Proceedings of the 2024 Winter Simulation Conference
H. Lam, E. Azar, D. Batur, S. Gao, W. Xie, S. R. Hunter, and M. D. Rossetti, eds.

ENHANCING DIGITAL TWINS WITH DEEP REINFORCEMENT LEARNING: A USE CASE IN MAINTENANCE PRIORITIZATION

Siyuan Chen[1], Paulo Victor Lopes[2], Silvan Marti[1], Mohan Rajashekarappa[1], Sunith Bandaru[3], Christina Windmark[4], Jon Bokrantz[1], and Anders Skoogh[1]

[1]Dept. of Industrial and Materials Science, Chalmers University of Technology, Gothenburg, SWEDEN
[2]Dept. of Computer Science, Aeronautics Institute of Technology (ITA), Sao Jose dos Campos, BRAZIL
[3]School of Eng. Science, University of Skövde, Skövde, SWEDEN
[4]Dept. of Mechanical Eng. Sciences, Lund University, Lund, SWEDEN

ABSTRACT

This paper introduces an innovative framework that enhances digital twins with deep reinforcement learning (DRL) to support maintenance in manufacturing systems. Utilizing a sophisticated artificial intelligence (AI) layer, this framework integrates real-time and historical production data from a physical manufacturing system to a digital twin, enabling dynamic simulation and analysis. Maintenance decisions are informed by DRL algorithms that analyze this data, facilitating smart maintenance strategies that adaptively prioritize tasks based on predictive analytics. The effectiveness of this approach is demonstrated through a case study in a lab-scale drone factory, where maintenance tasks are prioritized using a proximal policy optimization. This integration not only refines maintenance decisions but also aligns with the broader goals of operational efficiency and sustainability in Industry 4.0. Our results highlight the potential of combining DRL with digital twins to significantly enhance decision-making in industrial maintenance, offering a novel approach to predictive and prescriptive maintenance practices.

1 INTRODUCTION

As Industry 4.0 continues to evolve, manufacturing maintenance practices are increasingly adopting intelligent strategies, shifting towards predictive and even prescriptive maintenance modalities (Bokrantz et al. 2020). Digital twins (DT), virtual representations of physical systems, offer a promising solution by capturing real-time data and enabling virtual simulations (Schleich et al. 2017). These advancements highlight the critical role of technology in transforming traditional manufacturing landscapes into highly efficient, data-driven environments.

Despite these significant advancements, current solutions often fail to integrate real-time operational data with advanced AI methodologies within a cohesive, scalable framework (Aldoseri et al. 2023). Moreover, the application of DRL in digital twins for maintenance decision support remains underexplored, especially in the context of dynamic and complex manufacturing environments (Siraskar et al. 2023). This oversight limits the potential for digital twins to effectively support intelligent maintenance (Huang et al. 2021), suggesting a clear gap in the application of AI to optimize maintenance strategies.

This paper addresses these gaps by introducing a novel framework that integrates AI with DT for enhanced maintenance in manufacturing field. Our framework creates a closed-loop system for continuous optimization, populated with real-time data from manufacturing processes, such as event log. The AI layer employs DRL to analyze data, simulate production scenarios, recommend optimal maintenance actions, and relay decisions back to maintenance practitioners.

A case study involving a drone assembly line demonstrates the efficacy of our framework. The DT model, implemented in Plant Simulation software and enhanced by DRL, prioritizes maintenance based on critical factors. By merging AI with DT, our approach provides a robust tool for decision-making, enhancing

979-8-3315-3420-2/24/$31.00 ©2024

the predictive capabilities of maintenance systems. This integration not only optimizes operational efficiency but also contributes to sustainable manufacturing practices.

In Section 2, we reviewed the related work, exploring how AI enhances DT, the application of DRL in smart maintenance, and decision-support system in this field. Section 3 details the AI-enhanced DT framework and maps maintenance problems to DRL, including a case study using proximal policy optimization to optimize maintenance priorities. Section 4 compares the results of the DRL application with emprical and random selection of priorities. Section 5 discusses the applied work and outlines future research directions. Section 6 concludes the paper and summarizes key findings.

2 LITERATURE REVIEW

2.1 AI Enhanced Digital Twins

In contemporary industrial applications, digital twins and smart industrial systems, based on AI and ML, are crucial for enhancing operational flexibility, knowledge transference and data driven decision-making (Ciano et al. 2021). AI techniques arm digital twins with tools to create models based on observed behavior and historical data, improving efficiency of data analysis and prediction accuracy (Huang et al. 2021). AI-ML and big data significantly enhance DT applications by facilitating advanced simulations, real-time monitoring, predictive maintenance, and decision-making processes (Rathore et al. 2021).

A manufacturing DT provides data to train AI models, enabling decision-making, real-time monitoring, predictive maintenance, and the simulation of manufacturing processes (Kharchenko et al. 2020). In a practical context, the creation of virtual datasets via DT simulations addresses the challenges of collecting high-quality real-world data for training AI and ML models (Alexopoulos et al. 2020). For example, a modular AI trained in synthetic data generated by DTs can be used to dynamically adapt manufacturing systems, showing up to 10% improvement in the process time (Mo et al. 2023).

In the smart manufacturing of the future, the spread use of blockchain technology is expected to break information silos and release a large amount of previously inaccessible data, creating immeasurable potential for DT virtual workspaces development (Lv and Xie 2022). DT potential applications aligns to Industry 4.0 objectives in practical cases as product design and development, process design and optimization, quality control and predictive maintenance, empowerment and cross-functional collaboration, facility and asset management, production planning and control, training, and education (Attaran et al. 2023). Ultimately, the intersection of metaverse, DT, and AI in training and maintenance not only streamlines processes but also facilitates up-skilling in a more interactive and effective manner, potentially transforming maintenance practices aligned to the principles of Industry 5.0 (Bordegoni and Ferrise 2023).

