Operational Research in Engineering Sciences: Theory and Applications
Vol. 7, Issue 2, 2024, pp. 241-272
ISSN: 2620-1607
eISSN: 2620-1747
DOI: https://doi.org/10.5281/zenodo.15086586

OPTIMIZING LAST-MILE DELIVERY BY DEEP Q-LEARNING APPROACH FOR AUTONOMOUS DRONE ROUTING IN SMART LOGISTICS

Pannee Suanpang[1*], Pitchaya Jamjuntr[2]

[1]Department of Information Technology, Faculty of Science & Technology, Suan Dusit University, Bangkok, 10300, Thailand.
[2] Electronic and Telecommunication Engineering, King Mongkut's University of Technology Thonburi, Bangkok, 10140, Thailand.

Received: 14 January 2024
Accepted: 22 May 2024
First Online: 30 June 2024

Research Paper

Abstract: *The advancement technology of artificial intelligence and e-commerce has increased and this has called for new ways to improve last-mile transportation, which is regarded as an essential part of the logistics value chain, especially in smart logistics. This paper addresses the problem of developing effective routes for autonomous drones in last-mile logistics using deep Q-learning. This paper aims to improve the process of delivery by utilizing the flexibility and intelligence of self-driven autonomous drones in smart logistics transportation. The key challenge for the effective provision of last-mile delivery services remains the decision on the routing of many aerial drones in an indoor urban environment, concerning the restrictions of a time window for delivery, energy consumption and traffic. This paper implements a deep Q-learning paradigm that allows drones to relearn their flight paths and delivery strategy during the lifecycle, thereby reducing the cost in the long run while using the costing strategies as part of the reengineering process. The approach has been validated through extensive experimentation and simulations. Results obtained indicate that the delivery drones modified for the study attained the designed requirements of deep Q-learning, including optimal navigation and performance that attained 12.8% shorter delivery time, an increase in energy efficiency by 8.4%, and a route quality improvement of 20.1%. Furthermore, highlights the performance of the system in various situations where deep Q-learning and standard routing approaches are compared. This paper not only aids in the minimization of the last-mile delivery constraint by the use of shipping drones but also emphasizes the capacities of reinforcement learning strategies such as deep Q-learning in tackling the routing problems in smart logistics systems. At last, it advocates carrying on deeper into the application of reinforcement learning in the solving of complex optimization problems in various other fields.*

*Corresponding author: pannee_sua@ dusit.ac.th. (P. Suanpang)
pitchaya.jam@kmutt.ac.th (P. Jamjuntr)

Keywords: *Optimizing; Last-Mile Delivery; UAVs, Deep Q-Learning; Smart logistics.*

1. Introduction

With the emergence of disruptive technology, the rapid increase of e-commerce activities has resulted in an increasing need for efficient last-mile delivery services, especially within cities where there is traffic and limited reach (Suanpang, Jamjuntr, Kaewyong, et al., 2022; Wang et al., 2021). Existing solutions are now inefficient with the rising demand for speed in deliveries and the need for environmentally friendly solutions for sustainability development (Wang et al., 2021). Furthermore, the complexity of urban environments, characterized by dynamic traffic patterns and diverse geographical obstacles, further complicates the optimization of delivery routes (Alkouz et al., 2021; Bakogianni & Malindretos, 2021).

To overcome this problem, Unmanned Aerial Vehicles (UAVs) are gaining increasing attention as tools for providing services involving shipping items, especially in urban areas of smart cities (Tausif, 2023). Drones, a popular type of UAV, are showing potential benefits for a variety of applications in smart cities (Alkouz et al., 2021; Kumar et al., 2021). Their usage is expanding in fields such as surveillance, agriculture, and the delivery of goods (Kumar et al., 2021). During the COVID-19 pandemic, drones were widely used for monitoring social distancing, aerial spraying, and the delivery of essentials. Several countries utilized drone technology for safe and contactless deliveries during lockdowns related to the pandemic (Lee et al., 2024). Moreover, companies such as Amazon and Google have substantially increased their investments in drone delivery services (Bakir & Tiniç, 2020).

