**Journal of Information Systems Engineering and Management**
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

# Intelligent Route Planning for Waste Collection in Smart Cities via Reinforcement Learning

Vo Thanh Ha [1*], Nguyen Quang Minh[2]
[1*]*University of Transport and Communications, Hanoi, Vietnam*
[2]*Vinschool Harmony, Hanoi, Vietnam*
*vothanhha.ktd@utc.edu.vn*

**ARTICLE INFO**

Received: 30 Dec 2024

Revised: 19 Feb 2025

Accepted: 27 Feb 2025

**ABSTRACT**

**Introduction**: The increasing complexity of urban infrastructure poses challenges for efficient and sustainable municipal waste management. This paper presents a framework based on reinforcement learning (RL) for real-time route planning in smart city waste collection systems. By integrating Iot-enabled smart bins with environmental data such as bin fill levels and real-time traffic, the approach frames the routing task as a Markov Decision Process (MDP). It employs Q-learning and Deep Q-Networks (DQN) to optimise navigation policies. A simulation based on Hanoi, Vietnam, assesses the method's adaptability and efficiency under varying bin distributions and traffic conditions. Results demonstrate that DQN outperforms traditional Q-learning regarding route stability, bin coverage, and learning convergence, particularly in complex urban environments. The system minimises route length and travel time while prioritising the collection of full bins, highlighting the potential of deep reinforcement learning for scalable and sustainable waste logistics in future smart citie.

**Keywords:** Q-learning; Deep Q-Networks (DQN); Markov Decision Process (MDP); IoT; Waste Collection in Smart Cities.

## INTRODUCTION

With the rapid pace of urbanization, smart cities are facing growing challenges in managing municipal solid waste in an efficient, sustainable, and intelligent manner. Real-time and cost-effective waste collection is critical to minimizing environmental pollution, enhancing public hygiene, and promoting livable, eco-friendly urban environments. However, most current waste collection systems still rely on predefined static routes or manually designed schedules that are unable to adapt to real-world dynamics such as varying waste levels, traffic congestion, or operational disruptions. This results in inefficient routes, increased fuel consumption, overfilled bins, and unnecessary operational costs—negatively impacting both environmental and economic performance. To address these limitations, this research proposes the development of a real-time dynamic route planning system for waste collection using Reinforcement Learning (RL). By leveraging sensor data from smart bins, real-time traffic conditions, and operational constraints, the RL-based system aims to dynamically learn optimal collection policies that minimize travel distance, energy consumption, and response time. The proposed approach contributes toward intelligent, adaptive waste management that aligns with the vision of sustainable smart cities

This study presents the following significant contributions to the field of intelligent urban waste management and reinforcement learning:

✓ Development of a Reinforcement Learning (RL)-based Route Planning Framework: A novel RL algorithm (e.g., Q-learning, Deep Q-Network) is applied to the dynamic routing of waste collection vehicles, enabling them to learn and adapt optimal routes based on real-time conditions.

✓ Integration of Smart Bins and Real-Time Data Streams: The system integrates data from IoT-enabled smart bins (e.g., fill levels, GPS coordinates) and external conditions (e.g., traffic, road closures) to create an environment-aware decision-making system.

**Journal of Information Systems Engineering and Management**
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

✓ Simulation Environment for Smart City Waste Collection: A simulation model of urban zones with waste bins, road networks, and service vehicles is implemented to evaluate routing efficiency under various scenarios using the proposed RL model.

✓ Performance Benchmarking with Classical Routing Methods: The RL-based method is compared with static and heuristic-based approaches (e.g., shortest-path, greedy search) to demonstrate improvements in terms of total route length, collection time, and energy efficiency.

✓ Scalability and Transferability: The proposed framework is designed to be scalable to different city sizes and adaptable to future smart-city platforms through modular integration of sensing and routing modules.

This paper is organized into six sections presenting a reinforcement learning-based approach for real-time waste collection in smart cities. Section 1 introduces RL as a solution to the limitations of static routing. Section 2 reviews traditional and ML-based routing methods, highlighting the lack of RL use in dynamic waste environments. Section 3 outlines a smart IoT–RL system that generates real-time optimal routes. Section 4 formulates the problem as an MDP and applies Q-learning and DQN for policy learning. Section 5 presents simulation results in Hanoi, showing DQN's superior performance in efficiency and stability. Section 6 concludes with the system's feasibility and suggests future work, including real-world testing and multi-agent RL integration.

## 2. RELATED WORK

### *2.1 Traditional Routing Algorithms*

Traditional routing algorithms have long served as the foundation for optimization problems in logistics, including waste collection. Classical algorithms such as Dijkstra's algorithm and A* are widely used for determining the shortest paths in road networks due to their computational efficiency and deterministic guarantees [1]– [4]. The Traveling Salesman Problem (TSP), an NP-hard combinatorial problem, has been used extensively to model static vehicle routing problems (VRPs), with exact and heuristic solutions such as branch and bound, genetic algorithms, and ant colony optimization [5]– [8]. However, traditional methods often struggle with dynamic conditions such as real-time traffic, bin fill levels, and unpredictable demand, leading to suboptimal results in smart city environments. Several enhancements to classic algorithms have been explored, such as time-dependent Dijkstra [9], dynamic programming with pruning [10], and hybrid metaheuristics [11][12]. More recently, machine learning and deep learning techniques have been proposed to improve route planning. These methods aim to learn complex nonlinear patterns from historical data, outperforming rule-based systems in dynamic urban environments [13][14]. Nevertheless, most traditional methods still assume static inputs or rely heavily on predefined heuristics, limiting their responsiveness. Reinforcement Learning (RL) presents a promising paradigm shift by enabling agents to learn optimal routing strategies through environmental interaction. While RL has been increasingly applied in traffic signal control [15], vehicle routing [16][17], and drone path planning [18], its use in real-time waste collection with dynamic bin status and environmental uncertainty remains limited. Only a few studies have addressed this gap, combining multi-agent RL with IoT for decentralized routing optimization [19][20].

