Deep Reinforcement Learning Approach with Digital Twin Toward Smarter Warehouse

Kanita Jerin Tanha, Md Mahinur Alam, and Taesoo Jun
Pervasive Intelligent Computing Lab, Department of IT Convergence Engineering,
Kumoh National Institute of Technology, Gumi 39177, South Korea
(kanitajerin17, mahinuralam213, taesoo.jun)@kumoh.ac.kr

***Abstract*—The rapid evolution of smart warehouse logistics has accelerated the demand for intelligent, adaptive systems capable of optimizing operations in real time. Traditional static product placement strategies are becoming increasingly inadequate for handling dynamic order patterns, fluctuating demand, and complex material handling interactions. This study presents an adaptive slotting framework driven by a Deep Q-Network (DQN) integrated within a digital twin environment. This DQN continuously learns from operational feedback to generate optimal product placement decisions, aiming to minimize picking time and enhance overall warehouse performance. A warehouse logistics configuration within a Digital Twin virtual environment for product handling processes is modeled using the Asset Administration Shell (AAS), enabling standardized data representation and seamless interaction between the model and the virtual warehouse environment. Experimental results demonstrate significant operational improvements in key areas. The proposed approach reduces average picking time by 30.83%, increases throughput by 32.35%, and decreases average travel distance by 13.15%.**

***Index Terms*—Adapting slotting, asset administration shell (AAS), digital twin, reinforcement learning, smart warehouse logistics.**

I. Introduction

Modern warehouse operations is undergoing a profound transformation as contemporary logistics systems increasingly rely on automation technologies, IoT sensing devices, and data-driven decision mechanisms to sustain high levels of efficiency. The rapid growth of e-commerce, coupled with fluctuating customer demand and the diversification of product assortments, has imposed unprecedented pressure on warehouses to fulfill orders at greater speed, precision, and scalability [1]. Despite these advancements, many facilities still struggle with issues such as uneven stock-keeping unit (SKU) demand patterns, congestion in high-traffic zones, inefficient item placement, and the limitations of static slotting methods that cannot adapt to real-time operational changes.

Digital twin (DT) technology has emerged as a central enabler in addressing these challenges. This DT provides a synchronized virtual replica of warehouse assets, processes, and interactions, allowing organizations to analyze performance, simulate alternative configurations, and test optimization strategies without interrupting live operations [2]. Several DT simulation platforms, such as AnyLogic, FlexSim, and NVIDIA Omniverse, are increasingly used to build high-fidelity models of intralogistics systems; these tools support discrete-event, agent-based, and 3D physics-based simulations, thereby enabling the detailed representation of material flows, resource behaviors, and human–robot interactions under varying demand profiles and control policies [3]. Complementing DTs, the AAS serves as a core concept of Industry 4.0, standardizes the digital representation of each physical asset, enabling interoperable data exchange, structured lifecycle information, and seamless integration across heterogeneous warehouse systems [4]. Despite these capabilities, many DT and AAS implementations rely on fixed models or deterministic rules, limiting their ability to respond autonomously to dynamic changes, stochastic events, or real-time variations in operational load [5].

Alongside DT development, machine learning (ML), deep learning (DL), federated learning (FL), and reinforcement learning (RL) techniques have been increasingly explored for warehouse optimization tasks such as demand forecasting, routing, and resource scheduling [6]. While these methods provide strong predictive capabilities, their reliance on pre-collected datasets and absence of direct interaction with the operational environment restricts their adaptability when confronted with rapidly changing conditions. DL models require large labeled datasets, FL introduces communication and synchronization challenges, and RL struggles with large action spaces and unstable convergence and high variability in reward structures, especially within the complex and continuous slotting decisions [7], [8].