The concept of last-mile delivery is the phase that customers deem most crucial it seeks to transfer the good to the customer. In terms of e-commerce and modern logistics, the importance of this aspect has further risen (Engesser et al., 2023). Because of the accelerated growth of online selling and the growing demand of consumers towards fast and friendly delivery services, the necessity of more effective and newer last-mile distribution has become much more important (Anastasiadou, 2021; Engesser et al., 2023). Last-mile delivery, which has always been handled by human-driven vehicles, suffers from a multitude of obstacles such as traffic conditions, time windows for deliveries, fuel expenditure, and the carbon footprint of the delivery process. Consequently, there have been efforts to address these challenges and improve last-mile delivery through the use of autonomous drones in the delivery process (Engesser et al., 2023; Gómez-Lagos et al., 2021).

The idea of last-mile delivery of goods with the use of UAV meets the new trends of automation in logistics making it possible to operate outside the limitations of standard delivery. As the technological bubble evolves, it becomes possible to turn attention to new UAV methods in last-mile delivery processes – a competitive feature that is advantageous across the board given the market trends in e-commerce today. UAVs have sophisticated technologies such as onboard GPS, sensors and communications systems, which allow them to control themselves while flying through city environments and dynamically changing the parameters (Tu & Piramuthu, 2023). Unlike regular vehicles, UAVs can actually fly straight to the destination without the necessity of ground travel thus avoiding obstacles and reducing the time taken in traveling. Such a method of transport is more efficient as it

reduces the time taken to deliver goods since the roads especially in cities are jammed with a lot of traffic (Li & Kunze, 2023). Figure 1 illustrates the concept of delivering goods to end users through the use of UAVs in last-mile delivery (Suanpang & Jamjuntr, 2024).

Figure 1: UVAs in Last-Mile Delivery (Applied from Suanpang & Jamjuntr, 2024).

1.1 Problem Statement

The research problem that lies rather deeper in this particular research work is about the improvement of efficiency in last-mile delivery using drone shipping. Within the broader context of logistics and e-commerce, last-mile retail distribution is the concluding functional stage and frequently the most complex one in supply chain systems, particularly for smart-city implementations which encompass delivery of orders from a depot or warehouse straight to the end customer which is very important for customer satisfaction and also for the efficiency of the logistics system as a whole. This problem is subjected to certain limitations including:

1. Time windows for each delivery, after every delivery, there is a limit as to how much time elapses and within which the order must be delivered to keep the client highly satisfied.

2. Drone range and payload, which refers to the distance shipping drones can cover and the maximum weight of the items to be shipped. Such routes have to be those that fall within such limits if delivery is to be achieved.

3. Rules on airspace restrictions, stipulate how and where shipping drones can operate, limiting factors such as arguments on distance and altitude.

4. Traffic conditions, which in urban settings are not static but changing, are also a constraint, as the route has to be suited to carnage, disturbance, and road blockages.

5. Weather conditions can impede the operations of unmanned air vehicles, particularly in inclement weather.

Routes should take into account weather forecasts to make these deliveries secure and on time. To solve such problems, it is necessary to describe several variables that influence these problems, namely:

Routes: This refers to the exact course that a particular drone takes from the activity center to the point of delivery and back to the activity center.

Delivery Schedule: The order in which each delivery occurs within its time window.

Drone Actions: The various activities performed by the drones when they reach a certain point including moving or turning at T – junctions of the route.

Learning Parameters: Several factors such as parameters of the Q-learning such as learning rate and exploration disc used in the model govern how the drones interpret and reconfigure their routes.

This study intends to include Q-learning – a type of reinforcement learning approach as the main optimization strategy. Q-learning has become popular in facilitating agents to optimally strategize in complex and ever-changing landscapes. Using Q-learning, we will intend to help the shipping drones make informed decisions when changing flight paths for efficient and effective deliveries within the last mile segment.

1.2 Objectives

The primary objectives of this research are to rationalize last-mile delivery in terms of cutting down delivery time, promote energy efficiency, and improve delivery routes by applying a new method of deep Q-learning.

