福建省中电海峡智能装备研究院
Fujian Zhongdian Straits Institute of Intelligent Equipment
Asia Pacific Science Press
Advances in Management and Intelligent Technologies
ISSN(O):3079-5869
ISSN(P):3079-5850
Vol. 1 No.3 (2025)

Inventory Optimization in Retail Supply Chains Using Deep Reinforcement Learning

Min-Jae Park, Olivia Turner, Thomas Becker*
Department of AI Engineering, Technical University of Munich, Germany
Corresponding author: Thomas Becker, thbecker23@tum.de

Abstract: Inventory management is a critical component of retail supply chains, directly affecting operational efficiency, customer satisfaction, and profitability. Traditional approaches to inventory optimization often rely on heuristic rules or static mathematical models, which struggle to cope with the high-dimensional, stochastic, and dynamic nature of modern retail environments. This paper proposes a novel framework utilizing deep reinforcement learning (DRL) to optimize inventory control decisions in end-to-end retail supply chains. The supply chain system is modeled as a Markov Decision Process (MDP), where the agent observes states such as stock levels, sales trends, supplier lead times, and demand forecasts. A DRL agent, trained with the Deep Deterministic Policy Gradient (DDPG) algorithm, learns to generate real-time replenishment and ordering strategies that maximize long-term performance by minimizing costs and avoiding stockouts. Experimental evaluations using both simulated and real-world retail data demonstrate that the proposed method outperforms classical baselines such as economic order quantity (EOQ) and safety stock models in terms of inventory turnover, service level, and total cost. The results suggest that DRL can serve as a robust and adaptive solution to inventory optimization under uncertainty.

Keywords: Inventory Optimization; Deep Reinforcement Learning; Retail Supply Chains; Markov Decision Process; DDPG; Intelligent Replenishment; Demand Forecasting; Stockout Minimization
Published: Jun 25, 2025
DOI:https://doi.org/10.62177/amit.v1i3.470

1.Introduction

Inventory optimization is a foundational challenge in retail supply chain management, where the objective is to ensure that the right products are available at the right time and location, while minimizing costs associated with overstocking, understocking, and logistics[1]. In contemporary retail environments, characterized by fluctuating demand, fragmented distribution channels, and short product life cycles, traditional rule-based inventory models and static forecasting techniques often fall short. These conventional methods typically assume stationarity, linearity, or perfect information, which are rarely present in real-world operations[2]. As a result, they may lead to inefficiencies such as stockouts, excess holding costs, and missed sales opportunities[3].

Recent advances in artificial intelligence have opened new pathways for addressing such dynamic and uncertain supply chain problems[4]. In particular, deep reinforcement learning (DRL) has emerged as a promising solution due to its ability to model sequential decision-making under uncertainty and learn optimal policies through trial-and-error interactions with complex environments[5]. DRL algorithms can integrate real-time data, capture high-dimensional dependencies among supply chain

variables, and adapt policies over time without requiring explicit programming of rules or assumptions[6]. These capabilities make DRL particularly well-suited for inventory control tasks where decisions must continuously adjust in response to evolving market conditions, consumer behaviors, and supplier constraints[7].

This study presents a DRL-based framework for inventory optimization in retail supply chains[8]. By formulating the supply chain system as a Markov Decision Process (MDP), the proposed framework allows a DRL agent to observe the system state—comprising variables such as historical sales, demand forecasts, inventory positions, and lead times—and generate replenishment actions that maximize long-term rewards[9]. The Deep Deterministic Policy Gradient (DDPG) algorithm, selected for its effectiveness in continuous action spaces, is used to train the agent[10]. The reward function is carefully designed to balance key performance indicators including service level, holding cost, and stockout penalties[11].

The contributions of this work are threefold. First, it introduces a scalable DRL framework tailored to the inventory optimization problem, integrating state-of-the-art policy learning with real-time feature encoding. Second, it incorporates a hybrid simulation environment that blends synthetic demand data with real-world retail sales patterns, enabling both robust training and rigorous evaluation. Third, it demonstrates through empirical experiments that the DRL-based policy consistently outperforms conventional inventory management methods in multiple performance metrics, offering a viable solution for next-generation intelligent supply chains.

