Adaptive Maintenance in Intelligent Production Systems via the Synergistic Integration of Digital Twin and Reinforcement Learning

Amirhossein Khezri[1,*], Nathalie Klement[1], Belgacem Bettayeb[2], Richard Béarée[1], David Baudry[3]

[1]Arts et Métiers Institute of Technology, LISPEN, 59000 Lille, France
{amir_hossein.khezri,nathalie.klement,richard.bea-ree}@ensam.eu
[2]CESI, LINEACT, 59000 Lille, France
[3]CESI, LINEACT, 76800 Rouen, France
{bbettayeb,dbaudry}@cesi.fr

Abstract. The growing complexity of production systems demands intelligent maintenance strategies to enhance efficiency and reliability. This study proposes a Reinforcement Learning (RL)–based maintenance decision framework integrated within a digital twin environment for data-driven production management. The proposed approach combines real-time system state monitoring, digital twins, and RL to learn adaptive maintenance policies. A digital twin of the production system is used as a simulation environment where the learning agent evaluates operational actions, including operating, inspecting, calibrating, and performing maintenance. The decision process is guided by key indicators such as predicted production yield and machine health conditions, allowing the agent to balance production performance with maintenance interventions. The framework is demonstrated through an application to an exemplary stage of a battery manufacturing process, where the learned policy aims to control process degradation while maintaining stable production yield. Simulation results indicate that the proposed approach can effectively anticipate degradation and trigger maintenance actions at appropriate times. The study highlights the potential of combining digital twin technology and RL to support intelligent and adaptive maintenance strategies in smart manufacturing systems.

Keywords: Adaptive maintenance, Digital twin, Reinforcement Learning, Agent-based modeling.

1 Introduction

The increasing complexity of modern production systems has intensified the need for advanced monitoring, modeling, and decision-support tools capable of improving operational efficiency and system reliability. Among the emerging technologies supporting this transformation, digital twins have gained significant attention in recent years. Digital twins have emerged as a key enabling technology in Industry 4.0 and smart

… , p. 1, 2011.
© Springer-Verlag Berlin Heidelberg 2011

manufacturing [1]. They allow the integration of real-time operational data, simulation models, and analytical tools to monitor and analyze industrial systems. Digital twins enable data exchange between physical systems and their virtual counterparts, supporting performance monitoring, scenario evaluation, and data-driven decision-making in manufacturing environments. By integrating data from sensors, industrial control systems, and enterprise-level platforms such as Manufacturing Execution Systems (MES), digital twins enable real-time visibility of production systems and facilitate the analysis of complex interactions among machines, materials, and operational parameters. Furthermore, the ability of digital twins to simulate production processes under different operating conditions allows manufacturers to evaluate alternative operational strategies without disrupting the physical system.

Beyond production monitoring and optimization, digital twins also offer significant potential for maintenance management [2]. Maintenance planning is a critical aspect of industrial operations, as equipment failures can lead to costly production interruptions, reduced product quality, increased operational expenses, and major safety risks. Traditional maintenance strategies, such as corrective age- or time-based preventive maintenance, often fail to fully account for the dynamic operating conditions and degradation patterns of industrial equipment. Digital twins have therefore been increasingly explored to support advanced maintenance strategies based on equipment condition and operational data. Through this integration, digital twins can support condition-based and predictive maintenance strategies by estimating equipment degradation, forecasting failures, and recommending optimal maintenance interventions. The integration of digital twins with advanced analytical techniques, including machine learning and optimization methods, further enhances their capability to support maintenance decision-making. By simulating equipment degradation and operational scenarios, digital twins can evaluate the impact of different maintenance actions on production performance and system reliability. Such capabilities are particularly valuable in complex and highly automated industrial environments where multiple production processes interact and equipment reliability directly affects overall system performance [3]. One example is lithium-ion battery production, where large-scale production systems consist of interconnected processes such as slurry mixing, electrode coating, drying, calendaring, slitting, and cell assembly. In these environments, failures or performance degradation in critical equipment can lead to substantial production losses and quality issues. Digital twins therefore offer a promising approach for monitoring equipment conditions, simulating degradation behavior, and supporting maintenance planning in such complex production systems.