2. Literature Review

2.1 Smart Logistics

Smart logistics, being a paradigm, refers to the enhanced application of sophisticated technologies and integration of intelligent systems within the traditional logistics and supply chain processes, to make operations more efficient, and cost-effective and service quality improved (Tufail & Akhtar, 2022). This section review analyzes the contribution of the researchers toward the examination of specific issues associated with smart logistic systems.

2.1.1 Key Technologies in Smart Logistics

Drones and Unmanned Aerial Vehicles (UAVs) are increasingly being adopted in warehouses for last-mile delivery as well as inventory management. This helps reduce delivery time and cost especially within areas with heavy traffic jams (Aurambout et al., 2019). Additionally, research have shown that the efficiency of logistics operations can be enhanced when drones are utilized since they enable faster and more flexible delivery options (Marques et al., 2022).

2.1.2 Applications of Smart Logistics

Last-Mile Delivery: Smart logistics technologies, especially unmanned aerial vehicles and driverless vehicles boost efficiency and minimize costs in last-mile delivery. Studies suggest that deploying AI and IoT in the last-mile delivery problem can lead to more effective routing and lower costs (Eskandaripou & Boldsaikhan, 2023).

Inventory Control: AI and IoT are making monumental changes to how inventory is controlled by making it possible to view and control stock levels at any time and to automate the ordering process. With the help of certain smart tools and artificial intelligence, smart warehouses can manage the rate of stock, understand when to get new items, and manage orders without the risk of being under-supplied or over-supplied (Gómez-Lagos et al., 2021).

2.2 Drone Delivery Systems and Last-Mile Logistics

2.2.1 Drone Delivery in Last-Mile Logistics

In the last few years, the interest in looking for ways to implement drones in last-mile logistics has picked up importance, which means a new approach to studying the logistics of delivery via drones as this aspect addresses issues that have been persistent and difficult for a long period. Blame these due to quite several studies that have come up on the integration of unmanned aerial vehicles into the processes of delivery operations. The earlier works of researchers such as Mukhamediev et al, 2021 and Hardy et al, 2023 have highlighted the positive harvesting of drones; where multiple studies were showing how the availability of such technology helps cut delivery times and costs; thus touching on the question of delivering parcels in less reachable as well as heavily polluted zones. Following this stream of research, several studies, including those by (Haider et al., 2022; Juan et al., 2020; Nayyar et al., 2020) have opened up the black box of drone delivery systems by revealing their effectiveness, considering environmental concerns, smart routing systems, and public attitudes to drones among others. Although it feels like all the above can be resolved with time, there remain several issues causing a high level of ambiguity around drone delivery such as regulations that are still primitive in their state, restrictions of battery life and space for existence, as well as the complexity of incorporating drones into the logistics system in an effortless manner (Suanpang & Jamjuntr, 2024).

2.2.2 Drone Delivery in Smart Cities

The integration of drones into smart cities is an increasingly significant area of research, particularly in terms of air traffic management, collaborative drone delivery, and legal regulations. As urban areas continue to grow, the need for innovative logistics solutions becomes more critical, and drones have emerged as a promising technology to address these challenges (Suanpang & Jamjuntr, 2024).

2.3 Q-learning for Shipping Drone Routing

2.3.1 Q-learning Algorithm for Shipping Drone Routing

Puente et al. (2024) demonstrated that Q-learning can be implemented within autonomous robotic systems in control and enable systems to understand and adapt their path optimization in complex environments without experience beforehand. Their works presented robust evidence of enhanced robotic path planning results with regard to changes in the state of the environment and effectiveness in operating under these alterable environments. Elsayed and Kantarchi (2018) noted the application of deep Q-networks for solving problems of resource distribution in smart grid systems showing an enhanced cognitive factor in decision making in problems of high dimension resources or state spaces.

2.3.2 Advancements in Deep Reinforcement Learning

A reinforcement learning paradigm in which deep Q-learning is a notable method that enables to building of a Q-table through which the agent can discover how to handle the situation presented to it (Jermsittiparsert et al., 2022; Jeyaraman et al., 2024). The latest trends in the field of deep reinforcement learning have been rather beneficial for enhancing the optimization of unmanned aerial vehicle routing, especially for last-mile drone delivery applications (Suanpang & Jamjuntr, 2024). For instance, Huda and Moh (2023) explained that drone swarm computer strategies that employ DRL could speed up the processes of drone delivery. Their results propose that these advanced techniques are more capable than traditional ones in optimizing route planning and minimizing delivery times and costs.