2.2 Deep Reinforcement Learning Applied for Smart Maintenance

Reinforcement learning (RL) entails an agent's acquisition of behavior through iterative trial-and-error interactions with a dynamic environment (Kaelbling et al. 1996). Most DRL algorithms employed today leverage deep neural networks to approximate complex functions, enabling the learning of policies for sequential decision-making tasks. Among these algorithms, Deep Q-learning Network (DQN) utilizes a neural network to estimate action values, while Policy Gradient methods directly optimize policy parameters through gradient ascent (Huang 2020). Actor-Critic methods combine value-based and policy-based approaches, employing separate networks to estimate action values and policy distributions (Grondman et al. 2012). Proximal Policy Optimization (PPO) addresses stability concerns by constraining policy updates (Schulman et al. 2017), while Trust Region Policy Optimization (TRPO) ensures policy updates remain within a safe region (Schulman et al. 2015). Additionally, Deep Deterministic Policy Gradient (DDPG) extends DRL to continuous action spaces (Li et al. 2019), crucial for many real-world applications such as traffic control (Casas 2017).

Maintenance is becoming an increasingly prominent focus of DRL applications (Panzer and Bender 2022). A framework proposed based on value-decomposition multi-agent actor–critic algorithm for designing preventive maintenance policies in large-scale manufacturing systems, effectively addressing the challenges posed by system complexity and action space explosion (Su et al. 2022). A predictive maintenance (PdM) model is proposed based on a long short-term memory (PPO-LSTM) model which outperformed other DRL methods or humans in a simulated maintenance repair environment (Su et al. 2022). DQN is applied and deep RL's effectiveness is demonstrated in early fault detection for rotary machines, detecting errors without manual tuning or expert intervention (Dai et al. 2020). Due to the limited exploration of existing algorithms and the absence of hands-on guidelines, only few deep RL applications were being evaluated in real-world scenarios, therefore further refinement is imperative (Panzer and Bender 2022).

2.3 Decision Support System to Maintenance

Traditional maintenance methods often result in premature machine part replacements and production line downtime, leading to significant material, time, and financial waste (Taşcı et al. 2023). Recently, the adoption of AI and the Industrial Internet of Things (IIoT) in manufacturing has shifted focus towards PdM. This modern approach significantly boosts manufacturing efficiency and reliability by preempting equipment failures before they impact production. Furthermore, the synergy of AI with IIoT not only optimizes equipment lifecycles but also drives substantial cost reductions and operational enhancements (Anandan et al. 2022).

Advanced decision support systems, crucial for modern maintenance strategies, leverage real-time, dynamic, and interactive dashboards to monitor sensor readings and respond swiftly to anomalies, enhancing PdM and tool life estimation (Sdiri et al. 2023). Intelligent Decision Support Systems (IDSS) integrate various AI technologies, including machine learning and data analytics, to process and analyze the vast data generated from maintenance activities and machinery operation. This analysis provides critical insights that support maintenance managers in reducing unplanned downtime and extending equipment life. Furthermore, E-maintenance systems utilize AI to combine web-based technologies with maintenance functions, enabling remote monitoring, management, and predictive analytics to foresee failures and suggest preventative actions (Turner et al. 2019). To handle the challenges of real-time data processing in industrial setups, edge computing can be deployed. This approach places machine learning models and predictive maintenance algorithms directly on edge devices, enhancing operational efficiency in the demanding and dynamic environments typical of industrial IoT (Hafeez et al. 2021).

3 METHODOLOGY

3.1 Overview of Conceptual Framework for AI Enhanced Digital Twins Support Smart Maintenance

The proposed framework, illustrated in Figure 1, establishes a three-layered architecture for facilitating smart maintenance decision-making. The real system layer includes the physical manufacturing assets and sensors, continuously generating production data. The DT layer ingests this data, creating a virtual representation of the system that mirrors real-world conditions. The integrated AI layer employs advanced techniques to analyze this digital twin data and provide actionable insights. Maintenance practitioners interact with the framework, receiving recommendations from the AI layer and executing decisions that directly impact the real system. This closed-loop architecture promotes continuous learning and optimization of maintenance processes.

In detail, the framework begins with a generic physical manufacturing system equipped with sensors that collect real-time production data. This data, which includes processing times, cycle times, setup times, and failure events, each with timestamps, is transmitted to a centralized data platform that stores both historical and live data. Utilizing the Open Platform Communications Unified Architecture (OPC UA) protocol (Mahnke et al. 2009), the system synchronizes and imports this data into a simulation environment, creating an accurate and continuously updated digital twin of the physical manufacturing system. This digital twin facilitates detailed simulations and analyses based on actual production data.

