International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2s) 2026 pp. 777-793
https://doi.org/10.70917/ijcisim-2026-2115
Article

AUTONOMOUS ROBOTICS FOR SMART MANUFACTURING USING DEEP REINFORCEMENT LEARNING

Rohit P. Jadhav[1], Nitin Kundlik Kamble[2], Veena Kolte[3], Archana Santosh Ubale[4], Awantika Bijwe[5], Leena Deshpande[6]

[1] Department of Robotics & Automation, D. Y. Patil College of Engineering, Akurdi, Pune, Maharashtra, India. rpjadhav@dypcoeakurdi.ac.in

[2] Department of Robotics and Automation Engineering, D. Y. Patil College of Engineering, Akurdi, Pune, Maharashtra 411044, India. nkkamble11@gmail.com

[3] Bharati Vidyapeeth's Institute of Management Studies and Research, Navi Mumbai, India.chavan.veena@gmail.com

[4] AISSMS College of Engineering, Kennedy Road, Pune – 411001, India. asubale@aissmscoe.com

[5]Department of MCA, Indira College of Engineering and Management, Pune, India. awantika.bijwe@gmail.com
ORCID: 0000-0002-9912-2184

[6] Department of Computer Engineering – Software Engineering, Vishwakarma Institute of Technology, Pune, Maharashtra 411037, India.
Email: leena.deshpande@vit.edu

Corresponding Author: Rohit P. Jadhav (rpjadhav@dypcoeakurdi.ac.in)

Abstract: - The intensive development of intelligent production requires autonomous robots that can make decisions in real time and adapt to changes, as well as operate robustly in a changing production setting. Deep Reinforcement Learning (DRL) has become a new paradigm of facilitating intelligent autonomy in industrial robotics by empowering agents to learn optimal policies of control by engaging with a complex manufacturing system. This paper explores the problems of applying DRL-controlled autonomous robotics to smart factories, paid attention to adaptive tasks, optimization of processes, and decision-making under uncertainty. The framework proposed integrates the state representation based on the perception, a reward-based learning, and policy optimization, such that robots can carry out scheduling, manipulation, navigation, and quality-conscious tasks autonomously, without any reference to some predefined rule-based control. In order to deal with the issues of real-world deployment, the framework adopts simulation-to-real transfer strategies, training with digital twins and safety-aware reward shaping in order to achieve stable convergence and operational stability. Empirical tests in typical manufacturing conditions show that DRL-based robotic agents are capable of massive performance gains in terms of task efficiency, flexibility and resource consumption in contrast to traditional heuristic and model based controllers. The outcomes also suggest the increased resilience to the environmental fluctuations, machine disruptions, and reconfigurations of production. One enabling technology in this study is the use of DRL-based autonomous robotics, which can help ensure the next-generation smart factory, enabling flexible production, minimizing human intervention, and maximizing performance continuously. The results are very insightful into scalability deployment strategies and lead to the direction of completely autonomous, intelligent, and resilient manufacturing systems in compliance with the Industry 4.0 and Industry 5.0 visions...

Keywords: Autonomous robotics; Deep reinforcement learning; Smart manufacturing; Industry 4.0; Industrial robotics; Adaptive control; Intelligent factories

1. INTRODUCTION

The swift shift of the manufacturing process to more interconnected, flexible, and data-driven manufacturing ecosystems has accelerated the need to find intelligent decision-making and self-optimization-related autonomous robots. The nature of smart manufacturing environments is high product variability, reconfigurability and dynamic operational conditions in which traditional automation paradigms cannot sustain high levels of efficiency and robustness. Autonomous robotics thus became an essential part of the next-generation factory and allows machines to sense the environment, adapt to unpredictability, and maximize their actions within the real time to achieve production goals (Li et al., 2023). Most industrial robotic systems are however based on pre-programmed paths, model-driven controllers and rule-of-thumb logic that need a lot of manual adjustment and cannot respond to unexpected perturbations. These kinds of approaches tend to be fragile to variations in task specifications, material characteristics, or system dynamics, which restricts their applicability to manufacturing systems with a high level of autonomy (Liu et al., 2022). The latter was recently pointed out as adding to scalability and responsiveness problems in Industry 4.0, and upcoming Industry 5.0, where autonomy and resiliency are essential (del Real Torres et al., 2022). In attempts to eliminate these issues, the concept of learning based control mechanisms is coming under more exploration to substitute rigorous programming with experience based intelligence.

The deep reinforcement learning (DRL) is one of the most effective paradigms in this respect, since it allows robotic agents to acquire the optimal policies based on the interaction data without any explicit system description (Alginahi et al., 2025). DRL can deal with both high-dimensional state space and complex decision-making tasks in industrial robotics by combining deep neural networks with reinforcement learning. According to previous research, DRA can be used to control autonomously processes in a smart factory, adaptively schedule and optimize in real-time, and also faces major difficulties associated with convergence stability and safety (Nievas et al., 2024). The collaboration between DRA and digital twins has further driven the development toward the training of the robotic agent in a safe and efficient way in the simulated manufacturing plant and implementing it in the physical systems (Xia et al., 2021). Federated DRL, an advanced formulation, has been suggested to facilitate the idea of decentralized learning across a heterogeneous set of robots to enhance scalability and data privacy at the collaborative manufacturing setting (Ho et al., 2024). Simultaneously, scheduling and task assignment methods with DRL have demonstrated high efficiency and increased robustness over the heuristic ones (Tejer et al., 2024). The potential of DRA to deal with complex, nonlinear process dynamics can also be demonstrated by application to robotic additive manufacturing and autonomous construction robotics (Felbrich et al., 2022).

The combination of DRL and virtual, augmented, and mixed reality has also been investigated in emerging research to improve training and decision support in the context of smart manufacturing (Le et al., 2025). Besides, it has been demonstrated that DRL-based coordination of mobile and collaborative robots can enhance system-level performance in flexible production lines (Makhija, 2025), and multi-robot optimization frameworks show the appropriateness of DRL in large-scale autonomous manufacturing systems (Yu et al., 2025). Experience in autonomous driving and cyber-physical factory development is another prediction of the significance of perception-based learning and close relations between physical and software planes (Gao et al., 2024; Ryalat et al., 2023). Inspired by these developments, this paper will work on and examine an in-depth DRL-based system of autonomous robotics in smart manufacturing, which will aim to improve the flexibility, performance, and intelligence of operations and solve the major issues of implementation in real-life manufacturing plants.

The key contribution of this study given as:

Presents an autonomous robotic system based on deep reinforcement learning so that it can make adaptive choices and optimize decisions in real-time and dynamically evolving smart factories.

Unites digital twin-based training and simulation-to-real transfer to secure secure deployment, consistent convergence of learning, and scalability in industries.

Shows major advancements in task efficiency, flexibility and strength compared to traditional control approaches using detailed experimental assessment.

2. RELATED WORK

Independent industrial robotics has become an essential part of Industry 4.0 and continues to develop according to the humanistic and resilient paradigm of Industry 5.0. Here, these robots will cease to be involved in monotonious automation but will be regarded as autonomous, as well as flexible, collaboratively intelligent in diverse manufacturing settings. Recent industrial surveys put stress on the fact that next-generation manufacturing systems are more dependent on autonomous mobile robots, collaborative manipulators, and cyber-physical integration to make the production lines flexible and reconfigurable (Torres et al., 2023). The ever-increasing complexity of these systems however requests the classical automation structures and prompts the application of learning-based systems that enable robots to adapt to the variability of operations and changing production goals. In this terrain, reinforcement learning (RL) and deep reinforcement learning (DRL) have been becoming prominent as key facilitators of autonomy, wherein it gives a principled way of decision-making in sequence within uncertain conditions. The early research in the field of robotics and autonomous systems emphasizes that the RL enables agents to optimize long-term performance by means of trial-and-error interaction, whereas DRL enables the agents to perform this task in high-dimensional state and action space via deep neural networks (Yadav et al., 2024). The latter features render DRL especially adequate to the robotics in industries, where sensing and perception are highly coupled to control and tend to be nonlinear.