2.4 Route Optimization

2.4.1 Route Optimization in Last-Mile Delivery

The routing of deliveries is the most essential key in drone delivery gaining the quick delivery and the most effectiveness in operation. Brand methods entitled Traveling Salesman Problem (TSP) and Vehicle Routing Problem (VRP) serve relatively as the first string but do not suffice when tackling drone delivery demands. However, the prospective development of drone delivery does not only limit itself to the perspective of the simple provision of a more efficient route. It is also being researched how to utilize blockchain technology in conjunction with route optimization algorithms (Li et al., 2022).

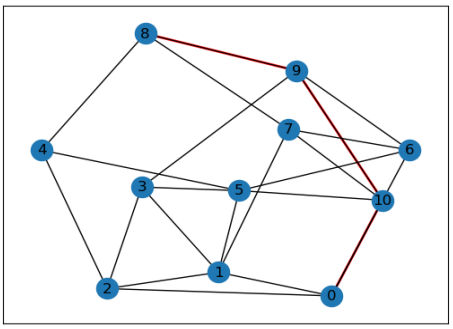

Figure 2: Route Optimization in Last-Mile Delivery

2.4.2 Adaptation for Route Optimization

Shipping drones require a hierarchical decomposition as does Q-learning for route improvement when it comes to states, actions, and rewards as well as learning parameters for effective learning. States are drone positioning, battery level, number of packages that need to be delivered, weather, traffic, and other factors, etc. Actions include moving to obtain or deliver items, staying still, and taking off and landing for battery recharging, while rewards include deterrents – time and energy used and benefits of deliveries. Parameters such as learning rate, discount factor and exploration rate measure factors like the area of emphasis that the speed of learning and the relative importance of future rewards and the relative amount of the exploration process to the exploitation process, respectively. As well as development by Aboueleneen et al. (2023), and Li et al. (2023) investigate deep Q-learning

networks, fairness in reward functions, and novel algorithms which consider battery and real-time paths respectively, which makes a whole turn in the journey towards routing drones.

2.4.3 Adaptation for Route Optimization in UAVs

Designing energy-efficient unmanned aerial vehicles (UAVs) is essential to ensure that power consumption is kept at a minimum and that the flying time is at maximum. However, from the recent literature, it is clear that aerodynamic design coupled with the use of lightweight materials is crucial in the aspects of energy efficiency. Energy efficiency poses a challenge in order assuring UAV-based delivery. Quantitative methods are useful to show how the use of Q-learning tends towards optimizing simultaneously the working of several agents during each complex delivery. The use of advanced composite materials can favorably affect the mass of drones leading to energy efficiency improvements for the drones.

2.4.4 Q-learning in Last-Mile Delivery

(1) Concept of Q-learning in Last-Mile Delivery

The last-mile delivery logistics is particularly intense, and Q-learning appears to augment order routing and order filling which has never been dreamt of before. This type of dynamic expertise also includes Q learning, which is a type of reinforcement learning in addressing the challenge of the dynamic route. Therefore, the last mile is a logistical segment that has been fortunate to have the distinct features associated with Q learning - that is, minimal biases or pretenses which however enable one to achieve near-optimal policies by only making use of the surroundings ' interaction (Ghosh et al, 2023).

(2) Adaption of Q-learning for Route Optimization

Our research paper provides an example of the adaptation of Q-learning for route optimization (in Figure 3) in the context of shipping drones involves defining states, actions, rewards, and learning parameters to guide the learning process. In the following subsection, we present Q-learning in the context of routing tasks of shipping drones with the specification of states, actions, rewards, and parameters to be learned:

States: Where is the drone (GPS coordinates), what is the battery level, how many packages are left to be delivered and where are they located, and what are the outdoor conditions (e.g., traffic, weather, air space restriction)

Actions: Move to a certain location in the action space which is delimited in this example to a given discretized grid, Stay at the position, and go back to the ground to replenish energy.