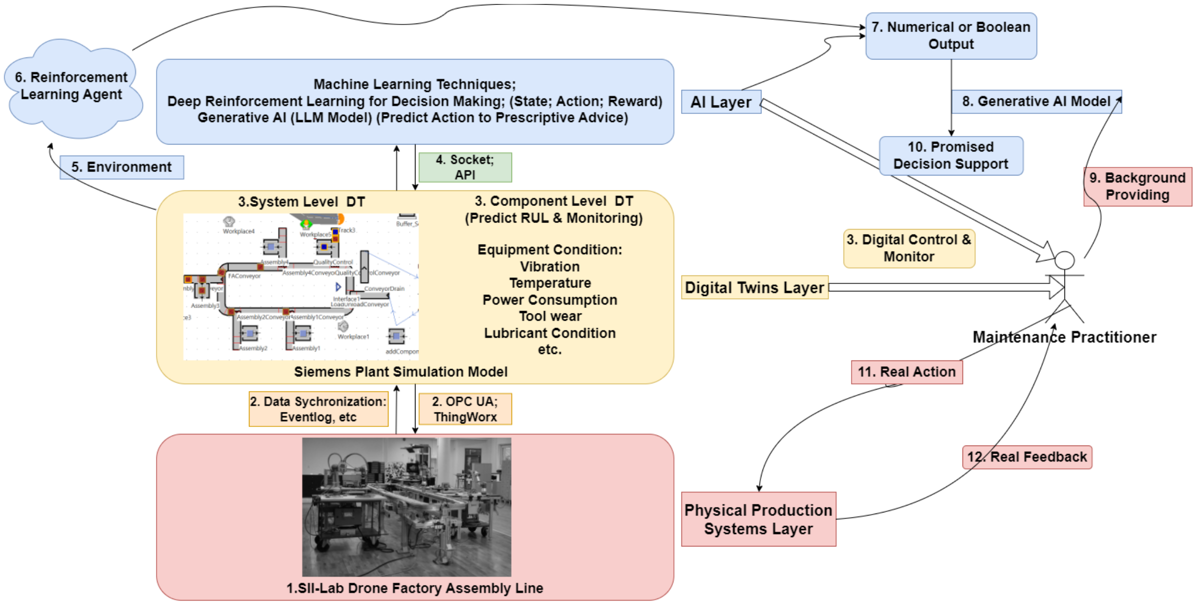

6. Reinforcement Learning Agent
Machine Learning Techniques;
Deep Reinforcement Learning for Decision Making; (State; Action; Reward)
Generative AI (LLM Model) (Predict Action to Prescriptive Advice)
7. Numerical or Boolean Output
AI Layer
8. Generative AI Model
10. Promised Decision Support
9. Background Providing
5. Environment
4. Socket; API
3.System Level DT
3. Component Level DT (Predict RUL & Monitoring)
Equipment Condition:
Vibration
Temperature
Power Consumption
Tool wear
Lubricant Condition
etc.
Siemens Plant Simulation Model
3. Digital Control & Monitor
Digital Twins Layer
Maintenance Practitioner
11. Real Action
2. Data Sychronization: Eventlog, etc
2. OPC UA; ThingWorx
12. Real Feedback
Physical Production Systems Layer
1.SII-Lab Drone Factory Assembly Line

Figure 1: AI-enhanced DT for PdM framework.

Real-time data interchange between the digital twin and the AI layer is facilitated through established APIs. The AI layer employs advanced machine learning techniques to process and analyze this data, simulating various production scenarios to generate actionable maintenance insights. To make these insights accessible, a Large Language Model (LLM) is applied to interpret the boolean or numerical outputs from machine learning or deep reinforcement learning models. The LLM utilizes context provided by maintenance practitioners to transcribe these outputs into comprehensible maintenance recommendations.

These recommendations, informed by the real-time state of the DT, are then displayed through a user interface, guiding personnel on the necessary actions. The maintenance team implements these actions in the physical system, and the results are immediately fed back to the digital twin. This creates a closed-loop system that continuously updates the data interchange between the physical system and the AI layer. Such a dynamic allows the AI algorithms to continuously refine their decision-making based on feedback from the physical system, thereby enhancing and optimizing maintenance strategies.

3.2 Conceptual Framework for RL Support Maintenance

DRL is a core tool in the AI layer and how it supports maintenance framework is illustrated in Figure 2. It utilizes the digital twin as simulation interface, enabling iterative interaction with a virtual model of the manufacturing equipment. The maintenance problem is formulated into a Markov Decision Process (MDP) and can be described by a six-element tuple (S, A, P, γ, R, π). Where S represents the set of possible states of the system or equipment, A represents the set of possible maintenance actions. $P(s', r|s, a)$ is the transition probability function, defining the probability of transitioning to state s' and receiving reward r after taking action a in state s; $\gamma \in [0, 1]$ is the discount factor, determining the importance of future rewards; $R(s, a)$represents the reward function, specifying the immediate reward received when taking action a in state s. $\pi(a|s)$ represents the policy, defining the probability of selecting action a in state s.

In the DRL framework for smart maintenance, states are captured by an array of production parameters, reflecting both the real-time operational status of equipment and key performance metrics. The operational status encompasses whether equipment is running, idle, under maintenance, or has failed, while performance metrics include throughput, efficiency, and quality. Additionally, health indicators such as vibration levels and temperature, maintenance history, resource availability, and current workload are integrated to provide a multi-faceted view of the manufacturing system's condition at each decision point (Siraskar et al. 2023).

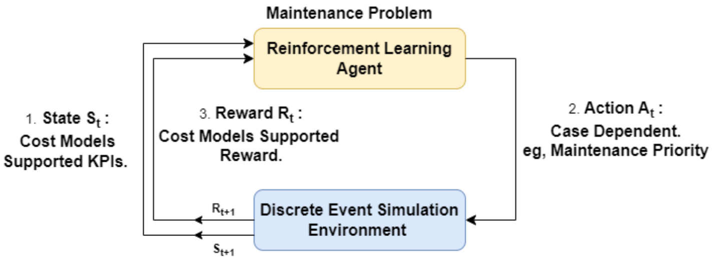

Maintenance Problem
Reinforcement Learning Agent
1. State S$_t$: Cost Models Supported KPIs.
3. Reward R$_t$: Cost Models Supported Reward.
2. Action A$_t$: Case Dependent. eg, Maintenance Priority
R$_{t+1}$
S$_{t+1}$
Discrete Event Simulation Environment

Figure 2: DRL support maintenance framework.

This rich state representation allows the DRL agent to strategically prioritize maintenance tasks, promoting an optimal balance between immediate functionality and long-term system health.

The action space of the DRL model is designed to be adaptable to various maintenance scenarios. Possible actions include scheduling maintenance, which involves deciding whether to perform immediate maintenance, defer it to a later time, or continue operations without intervention. Additionally, resource allocation actions allow for the assignment of maintenance personnel, spare parts, and tools to specific tasks. Preventive measures also form a part of the action space, where actions are taken based on predictive analytics to preemptively address potential failures.