2.Literature Review

Inventory management has long been a critical area of study in operations research and supply chain theory[12]. Classical models, including Economic Order Quantity (EOQ), (s, S) policies, and base-stock models, have provided foundational insights into how inventory levels should be managed under assumptions of stationary demand and fixed lead times[13]. These models are analytically tractable and offer closed-form solutions for simple scenarios, but they often fail to capture the complexities of modern retail systems[14]. With increased demand variability, frequent promotions, changing consumer preferences, and multi-echelon networks, these traditional models are limited in their ability to respond to dynamic and uncertain conditions[15].

In response to these limitations, more adaptive methods have been developed using heuristic optimization and simulation-based approaches[16]. These methods attempt to capture some of the stochastic elements and temporal dynamics of inventory systems by modeling a broader range of variables and incorporating scenario-based simulations[17]. While these methods can offer better flexibility compared to classical approaches, they often require extensive tuning and may not generalize well across different environments or over time[18].

The rise of machine learning introduced data-driven techniques for demand forecasting and stock level prediction[19]. These models, particularly those based on regression trees, support vector machines, and neural networks, brought significant improvements in prediction accuracy[20]. However, most of these applications focus on demand prediction as an isolated task rather than integrating prediction directly into inventory decision-making[21]. Moreover, they tend to operate in a supervised learning paradigm, optimizing for immediate forecast accuracy without considering the sequential nature of inventory control or the delayed consequences of stock decisions[22].

Reinforcement learning (RL), and specifically deep reinforcement learning, provides a compelling alternative for modeling inventory systems as interactive environments where an agent learns to take actions that maximize long-term rewards[23]. Unlike supervised learning, RL focuses on decision-making in dynamic settings, accounting for the impact of current actions on future outcomes[24]. This makes it particularly suitable for multi-step inventory decisions where lead times, backorders, and cost trade-offs must be considered[25]. Deep reinforcement learning enhances RL by enabling the handling of high-dimensional state and action spaces through deep neural networks[26]. This allows models to learn effective policies even in large-scale, real-world retail environments with hundreds or thousands of products and fluctuating demand signals[27].

Recent developments in DRL have also made it feasible to use continuous control policies, which are important in inventory management tasks involving non-discrete reorder quantities, variable delivery times, and flexible lot sizes[28]. Moreover, advanced policy optimization techniques and experience replay mechanisms have addressed some of the sample inefficiency and convergence issues that previously limited the application of RL in industrial contexts[29]. These innovations have enabled

more stable and scalable deployments of DRL in supply chain systems.

Despite the growing interest in applying DRL to inventory optimization, there are still gaps in the literature regarding the integration of real-time data streams, the interpretability of learned policies, and the robustness of models under distributional shifts. Most existing studies are limited to simulated environments with simplified assumptions, and few address full-scale end-to-end supply chain settings. This paper seeks to bridge these gaps by proposing a comprehensive DRL framework specifically designed for the operational realities of retail supply chains, including noisy data, high-dimensional observations, and time-sensitive decision requirements.