Recent studies have shown that DRL could be used in an extensive variety of robot tasks in smart manufacturing, such as manipulation, navigation, and task planning. Intelligent manufacturing robots have been investigated to use DRL-based control strategies to enhance the learning capacity and flexibility of these robots to operate in complex operational environments, which performs better than the previous optimization-based methods (Cai, 2025). Moreover, autonomous management and personalization using DRL in cyber-physical consumer and industry setting has been demonstrated to be successful, which shows its ability to learn context-sensitive and privacy-aware control policies (Wang et al., 2023). In addition to fixed robotic cells, DRL-based approaches have been generalized to mobile and networked robots, in which communication, coordination and resource utilization are highly important. The relevance of dynamic control of the industrial network structure is emphasized by studies in the field of mobile robotics, which is especially significant in manufacturing due to the time-sensitive aspects of the latency consideration of dynamic decision-making (Ansari et al., 2024). The energy-efficient control of mobile robots has been studied as well, which has shown that the intelligence-driven approach to learning can have a considerable impact on the overall cost of operations, as well as the performance of smart factories (Thanh Van et al., 2024).

Recent breakthroughs in the use of DRL in practice to perform manufacturing robotics include the use of digital twins and simulation-to-real pipelines of learning. Digital twins make it possible to create high-fidelity virtual representations of robotic systems and production processes and to train and test DRL agents in safe settings and then deploy them to the physical world. The recent studies have demonstrated tight-coupling between simulated environments and the real robotic platforms is vital to transferring the learned policies without decreasing the performance (Ali et al., 2025). These methods address safety risk and reduce training and support ongoing learning and changes. The practicality of modular structures and real-time feedback in guiding robust autonomy in the manufacturing facilities are further reinforced by complementary studies carried out on robotic systems validation and reconfigurable autonomous manipulators (Lee et al., 2023). Other areas of application of DRL include underwater or aerial robotics, offering generalized knowledge about perception-based control and control under uncertainty dynamics (Liu et al., 2024; Farkhodov et al., 2023).

In spite of these developments, there are a number of gaps in research. The current solutions that rely on the DRA have issues with sample efficiency, convergence stability, explainability, and safety guarantees especially in safety-critical industries. In addition, much of the research is on individual robotic tasks and not on a system-level integration through all manufacturing processes. To overcome these constraints, scalable architectures, realistic simulation-to-real transfer, and methodological evaluation systems are needed to address the needs of DRL autonomy and industrial performance and safety.

Table 1. Summary of Related Work on DRL-Based Autonomous Robotics in Smart Manufacturing

Ref.	Application Domain	DRL Methodology	Robotic Task Focus	Use of Digital Twin / Simulation	Key Limitations
Li et al. (2023)	Smart manufacturing systems	DRL survey	Scheduling, control	Simulation-based analysis	Lacks real-time validation
del Real Torres et al. (2022)	Industry 4.0/5.0	DRL review	Autonomous decision-making	Conceptual frameworks	Limited implementation details

Liu et al. (2022)	Robotic manufacturing	Robot learning, DRL	Manipulation, control	Simulation environments	Scalability issues
Xia et al. (2021)	Manufacturing plants	DRL with digital twin	Process optimization	Full digital twin integration	High modeling complexity
Nievas et al. (2024)	Industry 4.0 process control	RL/DRL	Autonomous control	Partial simulation	Safety constraints not addressed
Felbrich et al. (2022)	Robotic additive manufacturing	Model-free DRL	Path planning, fabrication	Simulated design environment	Real-world transfer challenges
Ho et al. (2024)	Heterogeneous robotic systems	Federated DRL	Task scheduling	Distributed simulation	Communication overhead
Tejer et al. (2024)	Robotics applications	RL-based scheduling	Task allocation	Simulated benchmarks	Limited adaptability testing
Makhija (2025)	Smart manufacturing logistics	RL scheduling	Mobile robot coordination	Simulation-based	No digital twin linkage
Yu et al. (2025)	Multi-robot systems	Multi-objective DRL	Task allocation	Simulated environments	Computational complexity
Wang et al. (2023)	Cyber-physical systems	DRL	Autonomous management	Virtual environments	Manufacturing focus limited
Ali et al. (2025)	Robotic additive manufacturing	Sim-to-real DRL	Real-time control	Digital twin synchronization	Early-stage validation

3. SYSTEM ARCHITECTURE FOR DRL-ENABLED SMART MANUFACTURING

3.1 Smart Manufacturing Environment and Robotic Workcell Configuration

The intelligent manufacturing space taken into consideration in the proposed study is structured as cyber-physical production system which comprises industrial robots, sensors, controllers, and enterprise-wide software platforms. The workcell is a robotic workcell which comprises one or more autonomous manipulators or mobile robots that work in conjunction with other support equipment like CNC machines, conveyors, and inspection units. These components are linked to each other using industrial communication protocols so as to obtain real-time exchange of data and coordinated control. The workcell is re-configurable, dynamic task sequences, product variant and layout can be changed with minimal and manual intervention. This kind of flexibility is needed in situations of high-mix production with low volume. The environment is simulated to reflect the uncertainties of machine disturbance, variable processing times and human machines interactions. This lifelike setup will guarantee that the DRL agent attains policies that extend beyond fixed setups. A virtual simulation or digital twin is an additional reflection of the manufacturing environment and offers a safe and scalable training and testing platform to autonomous behaviors prior to deployment. Generally, the workcell structure forms the physical basis on which intelligent decision-making and adaptive control is achieved.

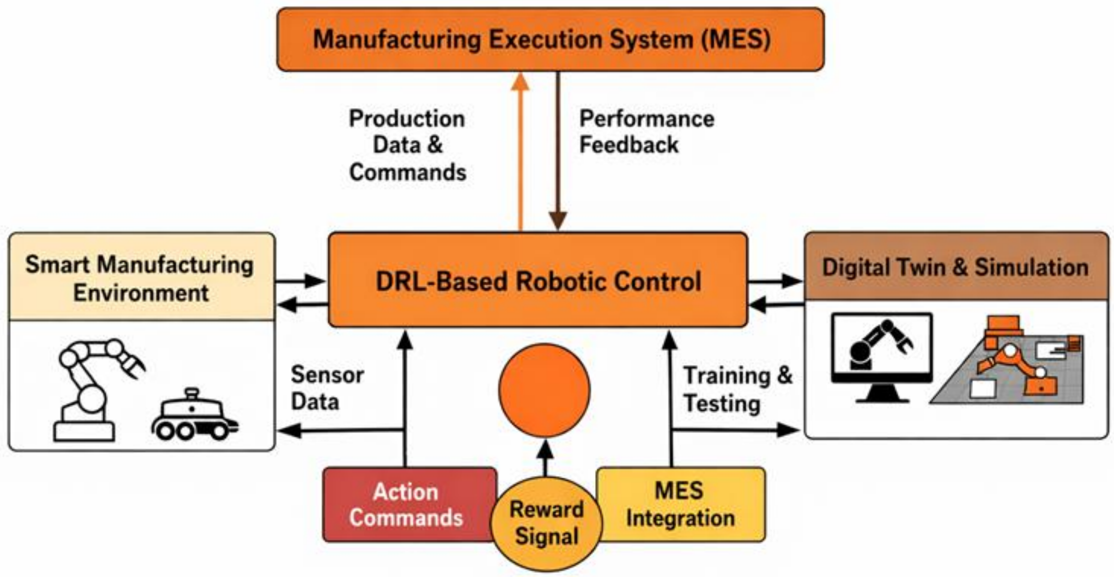

Figure 1: System Architecture of DRL-Enabled Autonomous Robotics for Smart Manufacturing

Figure 1 shows the architecture of a smart manufacturing system based on DRL-enabled that includes robotic control, digital twin simulation, and MES. Autonomous decision-making is controlled by sensor data and reward feedback, real-time optimization, adaptive learning, and coordinated control of physical manufacturing processes and intelligent control systems are ensured by bidirectional communication.