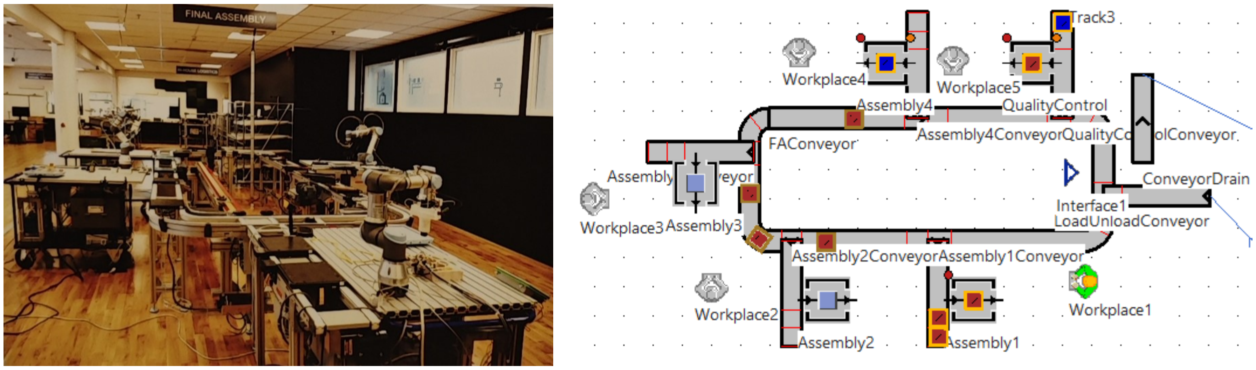
The effectiveness of the DRL agent is primarily rely on its reward function, which is designed for two objectives. Firstly, it incorporates the detailed cost models of the production environment, ensuring that financial aspects are thoroughly considered. Secondly, it aligns with broader sustainability goals, promoting environmentally conscious decisions. This sophisticated reward function, shown in the formula provided below (Ståhl and Windmark 2022), includes various elements of manufacturing costs, providing a comprehensive basis for decision-making.

$$
\begin{aligned}
r = {} & \underbrace{\frac{K_A}{N_0}\left[\frac{1}{n_{pA}}\right]_a}_{\text{Tool cost}} + \underbrace{\frac{k_B}{N_0}\left[\frac{N_0}{(1-q_Q)\cdot(1-q_B)}\right]_b}_{\text{Workpiece material cost}} + \underbrace{\frac{k_{CP}}{60N_0}\left[\frac{x_p\cdot t_0\cdot N_0}{(1-q_Q)\cdot(1-q_P)}\right]_{c1}}_{\text{Equipment and production cost}} \\
& + \underbrace{\frac{k_{CS}}{60N_0}\left[\frac{x_p\cdot t_0\cdot N_0}{(1-q_Q)\cdot(1-q_P)}\cdot\frac{q_S}{(1-q_S)} + x_{su}\cdot T_{su} + \frac{1-U_{RP}}{U_{RP}}\cdot T_{pb}\right]_{c2}}_{\text{Equipment cost during downtime}} + \underbrace{\frac{1}{N_0}(K_{AUH}+K_{CUH}+K_{GUH})_e}_{\text{Maintenance costs}} \\
& + \underbrace{\frac{n_{op}\cdot k_D}{60N_0}\left[\frac{x_p\cdot t_0\cdot N_0}{(1-q_Q)\cdot(1-q_S)\cdot(1-q_P)} + x_{su}\cdot T_{su} + \frac{1-U_{RP}}{U_{RP}}\cdot T_{pb}\right]_d}_{\text{Salary cost}} + \underbrace{\frac{1}{N_0}(K_{HL}+K_{Tno}+K_{RW})_h}_{\text{Handling, no operation, and rework costs}}
\end{aligned}
$$

Each term within the equation corresponds to a specific category of expenses, adjusted for the number of parts produced N_0, and is modified by factors representing the efficiency and quality of the production process such as scrap rates q_Q, q_P, q_S, cycle times t_0, setup times T_{su}, and machine uptime U_{RP}. By integrating these diverse cost factors into the reward function, the DRL agent is incentivized to find maintenance strategies that not only reduce immediate costs but also enhance overall production system sustainability.

3.3 Use Case: PPO Empower DT on Maintenance Priority

The use case was conducted at the SII-Lab in Lindholmen, Gothenburg, which serves as a Swedish national testbed for industrial digitalization. There is a drone assembly line that has a main conveyor belt with four assemble stations and one quality check station, and a digital twin model created in Siemens Plant Simulation which synchronized with the real system by Thingworx IIOT platform (Chávez et al. 2022). The real system and the digital twin model are shown in Figure 3.



(a) Drone assembly stations.

(b) PlantSim digital twin model.

Figure 3: Real system and virtual system.

For the purposes of this use case, certain operational parameters will be hypothetically adjusted to facilitate a more controlled experimental environment. Notably, the availability of each station is postulated at 60 percent, intentionally set below the real-world operational availability, to simulate a stress-test scenario. Within the operational dynamics of the system, simultaneous downtimes across multiple assembly stations may occur. In such instances, a maintenance prioritization strategy is experimented to address the most critical bottlenecks within the production system.

3.3.1 Problem Formulation

The problem is formulated by MDP that mentioned above.

Action Representation The action space, denoted by A, is a discrete and finite set, comprising 24 unique actions that represent all permutations of maintenance priority levels for three assembly stations (AS1,AS3,AS4) and quality control station (QC), calculated as 4!. Each action $a \in A$ is a permutation of the set $\{1,2,3,4\}$, with each element a_i signifying the assigned priority for the i-th station. A higher numeric value of a_i reflects a greater urgency for maintenance work. In instances of simultaneous downtimes, the action selected by the policy dictates that the station with the highest priority a_i receives immediate attention, ensuring operational continuity for the most critical processes.

State Representation The state representation is designed to capture the operational dynamics of the assembly system, enabling effective decision-making in the reinforcement learning framework. At any time t, the state s_t is a vector $s_t = (\mathrm{FPY}_t, \mathrm{Ava}_t, \mathrm{TPH}_t, \mathrm{CT}_t, \mathrm{AU}_i t)$, where each component is relevant to system performance:

1. FPY_t, first pass yield, indicating the proportion of products meeting quality standards without requiring rework. This KPI reflects the efficiency of the production process.
2. Ava_t stands for the availability of the system, representing the ratio of the operating time to the planned production time. It is crucial for assessing the reliability of the assembly stations.
3. TPH_t, throughput per hour, measures the number of units produced per hour, serving as a direct indicator of production capability.
4. CT_t, the cycle time, is the total time from the start to the end of the production process for a single product. This metric is vital for identifying bottlenecks in the production line.
5. $\mathrm{AU}_i t$, assembly utilization efficiency, expresses the degree to which the assembly capacity is used over a certain period. It is computed by the ratio of the active assembly time to the total available assembly time.