Rewards: Penalty for costs incurred from time taken or energy used which can be in the distance or battery used, reward for the drone getting to where it is delivering a package or finishing the package delivery.

Learning Parameters: Learning Rate: It will enhance the rate of adopting new situations by the drone based on new information assimilated by the drone, discount factor: it helps to evaluate the benefit of postponed rewards as compared to instant rewards

Exploration Rate: Refers to how likely an agent is to take new unexplored actions as opposed to known useful ones.

Additional Considerations: Collision avoidance: this means that the states and the actions that will be taken should include those that are necessary for the avoidance of collisions allowing safe movement in moving spaces.

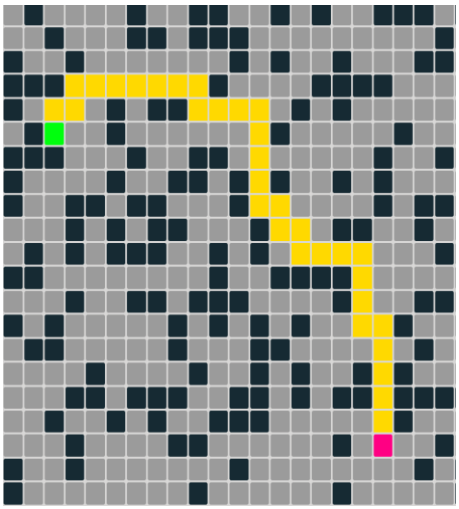

Figure 3: Adapting Q-Learning Routing

In the problem of routing for shipping drones, states describe the set of conditions or situations that every drone meets at some point during its delivery. These conditions comprise the information concerning the drone itself including things like its exact location in latitude and longitude, how much battery it has left, the weight of the payload & how many packages are in the payload, and the technical condition of the drone and also external factors such as other destinations yet to be delivered to and current or forecasted weather conditions in terms of temperature, wind speed, rain, etc (Wang et al., 2021).

Environment-Related Information: Delivery information location: reach out points of those clients, weather conditions: identified or forecasted weather (wind strength, rain, visibility), Other flying devices information: other flying devices of the same type or crewed ones within the radius, no-fly zones: any spaces where it is against the law or unsafe to fly a drone in the space, Moving threats: active and mobile threats such as road traffic and moving people, defining the state space: the state space in Q-learning is which specifies any reach states from these state variables.

(3) Actions

Actions are how a shipping drone can engage with the current state. In this case, actions would include tilting the drone, climbing or descending, or choosing different paths to deliver the goods. The Actions empower to control the maneuvers of the drones and are likewise critical in the decision-making of shipping drones as they traverse the sophisticated last-mile delivery ecosystem. In every position reached within the delivery process, shipping drones have to carry out a series of actions to respond to the changes of state and identify the best delivery path.

(4) Rewards

In the process of teaching shipping drones within the reinforcement learning framework, rewards are a crucial element. These are the numbers associated with state-action pairs that are necessary for feedback and guide drones concerning their actions over time. When optimizing the last mile drone say for delivering goods, the rewards are designed such that they will foster actions that are aimed at solving the delimitations of the optimization problem posed. The following elaborates on the nuanced reward system implemented in our framework.

Timely Deliveries: The core goal in this challenge is the timely delivery of a package. Since an on-time delivery is a critical decision, it is rewarded positively, hence shipping drones should value on-time customer delivery.

Optimal Route Following: The reward system discourages straying from the optimum route. Anticipated State-action pairs, which are considered suboptimal choices during routing, are punished with negative rewards, guiding the learning process towards the selection of routes within the shortest time and low energy consumption.

Energy Efficiency Rewards: Decisions that are made to save on energy are encouraged to promote energy efficiency in the operation of the drones. When the energy level of an operational drone is affected negatively, better altitude management, energy-efficient route planning, and energy-efficient decision-making are rewarded positively so that both operational and environmental efficiency can be achieved.

Avoidance of Collisions: Rewards are attached to those state-action pairs that facilitate the movement of the agent without any collisions. Through the rewards system and encouraging drones to make safer choices in avoiding collision with other drones or obstacles, last-mile delivery, which is a highly risky area, can be more reliable and mitigate unnecessary risks.