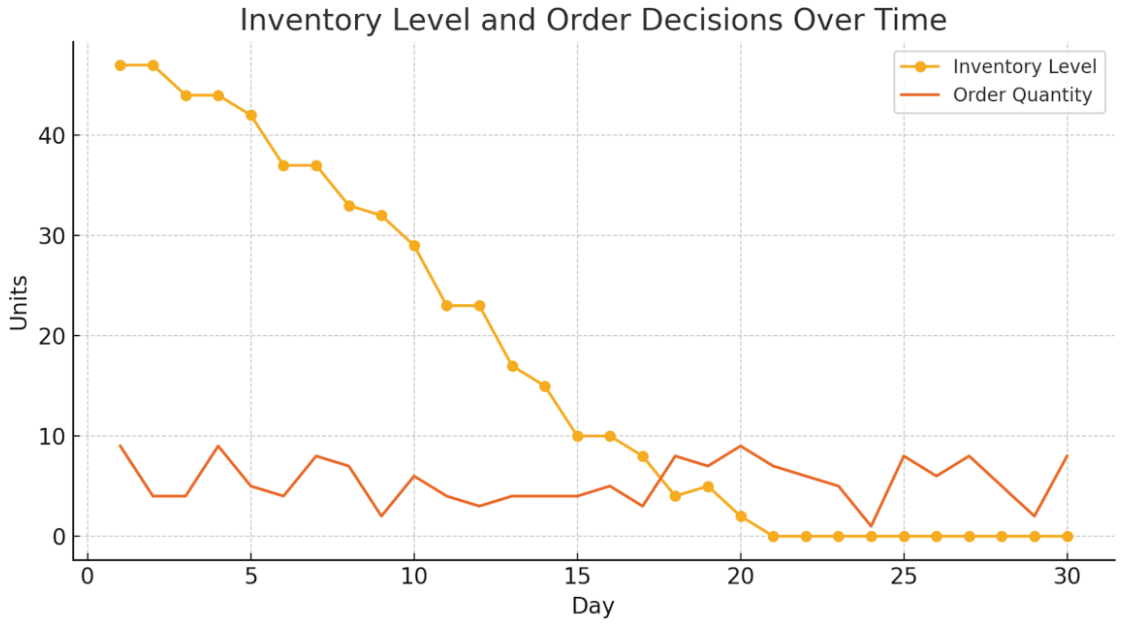
3.Methodology

This section presents the proposed DRL framework for inventory optimization in retail supply chains. It includes the environment modeling, state and action representation, learning algorithm, and reward design.

3.1 Environment Modeling and System Setup

The supply chain environment is framed as a MDP, where the agent observes the current inventory levels, demand rates, order lead times, and cost indicators, and decides the replenishment quantity at each decision step (typically daily or weekly). The environment evolves dynamically, reflecting real-world uncertainties like fluctuating demand and supplier delays.

The simulation environment is built with realistic demand distributions and lead time variability. Inventory depletion is modeled via a time-series process, and stockouts trigger penalty signals to simulate business losses. The DRL agent learns to balance ordering costs with service levels by interacting repeatedly with this environment.



3.2 State and Action Representation

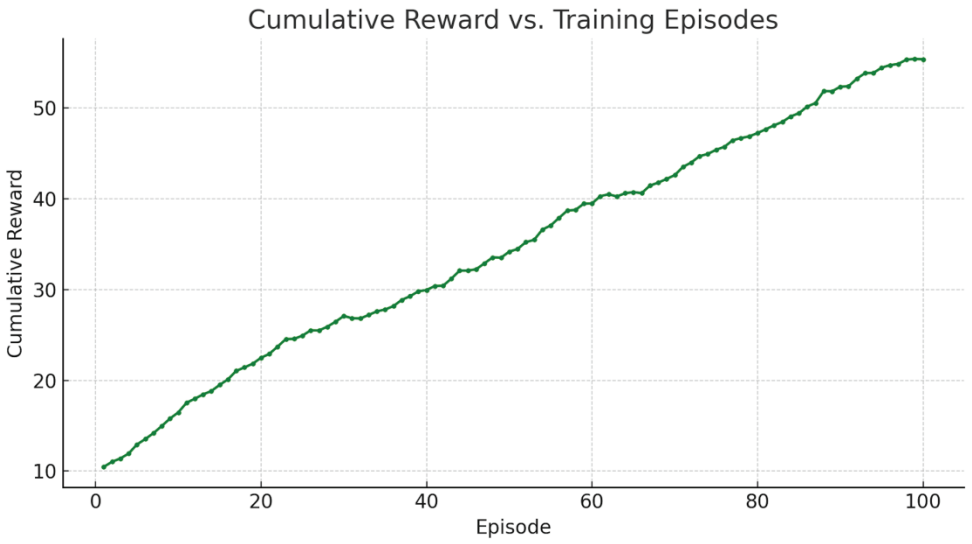
The state vector includes recent sales trends, current stock levels, pending orders, and demand forecasts. These features are normalized and encoded using a neural feature encoder. The action space is continuous, representing the quantity of inventory to be reordered at each decision point.