These KPIs are selected as state variables due to their capacity to provide a clear and quantitative snapshot of the system's current state. The calculation of each state variable is derived from underlying cost models.

Reward Function The reward function is strategically derived from select components of a comprehensive cost model. Given the partial availability of data within the model, the function integrates only those elements for which empirical values are ascertainable. To ensure uniformity in valuation, each contributing factor is normalized within the unit interval $[0,1]$. Consequently, the immediate reward at time t, R_t, is computed as a weighted sum of these normalized factors:

$$R_t = w_1 \cdot \mathrm{N_{OPE}} + w_2 \cdot \mathrm{N_{CT}} + w_3 \cdot \mathrm{N_{SR}} + w_4 \cdot \mathrm{N_{TPH}}$$

where:

- $\mathrm{N_{OPE}}$ is the Normalized Overall Process Effectiveness, $\mathrm{N_{OPE}} = \frac{OPE}{\max(OPE)}$.
- $\mathrm{N_{CT}}$ is the Normalized Cycle Time, $\mathrm{N_{CT}} = \frac{\max(CT)-CT}{\max(CT)-\min(CT)}$.
- $\mathrm{N_{SR}}$ is the Normalized Scrap Ratio, $\mathrm{N_{SR}} = \frac{\max(SR)-SR}{\max(SR)}$.
- $\mathrm{N_{TPH}}$ is the Normalized Throughput Per Hour, $\mathrm{N_{TPH}} = \frac{TPH-\min(TPH)}{\max(TPH)-\min(TPH)}$.
- w_1, w_2, w_3, and w_4 are the weights corresponding to each KPI, with constraint $w_1 + w_2 + w_3 + w_4 = 1$.

This ensures that the reward function is balanced, directing the RL agent towards optimal actions that enhance production effectiveness, cycle time efficiency, scrap minimization, and throughput maximization. This approach effectively balances sustainability and productivity.

3.3.2 Training Algorithm of PPO

PPO is a type of policy gradient algorithm in DRL.The policy gradient algorithm is very sensitive to the step size, but it is difficult to choose the appropriate step size, and the difference between the old and new policies during the training process is detrimental to the learning process if the difference between the old and the new policies is too large (Pirotta et al. 2013). PPO proposes a new objective function that can be updated in small batches in multiple training steps, which solves the problem of difficult to determine the step size in the policy gradient algorithm (Schulman et al. 2017). There are two general approaches of PPO, one is PPO-Penalty, where the limit of KL divergence is put into the objective function using Lagrange multipliers, and the coefficients in front of the KL divergence are constantly updated during the iterations (Wang et al. 2019). The other form is PPO-Clip, which puts restrictions in the objective function to ensure that the gap between the new parameters and the old ones is not too large (Huang et al. 2024). We selected the PPO-Clip and its optimization objective equation is shown at Equation (1).

$$L(s,a,\theta_k,\theta) = \min\left(\frac{\pi_\theta(a|s)}{\pi_{\theta_k}(a|s)}A_{\pi_{\theta_k}}(s,a), \mathrm{clip}\left(\frac{\pi_\theta(a|s)}{\pi_{\theta_k}(a|s)}, 1-\varepsilon, 1+\varepsilon\right)A_{\pi_{\theta_k}}(s,a)\right) \tag{1}$$

where clip(x, l, r) := max(min(x, r), l), and bounds l and r define the clipping range.

Our PPO model adopts a structure comprising five fully connected layers for both the Actor and Critic networks. With three hidden layers nestled between the input and output layers, our architecture is designed to extract hierarchical features from the state space, enabling effective decision-making. Leveraging Rectified Linear Unit (ReLU) activation and Adam optimization, we ensure robustness and adaptability during training. The Pseudocode of how PPO optimize maintenance priority is shown at Algorithm 1. Part of parameters of PPO we used is shown at Table 1.

3.3.3 Experiment Design

In our experiment, we first compared the PPO with both random and empirical selection methods. In a separate set of tests, we adjusted the KPI weightings to reflect their importance as perceived by different

Algorithm 1 PPO algorithm for optimizing maintenance priority for assembly stations

```
Initialize Actor network parameters (Actor network)
Initialize Critic network parameters (Critic network)
for each iteration over the environment do
    Initialize the memory buffer (PPO Memory)
    while not at the end of the assembly process do
        Observe the current state, $s_t$
        if now is decision point $t$ (System starts and finishes setting up assembly stations) then
            Obtain action $a_t$ and $\pi(a_t|s_t)$ based on the current policy $\pi$ with $s_t$ as input
            Obtain value $V(s_t)$ from Critic network with $s_t$ as input
            Execute action $a_t$ and choose maintenance priority for stations
            Get reward $r_t$ and next state $s_{t+1}$ from environment
            Store the trajectory $\tau_t = (s_t, a_t, r_t, V(s_t), \pi(a_t|s_t))$ into memory buffer
    for each epoch do
        Get samples from memory buffer
        for each step in the trajectory do
            Calculate advantages, $\hat{A}_t$, using Generalized Advantage Estimation (GAE)
        for each mini-batch do
            With gradient tape:
                Calculate ratio $\pi(a_t|s_t)/\pi_{\text{old}}(a_t|s_t)$
                Compute PPO-clip objective
                Calculate actor loss and critic loss
            Compute gradients and perform Adam optimization for actor and critic
    Clear memory buffer
    Optionally save and/or load models to/from the checkpoint directory
```

Table 1: Hyperparameters for PPO algorithm.

Hyperparameter	**Value**
Hidden Nodes	256 &128 &64
Policy Clip	0.15
Training Episodes	1000
Batch Size	64
Mini Batch Size	16
Update Epochs (n_epochs)	20
Discount Factor (gamma)	0.99
GAE Lambda	0.95
Actor & Critic Learning Rate	0.0003

stakeholders. We conducted five experiments in total: four prioritized one specific KPI with a dominant weight of 0.7, while assigning 0.1 to the others to emphasize strategic priorities. The fifth experiment served as a control with equal weights for all KPIs to establish baseline performance. We then analyzed the results to assess the impact of these weightings.