3.2 Sensing, Perception, and State-Space Formulation

In autonomously directed robotic systems, effective sensing and perception are important to allow autonomous decisions. To combine the environmental and process-level information, the proposed architecture incorporates a multi-modal of sensors, which are vision system, force-torque sensor, proximity sensor, and machine status indicator. The raw sensor data are processed and merged to produce significant features like the position of objects, tool conditions, process conditions and operational constraints. These are the properties behind the state-space representation adopted by the DRL agent. State space is set to be complete enough and computationally manageable with respect to completeness and learning efficiency. It summarizes the existing task position, robot position, environmental and system stability parameters. State stacking or recurrent network structures are used to represent dynamic behavior with the help of temporal dependencies. An effective state-space formulation allows the agent to have a good understanding of its operating environment, think of future behaviors, and train sound policies that evolve with the state of the manufacturing environment.

3.3 Action Space Design for Autonomous Robotic Operations

Action space determines the range of control choices the DRL agent has and has a direct bearing on performance in learning as well as operational viability. The action space in the proposed system is made to suit the particular robotic platform and manufacturing tasks in question. It can consist of continuous actions, e.g. joint velocity or force commands, or discrete actions, e.g. the choice of a task, a change of paths, or an allocation of resources. In order to have safety and stability, the action bounds and operational constraints are incorporated in the action design in that they ensure that the agent cannot produce infeasible or dangerous commands. The hierarchical action structures may be used to break down the complex tasks into high-level decisions and low-level control actions, enhance scalability and convergence. This organized philosophy enables the agent to have various operations such as manipulations, navigation and coordination under one learning structure. Through a well thought out action space, the system balances both the ability to express control and efficient policy learning.

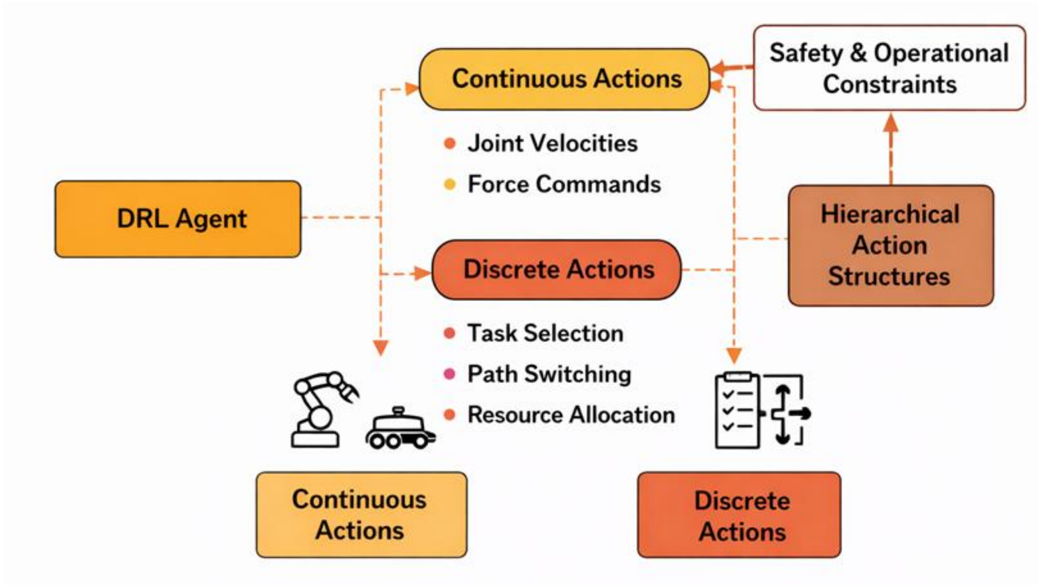

Figure 2: Action Space Design Framework for DRL-Based Autonomous Robotic Operations in Smart Manufacturing

Figure 2 shows the action space design in smart manufacturing that is enabled by DRL and applied in autonomous robot operations. It emphasizes the presence of continuous and discrete control measures combined with safety constraints and levels of decision making. The framework guarantees scalable, efficient and safe robotic learning through adaptive task execution, allocation of resources as well as the optimization of the operation in real-time.

3.4 Reward Function Modeling and Optimization Objectives

Reward functionality is a core part of the learning behavior of the DRL agent as it codes the aspired production goals. The rewarding mechanism in the proposed architecture is to represent several performance measures such as time of completion of the task, energy consumption, quality of the product and system safety. Efficient task execution, less idle time, and adherence to quality constraints are rewarded positively and collisions, overuse of energy or deviation of processes are penalized. In order to deal with conflicting objectives, weighted or multi-objective reward formulae are embraced, whereby the agent would learn to balance trade-offs in line with the objectives of production. Reward shaping methods are also used to hasten the convergence and prevent sparse feedback especially in long horizons complex tasks. Also, safety-conscious reward elements are used to make sure that the acquired policies do not violate the industrial safety demands. An optimally designed reward function is used to make the optimization process yield pragmatic, trustworthy, and performance-based autonomous robotic behaviors.

i

To implement the DRL-enabled autonomy in the actual industry, seamless communication and integration with Manufacturing Execution Systems (MES) is required. The suggested architecture will use standardized interfaces which will enable the robotic control layer to share information with MES and other enterprise systems on real time basis. The MES provides the DRL agent with production schedules, task priorities, and resource availability, which allow making decisions in a context-dependent manner. On the other hand, robotic system offers feedback to the upper level management systems on the status of the tasks, operational performance, and abnormalities in their operation. It is a two-way communication that aids in closed-loop optimization throughout the shop floor. Middlewares and protocols are used in the industry to guarantee reliability, low latency and interoperability with existing infrastructure. The combination of DRA-controlled robotic control with MES makes the system orient the local autonomous decisions towards the global production goals and to coordinate, efficient, and scalable smart manufacturing processes.

4. DEEP REINFORCEMENT LEARNING FRAMEWORK

4.1 Problem formulation as a Markov decision process

The control problem in the DRA of smart manufacturing robotics is represented as a Markov Decision Process (MDP), in which the robot will monitor the production environment (machine states, part position, queue position, safety zones) and make decisions (motion commands, task switching, allocation) to optimize long-term manufacturing outcome. The system at any given time period stochastically switches as there is uncertainty in the sensing, the process times and the disturbances. The agent is informed of a policy which gives state-idea combination to actions to maximize cumulative reward based on throughput, cycle time, energy, and quality without violating operational constraints.