The rest of the paper is organized as follows. Section 2 reviews the relevant literature on digital twins and intelligent maintenance strategies in production systems. Section 3 presents the proposed digital twin–driven adaptive maintenance framework and its mathematical formulation. Section 4 describes the model deployment and analyzes the obtained results. Finally, Section 5 concludes the paper and outlines future research directions.

2 Literature Review and Related Works

To better understand the current advances and limitations in this domain, it is necessary to review the existing literature on digital twins and their applications in production and maintenance management. A particular attention is given to recent developments in digital twin–based monitoring and predictive maintenance approaches, as well as their potential application in complex production environments such as lithium-ion battery production.

2.1 Digital Twins in Service of Production Systems

The concept of digital twins has emerged as a key enabling technology in the context of Industry 4.0 and smart production. A digital twin refers to a dynamic virtual representation of a physical system that is continuously updated using real-time operational data, simulation models, and advanced analytics. By establishing a bidirectional connection between the physical system and its digital counterpart, digital twins enable monitoring, analysis, and prediction of system behavior throughout the lifecycle of industrial assets [4]. In production systems, digital twins have been widely investigated for applications such as process monitoring, performance analysis, and production optimization. Early implementations of digital twins primarily focused on representing individual machines or components to support condition monitoring and fault diagnosis. These models typically relied on sensor data and physics-based simulations to provide insights into equipment performance and operational conditions [5].

Recent research has extended digital twin applications toward system-level representations, enabling the modeling of entire production lines. Such models integrate various aspects of the production process, including material flow, machine interactions, quality propagation, and energy consumption. By capturing the interdependencies among different production stages, system-level digital twins allow manufacturers to analyze complex operational scenarios and evaluate the impact of production strategies on system performance. Furthermore, digital twins have been increasingly combined with advanced data analytics techniques, including machine learning and artificial intelligence, to enhance predictive capabilities [6]. These approaches enable the identification of hidden patterns in operational data and support the development of predictive models for system performance and equipment degradation [7]. Despite these advancements, many digital twin implementations rely on computationally intensive simulation models, which may limit their applicability in large-scale production environments requiring rapid scenario evaluation or iterative optimization.

2.2 Digital Twin–Driven Maintenance and Intelligent Decision-Making

Maintenance management plays a crucial role in ensuring the reliability and efficiency of modern production systems. Traditional strategies such as corrective age- or time-based preventive maintenance often fail to address the dynamic operating conditions and complex degradation patterns of industrial equipment [8]. Consequently, data-driven approaches, including condition-based and predictive maintenance, have gained

increasing attention. In this context, digital twins provide a powerful framework by integrating sensor data, degradation models, and operational information to continuously monitor equipment health and anticipate potential failures [9]. This capability enables proactive maintenance planning, reducing unplanned downtime and improving system availability. Several studies have investigated the use of digital twins for predictive maintenance in industrial environments [2]. In such approaches, the digital twin serves as a virtual representation of the physical system, allowing simulation and analysis of equipment degradation under varying operating conditions. Predictive models—often based on machine learning or statistical techniques—are used to estimate key indicators such as remaining useful life, failure probabilities, and performance degradation. These predictions support maintenance planning, resource allocation, and production scheduling [10]. As a result, digital twin technologies have been increasingly adopted across various industrial sectors, including automotive manufacturing, aerospace systems, energy production, and semiconductor fabrication.