4 RESULT

The learning curve from the training of PPO as shown at Figure 4 illustrates the progression of average rewards per episode over 1000 episodes. The plot indicates an initial learning phase with high variability in rewards, stabilizing to a relatively consistent performance level near 0.65 as the episodes progress, suggesting that the model's policy is converging and improving in optimizing the decision-making process.

Due to empiricism and interviews with the technicians in charge of the laboratory, the maintenance priority of each machine is AS1, QC, AS3, AS4 in descending order, which shows the different importance of those machines. The comparison of maintenance priority selection strategies: Random Decision, Empirical Decision, and PPO, under the same weight conditions is presented in Figure 5. Analyzing the median and average values, it is evident that the Empirical Decision strategy outperforms the Random Decision. This finding emphasizes the value of maintenance practitioners' experience in making effective decisions. Conversely, the PPO strategy demonstrates superior performance compared to the Empirical Decision. This enhanced efficacy is attributed to the PPO agent's ability to timely detect system bottlenecks based on state changes.

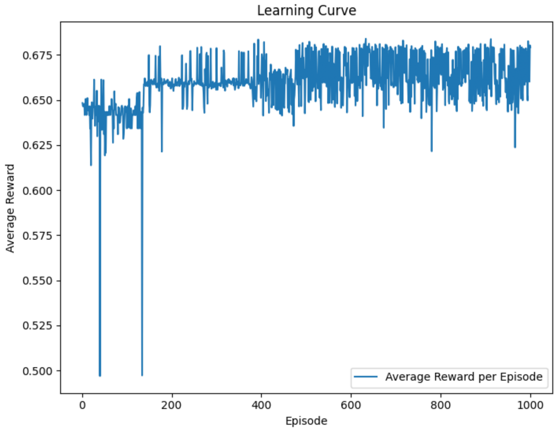


Figure 4: PPO agent learning curve.

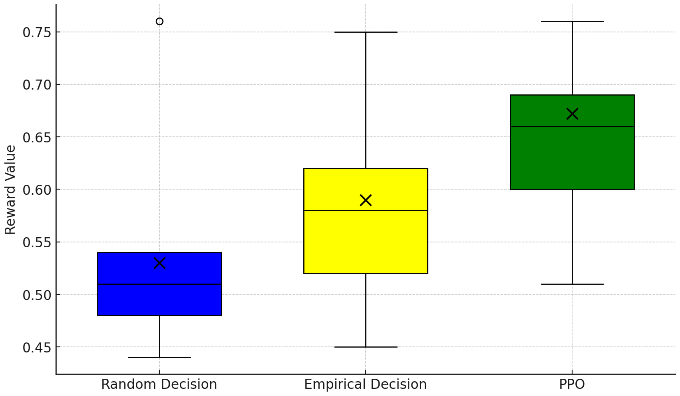


Figure 5: Comparison of maintenance priority selection.

Despite the varied performance across strategies, the maximum reward values for all three scenarios are remarkably similar, suggesting the presence of an optimization ceiling in this context. Regarding the minimum values, both Random and Empirical Decisions exhibit comparable results, whereas the minimum value achieved through PPO is significantly higher. This indicates that PPO can effectively elevate the lower performance threshold of the entire production system, thus ensuring more consistent outcomes.

The variations in reward functions, as presented in Table 2, illustrate how maintenance priorities shift based on the differing emphases within the production system. This table clearly demonstrates that changing the weight of each parameter that reflects different stakeholder requirements will significantly alter the resulting maintenance strategies. Such findings emphasize the importance of carefully designing the reward functions in RL to align with specific operational goals and stakeholder expectations. This adaptability is crucial for optimizing maintenance decisions to enhance efficiency and effectiveness in dynamic manufacturing environments.

5 DISCUSSION & FUTURE WORK

In the preceding sections, we introduced an AI-driven framework designed to enhance DT for informed maintenance decision-making. This framework incorporates RL, specifically tailored reward functions derived from production cost models, to facilitate smart maintenance strategies. A practical implementation using PPO demonstrated the framework's effectiveness in optimizing maintenance decisions within a

Table 2: Reward and maintenance priority based on weights.

w1	w2	w3	w4	Average Reward	Major Priority
0.25	0.25	0.25	0.25	0.672	AS3, AS4, AS1, QC
0.7	0.1	0.1	0.1	0.728	AS4, QC, AS1, AS3
0.1	0.7	0.1	0.1	0.647	AS4, QC, AS3, AS1
0.1	0.1	0.7	0.1	0.668	AS3, QC, AS4, AS1
0.1	0.1	0.1	0.7	0.580	AS1, AS3, QC, AS4

DT environment of a lab-scale drone factory. This case study highlighted the framework's capability to dynamically prioritize maintenance tasks, moving beyond traditional static scheduling approaches.

The results of this application are promising, showcasing significant enhancements in operational efficiency through the use of AI-enhanced DT. Comparatively, our approach advances the integration of reinforcement learning in industrial maintenance, which has been less emphasized in prior studies. However, it is important to acknowledge that while our simulation model performs well under controlled conditions, the assumed constant station availability may not accurately reflect the more complex dynamics of real-world manufacturing environments. Alternative explanations for the observed efficiencies might include the unique configurations of the simulated environment which may not be entirely replicable in different industrial contexts.

One of the limitations of our study is the simplified assumptions about station availability in our simulation based DT model, which could affect the generalizability of our findings. Future iterations of our research will need to address these simplifications by incorporating more complex and variable data that better represents real-world conditions.

Looking forward, our research will explore additional DRL techniques, such as Double Deep Q-Learning, to tackle more complex maintenance scheduling and path optimization challenges within the SII-Lab's digital twin model. To meet stakeholder's requirements from our previous research (Chen et al. 2023), we also plan to refine our existing framework to enhance its application in real-world settings. To make the results more interpretable to maintenance professionals, we aim to deploy LLM that will be fed with large industrial time-series data, to assist in decoding and elucidating the decisions made by the DRL agents, thus bridging the gap between advanced AI techniques and practical maintenance applications.