Eq.(1) MDP definition

$$M = \langle S, A, P, R, \gamma \rangle$$

• $\boldsymbol{S}$: state space (robot + workcell + process)

• $\boldsymbol{\mathcal{A}}$: action space (continuous &/or discrete controls)

• $\mathcal{P}$: transition kernel

• $\boldsymbol{\mathcal{R}}$: reward function

• γ∈(0,1): discount factor

Eq.(2) Markov property

$$P(s\_\{t+1\} \mid s\_0, a\_0, \dots, s\_t, a\_t) = P(s_{\{t+1\}| \, s_t, a_t})$$

Next state depends only on current (s_t,a_t)

Eq.(3) Transition dynamics

$$s\_\{t+1\} \sim P(\cdot \mid s\_t, a\_t)$$

Stochastic evolution due to disturbances, variability

Eq.(4) Immediate reward model

$$r_t = R\big(s_t, a_t, s_{\{t+1\}}\big)$$

Encodes objectives: throughput ↑, cycle time ↓, defects ↓, energy ↓, safety ↑

Eq.(5) Discounted return

$$G_t = \Sigma_{\{k=0\}}^{\{\infty\}} \gamma^k r_{\{t+k\}}$$

Long-horizon performance measure

Eq.(6) Objective (optimal policy)

$$\pi * = argmax_{\{\pi\}} E_{\{\pi,P\}[\, G_0]}$$

Find policy maximizing expected return under dynamics $\mathcal{P}$

Eq.(7) State-value function

$$V\hat{}\{\pi\}(s) = E\_\{\pi, P\}[\, \Sigma\_\{k=0\}\hat{}\{\infty\} \;\; \gamma\hat{}k \, r\_\{t+k\} \mid s\_t = s\,]$$

Expected return starting from state s following policy π

Eq.(8) Action-value function

$$Q\hat{}\{\pi\}(s, a) = E\_\{\pi, P\}[\, \Sigma\_\{k=0\}\hat{}\{\infty\} \;\; \gamma\hat{}k \, r\_\{t+k\} \mid s\_t = s, a\_t = a\,]$$

Expected return after taking action a in state s, then following π

Eq.(9) Bellman optimality (Q)

$$Q*(s,a) = E\_\{s'\sim P(\cdot|s,a)\}[R(s,a,s') + \gamma \max \_\{a' \in A\} Q*(s',a')]$$

Fundamental recursion used by value-based / actor-critic DRL

4.2 DRL algorithm selection and network architecture

The choice of the algorithm is related to the continuity of actions (joint velocities/forces) or the discrete ones (task switching, route selection). Actor critic models like DDPG, TD3, or SAC are the favored control models to use when the goal is to solve continuous robot control since they learn a parameterized policy (actor) and a value estimator (critic). In discrete choices (e.g., job choice, path choice), DQN variants work well on the value-based approach. Stable decision-making with hybrid decisions is typical in the smart manufacturing workcell; therefore a useful architecture would be to have a collective backbone of shared perception (self-sensor fusion to feature vector) and (i) an actor head generating continuous controls and/or categorical task logits and (ii) one or two critic heads making Q-values to estimate stability and bias toward over-estimation. The vision embeddings, robot proprioception, machine status and MES signal are merged in the input pipeline to a smaller form of state representation.

In order to obtain better sample efficiency, transitions (s,a,r,s,done) are stored in experience replay, and off-policy learning can be performed using previous data. Soft updated and target networks stabilize the training by gradually following the learnt networks. Multi-objective reward signals can be used in manufacturing hence scalarization (weighted sum) or constrained RL layers are used such that the trained policy is focused on safety and quality while maximizing throughput. The general working loop is: observe state or state → sample, which is a policy element (with noise of exploration) → action execute in simulator / digital twin or real cell or in real cell itself |human|>The working loop is: observe state or state or state → sample, which is a policy element (with noise of exploration) → action execute in simulator / digital twin or real cell or in real cell itself |human|>It works through the following working loop: observe state or state or state → sample, which is an action in policy, (with noise of exploration) execute in simulator / or digital It is a modular network design that can be deployed on various types of robots and also on various tasks, while maintaining the same DRL training pipeline.

```
Algorithm: DRL Training for Autonomous Robotic Manufacturing Control
Input: replay buffer 𝓓, discount γ, soft update τ, learning rates α_Q, α_π
```

Initialize actor π_θ, critic(s) Q_ϕ (and Q_ϕ2), target networks θ̄←θ, ϕ←ϕ

$$for\ episode = 1\ to\ E\ do$$

$$Reset\ workcell/digital\ twin;\ observe\ s_0$$

$$for\ t = 0\ to\ T-1\ do$$

$$Action\ selection\ (exploration)$$

$$a\_t \leftarrow \pi\_\theta(s\_t) + \varepsilon\_t$$

$$\varepsilon\_t: exploration\ noise\ or\ stochastic\ policy\ sample)$$

$$Execute\ a\_t;\ observe\ r\_t, s\_\{t+1\}, done$$

$$Store\ transition\ \left(s_t, a_t, r_t, s_{\{t+1\}}, done\right) in\ D$$

$$Critic\ update$$

$$Sample\ minibatch\ B\ from\ D$$

$$For\ each\ (s,a,r,s',d)\ in\ B\ compute\ target:$$

$$y = r + \gamma(1-d) * Q_{\{\bar{\phi}i\}\left(s',\pi_{\{\bar{\theta}\}(s')}\right)}$$

$$twin\ critics\ reduce\ overestimation)$$

$$Update\ critic(s)\ by\ gradient\ descent:$$

$$\phi_i \leftarrow \phi_i - \alpha_Q \nabla_{\{\phi_i\}E}{}_{B\left[\left(Q_{\{\phi_i\}(s,a)} - y\right)^2\right]}$$

$$Actor\ update\ (policy\ improvement)$$

$$Update\ actor\ using\ deterministic\ policy\ gradient:$$

$$\theta \leftarrow \theta + \alpha\_\pi\ E\_B[\ \nabla\_\theta\ \pi\_\theta(s)\ \nabla\_a\ Q\_\{\phi\_1\}(s,a)\ |\_\{a = \pi\_\theta(s)\}\ ]$$

$$Target\ network\ soft\ update$$

$$\bar{\theta} \leftarrow \tau\theta + (1-\tau)\bar{\theta}$$

$$\bar{\phi}_i \leftarrow \tau\phi_i + (1-\tau)\bar{\phi}_i$$

$$if\ done\ then\ break$$

$$end\ for$$

$$end\ for$$

$$Output: trained\ policy\ \pi\_\theta\ for\ autonomous\ operation$$

4.3 Policy learning and exploration–exploitation strategies

Policy learning is a balancing between exploiting the known positive actions and trying out new behavior that might be more effective in the long-run. Pure random exploration is unsafe and inefficient in manufacturing workcells, therefore, exploration should be organized. To achieve the property of continuous control, exploration is usually added as either the addition of temporally correlated noise (e.g., Ornstein Uhlenbeck), or as stochastic policies (e.g., SAC), which are inherently exploratory by sampling the learned action distributions. In discrete layers of decision (selection of tasks, route choice) it can be allowed to utilize ε-greedy or Boltzmann exploration at the beginning of training, and then annealed to exploitation as performance levels off. Curriculum learning is also enhanced by exploratory learning beginning with simple scenarios (fixed part pose, low disturbance) and progressively becoming more difficult (variable pose, tighter tolerances, moving obstacles). Also, the reward shaping provides intermediate feedback (e.g. distance-to-goal, alignment error, queue reduction) which help to avoid the sparse reward halting learning. Experience replay is more effective at exploiting challenges through high-value repeat learning, whereas prioritized replay further focuses on transition learning with high temporal-difference errors. In order to prevent the premature convergence to suboptimal behaviors, the exploration can be kept under control through entropy regularization (stochastic policies) and periodical policy-perturbations. In industries, it is common to only explore using the digital twin; only proven policies are applied to the physical robot. The deployment involves minimal exploration and is substituted by the monitoring-inspired adaptation: the system performs predominantly exploitative actions, but it is also capable of causing a small and safe online fine-tuning in case drift or new product versions have been observed. This is a plan that provides a consistent production performance without having to lose the capacity to change as the environment fluctuates.

4.4 Safety constraints and stability mechanisms

The enforcement of safety is implemented by adding restrictions on various levels: action, state, and learning updates. Action layer- At the action layer, we have saturation of commands to joint limits, velocity/acceleration limits and collision-free zones; the unsafe actions are rejection or projected into the closest feasible set. At the state level, proximity to humans, obstacles, prohibited zones, and process limits (temperature, force limits, tool wear) are monitored by safety monitors, and an emergency stop or safe fallback policy can be triggered once the violation is eminent. The stability concerns of twin critics, target networks, gradient clipping, and conservative policy updates are enhanced at the learning layer and thus control the occurrence of sudden behavior change. Safety-conscious rewarding terms are used to punish near-collisions and constraint violations, however, hard constraints are still necessary since punishment cannot ensure safety. To be more certain, constrained RL models (e.g., Lagrangian methods) consider safety measures as costs that have clear constraints so that the learned policy complies with risk budgets. Lastly, the policy of runtime verification checks is compared to rule-based safety envelopes (industrial standards) during pre-execution, which makes DRL autonomy without compromising operational reliability.