More recently, research has explored integrating intelligent decision-making methods within digital twin environments. Reinforcement Learning (RL) has emerged as a promising approach for optimizing operational and maintenance strategies through interaction with simulated systems. By using the digital twin as a training environment, RL agents can learn optimal decision policies without disrupting real-world production [11-13]. However, developing high-fidelity digital twins for large-scale production systems often involves significant computational costs. Moreover, many existing studies focus mainly on failure prediction at the machine level, with limited attention given to system-level maintenance decision-making. These challenges highlight the need for more efficient and integrated digital twin frameworks capable of supporting intelligent maintenance optimization in complex production environments.

3 Proposed Digital Twin–Driven RL for Adaptive Maintenance

The integration of digital twin technology with RL provides a promising approach for improving maintenance decision-making in complex production environments. By combining real-time operational data with simulation-driven system modeling and data-driven learning algorithms, it becomes possible to evaluate maintenance strategies and optimize decisions under uncertain system conditions. A structured framework is proposed to link the physical system, its digital counterpart, and the RL agent, as detailed in the following subsections.

3.1 Reinforcement Learning–Integrated Digital Twin Framework

An effective predictive maintenance strategy requires a framework capable of integrating real-time operational data, virtual system representations, and data-driven decision-making mechanisms. To address this requirement, the proposed framework combines a digital twin simulation environment with an RL agent in order to support adaptive inspection and maintenance decisions in production systems. Fig. 1 illustrates the overall framework of the proposed digital twin–driven adaptive maintenance model.

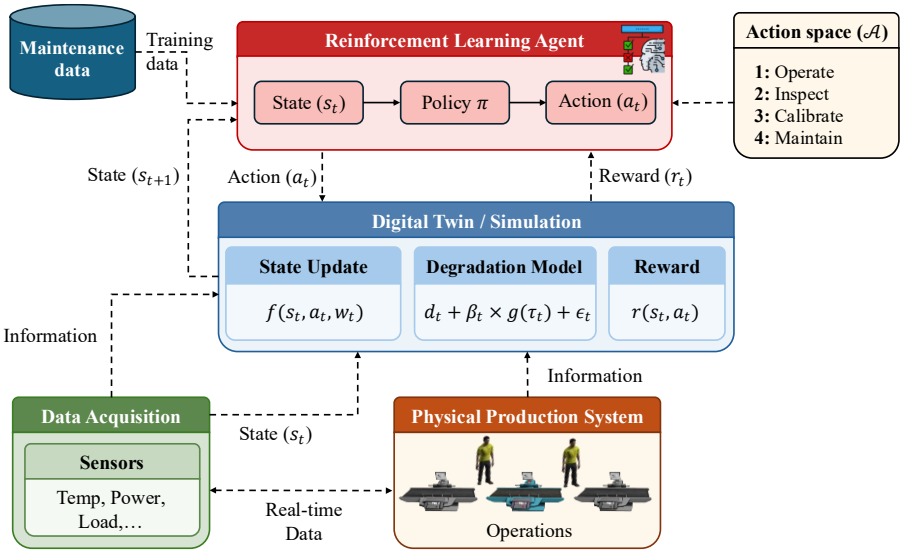

Fig. 1. Proposed digital twin–driven RL framework for adaptive maintenance

The proposed framework integrates several components that enable data-driven maintenance optimization. At the foundation lies the data acquisition layer, which collects operational data from production equipment through embedded sensors. These sensors monitor variables such as temperature, vibration, load, and power consumption, providing real-time information about equipment condition. The collected measurements define the system state x_t, representing the operational status of the production system at a given time. This state information serves as the primary input for both the digital twin and the RL agent. The physical production system represents the real manufacturing environment, including machines, production lines, and operational processes. During operation, the system continuously generates sensor data describing equipment behavior and process conditions. These observations ensure that the digital twin remains synchronized with the real system while providing the contextual information necessary for maintenance decision-making. At the center of the framework is the digital twin, which acts as a virtual representation of the production system. Using the observed system state, the digital twin reproduces equipment behavior and degradation dynamics through simulation models. This environment allows different maintenance strategies to be evaluated without interrupting the physical process. Performance indicators such as production efficiency, maintenance cost, and equipment reliability are computed within the simulation to assess the impact of alternative maintenance decisions.

