.rcim.2022.102488.

[18] C. Li, X. Zhao, H. Cao, L. Li, X. Chen, A data and knowledge-driven cutting parameter adaptive optimization method considering dynamic tool wear, Robotics and Computer-Integrated Manufacturing 81 (2023) 102491. doi:10.1016/j.rcim.2022.102491.

[19] J. Lu, Z. Chen, X. Liao, C. Chen, H. Ouyang, S. Li, Multi-objective optimization for improving machining benefit based on woa-bbpn and a deep double q-network, Applied Soft Computing 142 (2023) 110330. doi:10.1016/j.asoc.2023.110330.

[20] X. Zhang, T. Yu, Y. Dai, S. Qu, J. Zhao, Energy consumption considering tool wear and optimization of cutting parameters in micro milling process, International Journal of Mechanical Sciences 178 (2020) 105628.

[21] N. Xie, J. Zhou, B. Zheng, Selection of optimum turning parameters based on cooperative optimization of minimum energy consumption and high surface quality, in: Proceedings of the 51st CIRP Conference on Manufacturing Systems, volume 72, Procedia CIRP, 2018, pp. 1469–1474. doi:10.1016/j.procir.2018.03.099.

[22] C. Prieto, C. Barreneche, M. Martínez, L. F. Cabeza, A. I. Fernández, Thermomechanical testing under operating conditions of a516gr70 used for csp storage tanks, Solar Energy Materials and

Solar Cells 174 (2018) 509–514. doi:10.1016/j.solmat.2017.09.029.

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, Édouard Duchesnay, Scikit-learn: Machine learning in python, Journal of Machine Learning Research 12 (2011) 2825–2830. doi:10.5555/1953048.2078195.

[24] M. Towers, A. Kwiatkowski, J. Terry, J. U. Balis, G. D. Cola, T. Deleu, M. Goulão, A. Kallinteris, M. Krimmel, A. KG, R. Perez-Vicente, A. Pierré, S. Schulhoff, J. J. Tai, H. Tan, O. G. Younis, Gymnasium: A standard interface for reinforcement learning environments, 2024. URL: https://arxiv.org/abs/2407.17032. arXiv:2407.17032.

[25] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms, 2017. URL: https://arxiv.org/abs/1707.06347. arXiv:1707.06347.

[26] T. Haarnoja, A. Zhou, P. Abbeel, S. Levine, Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor, 2018. URL: https://arxiv.org/abs/1801.01290. arXiv:1801.01290.

[27] T. P. Lillicrap, J. J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, Continuous control with deep reinforcement learning, 2019. URL: https://arxiv.org/abs/1509.02971. arXiv:1509.02971.

[28] S. Fujimoto, H. van Hoof, D. Meger, Addressing function approximation error in actor-critic methods, 2018. URL: https://arxiv.org/abs/1802.09477. arXiv:1802.09477.

[29] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, N. Dormann, Stable-baselines3: Reliable reinforcement learning implementations, Journal of Machine Learning Research 22 (2021) 1–8.

[30] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. URL: https://www.tensorflow.org/, software available from tensorflow.org.