6 CONCLUSION

In conclusion, this paper has presented an AI-enhanced digital twin framework designed for smart maintenance in the manufacturing sector, embodying the progressive aspirations of Industry 4.0. Through the integration of DRL with DT, we showcased in a detailed case study focused on a drone assembly line how real-time data can be effectively leveraged to refine maintenance decisions, thereby boosting operational efficiency and sustainability. Notably, our work contributes a novel approach by mapping the maintenance problem to RL, specifically through the development of cost models for the reward function, enhancing the decision-making process. The results emphasize the framework's robustness in providing actionable insights, significantly contributing to predictive and prescriptive maintenance practices.

ACKNOWLEDGMENTS

The study was supported by the Swedish innovation agency VINNOVA under grant no.2021-02537 (Integrated Manufacturing Analytics Platform, IMAP project) and grant no. 2021-02421 (Digitala Stambanan). The computation was enabled by resources provided by Chalmers e-Commons at Chalmers. The work was carried out within Chalmers' Area of Advanced Production whose support is greatly acknowledged. During the preparation of this work the authors used ChatGPT 4.0 in order to proofread and enhance readability. After using this tool, the authors reviewed and edited the content and take full responsibility.

REFERENCES

Aldoseri, A., K. Al-Khalifa, and A. Hamouda. 2023. “A Roadmap for Integrating Automation With Process Optimization for AI-powered Digital Transformation”. *Preprints.org preprint preprints:202310.1055.v1*.

Alexopoulos, K., N. Nikolakis, and G. Chryssolouris. 2020. “Digital Twin-Driven Supervised Machine Learning for the Development of Artificial Intelligence Applications in Manufacturing”. *International Journal of Computer Integrated Manufacturing* 33(5):429–439.

Anandan, R., S. Gopalakrishnan, S. Pal, and N. Zaman. 2022. *Industrial Internet of Things (IIOT): Intelligent Analytics for Predictive Maintenance*. Hoboken, New Jersey: John Wiley & Sons.

Attaran, M., S. Attaran, and B. G. Celik. 2023. “The Impact of Digital Twins on the Evolution of Intelligent Manufacturing and Industry 4.0”. *Advances in Computational Intelligence* 3(3):11.

Bokrantz, J., A. Skoogh, C. Berlin, T. Wuest and J. Stahre. 2020. “Smart Maintenance: an Empirically Grounded Conceptualization”. *International Journal of Production Economics* 223:107534.

Bordegoni, M. and F. Ferrise. 2023. “Exploring the Intersection of Metaverse, Digital Twins, and Artificial Intelligence in Training and Maintenance”. *Journal of Computing and Information Science in Engineering* 23(6):060806.

Casas, N. 2017. “Deep Deterministic Policy Gradient for Urban Traffic Light Control”. *arXiv preprint arXiv:1703.09035*.

Chávez, C. A. G., M. Bärring, M. Frantzén, A. Annepavar, D. Gopalakrishnan and B. Johansson. 2022. “Achieving Sustainable Manufacturing By Embedding Sustainability KPIs in Digital Twins”. In *2022 Winter Simulation Conference (WSC)*, 1683–1694 https://doi.org/10.1109/WSC57314.2022.10015336.

Chen, S., J. P. G. Sánchez, E. T. Bekar, J. Bokrantz, A. Skoogh and P. V. Lopes. 2023. “Understanding Stakeholder Requirements for Digital Twins In Manufacturing Maintenance”. In *2023 Winter Simulation Conference (WSC)*, 2008–2019 https://doi.org/10.1109/WSC60868.2023.10408657.

Ciano, M. P., R. Pozzi, T. Rossi, and F. Strozzi. 2021. “Digital Twin-enabled Smart Industrial Systems: a Bibliometric Review”. *International Journal of Computer Integrated Manufacturing* 34(7-8):690–708.

Dai, W., Z. Mo, C. Luo, J. Jiang, H. Zhang and Q. Miao. 2020. “Fault Diagnosis of Rotating Machinery Based on Deep Reinforcement Learning and Reciprocal of Smoothness Index”. *IEEE Sensors Journal* 20(15):8307–8315.

Grondman, I., L. Busoniu, G. A. Lopes, and R. Babuska. 2012. “A Survey of Actor-critic Reinforcement Learning: Standard and Natural Policy Gradients”. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42(6):1291–1307.

Hafeez, T., L. Xu, and G. Mcardle. 2021. “Edge Intelligence for Data Handling and Predictive Maintenance in IIOT”. *IEEE Access* 9:49355–49371.

Huang, N.-C., P.-C. Hsieh, K.-H. Ho, and I.-C. Wu. 2024. “PPO-Clip Attains Global Optimality: Towards Deeper Understandings of Clipping”. In *Proceedings of the AAAI Conference on Artificial Intelligence*, edited by D. Jennifer and N. Sriraam, 12600–12607. Menlo Park: The Association for the Advancement of Artificial Intelligence Conference Press.

Huang, Y. 2020. “Deep Q-networks”. In *Deep Reinforcement Learning: Fundamentals, Research and Applications*, edited by H. Dong, Z. Ding, and S. Zhang, 135–160. Singapore: Springer.

Huang, Z., Y. Shen, J. Li, M. Fey and C. Brecher. 2021. “A Survey on AI-driven Digital Twins in Industry 4.0: Smart Manufacturing and Advanced Robotics”. *Sensors* 21(19):6340.

Kaelbling, L. P., M. L. Littman, and A. W. Moore. 1996. “Reinforcement Learning: A Survey”. *Journal of Artificial Intelligence Research* 4:237–285.

Kharchenko, V., O. Illiashenko, O. Morozova, and S. Sokolov. 2020. “Combination of Digital Twin and Artificial Intelligence in Manufacturing Using Industrial IoT”. In *2020 IEEE 11th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, edited by A. Derek, 196–201. New York: Institute of Electrical and Electronics Engineers.