The decision-making process is carried out by an RL agent that interacts with the digital twin environment. The agent observes the current system state and determines an appropriate action according to a decision policy. The action space considered in this framework includes four possible decisions: continuing operation without intervention, performing an inspection to assess equipment condition, applying a calibration procedure, or executing a maintenance operation to restore system performance.

Continuing operation represents the default strategy when the system operates within acceptable performance limits, and no corrective action is required. Inspection actions are triggered when early signs of performance degradation are detected in the production system. Through inspection, the current condition of the equipment can be evaluated, and potential faults can be identified before they evolve into critical failures. If the inspection confirms that system performance is deteriorating, a calibration procedure may be applied. Calibration represents a corrective intervention aimed at adjusting machine parameters or operational settings in order to partially recover system performance. Compared with full maintenance operations, calibration generally involves lower costs and shorter downtime, although it may not completely eliminate the underlying source of degradation. When the deterioration persists or the calibration proves insufficient to restore acceptable operating conditions, a maintenance action is executed. Maintenance operations correspond to more comprehensive repair or replacement interventions intended to restore the system to a near-optimal operational state.

Maintenance decisions are generated by a RL agent interacting with the digital twin environment. Based on the current system state, the agent selects actions including continuing operation, performing inspection, applying calibration, or executing maintenance. Through repeated interactions with the simulated environment, the agent learns policies that balance production continuity, intervention costs, and system reliability. Historical maintenance data can also be incorporated to calibrate the digital twin and improve the representation of degradation patterns.

3.2 Mathematical Formulation and Learning Model Representation

The maintenance optimization problem can be formulated as a stochastic sequential decision process in which the objective is to determine optimal maintenance actions based on the current operational condition of the production system. Let $s_t \in S$ denote the observable system state at time t. In the proposed framework, the state represents the operational health of the equipment. The health state is derived from the production degradation rate, which serves as a practical indicator of equipment degradation. Let d_t denote the degradation rate at time t, representing the proportion of defective or rejected production units generated during operations. The observable system state is therefore defined as

$$s_t = H_{\max} - d_t \tag{1}$$

where $H_{\max}$ represents the maximum achievable health level of the equipment. In the proposed model, $H_{\max} = 100$, and higher values of s_t correspond to better operational conditions. At each decision step, a maintenance action $a_t \in \mathcal{A}$ is selected from a finite set of possible actions:

$$\mathcal{A} = \{Operate, Inspect, Calibrate, Maintain\} \tag{2}$$

These actions represent operational decisions that can be applied to the production system. Continuing operation corresponds to normal production without intervention. Inspection allows operators to assess the equipment condition when early degradation signals are detected. Calibration represents a corrective intervention aimed at adjusting

machine parameters in order to partially restore production performance. Maintenance corresponds to a full intervention intended to recover the equipment to a near-optimal operational state. The evolution of the system state follows a stochastic transition process that captures both degradation dynamics and the effects of maintenance actions. The state transition can therefore be expressed as

$$s_{t+1} = f(s_t, a_t, w_t) \tag{3}$$

where $f(\cdot)$ represents the system transition dynamics and w_t denotes stochastic disturbances associated with uncertain degradation processes and operational variability. Within the proposed framework, degradation is modelled through the evolution of the degradation rate d_t. In many production systems, degradation tends to increase progressively with machine usage rather than following a linear pattern. This behavior can be approximated through a degradation function expressed as

$$d_{t+1} = d_t + \beta_t \times g(\tau_t) + \epsilon_t \tag{4}$$

where τ_t denotes the time elapsed since the last major maintenance intervention, $g(\tau_t)$ is a generic degradation growth function capturing the effect of machine usage over time, β_t is a scaling factor that may vary depending on the selected maintenance action, and ϵ_t represents stochastic disturbances associated with operational variability. The function $g(\cdot)$ can take different forms depending on the characteristics of the production process, such as linear, logarithmic, exponential, or other nonlinear degradation patterns. Maintenance actions influence the degradation process through recovery effects. Calibration partially reduces the accumulated degradation rate, while maintenance resets the degradation process. The effect of maintenance actions on the degradation rate can therefore be expressed as