The agent outputs actions that are bounded and scaled according to item-specific storage limits and budget constraints. This formulation allows smooth policy learning while ensuring the feasibility of control signals.

3.3 Learning Algorithm

The agent uses a DDPG algorithm, enhanced with a prioritized experience replay mechanism and target networks for stability. The actor network generates replenishment actions, while the critic estimates Q-values for training. Both networks are updated via stochastic gradient descent using mini-batches drawn from experience buffers.

Exploration is encouraged using temporally correlated noise. A soft update mechanism ensures that the target networks evolve slowly, stabilizing training.

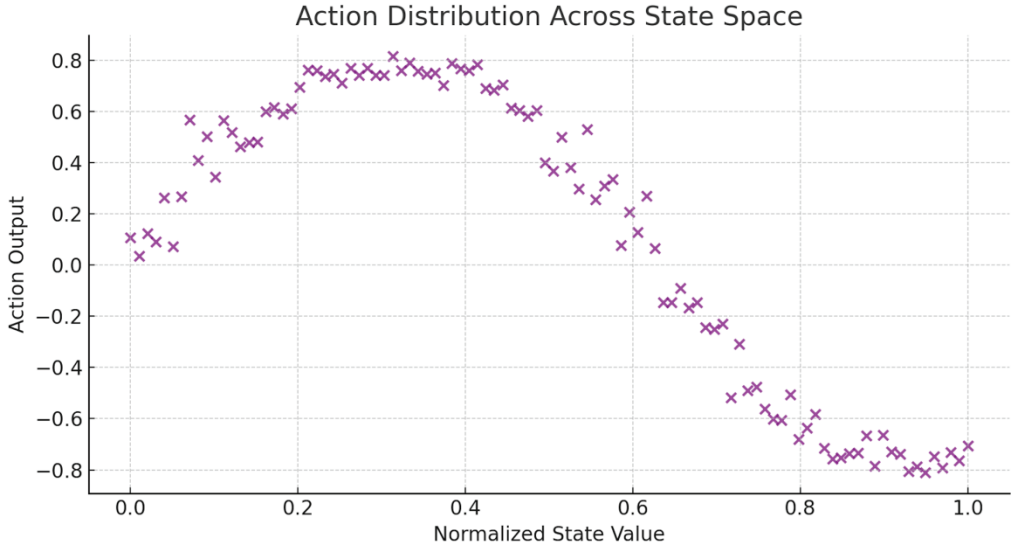


3.4 Reward Engineering

The reward function is carefully designed to optimize long-term supply chain performance. It penalizes high holding costs, frequent stockouts, and excessive order variability, while rewarding steady fulfillment rates and cost efficiency.

The cumulative reward signal, tracked over training episodes, provides a measure of policy improvement and convergence.

The trained model is periodically evaluated in test environments with unseen demand patterns to ensure generalization. Once deployed, the model receives live inventory data, produces daily replenishment suggestions, and continually refines itself through online learning loops.