To overcome these multifaceted constraints, the present study introduces a Deep Reinforcement Learning (DRL) approach embedded within a DT environment to achieve adaptive product slotting in smart warehouse logistics [9]. By integrating a high-fidelity DT with a DRL model, the proposed framework enables dynamic, data-driven slotting decisions that continuously adapt to real-time demand fluctuations and operational constraints [10]. This unified approach enhances efficiency, reduces travel distance, and provides a scalable mechanism for autonomous warehouse optimization. The contributions of this paper are as follows:

- We have developed a DQN that learns zone-assignment policies from streaming warehouse states, the model uses action masking, experience replay, and a greedy schedule to reduce pick latency, shorten travel distance and improve throughput.
- We proposed a DT simulation with training framework

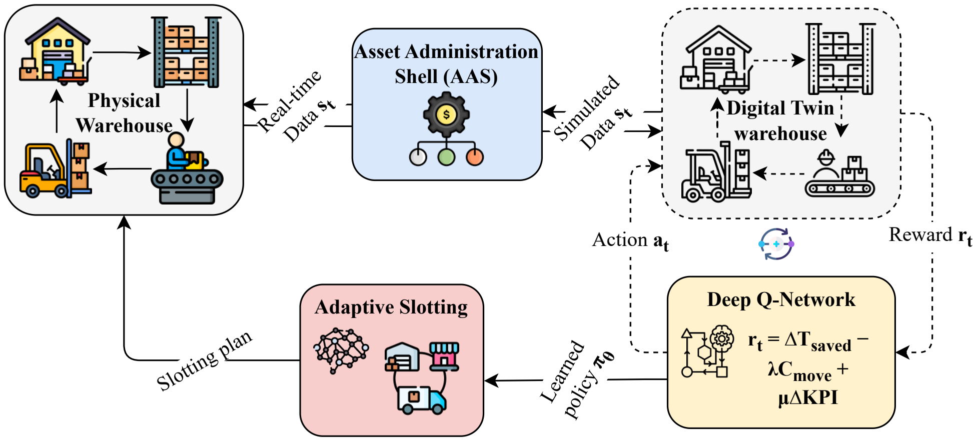

Fig. 1. Proposed adaptive slotting workflow combining physical warehouse data, AAS, DT simulation, and DQN Model.

and evaluation environment tightly integrated with an AAS to standardize asset metadata, enable closed-loop simulation, and support safe policy synchronization with live systems.

- Empirical validation on an open-source warehouse dataset with an extended evaluation protocol, demonstrating consistent operational gains.

The remainder of this paper is organized as follows. **Section II** reviews the related work on adaptive slotting, RL, and DT technologies. **Section III** introduces the proposed overall framework. **Section IV** describes the experimental performance analysis and discussion. Finally, **Section V** concludes the paper and outlines potential future directions.

II. Related Work

For the rapid expansion occurring in warehouse operations and multi-channel logistics has increasingly highlighted the inherent limitations of traditional warehouse slotting practices, because static product placement strategies and infrequent reconfiguration routines consistently struggle to keep pace with dynamic order patterns, fluctuating demand behaviors, and the constantly changing nature of operational workflows, ultimately leading to congestion, increased picking inefficiency, and declining performance in real-world environments. While classic approaches rely heavily on predetermined heuristics or offline optimization models, these techniques often become ineffective as operational complexity grows, resulting in bottlenecks and progressively diminishing efficiency over time when they fail to capture the real-time variability of warehouse activities [11].

Recent advances have therefore focused on leveraging artificial intelligence (AI) and ML to transform slotting optimization, as these learning-driven systems are capable of analyzing high-velocity SKUs movement, identifying item affinity relationships, discovering path utilization patterns, and incorporating historical picking data in order to recommend inventory placements that increase throughput, reduce the distance traveled by workers, and significantly enhance overall accuracy in daily operations. These data-intensive systems offer warehouse managers the ability to make more frequent, informed, and adaptive slotting adjustments based on evidence gathered from ongoing warehouse behaviors, rather than depending on expensive and infrequent large-scale layout revisions that do not reflect real-time operational changes [12], [13]. DT technology has emerged as a crucial enabler, providing real-time, dynamic simulations of warehouse layouts, material flows, and operational processes [14]. It allows for virtual experimentation and scenario analysis, making it possible to test slotting strategies rapidly and visualize the impact of layout, labor, and process changes before deployment. Such frameworks are now widely used for predictive analytics, bottleneck identification, and rapid operational optimization. Now RL and DRL have gained favor in this space for their ability to learn optimal slotting, picking, and routing policies in high-dimensional, dynamic environments. These RL model often trained within DT environments systematically test possible configurations, evaluate key performance indicators (KPIs) like picking speed and error rate, and adjust strategies in response to simulated feedback [15].