$$d_{t+1} = \begin{cases} 0, & a_t = a_3 \\ \max(d_t - \eta, 0), & a_t = a_2 \\ d_t + \beta_t \times g(\tau_t) + \epsilon_t, & a_t \in \{a_0, a_1\} \end{cases} \tag{5}$$

where η represents the reduction in degradation rate achieved through calibration. Continuing operation and inspection do not directly restore the system condition and therefore allow degradation to continue. The relationship between equipment degradation and production performance is evaluated through the digital twin environment. Let y_t denote the production yield ratio predicted by the digital twin model. This quantity is estimated using the digital twin model developed from production data and expressed as

$$y_t = \Phi(d_t) \tag{6}$$

where $\Phi(\cdot)$ represents the digital twin model mapping equipment degradation to production performance indicators. At each decision step, the system receives a reward reflecting the trade-off between production performance, maintenance cost, and equipment reliability. The reward function can be expressed as

$$r(s_t, a_t) = Reward_{production}(s_t) - Cost_{maintenance}(a_t) - Penalty_{failure}(s_t) \tag{7}$$

where $Reward_{production}(s_t)$ represents the production benefit associated with the current system condition, $Cost_{maintenance}(a_t)$ denotes the cost associated with the selected maintenance action, and $Penalty_{failure}(s_t)$ represents the penalty associated with degraded or failure states of the equipment. The objective of the maintenance strategy is to determine a policy π that maps system state to maintenance actions in order to maximize the expected cumulative reward over a given decision horizon. The optimal maintenance policy π^* can therefore be defined as [14]

$$\pi^* = arg\max_{\pi} \mathbb{E}\left[\sum_{t=0}^{T} \gamma^t r(s_t, \pi(s_t))\right] \tag{8}$$

where $\gamma \in [0,1]$ is a discount factor representing the relative importance of future rewards and Tdenotes the decision horizon. RL methods can be employed to approximate this optimal policy by iteratively learning the value of state–action pairs through interactions with the digital twin environment. In particular, the action-value function $Q(s_t, a_t)$ can be updated using a temporal-difference learning rule given by [14]

$$Q^{new}(s_t, a_t) = Q(s_t, a_t) + \alpha \times \left[r(s_t, a_t) + \gamma \times \max_{a'} Q(s_{t+1}, a') - Q(s_t, a_t)\right] \tag{9}$$

where α denotes the learning rate and the term $\max_{a'} Q(s_{t+1}, a')$ represents the estimated value of the best future action. Through repeated updates, the learning algorithm converges toward a policy that balances production continuity, maintenance costs, and failure risks.

4 Model Deployment and Performance Analyses

To illustrate the practical application of the proposed framework, the RL policy is deployed within a representative stage of the battery manufacturing process. The objective is to evaluate how the learned maintenance strategy supports operational decision-making while maintaining production performance and equipment reliability. As shown in Fig. 2, the RL agent is integrated into the stacking stage of the battery production line.

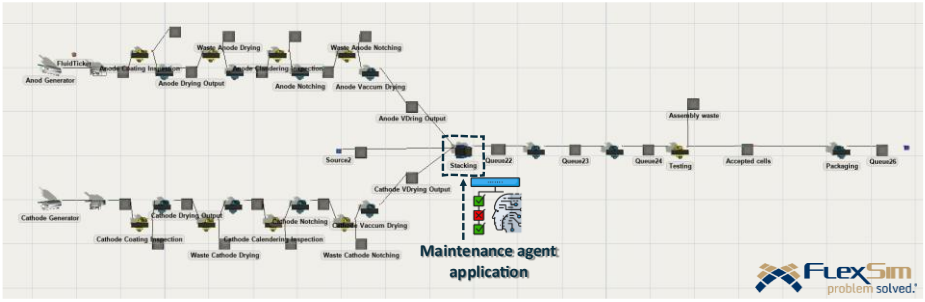

Fig. 2. Application of the RL–based maintenance framework in a digitalized battery production (stacking stage example)