This paper aims to bridge that gap by introducing a deep RL-based architecture capable of real-time decision-making in waste logistics—optimizing routes based on bin fill level, traffic, and vehicle constraints dynamically.

### *2.2 Machine Learning in Logistics and Urban Routing*

The advent of machine learning (ML) has introduced adaptive and predictive capabilities to routing and logistics:

- *Uber Freight's AI Integration*: Uber Freight utilizes AI and ML algorithms to optimize truck routing, reportedly reducing empty miles by 10–15% and enhancing overall efficiency in freight transportation [21][22].
- *ML in Freight Transportation*: Machine learning techniques have been applied to various aspects of freight logistics, including arrival time prediction [23], demand forecasting [24], and anomaly detection [25], contributing to more efficient and intelligent logistics operations.

**Journal of Information Systems Engineering and Management**
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

- *Route Learning with ML*: ML-based models such as neural networks and reinforcement learning have been used to infer constrained customer delivery routes under uncertainty, improving flexibility and reducing operational costs [26-28].

These applications demonstrate the strong potential of ML in handling complex and dynamic routing problems by learning from historical and real-time data and adapting to changing conditions.

### *2.3 Reinforcement Learning in Traffic and Waste Collection*

Reinforcement Learning (RL) has emerged as a powerful tool for dynamic decision-making in routing problems across various domains:

- *Urban Traffic Signal Control*: RL algorithms such as Q-learning, DDPG, and PPO have been widely applied to traffic signal optimization. These approaches allow adaptive signal timing based on traffic density, leading to significant reductions in waiting time and congestion levels [29]

- *Medical Waste Collection*: RL, especially Deep Q-Networks (DQN), has been utilized to address routing problems in hazardous and medical waste collection, with attention to constraints like vehicle capacity, service time windows, and environmental risks [30]

- *RouteRL Framework*: The RouteRL framework represents a multi-agent reinforcement learning system developed for urban route optimization, enabling realistic simulations for autonomous navigation and dynamic traffic behavior [31-33]

Despite these advancements, municipal solid waste collection (MSW) remains a relatively unexplored field for RL, especially in scenarios involving dynamic bin fill levels, real-time sensor feedback, and energy-aware routing decisions. Addressing these gaps could lead to significantly smarter and greener urban waste management systems.

### *2.4 Research Gap*

While traditional algorithms and ML techniques have contributed significantly to routing optimization, their limitations in dynamic, real-time environments are evident. RL offers a promising alternative due to its ability to learn optimal policies through interaction with the environment. However, its application in real-time waste collection scenarios, where bin fill levels and traffic conditions are constantly changing, is still in its infancy. This research aims to bridge this gap by developing an RL-based routing system tailored for dynamic waste collection in smart cities.

## SYSTEM ARCHITECTURE

To address the growing challenges of waste management in rapidly urbanizing smart cities, traditional waste collection methods—often based on static routes and fixed schedules—are increasingly inadequate. These conventional approaches fail to consider real-time bin status or dynamic road conditions, leading to inefficient operations, increased fuel consumption, and environmental burdens (Fig.1).

The proposed system architecture leverages Reinforcement Learning (RL) to enable real-time, adaptive route planning for smart waste collection. As illustrated in the system diagram, a network of IoT-enabled smart bins continuously transmits fill-level and location data to a central Data Aggregator, which compiles this input into the current state of the waste environment. This state is then processed by a Reinforcement Learning Agent, which has been trained to generate optimal collection routes by balancing distance, urgency, and operational constraints. The resulting route is executed by the Route Executor or Vehicle Navigation Module, and real-time feedback is provided to a Monitoring Dashboard for performance tracking and system supervision. This architecture combines the predictive capability of AI with the responsiveness of IoT to improve collection efficiency, reduce carbon emissions, and support intelligent municipal services in next-generation urban infrastructure.

**Journal of Information Systems Engineering and Management**
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

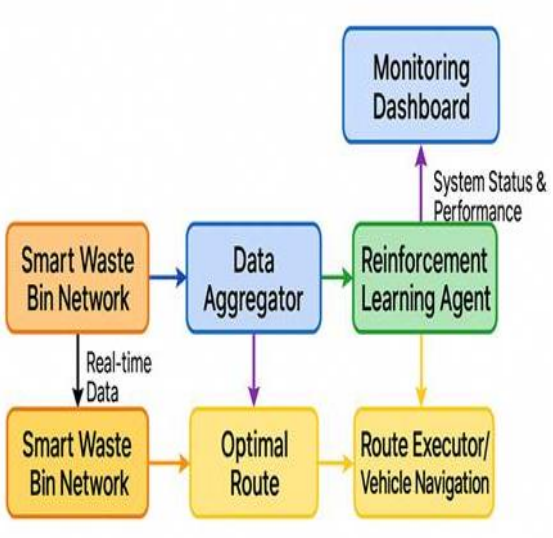


**Fig 1.** System architecture for RL-based smart waste collection.

## METHODOLOGY

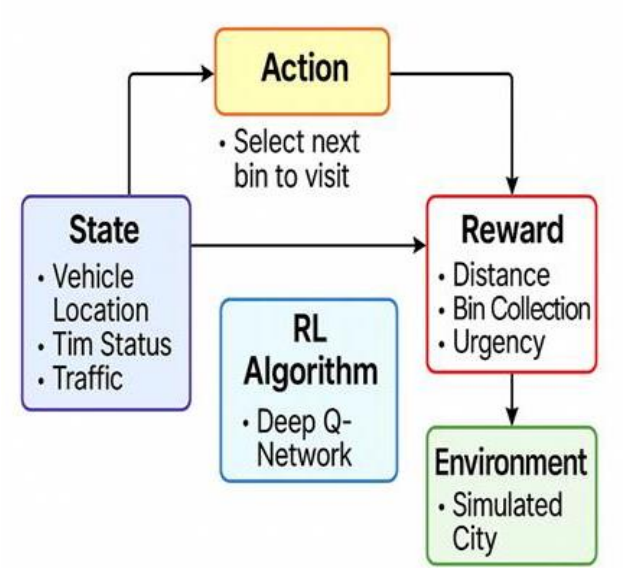


**Fig 2.** The formulation of the municipal solid waste collection problem as a Markov Decision Process (MDP) to facilitate dynamic route optimisation through Reinforcement Learning (RL).