Empirical case studies reveal RL-driven slotting and robot movement can decrease travel time by as much as 20%, and order accuracy mistakes by 25%. Hybrid models combining deep learning with RL further advance complex scenario reasoning and multi-objective optimization. Beyond improving slotting and picking, RL integrated with DT facilitates automated adaptation to real-world events such as seasonality, labor shortages, or structural changes, and offers scalable, rule-free, continuous improvement. Industry reviews anticipate that warehouses not investing in DT and RL-based AI solutions risk losing their competitive edge as customer expectations for order speed, accuracy, and transparency [16].

III. Proposed System

We proposed a framework that establishes an adaptive slotting architecture that couples an RL model with a DT-based representation of warehouse operations, thereby allow-

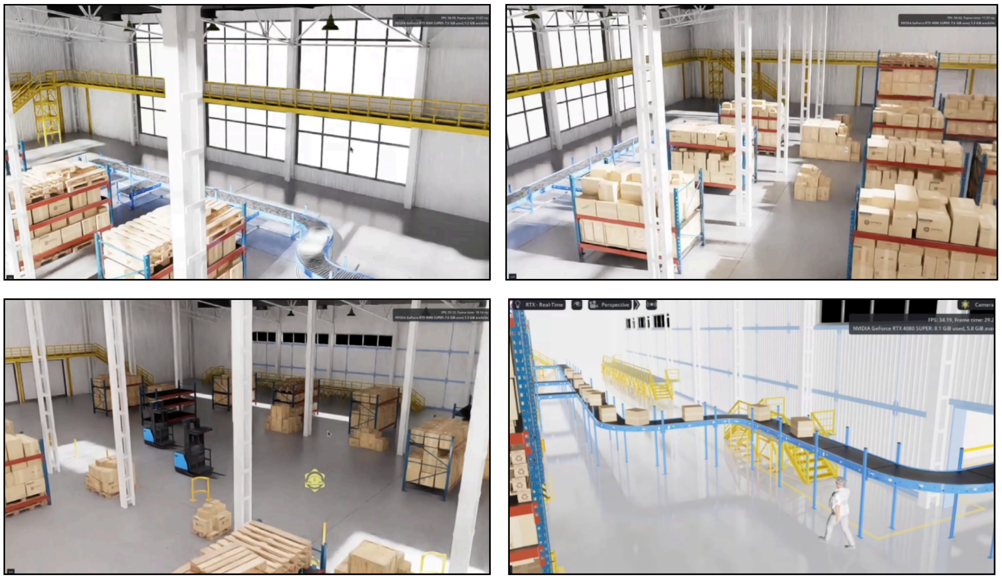

Fig. 2. Digital Twin simulation of warehouse logistics operations using Omniverse Isaac Sim.

ing learning-based decision making to be tightly bound to a continuously synchronized virtual mirror of the physical warehouse and its ongoing processes. In this design, real-time asset information, structured digital descriptions, and learning-based decision mechanisms are unified into a single cohesive pipeline in order to facilitate continuous optimization of storage allocation, ensuring that item placement decisions remain aligned with current demand profiles and operational constraints rather than static historical configurations. As we illustrated in Fig. 1, the physical warehouse continuously streams operational data s_t to the AAS, which standardizes each asset's identification, capabilities, and current state. This enriched and structured information is then propagated to the DT environment, where warehouse behaviors such as storage-zone states, worker and forklift movements, and item-handling flows are simulated to generate predictive and context-aware system responses. Here DQN receives both real-time and simulated feedback, enabling it to evaluate actions a_t, compute rewards r_t, and update its slotting policy π_θ. This closed interaction loop between the physical warehouse, AAS, DT, and DQN ensures that every warehouse component maintains an up-to-date virtual counterpart, allowing the model to generate accurate, data-driven slotting decisions that continuously adapt to the evolving operational environment rather than remaining fixed to an initial configuration.