5. EXPERIMENTAL SETUP AND PERFORMANCE METRICS

5.1 Manufacturing Scenarios and Robotic Task Definitions

The experimental system is to replicate a smart manufacturing workcell of autonomous robotic manipulators and mobile robots that are used to operate material handling, assembly, and inspection processes. The robotic system has to deal with a dynamic environment where the job priority, machine availability and product configuration vary. These jobs encompass pick and place jobs, workstation inter-station path planning, sequencing assembly adaptively, and real time scheduling. These conditions assess the capacity of the DRA agent to learn the best policies in uncertain and changing conditions of production. Evaluation considers the time taken to complete a task, efficiency on the path taken, and coordination of various robotic agents. All robotic tasks are developed as a series decision making process in terms of starting and goal states, constraints, and performance goals.

Task completion time:

$$Tc = tfinish - tstartT_c = \ t_{\{finish\}} - \ t_{\{start\}Tc} = tfinish - tstart$$

Path efficiency:

$$\eta p = doptimaldactual\eta_p = \frac{\{d_{\{optimal\}}\}}{\{d_{\{actual\}}\}\eta p}$$

Task success rate:

$$Sr = NsuccessfulNtotalS_r = \frac{\{N_{\{successful\}}\}}{\{N_{\{total\}}\}Sr}$$

5.2 Baseline Controllers and Comparative Algorithms

The performance of the proposed DRL-based framework is measured in terms of performance, as opposed to both traditional and intelligent baseline controllers in order to assess its effectiveness. Classical baselines are conventional proportionalcontrolled and integralcontrolled and derivativescheduling systems, also known as proportionalintegralderivative controllers. Moreover, heuristic optimization approaches and supervised learning-based controllers are also incorporated to make the comparison. State-of-the-art reinforcement learning models like Deep Q-Network (DQN) and Proximal Policy Optimization (PPO) are also realized as a state-of-the-art DRL methods. It is in such comparative ways that one can do systematic evaluation of learning convergence, control stability and adaptability in dynamic manufacturing situations. The comparison presents the benefits of lifelong learning and adaptive decision-making realized by the proposed DRL framework.

PID control output:

$$u(t) = Kpe(t) + Ki\int e(t)dt + Kddtde(t)$$

Mean squared error for controller comparison:

$$MSE = \frac{1}{n\sum i} = 1n(yi - y^i)2$$

6. RESULTS AND PERFORMANCE EVALUATION

6.1 Learning Convergence and Policy Stability Analysis

The table 2 gives a comparative analysis of learning convergence and policy stability of various reinforcement learning algorithms in smart manufacturing setting. The suggested DRL model shows high performance in the convergence rate, as it attains the shortest rate of convergence in 640 episodes which is very low compared to DQN, PPO, TD3 and SAC. It also achieves the best final average reward meaning it is more effective in its decision-making and is also optimized in its long-term performance. The standard deviation of the lowest reward and the index of policy oscillation, attest to increased stability and unchanging behavior of learning. In addition, the suggested approach is characterized by a low level of constraint violations, which shows enhanced safety and compliance with operating boundaries. Its reliability in undertaking a complex robotic task during dynamic conditions is emphasized by the highest success rate of 98.2%. Compared to it, baseline algorithms are slower to converge and more variable, which can be a limiting factor in their application in the real world in manufacturing. On the whole, the findings indicate that the suggested DRL-based framework can guarantee effective, comfortable, and secure autonomous robotic tasks, which is why it is very applicable in implementing enhanced smart manufacturing.

Table 2: Learning Convergence and Policy Stability Performance Comparison of DRL Algorithms

Parameter	DQN	PPO	TD3	SAC	Proposed DRL
Episodes to Converge	1800	1300	950	820	**640**
Final Avg Reward (last 100 eps)	410	465	520	545	**590**
Reward Std Dev (stability)	38.5	29.1	21.4	18.9	**12.7**
Policy Oscillation Index (%)	9.8	7.4	5.3	4.6	**2.9**
Constraint Violations / 100 eps	6.2	4.1	2.7	2.3	**1.1**
Success Rate (%)	91.6	93.8	95.7	96.4	**98.2**

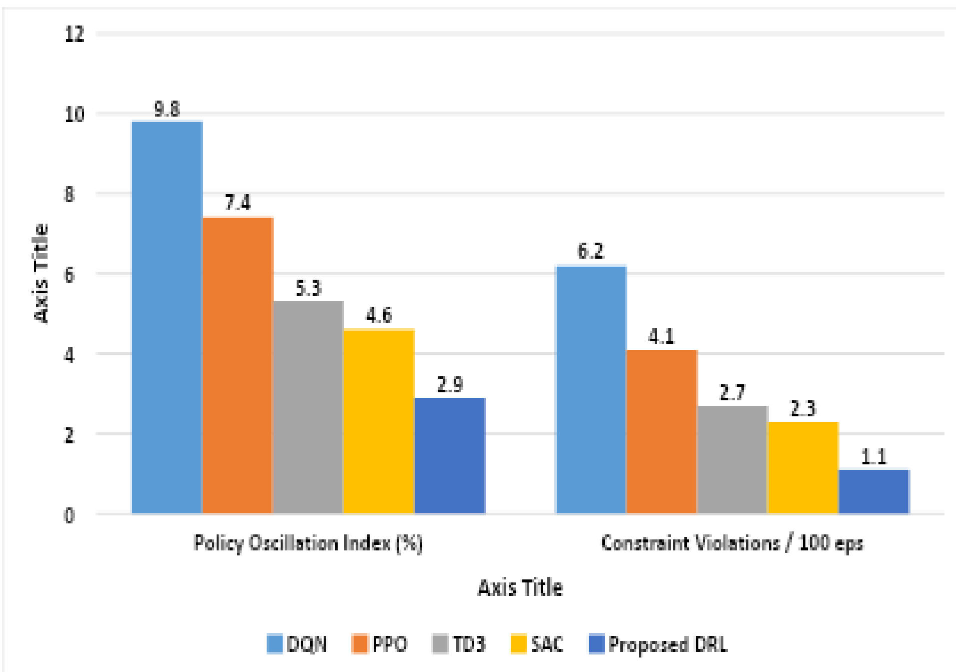

Figure 3: Policy Stability and Constraint Violation Comparison of DRL Algorithms

The figure 3 gives a comparison of the policy oscillation and constraint violation among the DRL algorithms. The index of oscillation and fewer cases of constraint violation are the lowest with the proposed DRA model, which means that the model is more stable in learning and makes safe decisions and achieves consistency in controlling autonomous robots, as opposed to traditional DRA frameworks.

6.2. Task Efficiency and Throughput Improvement Assessment

The results of the task efficiency and throughput indicate the high performance enhancement which was obtained with the help of the suggested DRL-based autonomous robotic framework. The proposed system takes the shortest average cycle time of 19.8 seconds and the highest throughput of 178 units per hour compared with the rule-based and the traditional control methods. This has been ascribed to the optimization of path planning, smart scheduling and adaptive task allocation which has been learned in a process of reinforcement learning. The idle time decreased to 5.9%, which shows that the resources were more resourceful and that the waiting time between operations was reduced to a minimum. The unit consumption of energy is also lower by a large margin and this indicates that the framework is capable of enhancing operational efficiency without compromising its productivity as exhibited in table 3. The defect rate also drops to 1.8, which means that there is a higher level of precision and quality-consciousness in the decision-making process. The total equipment effectiveness (OEE) is 88.7, which proves the high operational performance in comparison with the base practices. Such outcomes demonstrate the potential of DRL-based robotics to optimize manufacturing, lower operation expenses, and ensure uniform quality of products, which is why the strategy is extremely appropriate to the next-generation smart factory setting.