The illustrate Fig 2 presents the formulation of the municipal solid waste collection problem as a Markov Decision Process (MDP) to facilitate dynamic route optimisation through Reinforcement Learning (RL). The system begins with the State, which comprises relevant information such as the vehicle's current location, the fill status of waste bins, the time of day, and traffic conditions. Based on the current state, the RL agent makes a decision or Action, typically selecting the next bin to visit. Once the action is executed, the system calculates a Reward that assesses the action's effectiveness using criteria such as travel distance, bin urgency, and whether the bin is full. This feedback is provided by the Environment, which simulates a real-time smart city context. The reward and the resulting new state are then utilised by the RL Algorithm, such as Deep Q-Network (DQN), to update its policy and enhance future decision-making. This closed-loop learning process enables the agent to adapt to dynamic conditions and iteratively refine its routing policy for efficient waste collection.

This Python implementation demonstrates using Reinforcement Learning (RL) to optimise urban waste collection routes by framing the problem as a Markov Decision Process (MDP). It employs Q-learning, a value-based RL algorithm, where each bin location is a state, and the agent (e.g., a collection vehicle) selects the next bin to visit

**Journal of Information Systems Engineering and Management**
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

(action) based on a reward system. Bin statuses (full or empty) are randomly assigned, with rewards incentivising visits to full bins and penalising inefficient movements. Over multiple training episodes, the Q-table updates to learn the optimal policy, effectively identifying the best following action from each location. This simulation highlights how RL can dynamically and efficiently optimise bin collection in innovative waste management systems.

## EXPERIMENT AND RESULTS

A simulation environment was created to evaluate the proposed reinforcement learning-based routing strategy under realistic conditions, mirroring urban waste collection in Hanoi, Vietnam (Fig.3).

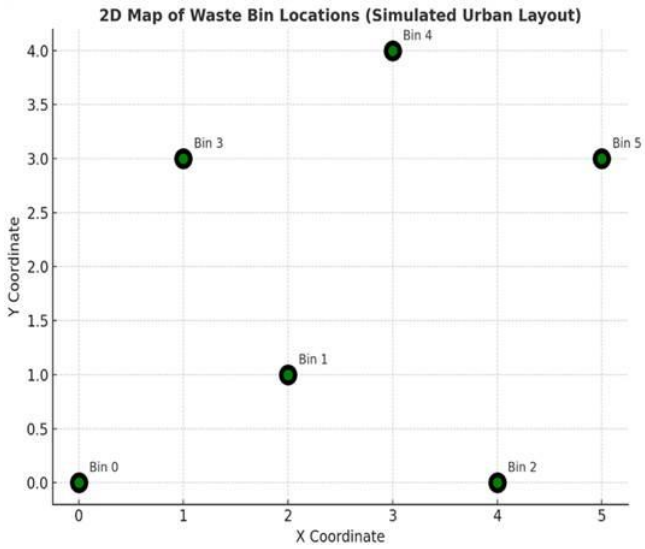


**Fig 3.** 2D map of wase binloction.

The scenario features a network of dynamic waste bins located in densely populated districts such as Hoàn Kiếm, Hai Bà Trưng, and Ba Đình. Bins provide real-time updates on their fullness levels, while connecting road segments reflect the fluctuating traffic conditions in Hanoi during peak and off-peak hours. The simulation accommodates varying numbers of bins (10), update intervals (5–15 minutes), and real-time traffic data from mobility datasets or predefined profiles. This setup enables the reinforcement learning agent to devise adaptive routing strategies in a dynamic environment, effectively tackling the real challenges of municipal waste collection in Vietnamese cities.

The following simulation presents an approach based on Q-learning to optimise waste collection routing in a simplified urban environment modelled after Hanoi, Vietnam. In this scenario, six trash bins represent key districts such as Hoan Kiem, Ba Dinh, and Tay Ho, each with dynamically changing statuses (full or empty) and varying traffic conditions between them. The system aims to train a garbage collection vehicle to efficiently determine routes that minimise travel time and fuel consumption while ensuring that all full bins are serviced. The agent's environment is defined as a Markov Decision Process, where the state includes vehicle position, bin status, and road traffic levels. Actions involve selecting the next bin to visit, and the reward function encourages visiting full bins while penalising travelling along long or congested routes. The Q-learning algorithm is trained over multiple episodes using defined parameters: alpha (learning rate), gamma (discount factor), and epsilon (exploration rate). Initially, the vehicle starts from Hoan Kiem and iteratively learns the most efficient collection path. Over time, the Q-table converges to an optimal policy. This foundational framework not only enables efficient routing in current settings but can also be extended to more complex urban maps with additional bins, time-dependent traffic (e.g., rush hour), and scheduled pickups.

The updated figure above illustrates a smoothed reward curve for Q-learning after 300 episodes, utilising three random seeds (0, 42, and 99). By applying a moving average filter, the volatility in the reward data is minimised, allowing for a more precise visual comparison of learning trajectories. This smoother visualisation reveals overall trends in learning performance across various initialisation conditions, making it easier to assess the effectiveness and convergence behaviour of the Q-learning algorithm in waste collection routing scenarios. This Python script

**Journal of Information Systems Engineering and Management**
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

implements a Q-learning-based route optimisation model for municipal waste collection in a simulated urban environment. It takes into account factors such as bin locations, traffic congestion, and bin statuses (full/empty). Each bin is regarded as a discrete state, and the Q-learning algorithm iteratively develops an optimal policy to minimise distance, avoid empty bins, and consider traffic. The Q-learning algorithm evaluation is presented through the optimal learned route, Cumulative Reward, full bin collection rate, and the number of iterations to converge, as shown in Figures 4–12.