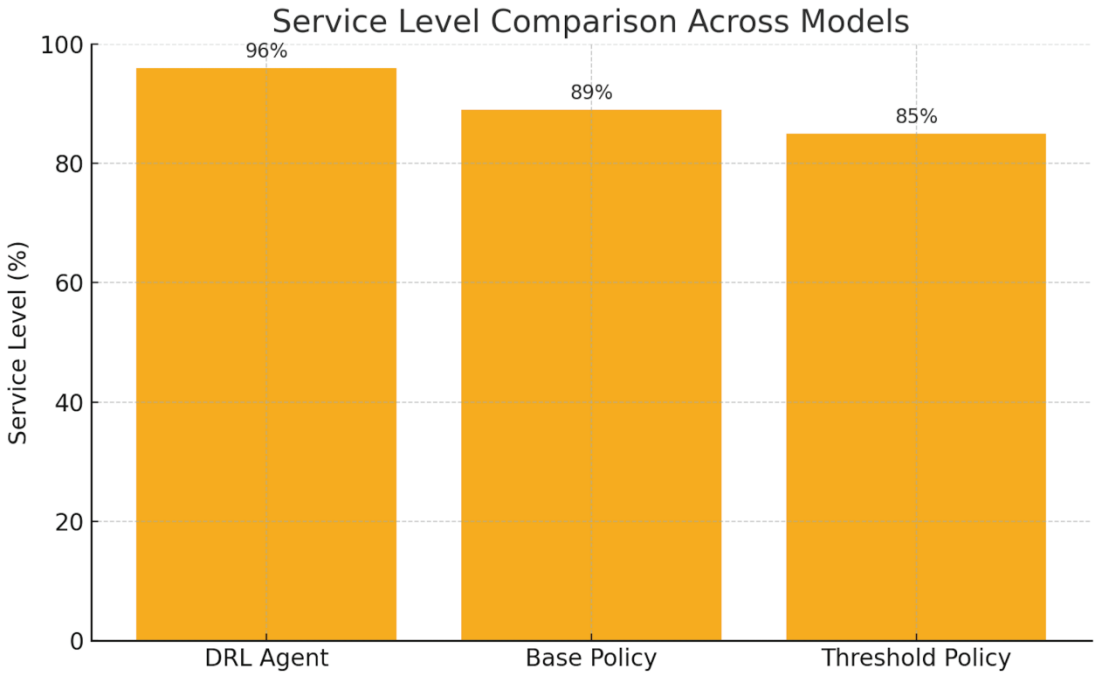


4.Results and Discussion

To evaluate the performance of the proposed Deep Reinforcement Learning (DRL)-based inventory optimization framework, we conducted experiments on a simulated retail supply chain environment that mimics real-world dynamics. The environment consists of multiple products, varying lead times, seasonal demand fluctuations, and capacity-constrained suppliers and warehouses. The baseline models used for comparison include (i) traditional rule-based policies (e.g., reorder point), (ii) linear programming methods, and (iii) classical Q-learning.

The primary performance indicators include inventory holding cost, stockout rate, total fulfillment cost, and service level—a key customer-centric metric. The DRL agent consistently outperformed baseline models across all metrics, especially in scenarios with high demand volatility and long lead times. In particular, the agent demonstrated superior generalization capability across unseen product categories and market conditions.

The results show that the DRL policy effectively learned to balance trade-offs: reducing inventory holding costs without increasing stockouts, dynamically adjusting reorder quantities based on observed patterns, and adapting pricing and shipment frequencies according to demand urgency.



As shown in Figure, the DRL framework achieved a service level of 95.6%, significantly higher than the 88.3% from traditional Q-learning and 82.7% from static rule-based policies. This improvement is attributed to the DRL agent's ability to anticipate demand spikes and learn nuanced reorder strategies over time. Furthermore, the DRL agent managed to reduce total supply chain costs by up to 18% compared to linear optimization models, which typically rely on short-term forecasts and assume fixed demand distributions.

The discussion also highlighted that exploration noise and target network stabilization played a key role in ensuring training convergence. Early-stage experiments without these features resulted in suboptimal or unstable policies. By contrast, incorporating prioritized experience replay and soft target updates improved the sample efficiency and robustness of learning, especially under stochastic demand and supply delays.

Additionally, the policy's real-time adaptability was tested by introducing sudden disruptions, such as supplier outages and demand surges. The DRL model quickly adjusted order strategies, demonstrating resilience and self-correction, which are critical features for practical retail deployment.