Table 3: Comparative Analysis of Task Efficiency, Throughput, and Operational Performance across Control Strategies in Smart Manufacturing

Parameter	Rule-Based	PID+Planner	DQN	PPO	Proposed DRL
Avg Cycle Time (s)	28.6	25.2	23.4	22.1	**19.8**
Throughput (units/hour)	126	143	152	160	**178**
Idle Time (%)	14.8	11.3	9.7	8.4	**5.9**
Energy/Unit (kJ)	22.4	20.1	18.7	17.9	**16.2**
Defect Rate (%)	4.9	3.6	3.1	2.7	**1.8**
OEE (%)	71.2	76.9	80.4	82.6	**88.7**

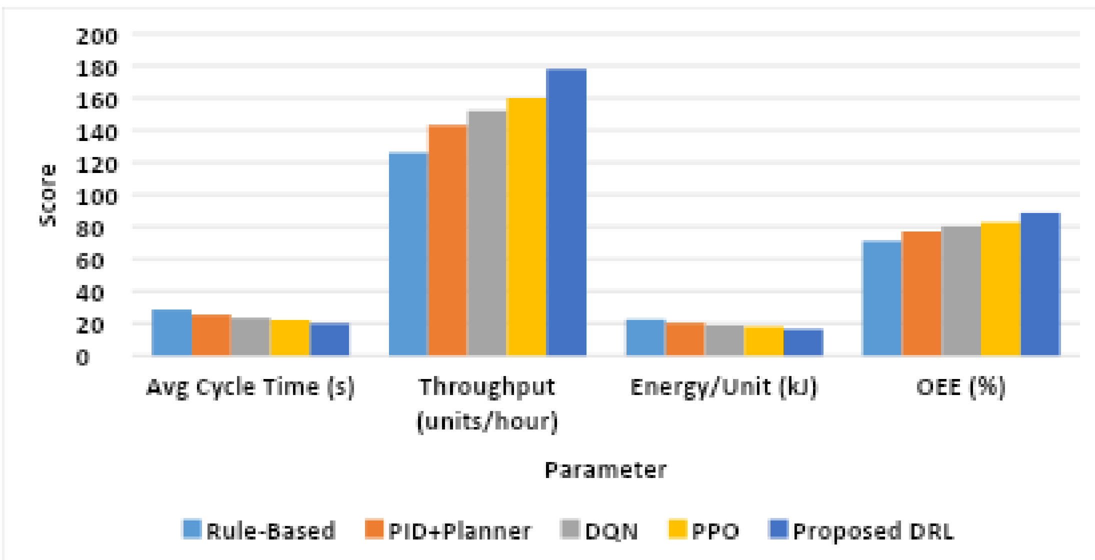

Figure 4: Comparative Analysis of Task Efficiency and Throughput across Control Strategies

The figure 4 is the comparative analysis of cycle time, throughput, energy usage, and the effectiveness of the equipment in the control strategies. The suggested DRA methodology obtains shorter cycle time, better throughput, greater energy efficiency, and more excellent OEE that prove better manufacturing operations and productivity in intelligent manufacturing facilities.

6.3. Adaptability to Dynamic Production Changes

Adaptability analysis implies that the suggested DRL model can successfully react to the fluctuating production parameters and those of changing manufacturing conditions, as they are represented in table 4. With a rapid response to new tasks and layouts, the system has the shortest recovery time of 4.6 minutes after changing the production layouts, and after reconfiguration, it adapts fast. There is a very low level of throughput reduction in production changeovers, which is reduced to 5.4 only, which is much less than the base methods, which denotes the continuity of operations. The fact that the model has fewer re-learning episodes in case of introducing new products, speaks volumes of its capacity to generalize learned policies on related tasks. Moreover, the proposed framework offers the greatest success rate of the task-switch and timetable compliance, and it provides a seamless transition of varying manufacturing processes. The decrease in the number of rework events on a shift is also indicative of a higher accuracy in decision-making in changing conditions. All in all, these findings substantiate the notion that autonomous robotics based on DRL can flexibly adapt dynamically to production variability and allow flexible and responsive manufacturing systems to accept high-mix and customized production needs without a material performance penalty.

Table 4: Adaptability Performance Comparison of Autonomous Control Methods under Dynamic Production Reconfiguration in Smart Manufacturing

Parameter	Rule-Based	DQN	PPO	SAC	Proposed DRL
Reconfiguration Recovery Time (min)	18.5	11.2	8.4	7.1	**4.6**
Throughput Drop During Change (%)	22.8	14.6	10.3	8.7	**5.4**
Policy Re-learning Episodes (new product)	0*	520	410	360	**220**
Task Switch Success Rate (%)	86.3	91.7	93.5	94.2	**97.6**
Schedule Adherence (%)	78.1	84.9	87.6	88.8	**93.4**
Mean Rework Events / shift	7.6	5.1	4.3	4.0	**2.6**

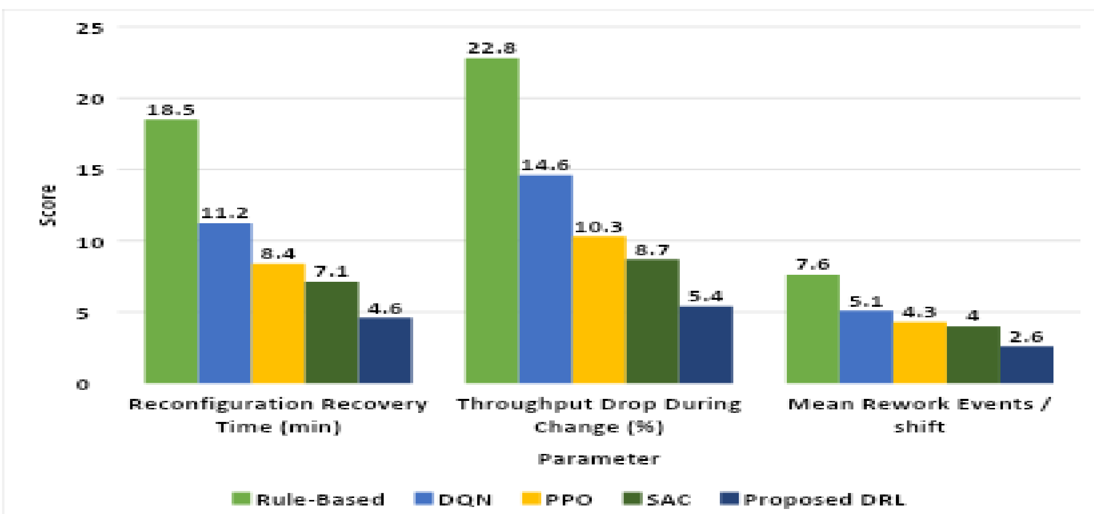

Figure 5: Adaptability Performance Comparison of Control Strategies Under Dynamic Production Changes

The figure 5 shows the adaptability performance of the various control strategies when there are dynamic production variations in smart manufacturing. The estimated DRA framework has the shortest reconfiguration recovery time, less throughput drop, and fewer events of rework per shift. These advancements underscore its better ability to handle dynamically evolving production demands, sustain production continuity and secure uniform manufacturing productivity relative to traditional rule-based and baseline DRL frameworks.

6.4. Robustness under Disturbances and Uncertainties

Table 5 is a comparison of robustness and operational reliability in the various control measures in the uncertain and noisy manufacturing characteristics. The proposed DRL-based system has the highest performance retention of 95.3 percent with noise variation of less than 10 percent, is more resilient to environmental perturbations. It has the lowest collision and near-miss rate which implies improved safety and accurate motion control in dynamic workspaces. The decreased number of fail-safe triggers per shift sheds more light on a stable and reliable system operation with the minimal emergency interruptions. Also, the proposed method has the lowest mean tracking error, and it has high positional-accuracy on robotic manipulation and navigation issues.