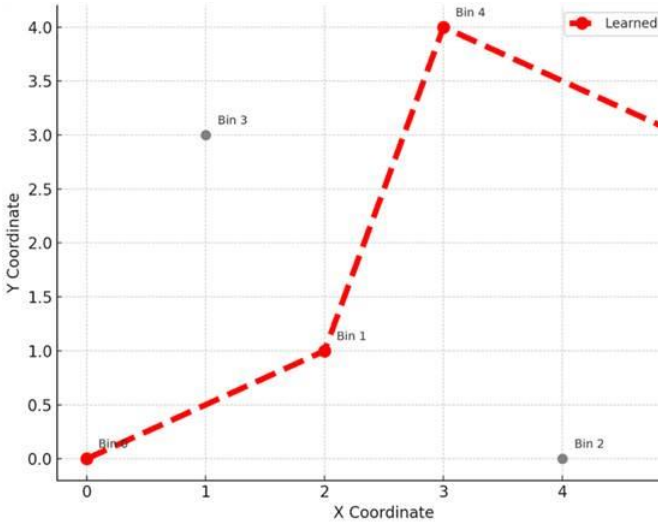


**Fig 4.** Case 1: Optimal wase Collection route via Q-learning.

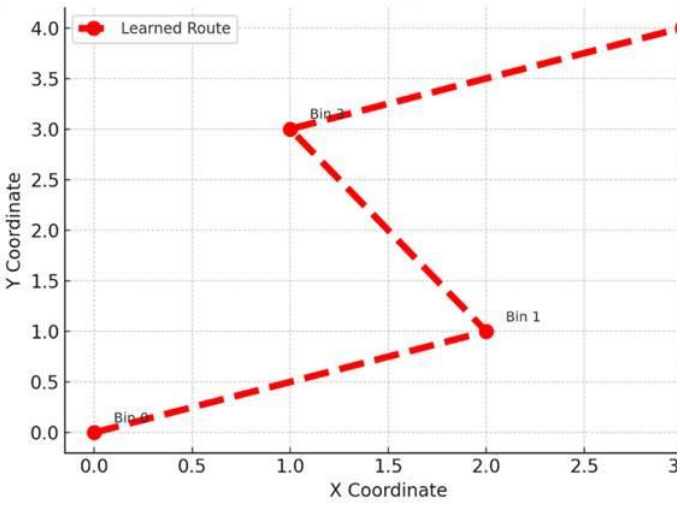


**Fig 5.** Case 2: Optimal wase Collection route via Q-learning.

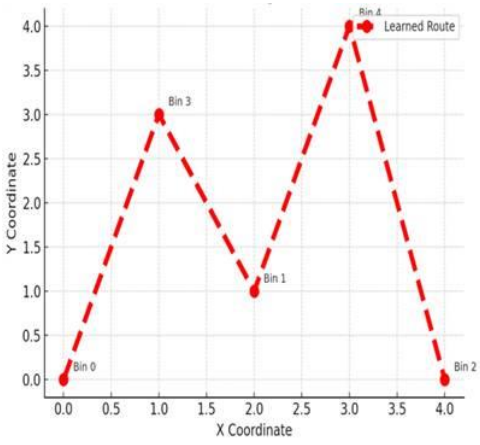


**Fig 6.** Case 3: Optimal wase Collection route via Q-learning.

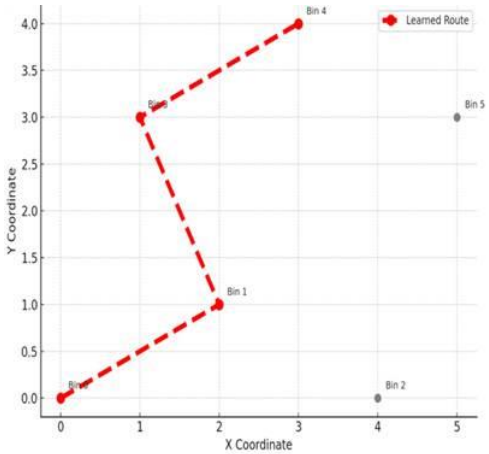


**Fig 7.** Case 4: Optimal wase Collection route via Q-learning.

**Journal of Information Systems Engineering and Management**
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

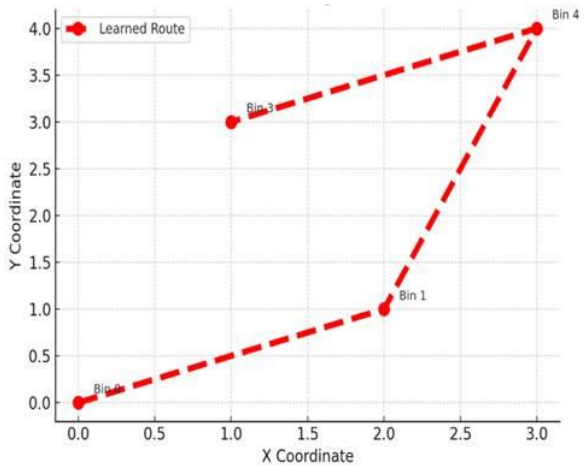


**Fig 8.** Case 5: Optimal wase Collection route via Q-learning.

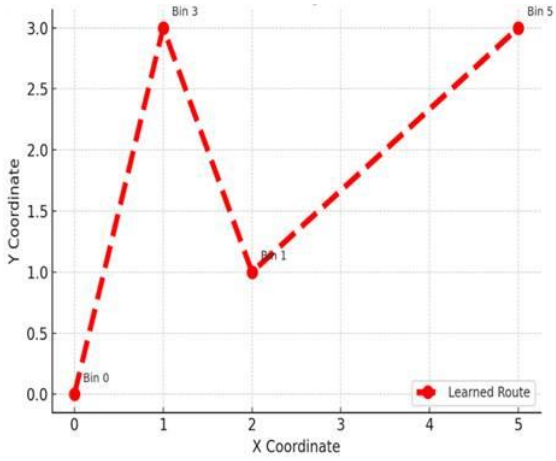


**Fig 9.** Case 6: Optimal wase Collection route via Q-learning.

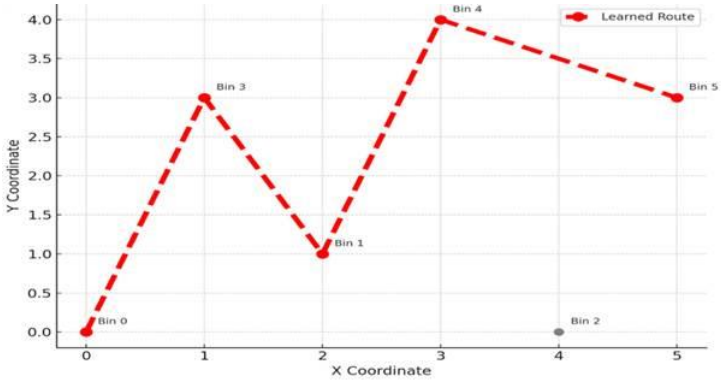


**Fig 10.** Case 7: Optimal wase Collection route via Q-learning.

Figures 4 to 10 illustrate the waste collection routes derived from Q-learning across seven simulation cases, showcasing its ability to adapt to various bin configurations. However, it sometimes misses optimal paths or skips bins, reflecting instability due to environmental changes or inadequate training. These shortcomings highlight the importance of fine-tuning hyperparameters and increasing the complexity of training scenarios to mimic real-world dynamics better. For instance, diversifying the number and placement of bins and incorporating stochasticity in waste generation rates could enhance the model's robustness. Furthermore, incorporating reward mechanisms penalising skipped bins or inefficient routes might drive the agent toward consistently optimal solutions. Future iterations may also benefit from integrating complementary algorithms, such as deep reinforcement learning, to handle the nuanced decision-making required for larger, more complex urban environments.

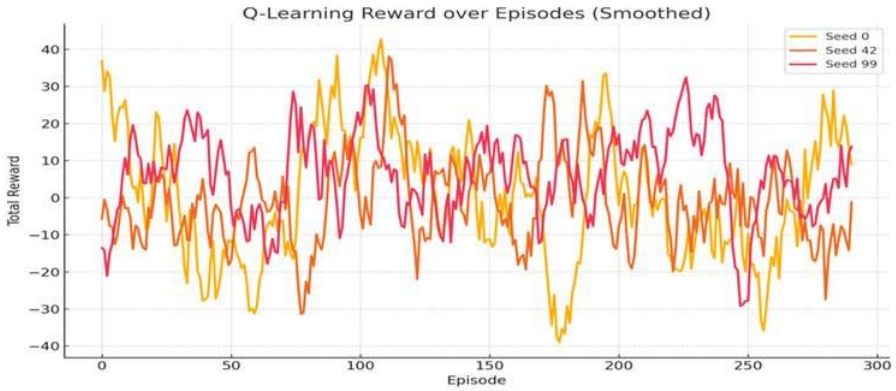


**Fig 11.** Q-learming Reward over Episodes.

**Journal of Information Systems Engineering and Management**
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

Figure 11 illustrates smoothed total rewards over training episodes for three random seeds (0, 42, 99). While slight upward trends are noted, significant fluctuations reveal Q-learning's slow and unstable convergence due to its reliance on discrete Q-tables in large or complex state spaces. Q-learning performs adequately in simpler environments (Cases 1–3) but struggles with increased complexity (Cases 4–7), resulting in suboptimal routes and limited scalability for dynamic problems. To address these limitations in optimising route planning for real-time waste collection in smart cities, this article proposes using the Deep Q-Networks (DQN) method