A. Reinforcement Learning-Based Slotting Optimization

Here we designed DQN to optimize dynamic product placement across four warehouse zones by minimizing operational metrics such as picking time and travel distance, while implicitly accounting for variations in demand intensity and spatial accessibility that shape the efficiency of retrieval operations. At each decision step t, the warehouse state is encoded as a feature vector s_t = [zone, item_demand, picking_time, stock_level, . . .]. Based on this state, the model selects an action $a_t \in \{A, B, C, D\}$, where each action corresponds to assigning an item to one of the four available storage zones. The expected value of selecting action a_t in state s_t is estimated by the action-value function $Q(s, a)$, which provides the predicted return for placing an item in a given zone. The immediate reward is computed using the inline function $r_t = \Delta T_{\text{saved}} - \lambda C_{\text{move}} + \mu \Delta \text{KPI}$, where ΔT_{saved} represents the improvement in picking time achieved by the action, C_{move} denotes the movement cost, and ΔKPI captures variations in key operational indicators such as throughput and zone efficiency. A composite key performance indicator (KPI) score is produced from normalized metrics combined through a weighted summation. The action-value function is approximated through a neural network $Q(s, a; \theta)$, which is trained using experience replay and a periodically updated target network. Each transition (s_t, a_t, r_t, s_{t+1}) is stored in a replay buffer and sampled uniformly for mini-batch updates. The training objective minimizes the temporal-difference loss $L(\theta) = \frac{1}{N}\sum_{j=1}^{N}(y_j - Q(s_j, a_j; \theta))^2$, where the target value is defined as $y_j = r_j + \gamma Q'(s_{j+1}, a'; \theta^-)$, and Q' denotes the target network with parameters θ^-.An ϵ-greedy exploration scheme is used, where ϵ gradually decreases during training to balance exploration and exploitation. This model is trained across thousands of episodes until convergence is observed in reward stabilization and consistent improvements in KPI performance, indicating that the learned policy has internalized robust slotting strategies that generalize across diverse warehouse conditions.

B. AAS–Digital Twin Interaction Cycle

For communication between the DQN and the warehouse environment is facilitated through the DT and the AAS, which

TABLE I
IMPROVEMENT GAIN USING THE PROPOSED DQN MODEL ON THROUGHPUT, PICKING TIME, AND TRAVEL DISTANCE.

Metrics	**Baseline**	**DQN Model**	**Changes (%)**
Average Picking Time (s)	172.22	119.12	**↓30.83%**
Throughput (orders/hour)	25.66	33.96	**↑32.35%**
Average Traveling Distance (units)	0.9201	0.7991	**↓13.15%**

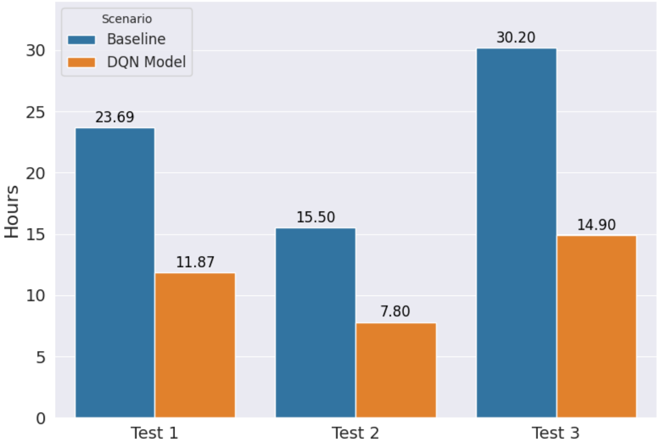

Fig. 3. Picking time optimization over different testing scenarios baseline against Proposed DQN model.