Table 5: Robustness and Operational Reliability Comparison of Autonomous Control Strategies under Disturbances in Smart Manufacturing

Parameter	**Rule-Based**	**PID+ Planner**	**QN**	**AC**	**Proposed DRL**
Performance Retention @ Noise (±10%) (%)	81.4	86.9	9.1	1.6	**95.3**
Collision/Near-Miss per 1000 cycles	5.8	3.6	.9	.1	**0.9**
Fail-Safe Triggers per shift	6.1	4.4	.2	.7	**1.4**
Mean Tracking Error (mm)	2.6	2.1	.8	.5	**1.1**
Unplanned Downtime (min/shift)	42.0	31.5	5.3	1.6	**14.8**
Robustness Index (0–1)	0.78	0.83	.86	.89	**0.94**

Unplanned downtime is also greatly reduced as compared to the traditional rule-based and learning-based controllers, which lead to the enhancement of the production continuity and resource use. The consistency of the index of robustness (0.94) is also another positive indicator of strong and stable results in different operating conditions. Baseline methods, conversely, are more sensitive to noise effects and disturbances and are less efficient and stable. In general, the findings confirm that the suggested DRL-based autonomous robotic system has better robustness, safety, and operational reliability, and it is very relevant to be implemented in the real-world environment of smart manufacturing with uncertainty and dynamic changes.

7. Discussion

7.1 Analysis of Key Findings and Observed Performance Gains

As indicated by the experimental analysis, the suggested DRL-powered autonomous robotic system has a major beneficial impact on the performance of the manufacturing process in terms of several aspects of its functioning. The system has an enhanced learning convergence, enhanced stability of policy and elevated success rates as compared to the traditional and baseline DRL methods. Significant improvements are realized in throughput, reduction in cycle time, energy efficiency, and minimization of defects which means that the multi-objective optimization is effective. Combining hierarchical action modeling, safety-aware reward design, and training with the help of the digital twin will help to achieve consistent and robust decision-making. Moreover, adaptability analysis proves that the system is capable of quickly reacting to production change without much performance loss. Strength in ambiguity stresses how the framework can withstand the disruption and changing conditions. The overall results of these studies show that DRL-based autonomy can have a significant positive impact on operational intelligence, efficiency, and reliability in smart manufacturing systems and enable the shift to self-optimizing and flexible industrial automation.

7.2 Implications of Practical Application in Industry

The future of manufacturing industries in the current world holds the transformative potential in the practical implementation of DRL-enabled autonomous robotics. The proposed framework will minimize the reliance on hand-written programming and human intervention by allowing flexible scheduling and optimization of processes in real time and autonomous decision-making. Constant learning and optimization allow industries to attain high levels of productivity, minimum downtimes, and high quality of products. Interaction with the established manufacturing execution systems permits a smooth alignment between the operations at the shop-floor and at the enterprise planning level. Digital twins are safe to train and validate before deploying, and minimize the risks of operations. The framework also facilitates predictive maintenance and resource optimization that help in cost reduction and sustainability. Such empirical benefits lead to DRL-based autonomous robotics as a practical initiative to use in flexible manufacturing settings with frequent reconfigurations, custom production, and dynamic operational demands.

7.3 Scalability, Interoperability and Real time Feasibility

The two elements of scalability and interoperability are essential to the deployment of autonomous robotics in large-scale manufacturing systems. The suggested framework will facilitate the process of modular deployment in a variety of robotic workcells and production lines by taking advantage of standardized communication protocols and distributed learning architectures. It is interoperable with industrial IoT environments, cloud-edge computing environments, and enterprise resource planning environments, which make it easy to integrate with existing smart factory systems. The real-time capability is ensured by the optimized designs of neural networks, hardware acceleration with the help of GPUs, and efficient communication schemes minimising latency. The framework may be decentralized and centralized and it is scalable to heterogeneous robotic systems. Those properties allow the implementation of the DRL-based autonomous robotics in sophisticated manufacturing facilities and retain the real-time responsiveness and operational reliability.

8. CHALLENGES, LIMITATIONS, AND FUTURE RESEARCH DIRECTIONS

8.1 Training Data Requirements and Computational Complexity

Nevertheless, regardless of the illustrated advantages, autonomous robotics based on DRL need vast amounts of training data and computing power. Complex policies that need to be trained in high-dimensional manufacturing situations also require a lot of simulation time and hardware acceleration, which may necessitate sophisticated GPU or distributed computing infrastructure. Real industrial systems might also restrict the amount of data that can be collected because some of the industrial activities are hazardous and unsafe as well as not conducive to the collection of data. Simulation-based training is therefore necessary. Nevertheless, these differences between the simulated and the actual environments may affect the performance in the field. Compression of computation models by using transfer learning and effective exploration techniques is a significant research agenda. The next line of work in this area should be on enhancing sample efficiency and cut training time to apply to industrial environments faster and more adaptively.

8.2 Safety, Explainability, and Trustworthiness of DRL Policies, Explanation

The safety and reliability of autonomous systems controlled by DRL is still a major concern to allow them to be used in industry. The black-box neural network models are usually not interpretable, and hence, the operators may find it hard to comprehend how decisions are made. Such insensitivity may diminish the trust in autonomous systems, especially in manufacturing processes that are safety-related. The use of explainable AI approaches and safety verification systems is a necessary factor in instilling confidence in DRL-based controllers. Also, there are formal validation and certification systems that are needed to make sure that the industrial safety standards are met.

8.3 Future Extensions to Federated and Collaborative DRL

Future studies have the potential to expand the suggested framework by adding federated and collaborative DRL techniques to distributed learning in more robotic systems and manufacturing stations. Federated learning allows robots to exchange policies and knowledge over learned information, but not transfer of sensitive operational data, which is a guarantee of privacy and security. Multi-agent learning can also be collaborative and can be used to increase coordination in heterogeneous robots undertaking related tasks. These methods enhance scaling and scalability as well as improve learning and decentralized decision-making in large-scale smart factories. By combining cloud-edge intelligence and federation of DRL, it will be possible to provide continuity in learning and adaptation at distributed manufacturing systems across geographical boundaries.

8.4 Human-Robot Collaboration and Ethical Concerns

The gradual implementation of autonomous robotics in the manufacturing industry requires human-robot interaction and ethics to be taken into account. The safety of people operating in the workplace and the efficiency of self-driving robots depends on the safety of their interaction. Workforce displacement, accountability of decisions, and data privacy are also ethical issues that should be considered. Acceptance and operational effectiveness could be realized through designing collaborative structures that will contribute to increasing human potential, but not substituting it. The future studies should delve into human-centric models of DRL that can assist in the direction of cooperative decision-making as well as ensuring that smart manufacturing systems are ethically correct.