Li, S., Y. Wu, X. Cui, H. Dong, F. Fang and S. Russell. 2019. “Robust Multi-Agent Reinforcement Learning via Minimax Deep Deterministic Policy Gradient”. *Proceedings of the The Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence* 33(01):4213–4220.

Lv, Z. and S. Xie. 2022. “Artificial Intelligence in the Digital Twins: State of the Art, Challenges, and Future Research Topics”. *Digital Twin* 1:12.

Mahnke, W., S.-H. Leitner, and M. Damm. 2009. *OPC Unified Architecture*. New York: Springer Science & Business Media.

Mo, F., H. U. Rehman, F. M. Monetti, J. C. Chaplin, D. Sanderson, A. Popov *et al*. 2023. “A Framework for Manufacturing System Reconfiguration and Optimisation Utilising Digital Twins and Modular Artificial Intelligence”. *Robotics and Computer-integrated Manufacturing* 82:102524.

Panzer, M. and B. Bender. 2022. “Deep Reinforcement Learning in Production Systems: a Systematic Literature Review”. *International Journal of Production Research* 60(13):4316–4341.

Pirotta, M., M. Restelli, and L. Bascetta. 2013. “Adaptive Step-size for Policy Gradient Methods”. In *Advances in Neural Information Processing Systems*, edited by C. Burges, 26. California: Neural Information Processing Systems Foundation, Inc.

Rathore, M. M., S. A. Shah, D. Shukla, E. Bentafat and S. Bakiras. 2021. “The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities”. *IEEE Access* 9:32030–32052.

Schleich, B., N. Anwer, L. Mathieu, and S. Wartzack. 2017. “Shaping the Digital Twin for Design and Production Engineering”. *CIRP Annals* 66(1):141–144.

Schulman, J., S. Levine, P. Abbeel, M. Jordan and P. Moritz. 2015. “Trust Region Policy Optimization”. *arXiv preprint arXiv:1502.05477*.

Schulman, J., F. Wolski, P. Dhariwal, A. Radford and O. Klimov. 2017. “Proximal Policy Optimization Algorithms”. *arXiv preprint arXiv:1707.06347*.

Sdiri, B., L. Rigaud, R. Jemmali, and F. Abdelhedi. 2023, Jul. “The Difficult Path to Become Data-Driven”. *SN Computer Science* 4(4):385.

Siraskar, R., S. Kumar, S. Patil, A. Bongale and K. Kotecha. 2023. “Reinforcement Learning for Predictive Maintenance: A Systematic Technical Review”. *Artificial Intelligence Review* 56(11):12885–12947.

Ståhl, J.-E. and C. Windmark. 2022. *Hållbara Produktionssystem*. Lund: Studentlitteratur AB. Artikelnummer: 39931-01.

Su, J., J. Huang, S. Adams, Q. Chang and P. A. Beling. 2022. “Deep Multi-agent Reinforcement Learning for Multi-level Preventive Maintenance in Manufacturing Systems”. *Expert Systems with Applications* 192:116323.

Taşcı, B., A. Omar, and S. Ayvaz. 2023. “Remaining Useful Lifetime Prediction for Predictive Maintenance in Manufacturing”. *Computers and Industrial Engineering* 184:109566.

Turner, C. J., C. Emmanouilidis, T. Tomiyama, A. Tiwari and R. Roy. 2019. “Intelligent Decision Support for Maintenance: an Overview and Future Trends”. *International Journal of Computer Integrated Manufacutring* 32(10):936–959.

Wang, Y., H. He, X. Tan, and Y. Gan. 2019. “Trust Region-Guided Proximal Policy Optimization”. *arXiv preprint arXiv:1901.10314*.

AUTHOR BIOGRAPHIES

SIYUAN CHEN is a PhD student in Production Systems division, Industrial and Materials Science department of Chalmers University of Technology. His research interest includes digital twins, deep reinforcement learning, generative AI and smart maintenance. He is building an AI-enhanced data-driven digital twin model to support decision-making of smart maintenance. His email address is siyuan.chen@chalmers.se.

PAULO VICTOR LOPES is a PhD student in the Operations Research Program at Aeronautical Institute of Technology and Federal University of Sao Paulo. His research interests include data driven modelling of Digital Twins, what-if experiments design and data-driven techniques to improve production lines performance. He currently is in a guest period at Industrial and Materials Science Department of Chalmers University of Technology. His email address is paulo.lopes@ga.ita.br.

SILVAN MARTI is a PhD student in the Production Systems division of the Industrial and Materials Science Department at Chalmers University of Technology in Gothenburg, Sweden. His primary research interest is in the area of deep learning based multivariate time-series analytics in operations research for an industrial setting. His email address is silvan@chalmers.se.

MOHAN RAJASHEKARAPPA is a PhD student in the Production Systems division at Chalmers University of Technology. His research focuses on applying Industrial AI to improve smart maintenance systems. By integrating AI techniques, he aims to enhance the reliability and efficiency of industrial operations. His email address is rmohan@chalmers.se.

SUNITH BANDARU is an Associate Professor of Production Engineering, University of Skövde. His research interests are knowledge discovery, data-mining and machine learning, multi-objective optimization, evolutionary algorithms, simulation-based optimization and meta-modeling. His email address is sunith.bandaru@his.se.

CHRISTINA WINDMARK is an Associate senior lecturer in Production and Materials Engineering, Lund University. Her research mainly revolves around monetary analyzes and models that support resource efficiency and informative industrial decisions regarding manufacturing development and sustainable development. Her email address is christina.windmark@iprod.lth.se.

JON BOKRANTZ is a researcher at the Department of Industrial and Materials Science at Chalmers University of Technology. His research focuses on production and operations management with a special emphasis on industrial maintenance. His research interests include the interplay of technology, people, and organization. His email address is jon.bokrantz@chalmers.se.

ANDERS SKOOGH is a Professor at Industrial and Materials Science, Chalmers University of Technology. He is a research group leader for Production Service & Maintenance Systems. Anders is also the director of Chalmers' Masters' program in Production Engineering and board member of the think-tank Sustainability Circle. His email address is anders.skoogh@chalmers.se.