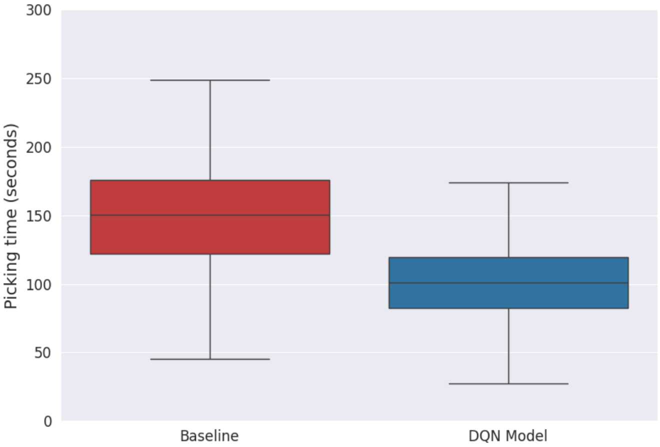

Fig. 4. Picking time distribution before and after adaptive re-slotting using DQN model.

together provide the standardized interface and virtual testbed needed to translate learning decisions into operational consequences and back into training signals. When the agent outputs an action a_t, an execution module updates the digital twin's internal state to reflect the newly selected slotting decision. The digital twin then evaluates the operational consequences of this action, producing updated metrics such as travel distance, picking time, and variations in KPIs. These outputs are converted into the next state s_{t+1} and the corresponding scalar reward r_t, which are subsequently stored in the replay buffer. This creates a closed interaction cycle agent decision $\rightarrow$ digital-twin update $\rightarrow$ KPI extraction $\rightarrow$ reward computation $\rightarrow$ replay storage $\rightarrow$ network update allowing the model to iteratively refine and improve the learned slotting policy through repeated interaction with a safe yet realistic virtual environment.

C. Dataset Description & Simulation

To evaluate our proposed framework, we utilized an open-source real-world warehouse logistics dataset, comprising detailed operational records for inventory levels, stock movements, product categories, location assignments, demand forecasts, and KPI across diverse warehouse zones, which together provide a rich empirical basis for modeling realistic slotting scenarios. We have utilized a key extension of DT technology, Omniverse Isaac Sim, which is NVIDIA's high-fidelity simulation platform designed for creating sophisticated digital representations of industrial environments, to instantiate a virtual warehouse that directly reflects the structure and dynamics encoded in the dataset. It enables the creation of DT for complex environments of warehouse logistics, offering realistic physics, agent-based behavior, and real-time interaction. We simulated the warehouse operations present in the dataset in a 3D environment, enhancing predictive analytics and enabling detailed scenario testing as demonstrated in Fig. 2. The simulation includes features for item identification, category segmentation, stock levels, storage location, picking times, demand rates, cost factors, and fulfillment performance. A selected subset of these features including item_id, category, stock_level, storage_location_id, zone, picking_time_seconds, daily_demand, KPI_score, order_fulfillment_rate, and relevant cost and efficiency metrics, was used for training and evaluation of our slotting optimization reinforcement learning model. We used it for supporting state representation, reward formulation, and performance benchmarking, providing a robust foundation for modeling adaptive slotting strategies in modern warehouse environments. This structured representation enabled the DQN model to process well-defined state vectors s_t and learn reliable action-value mappings $Q(s_t, a_t)$ based on accurate and consistent feedback from the environment.

IV. PERFORMANCE ANALYSIS

A. Evaluation Results

For consolidated performance improvements achieved by the proposed DQN model are summarized in Table I, which shows substantial gains across all evaluation metrics. Relative to the baseline configuration, average picking time is reduced from 172.22 s to 119.12 s, representing a 30.83% decrease, while throughput increases from 25.66 to 33.96 orders/hour, equivalent to a 32.35% improvement. In addition, the learned

TABLE II
TOP 10 SLOTTING RELOCATIONS RANKED BY TIME SAVED.

ID	**Item**	**Category**	**Previous Zone**	**Relocated Zone**	**Daily Demand**	**Orders (hr)**	**Baseline Pick Time (s)**	**Improved Pick Time (s)**	**Time Saved (s)**
384	ITM11478	Pharma	C	A	47	1.67	228.0	232.5	363.7
359	ITM12063	Electronics	C	D	43	1.81	155.7	149.8	321.1
227	ITM10683	Apparel	C	A	41	1.93	110.0	104.2	332.8
362	ITM10687	Automotive	C	D	42	1.56	39.3	33.7	264.9
579	ITM11572	Apparel	A	C	48	1.96	53.7	51.0	112.4
456	ITM12063	Pharma	A	C	47	1.99	52.7	49.3	101.5
228	ITM11268	Apparel	C	A	43	1.81	263.0	262.8	110.4
154	ITM12728	Automotive	A	C	49	2.04	112.5	110.3	125.9
56	ITM11478	Electronics	A	D	50	2.17	64.8	62.8	82.9