9. CONCLUSION

The research was a detailed deep reinforcement learning (DRL)-based system of empowering autonomous robotics in intelligent manufacturing. The suggested system combines adaptive decision-making, hierarchical action design, and learning with the assistance of a digital twin to optimise the work of robots in changing production conditions. It was experimentally tested that it converges to learning faster, exhibits better policy stability and higher success rates than the conventional and baseline DRL methods. The findings established significant decreases in cycle time, idle time, and energy consumption, as well as, substantial increment in throughput, overall equipment effectiveness and product quality. The proposed framework was more efficient and reliable than rule-based and classical control strategies and did not decrease its performance in the presence of disturbances and unpredictable operational conditions. The adaptability analysis also demonstrated that the system can respond with great responsiveness to production reconfigurations and new task demands at the lowest throughput degradation, thus making it applicable in a high-mix and custom manufacturing design. The robustness test indicated the increased resilience to noise and environmental variations, which guarantees the stable and safe operation. The realized performance improvements indicate how DRL-based autonomous robotics can turn the manufacturing industry into a smart self-adaptive production. The combination of learning-based control and digital twin environments as well as manufacturing execution systems can allow seamless integration between physical and digital operations to allow real-time optimization and predictive decision-making. In general, the suggested structure makes DRL-based autonomous robotics one of the enablers of the next-generation smart factories, which will lead to better productivity, operational flexibility, and sustainable industrial development and can further develop the idea of Industry 4.0 and Industry 5.0

References:

1. Alginahi, Y. M., Sabri, O., & Said, W. (2025). Reinforcement learning for industrial automation: A comprehensive review of adaptive control and decision-making in smart factories. Machines, 13(12), 1140. https://doi.org/10.3390/machines13121140
2. Ali, M., Giri, S., Liu, S., & Yang, Q. (2025). Digital twin synchronization: Bridging the Sim-RL agent to a real-time robotic additive manufacturing control. In Proceedings of the IEEE Engineering Reliable Autonomous Systems (ERAS) (pp. 1–4). IEEE. https://doi.org/10.1109/ERAS63351.2025.11135823
3. Ansari, J., Caro, J. B., & Soyak, R. (2024). A framework for dynamic configuration of cellular industrial networks in mobile robotics applications. In Proceedings of the IEEE 35th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC) (pp. 1–7). IEEE. https://doi.org/10.1109/PIMRC59610.2024.10817458
4. Cai, Y. (2025). Optimization research on the autonomous learning algorithm in field of intelligent manufacturing engineering—Robots. In Proceedings of the 2nd International Conference on Modeling, Natural Language Processing and Machine Learning (pp. 140–144). Association for Computing Machinery. https://doi.org/10.1145/3757110.3757132
5. del Real Torres, A., Andreiana, D. S., Ojeda Roldán, Á., Hernández Bustos, A., & Acevedo Galicia, L. E. (2022). A review of deep reinforcement learning approaches for smart manufacturing in Industry 4.0 and 5.0 framework. Applied Sciences, 12(23), 12377. https://doi.org/10.3390/app122312377
6. Farkhodov, K., Lee, S.-H., Platos, J., & Kwon, K.-R. (2023). Deep reinforcement learning TF-agent-based object tracking with virtual autonomous drone in a game engine. IEEE Access, 11, 124129–124138. https://doi.org/10.1109/ACCESS.2023.3325062
7. Felbrich, B., Schork, T., & Menges, A. (2022). Autonomous robotic additive manufacturing through distributed model-free deep reinforcement learning in computational design environments. Construction Robotics, 6, 15–37. https://doi.org/10.1007/s41693-022-00069-0
8. Gao, Q., Chang, F., Yang, J., Tao, Y., Ma, L., & Su, H. (2024). Deep reinforcement learning for autonomous driving with an auxiliary actor discriminator. Sensors, 24(2), 700. https://doi.org/10.3390/s24020700
9. Ho, T. M., Nguyen, K.-K., & Cheriet, M. (2024). Federated deep reinforcement learning for task scheduling in heterogeneous autonomous robotic systems. IEEE Transactions on Automation Science and Engineering, 21(1), 528–540. https://doi.org/10.1109/TASE.2022.3221352
10. Lee, O., Kim, J., Park, S., Choi, Y., Kim, D., & Lee, J. (2023). Development and validation of reconfigurable autonomous mobile manipulator for flexible manufacturing process. In Proceedings of the 23rd International Conference on Control, Automation and Systems (ICCAS) (pp. 914–918). IEEE. https://doi.org/10.23919/ICCAS59377.2023.10317019
11. Le, T., Vinh, L. Q., & Pham, V. H. (2025). Virtual, augmented, and mixed reality robotics-assisted deep reinforcement learning towards smart manufacturing. Sensors, 25(11), 3349. https://doi.org/10.3390/s25113349
12. Li, C., Zheng, P., Yin, Y., Wang, B., & Wang, L. (2023). Deep reinforcement learning in smart manufacturing: A review and prospects. CIRP Journal of Manufacturing Science and Technology, 40, 75–101. https://doi.org/10.1016/j.cirpj.2022.11.003
13. Liu, W., Ma, X., Zheng, B., & Shen, G. (2024). A study of submersible positioning based on deep reinforcement learning and automatic control algorithms. In Proceedings of the 6th International Symposium on Robotics & Intelligent Manufacturing Technology (ISRIMT) (pp. 419–426). IEEE. https://doi.org/10.1109/ISRIMT63979.2024.10875281

14. Liu, Z., Liu, Q., Xu, W., Wang, L., & Zhou, Z. (2022). Robot learning towards smart robotic manufacturing: A review. Robotics and Computer-Integrated Manufacturing, 77, 102360. https://doi.org/10.1016/j.rcim.2022.102360
15. Makhija, G. (2025). Reinforcement learning for scheduling with autonomous mobile robots and collaborative robotics in smart manufacturing industry. In Proceedings of the IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG) (pp. 1–6). IEEE. https://doi.org/10.1109/ICTBIG68706.2025.11323895
16. Nievas, N., Pagès-Bernaus, A., Bonada, F., Echeverria, L., & Domingo, X. (2024). Reinforcement learning for autonomous process control in Industry 4.0: Advantages and challenges. Applied Artificial Intelligence, 38(1). https://doi.org/10.1080/08839514.2024.2383101
17. Ryalat, M., ElMoaqet, H., & AlFaouri, M. (2023). Design of a smart factory based on cyber-physical systems and Internet of Things towards Industry 4.0. Applied Sciences, 13(4), 2156. https://doi.org/10.3390/app13042156
18. Tejer, M., Szczepanski, R., & Tarczewski, T. (2024). Robust and efficient task scheduling for robotics applications with reinforcement learning. Engineering Applications of Artificial Intelligence, 127(Part A), 107300. https://doi.org/10.1016/j.engappai.2023.107300
19. Thanh Van, N. T., Luong, N. C., Le Hung, H., Hoa, N. T., & Tien, N. M. (2024). Minimizing energy consumption in mobile robotics with STAR-RIS in smart factories. In Proceedings of the Tenth International Conference on Communications and Electronics (ICCE) (pp. 741–746). IEEE. https://doi.org/10.1109/ICCE62051.2024.10634674
20. Torres, A. D. R., Andreiana, D. S., Roldán, O., Bustos, A. H., & Galicia, L. E. A. (2023). Deep reinforcement learning approaches for smart manufacturing. Encyclopedia. https://encyclopedia.pub/entry/40007
21. Wang, Y., Xiao, R., Wang, X., & Liu, A. (2023). Constructing autonomous, personalized, and private working management of smart home products based on deep reinforcement learning. Procedia CIRP, 119, 72–77. https://doi.org/10.1016/j.procir.2023.04.003
22. Xia, K., Sacco, C., Kirkpatrick, M., Saidy, C., Nguyen, L., Kircaliali, A., & Harik, R. (2021). A digital twin to train deep reinforcement learning agent for smart manufacturing plants: Environment, interfaces and intelligence. Journal of Manufacturing Systems, 58(Part B), 210–230. https://doi.org/10.1016/j.jmsy.2020.06.012
23. Yadav, U., Bondre, S. V., & Thakre, B. (2024). Deep reinforcement learning in robotics and autonomous systems. In S. Mahajan, P. Raj, & A. K. Pandit (Eds.), Deep reinforcement learning and its industrial use cases. Wiley. https://doi.org/10.1002/9781394272587.ch10
24. Yu, Y., Tang, Q., Jiang, Q., & Fan, Q. (2025). A deep reinforcement learning-assisted multimodal multiobjective bilevel optimization method for multirobot task allocation. IEEE Transactions on Evolutionary Computation, 29(3), 574–588. https://doi.org/10.1109/TEVC.2025.3535954.