Sensor data collected from the process are used to estimate the system state, including the predicted production yield and machine health condition. Based on this information, the agent selects an operational action—operate, inspect, calibrate, or maintain—to control the process and reduce the generation of anode waste while maintaining stable production performance. Although demonstrated for the stacking operation, the same framework can be applied to other stages of the battery manufacturing process. For the deployment of the learning model, several operational assumptions are introduced to guide the decision-making policy. At each decision step, the agent selects one of four actions: operate (a_0), inspect (a_1), calibrate (a_2), or maintain (a_3). The decision process is primarily guided by the predicted yield y_t, which determines the preferred operational response, and the machine health state s_t, which influences reward adjustments. High yield levels encourage continued operation, moderate degradation triggers inspection or calibration actions, and low yield conditions result in maintenance interventions. Additionally, operating in severely degraded health states is penalized, while operating under optimal conditions receives a small reward bonus. These boundaries guide the agent's learning process and ensure that maintenance decisions balance production efficiency, equipment reliability, and intervention costs.

4.1 Experimental Results and Policy Evaluation

To evaluate the effectiveness of the learned maintenance policy, the following analysis examines its impact on production performance and machine degradation over a 50-day horizon. The discussion focuses on the evolution of degradation rates and yield ratios, highlighting how the RL agent balances maintenance decisions with operational efficiency. Based on the simulation data, **Fig. 3** and **Fig. 4** depict the behavior of the learned maintenance policy over a 50-day production horizon. The trajectory of the selected actions in **Fig. 3** shows how the agent adapts its decisions as machine degradation evolves

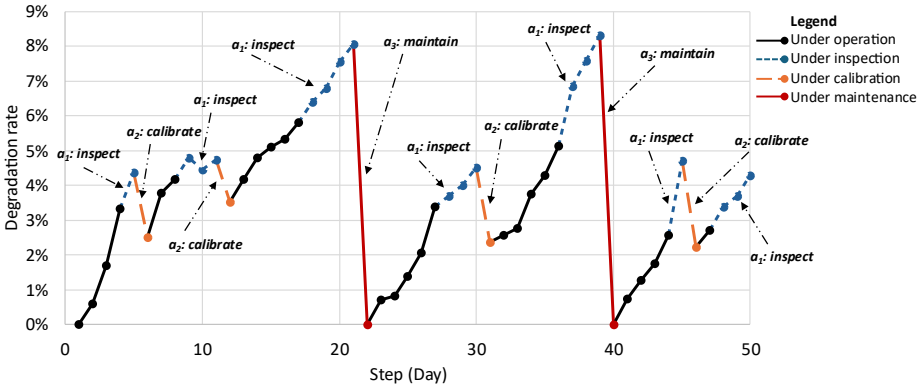

Fig. 3. Evolution of degradation rate and maintenance actions under the learned RL policy

At the beginning of each operational cycle the agent mainly selects operation and inspection actions, indicating that the machine is initially considered healthy and capable of sustaining production without major interventions. As the process continues and degradation accumulates, the policy progressively introduces calibration actions to partially mitigate deterioration. However, once the system approaches a critical degradation level, the agent triggers a full maintenance intervention. This behavior can be clearly observed around day 21 and day 39, where the action plot exhibits distinct spikes corresponding to maintenance decisions. These interventions reset the system condition and allow the production cycle to restart from a healthier state.