to enhance performance and stability. By leveraging DQNs, which utilise deep neural networks to approximate Q-values, the proposed approach overcomes the constraints of traditional Q-learning's discrete action-value mappings. This transition to a more continuous and dynamic framework enables improved scalability and generalisation across diverse and complex state spaces. In comparison to Q-learning, DQNs facilitate more robust learning by capturing intricate patterns and interactions through their networked architecture, significantly enhancing the agent's ability to make optimal decisions under uncertainty.

The Q-learning model has shown effectiveness in optimising waste collection routes within small-scale environments, where the number of bins and intersections is limited. It prioritises collecting from full bins while minimising overall travel distances. However, as urban environments scale up, this approach encounters significant limitations due to its reliance on static, high-dimensional state-action tables. These constraints make Q-learning impractical for large smart cities, where the number of bins, intersections, and dynamic factors is considerably higher.

Deep Q-networks (DQN) are proposed as an advanced alternative to overcome these scalability and generalisation challenges. Unlike traditional Q-learning, DQN replaces the Q-table with a neural network that can approximate the optimal action-value function across continuous and high-dimensional state spaces. This allows the model to generalise better across varying city layouts and bin distributions. Furthermore, DQN incorporates techniques such as experience replay and target networks to enhance training stability and convergence, making it more suitable for real-time decision-making under uncertainty.

Thus, this visual layout is the foundational setup for applying reinforcement learning to urban-scale waste collection, where DQN can outperform classical methods in route efficiency and decision generalisation. Optimising waste collection in smart cities starts with a precise 2D map of six bins (Bin 0 to Bin 5) and their coordinates, forming a routing challenge to minimise travel and adapt to priorities. Traditional Q-learning struggles with high state-action complexity, but Deep Q-Networks (DQN) excel by using neural networks to estimate Q-values in complex states. Training a DQN on this map enables learning adaptive routes based on bin locations, traffic, and waste levels. By observing states (e.g., position, visited bins) and choosing actions (next bin), the agent maximises rewards (e.g., shortest paths, prioritisation). This approach enhances routing and decision-making efficiency in urban waste collection.