policy also shortens average travel distance from 0.9201 to 0.7991 units (a 13.15% reduction). These combined results indicate that the model successfully accelerates individual pick operations and enhances overall processing capacity, reflecting more efficient placement of high-turnover SKUs in low-travel-cost locations.

We evaluate the consistency of these improvements across different operating conditions, which is further demonstrated in Fig. 3, where the total picking hours for three representative test scenarios are compared before and after applying the DQN-based slotting policy. In each scenario, total picking hours are reduced by approximately half: from 23.69 to 11.87 hours in Test 1, from 15.50 to 7.80 hours in Test 2, and from 30.20 to 14.90 hours in Test 3. These results show that the learned policy generalizes effectively across a variety of demand profiles and spatial configurations, reinforcing the aggregate performance gains highlighted earlier in Table I.

We conduct a distributional analysis of picking times, presented in Fig. 4, provides additional insight into the operational impact of the slotting decisions. After re-slotting, the median picking time shifts noticeably downward, the interquartile range contracts, and the number of extreme high-duration picks is markedly reduced. This tighter and more compact distribution is operationally significant because it produces more predictable throughput, reduces the likelihood of sudden congestion, and simplifies labor and resource planning.

To further understand which specific relocations contribute most to the efficiency improvements, Table II lists the top-ranked item movements ordered by time saved. Items with high daily demand and significant reductions in per-pick time dominate the upper portion of the list, indicating that these relocations account for a substantial portion of the overall performance gain and may serve as valuable candidates for targeted pilot testing in real deployments.

B. Discussion

The observed improvements in mean picking time, throughput, and travel distance collectively suggest that the proposed DQN model learns slotting configurations that reduce both movement inefficiencies and queuing delays. These benefits, however, must be considered alongside practical implementation factors. Frequent re-slotting can impose operational costs, such as labor effort, AGV utilization, and temporary disruptions to replenishment processes. Additionally, while the current evaluation demonstrates consistent gains across scenarios, further analysis is needed to quantify statistical significance at a per-pick level and to assess sensitivity to movement-cost parameters and seasonal demand fluctuations.We provided the evidence presented in Table I, Fig. 3, and Fig. 4 indicates that the DQN-based adaptive slotting approach delivers meaningful and reliable performance improvements while stabilizing pick-time variability.

V. CONCLUSION

This paper presented a novel approach for adaptive warehouse slotting using a DQN integrated within a DT framework, supported by the AAS. By combining real-time warehouse data, a high-fidelity DT and RL, this proposed system dynamically optimizes product placement to minimize picking time, reduce travel distance, and enhance throughput. We assess experimental results that the DQN model achieved a 30.83% reduction in average picking time, a 32.35% increase in throughput, and a 13.15% decrease in travel distance, showcasing the effectiveness of the approach in real-world logistics environments. This proposed system also proved to be adaptable across various warehouse configurations and demand profiles, making it a scalable solution for smart logistics systems. The integration of the DT environment with AAS provides a strong foundation for real-time monitoring, performance analysis, and continuous optimization. These findings underline the potential of combining RL with DT technology for next-generation warehouse management. Future work will focus on extending the framework to multi-agent systems, testing its performance under dynamic constraints, and exploring real-world deployment in live warehouse environments.

ACKNOWLEDGMENT

This work was supported in part by IITP (MSIT) under the Innovative Human Resource Development for Local Intellectualization (IITP-2025-RS-2020-II201612; 34%) and the ITRC program (IITP-2025-RS-2024-00438430; 33%), and by the NRF Basic Science Research Program (Ministry of Education, 2018R1A6A1A03024003; 33%).

References

[1] M. Dotoli, N. Epicoco, M. Falagario, N. Costantino, and B. Turchiano, "An integrated approach for warehouse analysis and optimization: A case study," *Computers in Industry*, vol. 70, pp. 56–69, 2015.