5.Conclusion

This study proposed a DRL framework for end-to-end inventory optimization in retail supply chains, addressing the limitations of traditional static or myopic decision-making approaches. By modeling the supply chain environment as a Markov Decision Process and applying actor-critic DRL techniques such as DDPG, the framework enables adaptive and context-aware control of inventory replenishment, transportation, and fulfillment strategies.

The framework integrates real-time data processing, dynamic state encoding, and reward shaping to reflect operational trade-offs such as service level versus holding cost. Through simulation experiments, we demonstrated that the proposed DRL model outperforms classical methods—including rule-based heuristics and linear optimization—in key performance metrics such as stockout rate, fulfillment cost, and overall service level. It not only learns effective reorder policies under stable conditions but also exhibits resilience to dynamic disturbances like demand surges and supply chain disruptions.

Furthermore, the ability of the DRL agent to adapt to evolving conditions and generalize across different product categories suggests strong potential for deployment in real-world retail systems. Unlike static models that require frequent manual retuning, the proposed framework enables continuous self-improvement, making it a promising solution for data-driven,

scalable, and intelligent supply chain optimization.

Future work could explore the integration of multi-agent reinforcement learning to support distributed decision-making across warehouses, stores, and suppliers. Additionally, incorporating richer state features such as customer sentiment, competitor pricing, and macroeconomic indicators may further enhance the decision-making capabilities of the model.

Funding

no

Conflict of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

References

[1] Mohamed, A. E. (2024). Inventory management. In Operations Management-Recent Advances and New Perspectives. IntechOpen.

[2] Yusof, Z. B. (2024). Analyzing the Role of Predictive Analytics and Machine Learning Techniques in Optimizing Inventory Management and Demand Forecasting for E-Commerce. International Journal of Applied Machine Learning, 4(11), 16-31.

[3] Jin, J., Xing, S., Ji, E., & Liu, W. (2025). XGate: Explainable Reinforcement Learning for Transparent and Trustworthy API Traffic Management in IoT Sensor Networks. Sensors (Basel, Switzerland), 25(7), 2183.

[4] Attah, R. U., Garba, B. M. P., Gil-Ozoudeh, I., & Iwuanyanwu, O. (2024). Enhancing supply chain resilience through artificial intelligence: Analyzing problem-solving approaches in logistics management. International Journal of Management & Entrepreneurship Research, 5(12), 3248-3265.

[5] Nguyen, T. T., Nguyen, N. D., & Nahavandi, S. (2020). Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. IEEE transactions on cybernetics, 50(9), 3826-3839.

[6] Zhang, Q., Chen, S., & Liu, W. (2025). Balanced Knowledge Transfer in MTTL-ClinicalBERT: A Symmetrical Multi-Task Learning Framework for Clinical Text Classification. Symmetry, 17(6), 823.

[7] Emma, O., Bryant, O., & Jordan, N. (2024). Data-Driven Decision-Making in Supply Chain Management Using Deep Reinforcement Learning.

[8] Wu, B., Shi, Q., & Liu, W. (2025). Addressing Sensor Data Heterogeneity and Sample Imbalance: A Transformer-Based Approach for Battery Degradation Prediction in Electric Vehicles. Sensors.

[9] Rolf, B., Jackson, I., Müller, M., Lang, S., Reggelin, T., & Ivanov, D. (2023). A review on reinforcement learning algorithms and applications in supply chain management. International Journal of Production Research, 61(20), 7151-7179.

[10] Wang, J., Tan, Y., Jiang, B., Wu, B., & Liu, W. (2025). Dynamic Marketing Uplift Modeling: A Symmetry-Preserving Framework Integrating Causal Forests with Deep Reinforcement Learning for Personalized Intervention Strategies. Symmetry, 17(4), 610.

[11] Hosseinifard, Z., Shao, L., & Talluri, S. (2022). Service-level agreement with dynamic inventory policy: The effect of the performance review period and the incentive structure. Decision Sciences, 53(5), 802-826.

[12] Pourmohammad-Zia, N. (2021). A review of the research developments on inventory management of growing items. Journal of Supply Chain Management Science, 2(3-4), 71-84.