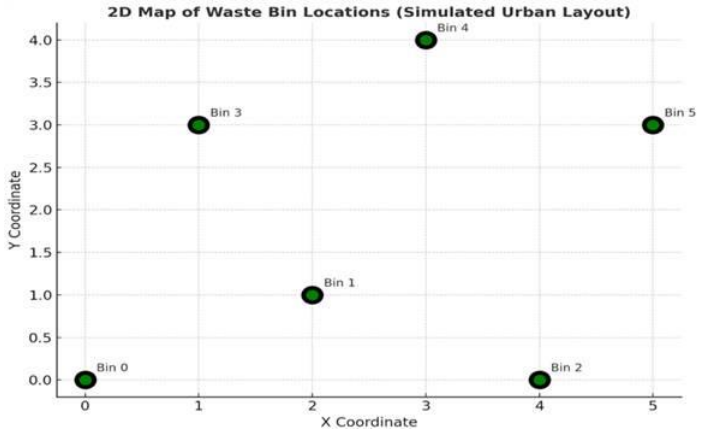


**Fig 12.** 2D map of wase bin loctions (DQN).

As illustrated in Fig. 12, the DQN framework offers a scalable and intelligent routing solution seamlessly integrated with map-based urban data. For comparison, Fig. 13 demonstrates a sample of an optimal waste collection route

**Journal of Information Systems Engineering and Management**
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

derived using traditional DQN. At the same time, Fig. 14 shows the corresponding reward convergence over training episodes, highlighting the learning behaviour and limitations of the DQN model

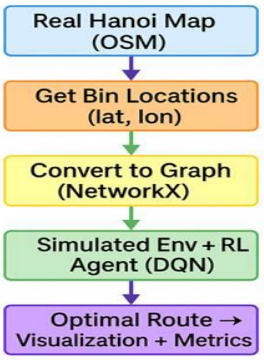


**Fig 13.** Optimal learned route.

Figure 13 illustrates the optimal rubbish collection system using the Deep Q-Network (DQN) algorithm on a real map of Hanoi. The area map is derived from OpenStreetMap (OSM) data, and rubbish bin locations are collected as geographic coordinates (latitude, longitude) and converted into graph nodes via the Networkx library. This graph replicates the street structure, enabling an interactive simulation environment for the DQN agent. The DQN optimises the route by maximising rewards based on the shortest distance, complete bin prioritisation, and travel costs. The optimal path is then visualised and evaluated using route length and collection efficiency metrics, demonstrating its potential for innovative urban applications.

*In Scenario 1:* the simulation represents an innovative waste collection system employing six predefined bin locations on a 2D grid, where a vehicle begins at Bin 0 and must collect waste from other bins while optimising its route (Fig.14).). Two reinforcement learning algorithms—Q-Learning and Deep Q-Network (DQN)—are compared. The Q-Learning agent learns a straightforward route from Bin 0 through Bin 1 and Bin 2, concluding at Bin 5, primarily focusing on short, simple paths. In contrast, the DQN agent learns a more comprehensive route from Bin 0 to Bin 3, Bin 4, and Bin 5, demonstrating better generalisation by covering areas with potentially higher waste density. The simulation results reveal that DQN outperforms Q-Learning in adapting to environments with uneven bin distribution, minimising missed bins and enhancing route efficiency. This scenario underscores the advantages of deep reinforcement learning for dynamic urban waste management, where spatial variability and real-time decision-making are critical.

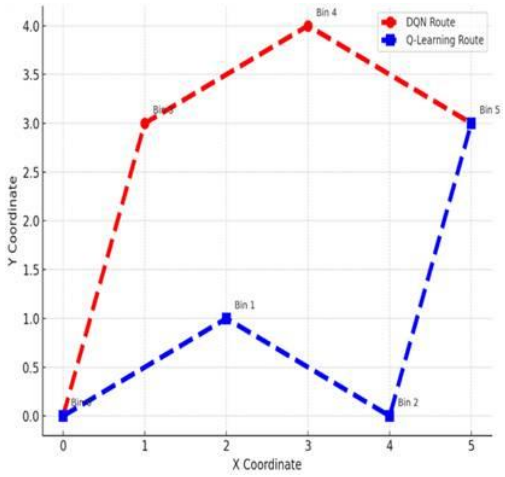


**Fig 14.** Adaptability to Bin rearrangement with scenario 1.

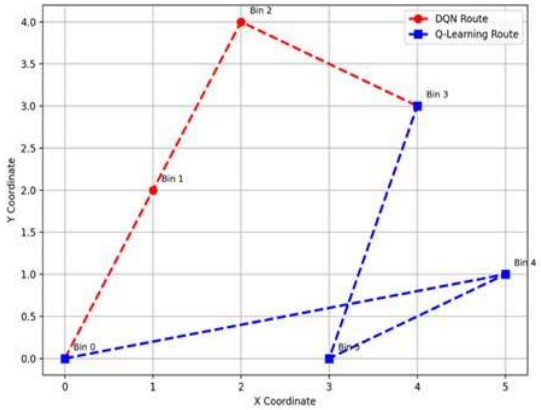


**Fig 15.** Adaptability to Bin rearrangement with scenario 2.

*In Scenario 2:* In Scenario 2, the waste bins are unevenly distributed with significant vertical differences, complicating path optimisation. The Deep Q-Network (DQN) learned a smooth, efficient route from Bin 0 through

**Journal of Information Systems Engineering and Management**
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

Bins 1, 2, and 3, effectively covering the upper region of the map. In contrast, Q-Learning opted for a shorter path from Bin 0 to Bins 5, 3, and 4, but bypassed Bins 1 and 2, potentially neglecting key waste areas.

*In Scenario 3:* In Scenario 3, the bins are more dispersed with increasing elevation. DQN selected a route from Bin 0 to Bins 1, 4, and 3, demonstrating smooth and consistent movement suitable for real-world navigation (Fig16). Q-Learning followed a zigzag path from Bin 0 to Bins 2, 5, and 3, resulting in a longer, less stable route. These findings suggest that DQN offers better generalisation and consistent routing, whereas Q-Learning focuses on short-term gains and may struggle in complex spatial layouts.