[2] K. J. Tanha, M. M. Alam, M. R. Subhan, and T. Jun, "Detecting threats in edge iot networks using federated learning and digital twin," *Proceedings of the Korean Institute of Communications and Information Sciences Summer Conference*, pp. 489–490, 2025.

[3] N. Ahmed, I. Afyouni, H. Dabool, and Z. Al Aghbari, "A systemic survey of the omniverse platform and its applications in data generation, simulation and metaverse," *Frontiers in Computer Science*, vol. 6, p. 1423129, 2024.

[4] Z. U. Rizqi, S.-Y. Chou, and W. N. Cahyo, "A simulation-based digital twin for smart warehouse: Towards standardization," *Decision Analytics Journal*, vol. 12, p. 100509, 2024.

[5] C. I. Acosta-Acosta, R. O. Andrade-Paredes, and S. V. Avilés-Sacoto, "Bridging the physical and virtual: Digital twin solutions with arduino and flexsim," in *2025 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT)*. IEEE, 2025, pp. 249–256.

[6] Q. Guo, F. Tang, and N. Kato, "Federated reinforcement learning-based resource allocation for d2d-aided digital twin edge networks in 6g industrial iot," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 5, pp. 7228–7236, 2022.

[7] W. Yang, W. Xiang, Y. Yang, and P. Cheng, "Optimizing federated learning with deep reinforcement learning for digital twin empowered industrial iot," *IEEE Transactions on Industrial Informatics*, vol. 19, no. 2, pp. 1884–1893, 2022.

[8] M. Liu, S. Fang, H. Dong, and C. Xu, "Review of digital twin about concepts, technologies, and industrial applications," *Journal of manufacturing systems*, vol. 58, pp. 346–361, 2021.

[9] L. Zhang, C. Yang, Y. Yan, Z. Cai, and Y. Hu, "Automated guided vehicle dispatching and routing integration via digital twin with deep reinforcement learning," *Journal of Manufacturing Systems*, vol. 72, pp. 492–503, 2024.

[10] X. Wang, X. Hu, and J. Wan, "Digital-twin based real-time resource allocation for hull parts picking and processing," *Journal of Intelligent Manufacturing*, vol. 35, no. 2, pp. 613–632, 2024.

[11] J. Leng, D. Yan, Q. Liu, H. Zhang, G. Zhao, L. Wei, D. Zhang, A. Yu, and X. Chen, "Digital twin-driven joint optimisation of packing and storage assignment in large-scale automated high-rise warehouse product-service system," *International Journal of Computer Integrated Manufacturing*, vol. 34, no. 7-8, pp. 783–800, 2021.

[12] J. C. Duque-Jaramillo, J. M. Cogollo-Flórez, C. G. Gómez-Marín, and A. A. Correa-Espinal, "Warehouse management optimization using a sorting-based slotting approach," *Journal of Industrial Engineering and Management*, vol. 17, no. 1, pp. 133–150, 2024.

[13] L. Van Der Hagen, N. Agatz, R. Spliet, T. R. Visser, and L. Kok, "Machine learning–based feasibility checks for dynamic time slot management," *Transportation Science*, vol. 58, no. 1, pp. 94–109, 2024.

[14] M. M. Alam, M. Golam, E. A. Tuli, M. R. Subhan, D.-S. Kim, and T. Jun, "Dcfl-chain: Digital-twin-based collaborative fl-integrated energy consumption prediction for smart factory," *Proceedings of the Korean Institute of Communications and Information Sciences Fall Conference*, pp. 310–311, 2024.

[15] K. Elmazi, D. Elmazi, and J. Lerga, "Digital twin-driven federated learning and reinforcement learning-based offloading for energy-efficient distributed intelligence in iot networks," *Internet of things*, p. 101640, 2025.

[16] M. M. Alam, G. Mohtasin, M. R. Subhan, D.-S. Kim, and T. Jun, "Federated semi-supervised digital twin for enhanced human-machine interaction in industry 5.0," in *2024 15th International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2024, pp. 1270–1275.