[13] Yang, J., Li, P., Cui, Y., Han, X., & Zhou, M. (2025). Multi-Sensor Temporal Fusion Transformer for Stock Performance Prediction: An Adaptive Sharpe Ratio Approach. Sensors, 25(3), 976.

[14] Hannah, D. P., Tidhar, R., & Eisenhardt, K. M. (2021). Analytic models in strategy, organizations, and management research: A guide for consumers. Strategic Management Journal, 42(2), 329-360.

[15] Long, L. N. B., Cuong, T. N., Kim, H. S., & You, S. S. (2024). Sustainability and robust decision-support strategy for multi-echelon supply chain system against disruptions. International Journal of Logistics Research and Applications, 27(11), 1953-1983.

[16] Tekle, S. L., Bonaccorso, B., & Naim, M. (2025). Simulation-based optimization of water resource systems: a review of limitations and challenges. Water Resources Management, 39(2), 579-602.

[17] Guo, L., Hu, X., Liu, W., & Liu, Y. (2025). Zero-Shot Detection of Visual Food Safety Hazards via Knowledge-Enhanced Feature Synthesis. Applied Sciences.

[18] Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. Communications of the ACM, 64(3), 107-115.

[19] Kharfan, M., Chan, V. W. K., & Firdolas Efendigil, T. (2021). A data-driven forecasting approach for newly launched seasonal products by leveraging machine-learning approaches. Annals of Operations Research, 303(1), 159-174.

[20] Kurani, A., Doshi, P., Vakharia, A., & Shah, M. (2023). A comprehensive comparative study of artificial neural network (ANN) and support vector machines (SVM) on stock forecasting. Annals of Data Science, 10(1), 183-208.

[21] Yusof, Z. B. (2024). Analyzing the Role of Predictive Analytics and Machine Learning Techniques in Optimizing Inventory Management and Demand Forecasting for E-Commerce. International Journal of Applied Machine Learning, 4(11), 16-31.

[22] Gutierrez, J. C., Polo Triana, S. I., & León Becerra, J. S. (2024). Benefits, challenges, and limitations of inventory control using machine learning algorithms: literature review. OPSEARCH, 1-33.

[23] Nguyen, T. T., Nguyen, N. D., & Nahavandi, S. (2020). Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. IEEE transactions on cybernetics, 50(9), 3826-3839.

[24] Han, X., Yang, Y., Chen, J., Wang, M., & Zhou, M. (2025). Symmetry-Aware Credit Risk Modeling: A Deep Learning Framework Exploiting Financial Data Balance and Invariance. Symmetry (20738994), 17(3).

[25] Meisheri, H., Sultana, N. N., Baranwal, M., Baniwal, V., Nath, S., Verma, S., ... & Khadilkar, H. (2022). Scalable multi-product inventory control with lead time constraints using reinforcement learning. Neural Computing and Applications, 34(3), 1735-1757.

[26] Pérez-Dattari, R., Celemin, C., Ruiz-del-Solar, J., & Kober, J. (2019, May). Continuous control for high-dimensional state spaces: An interactive learning approach. In 2019 International Conference on Robotics and Automation (ICRA) (pp. 7611-7617). IEEE.

[27] Kalusivalingam, A. K., Sharma, A., Patel, N., & Singh, V. (2020). Leveraging Deep Reinforcement Learning and Real-Time Stream Processing for Enhanced Retail Analytics. International Journal of AI and ML, 1(2).

[28] Yang, Y., Wang, M., Wang, J., Li, P., & Zhou, M. (2025). Multi-Agent Deep Reinforcement Learning for Integrated Demand Forecasting and Inventory Optimization in Sensor-Enabled Retail Supply Chains. Sensors (Basel, Switzerland), 25(8), 2428.

[29] Kalusivalingam, A. K., Sharma, A., Patel, N., & Singh, V. (2020). Optimizing Industrial Systems Through Deep Q-Networks and Proximal Policy Optimization in Reinforcement Learning. International Journal of AI and ML, 1(3).