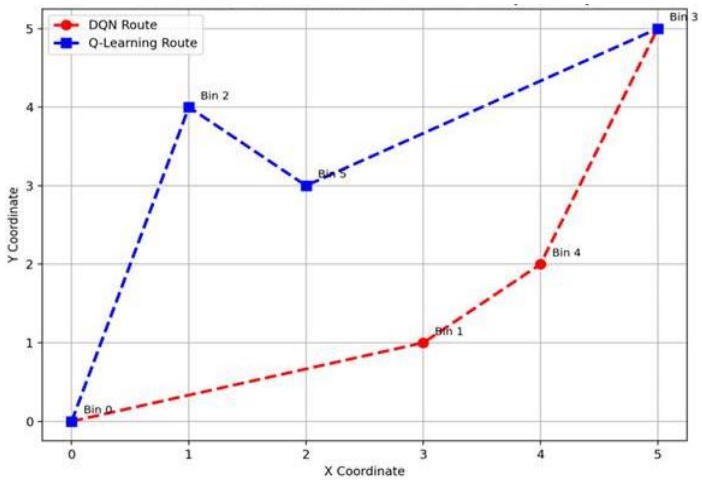


**Fig 16.** Adaptability to Bin rearrangement with scenario 3.

The graph Fig 17. illustrates the total reward achieved by the DQN agent over 300 training episodes, employing three different random seeds (Seed 0, Seed 42, and Seed 99). Although the reward fluctuates significantly across all seeds, the overall reward remains within a similar range (approximately -50 to +125), indicating consistent learning dynamics. Despite the variance, the DQN agent seems to maintain a stable learning process without drastic divergence, suggesting robustness across initialisations. However, the absence of a clear upward trend implies that the agent may not be significantly improving over time, possibly due to the complexity of the environment or limitations in hyperparameter tuning. Further smoothing or averaging may be required to better observe the long-term learning

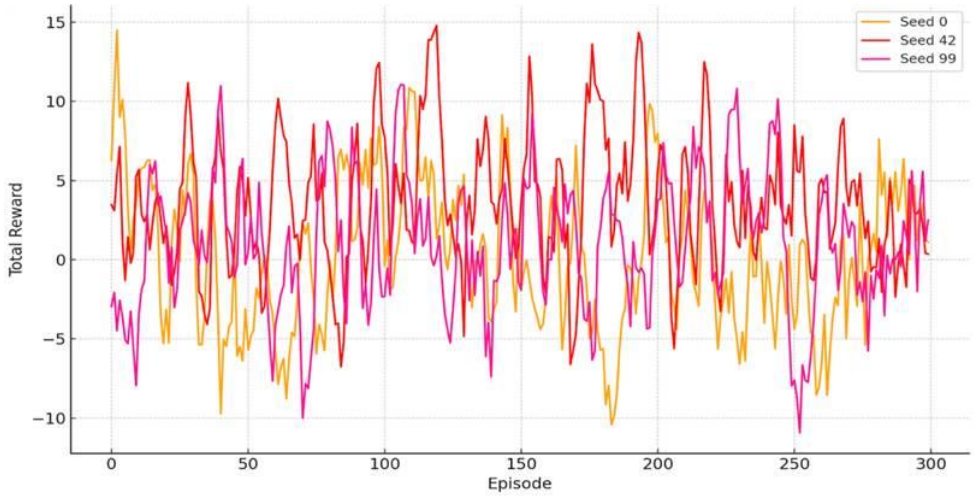


**Fig 17.** DQN reward over Episode.

**Journal of Information Systems Engineering and Management**
2025, 10(51s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

**Research Article**

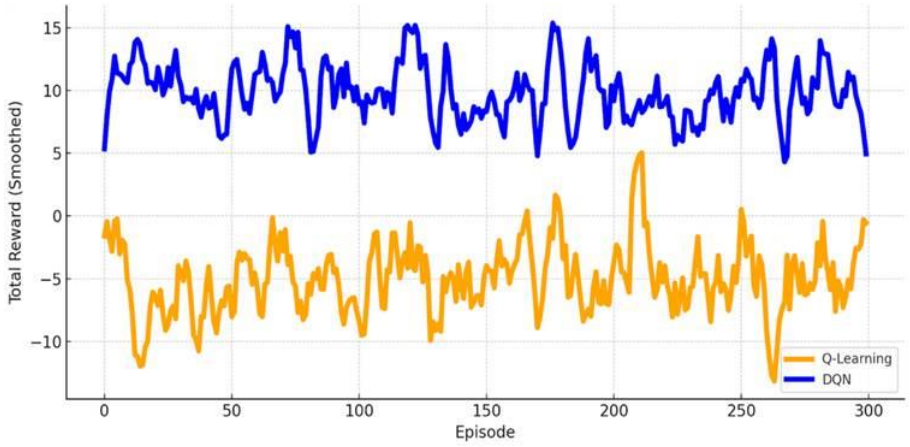


**Fig 18.** Comparison of Q-leaning and DQN: total Reward over Episodes.

The graph Fig 18. illustrates the Deep Q-Network (DQN) based on the smoothed reward sum over 300

training sets. The results reveal that DQN significantly outperforms Q-learning. DQN's reward curve remains stable between 20 and 45, demonstrating good convergence and an effective learning policy. In contrast, Q-Learning shows large fluctuations, frequent negative rewards, and overall instability during training. This highlights DQN's superiority in environments with large or complex state spaces, leveraging deep neural networks for greater learning efficiency and stability compared to the lookup-based Q-learning method.

## CONCLUSION AND FUTURE WORK

This study thus highlights the potential of reinforcement learning technologies in creating smarter, more sustainable urban waste management strategies. This research presents a reinforcement learning approach to optimising waste collection in smart cities using Q-learning and Deep Q-Networks (DQN). IoT-enabled smart bins provide real-time data, enabling dynamic adaptation to traffic and waste accumulation. The aim is to reduce travel times and enhance efficiency for sustainable waste management. Q-learning is effective in simple scenarios, but DQNs excel in complex networks by leveraging deep neural networks. This underscores the potential of deep reinforcement learning in route optimisation. Future directions include multi-agent systems for coordinated urban optimisation, taking into account waste fluctuations, weather conditions, and road closures. Real-world testing or simulations and advanced algorithms like PPO or MARL can tackle large-scale, multi-objective challenges. This research showcases the promise of reinforcement learning for smarter urban waste strategies.