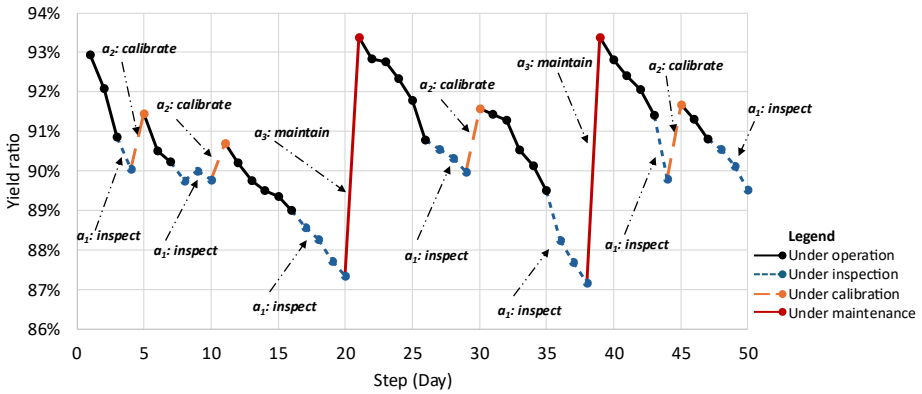

Fig. 4. Impact of degradation and maintenance actions on yield ratio during the 50-day simulation

Fig. 4 highlights the direct impact of degradation on production performance. When the degradation rate is low at the beginning of each cycle, the yield ratio remains high, typically around 92–93%. As degradation progresses and the degradation rate approaches its higher levels near 7–8, the yield gradually decreases to approximately 87–88%. This decline reflects the negative effect of machine deterioration on production quality. Immediately after maintenance resets the degradation rate, the yield ratio recovers sharply back to values close to 93%, demonstrating the restoration of production efficiency. When these figures are interpreted together, the learned policy reveals a clear operational strategy. The agent allows the machine to operate while degradation gradually increases from 0 to roughly 7–8 units of degradation, during which the yield decreases from about 93% to around 87–88%. Once the degradation reaches this region, the policy decides that the loss in yield outweighs the maintenance cost and performs a maintenance action. The resulting reset of the degradation rate immediately restores the yield to the higher 92–93% range, allowing production to continue efficiently.

Overall, the numerical trends observed in the results demonstrate that the RL agent has learned to balance short-term operational gains with long-term system health. Maintenance is not triggered continuously but instead occurs only when degradation begins to significantly affect production performance. This behavior shows that the learned policy effectively maintains yield within a relatively stable range of

approximately 87% to 93% while preventing the degradation rate from growing beyond critical levels. As a result, the agent successfully captures the dynamic relationship between machine degradation, maintenance decisions, and production efficiency over the simulated 50-day production horizon.

5 Conclusion and Future Perspectives

This study proposed a RL–based framework integrated with a digital twin environment to support maintenance decision-making in manufacturing systems. The approach combines real-time system state monitoring, digital twin, and RL to learn adaptive maintenance policies. By interacting with the digital twin, the agent selects operational actions—operate, inspect, calibrate, or maintain—based on predicted production yield and machine health conditions. The results indicate that the learned policy effectively balances production continuity, equipment reliability, and maintenance interventions, maintaining stable yield while preventing excessive degradation.

Future works will focus on extending the framework to more complex production environments, including multi-machine and production line-level optimization where interactions between processes must be considered. In such settings, multi-agent RL could be explored to coordinate maintenance decisions across multiple machines or process stages. In addition, future research may investigate the integration of advanced surrogate modeling techniques to approximate complex production dynamics and reduce computational costs within the digital twin environment. Such surrogate-assisted learning approaches could enable faster policy training and more efficient evaluation of maintenance strategies in large-scale industrial systems.

Acknowledgment. This project has received funding from the European Union's Horizon Europe program under grant agreement No. 101137954. The authors would like to thank the rest of the BATTwin consortium for supporting this research.