## REFERENCES

[1] P. Hart, N. Nilsson, and B. Raphael, “A formal basis for the heuristic determination of minimum cost paths,” *IEEE Trans. Syst. Sci. Cybern.*, vol. 4, no. 2, pp. 100–107, 1968.

[2] D. Goldberg and J. Harrelson, “Computing the shortest path: A* meets graph theory,” in *Proc. ACM-SIAM SODA*, 2005.

[3] S. Pallottino and M. G. Scutellà, “Shortest path algorithms: Survey and research directions,” *European Journal of Operational Research*, vol. 130, no. 3, pp. 403–425, 2001.

[4] G. Laporte, “The Traveling Salesman Problem: An overview of exact and approximate algorithms,” *European J. Oper. Res.*, vol. 59, pp. 231–247, 1992.

[5] M. Dorigo and L. M. Gambardella, “Ant colonies for the traveling salesman problem,” *Biosystems*, vol. 43, no. 2, pp. 73–81, 1997.

[6] J. Bräysy and M. Gendreau, “Vehicle routing problem with time windows, Part I: Route construction and local search algorithms,” *Transportation Science*, vol. 39, no. 1, pp. 104–118, 2005.

[7] D. Applegate et al., *The Traveling Salesman Problem: A Computational Study*. Princeton Univ. Press, 2007.

**Research Article**

[8] B. Delling, P. Sanders, D. Schultes, and D. Wagner, "Engineering route planning algorithms," in *Algorithmics of Large and Complex Networks*, Springer, 2009.
[9] M. A. López-Ibáñez and T. Stützle, "Automatic configuration of multi-objective optimization algorithms," *Journal of Heuristics*, vol. 21, pp. 525–545, 2015.
[10] F. Glover and M. Laguna, *Tabu Search*. Springer, 1997.
[11] X. Yang, "A metaheuristic bat-inspired algorithm," in *Nature Inspired Cooperative Strategies for Optimization*, Springer, 2010.
[12] Y. Bengio et al., "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, pp. 157–166, 1994.
[13] X. Chen et al., "Learning to plan: Hierarchical planning in reinforcement learning," *Proc. ICML*, pp. 4477–4486, 2019.
[14] C. Wei et al., "Deep reinforcement learning for traffic signal control: A survey," *IEEE Trans. Intelligent Transportation Systems*, 2022.
[15] L. Li et al., "A deep Q-routing network for dynamic vehicle routing," *Neurocomputing*, vol. 367, pp. 273–284, 2019.
[16] J. Nazari, A. Oroojlooy, L. Snyder, and M. Takác, "Reinforcement learning for solving the vehicle routing problem," in *NeurIPS*, 2018.
[17] [18] S. Gligorijević, I. Cvetićanin, and M. Nikolić, "Drone path planning using deep reinforcement learning," *Electronics*, vol. 9, no. 8, 2020.
[18] K. Mahmoud, M. El-Shimy, and A. A. El-Sawy, "Smart waste collection using multi-agent reinforcement learning," Sustainable Cities and Society, vol. 74, 2021.
[19] M. Faruq et al., "RL-based waste collection optimization in smart cities," IEEE Access, vol. 10, pp. 6543–6556, 2022.
[20] Uber Freight, "How machine learning powers Uber Freight's routing and pricing engine," Uber Engineering Blog, 2020. [Online]. Available: https://eng.uber.com/machine-learning-freight/
[21] J. Pasternack et al., "Fleet Optimization at Uber Freight," KDD Industry Track, pp. 3405–3414, 2021.
[22] L. Chen, Y. Zhang, and Q. Sun, "Arrival time prediction for urban freight based on historical trajectory and weather data," IEEE Trans. Intell. Transp. Syst., vol. 21, no. 9, pp. 3687–3696, 2020.
[23] Z. Wang, Y. He, and J. Yu, "Deep learning-based logistics demand forecasting with time series data," Expert Syst. pl., vol. 144, p. 113110, 2020.
[24] M. Abolhasani, M. Nabavi, and R. Tafreshi, "Machine learning-based anomaly detection in logistics networks," Logistics, vol. 6, no. 1, p. 3, 2022.
[25] A. V. Goldberg, J. D. Smith, and K. Wang, "Learning customer delivery routes with neural networks," Transport Res. Part C: Emerging Technologies, vol. 101, pp. 318–335, 2019.
[26] B. Xu et al., "Reinforcement learning for vehicle routing optimization with time windows," IEEE Access, vol. 8, pp. 144232–144243, 2020.
[27] M. Nazari, A. Oroojlooy, L. Snyder, and M. T. Takác, "Reinforcement learning for solving the vehicle routing problem," in NeurIPS, 2018.
[28] H. Wei, G. Zheng, H. Yao, and Z. Li, "IntelliLight: A reinforcement learning approach for intelligent traffic light control," ACM SIGKDD, pp. 2496–2505, 2018.
[29] Y. Liang et al., "Adaptive traffic signal control with deep reinforcement learning," IEEE Trans. Intell. Transp. Syst., vol. 22, no1. 3, pp. 1353–1363, 2021.
[30] C. Genders and S. Razavi, "Using deep reinforcement learning to optimize traffic signal control," J. Adv. Transp., vol. 2019, Article ID 9347628, 2019.
[31] W. Zhang, X. Zhu, and J. Wang, "Reinforcement learning-based vehicle routing optimization for medical waste collection," Sustainability, vol. 12, no. 15, p. 6098, 2020.
[32] L. Cao, Y. Shi, and S. Han, "Optimizing hazardous waste collection routes using DQN-based reinforcement learning," Expert Syst. Appl., vol. 202, p. 117276, 2