Author Contributions. **Amirhossein Khezri:** Writing – Original draft, Writing – Review & Editing, Data curation, Conceptualization, Methodology, Software, Formal analysis, Investigation. **Nathalie Klement**: Writing – Review & Editing, Supervision, Investigation. **Belgacem Bettayeb**: Writing – Review & Editing, Supervision, Investigation. **Richard Béarée**: Writing – Review & Editing, Project administration, Funding acquisition. **David Baudry:** Writing – Review & Editing, Supervision, Project administration.

References

1. Thelen A, Zhang X, Fink O, Lu Y, Ghosh S, Youn BD, et al. A comprehensive review of digital twin — part 1: modeling and twinning enabling technologies. Struct Multidiscip Optim. 2022;**65**(12):354. https://doi.org/10.1007/s00158-022-03425-4

2. Chen S, Bekar ET, Bokrantz J, Skoogh A. AI-enhanced digital twins in maintenance: Systematic review, industrial challenges, and bridging research–practice gaps. J Manuf Syst. 2025;**82**:678–99. https://doi.org/10.1016/j.jmsy.2025.07.006
3. Errandonea I, Beltrán S, Arrizabalaga S. Digital Twin for maintenance: A literature review. Comput Ind. 2020;**123**:103316. https://doi.org/10.1016/j.compind.2020.103316
4. Overbeck L, Graves SC, Lanza G. Development and analysis of digital twins of production systems. Int J Prod Res. 2024;**62**(10):3544–58. https://doi.org/10.1080/00207543.2023.2242525
5. Tliba K, Diallo TML, Penas O, Ben Khalifa R, Ben Yahia N, Choley J-Y. Digital twin-driven dynamic scheduling of a hybrid flow shop. J Intell Manuf. 2023;**34**(5):2281–306. https://doi.org/10.1007/s10845-022-01922-3
6. Aivaliotis P, Georgoulias K, Chryssolouris G. The use of Digital Twin for predictive maintenance in manufacturing. Int J Comput Integr Manuf. 2019;**32**(11):1067–80. https://doi.org/10.1080/0951192X.2019.1686173
7. Pandhare V, Negri E, Ragazzini L, Cattaneo L, Macchi M, Lee J. Digital twin-enabled robust production scheduling for equipment in degraded state. J Manuf Syst. 2024;**74**:841–57. https://doi.org/10.1016/j.jmsy.2024.04.027
8. Van Dinter R, Tekinerdogan B, Catal C. Predictive maintenance using digital twins: A systematic literature review. Inf Softw Technol. 2022;**151**:107008. https://doi.org/10.1016/j.infsof.2022.107008
9. Li Y, Tao Z, Wang L, Du B, Guo J, Pang S. Digital twin-based job shop anomaly detection and dynamic scheduling. Robot Comput-Integr Manuf. 2023;**79**:102443. https://doi.org/10.1016/j.rcim.2022.102443
10. Chen J, Zhou X. Reinforcement learning based maintenance scheduling of flexible multi-machine manufacturing systems with varying interactive degradation. Reliab Eng Syst. 2025;**260**:111018. https://doi.org/10.1016/j.ress.2025.111018
11. Yan Q, Wang H, Wu F. Digital twin-enabled dynamic scheduling with preventive maintenance using a double-layer Q-learning algorithm. Comput Oper Res. 2022;**144**:105823. https://doi.org/10.1016/j.cor.2022.105823
12. Ogunfowora O, Najjaran H. Reinforcement and deep reinforcement learning-based solutions for machine maintenance planning, scheduling policies, and optimization. J Manuf Syst. 2023;**70**:244–63. https://doi.org/10.1016/j.jmsy.2023.07.014
13. Hu Y, Miao X, Zhang J, Liu J, Pan E. Reinforcement learning-driven maintenance strategy: A novel solution for long-term aircraft maintenance decision optimization. Comput Ind Eng. 2021;**153**:107056. https://doi.org/10.1016/j.cie.2020.107056
14. Sutton RS, Barto AG. Reinforcement learning: An introduction: MIT press Cambridge; 1998.