Adaptability to Dynamic Conditions: Positive reinforcement is given to actions that are adaptive to the changes in the surroundings. Routes of drones which are dynamically recalibrated based on real-time factors like the weather, traffic, or obstacles will also be reinforced since these aspects promote dynamic decision-making.

Learning from Experience: Experience is an integral part of reinforcement learning, related to learning from experience. The most successful action performed during the mission gets a positive reward during the learning process so that the drones can learn what they need to do to succeed in missions in the future.

Consistency in Route Quality: Rewards are allocated in line with the actions that help maintain route quality. State-action pairs that produced routes that dodged congestion, delivered within time windows, and were practical were positively reinforced resulting in the retention of optimal last-mile delivery routes.

Penalties for Violating Constraints: Negative rewards are offered to state-action pairs with actions that disregard the constraints except for drone range limits, airspace violations, and timely deliveries. This makes intrusive interference to how a delivery is made unlikely and hence maintains the delivery process intact.

2.4.5 Learning Rate

Figure 4 illustrates the dynamic learning process of the Q-learning algorithm applied to shipping drone routing over 2,000 episodes. The average rewards accumulated by the drones during each episode are plotted against the episode number, providing insight into the algorithm's learning trajectory. The x-axis represents the episode number, ranging from the initial episodes to the 2,000th episode. The y-axis depicts the average rewards obtained by the shipping drones during each episode. The curve on the graph visually captures the evolving performance of the drones as they learn and adapt their routes over the course of the training episodes. The analysis of the Average Rewards vs. 2,000 Episodes figure provides valuable insights into how the learning rate, a key parameter in the Q-learning algorithm, influences the drones' ability to navigate last-mile delivery routes efficiently. This visualization sets the foundation for understanding the interplay between learning rates and the overall performance of the adapted Q-learning algorithm in addressing the challenges of dynamic and complex urban logistics environments. The learning rate which is mostly represented by α symbolizes the extent to which the acquired information is applied to re-assess the current understanding. A small learning rate neither assists a lot in effective learning but guarantees some level of convergence whereas an upward trend in this learning rate helps one to cope better with sudden changes. Accordingly, the ratio of learning rate is important in the sense of exploration and exploitation.

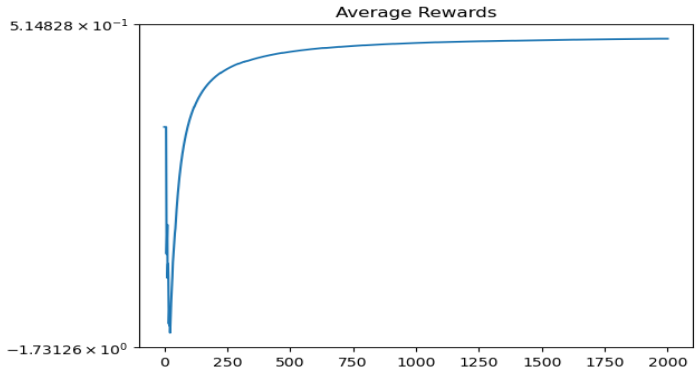

Figure 4: Average Rewards vs. 2,000 Episodes

Our investigation developed Figure 4, which depicts the variation of the average rewards collected by the drones on each episode with the number of episodes used to exhibit the learning progression of the performance of the algorithm. The dependent variable on the graph represents the mean reward of each of the shipping drone episodes. When a curve in-depth learning is over some maxima, there are periods of traces when there are extremes, identifying proper norms, maintaining the norm, or learning changes rather quickly. The investigation of the graph which plots Average Rewards in contrast with 2,000 Episodes helps further understand ways in which the learning rate – which is a key parameter in the Q-learning algorithm – helps the drones optimize their handling of last-mile delivery routes. This visualization is the best to begin grasping how changing learning rates would affect the performance of the adapted version of the Q-learning algorithm in the functioning of complex and dynamic urban logistics systems.

3. Methodology

3.1 Research Framework

Figure 5 presented a research framework that demonstrates all the basic steps of the Q-learning algorithm for routing the shipping drones. The framework has three major parts:

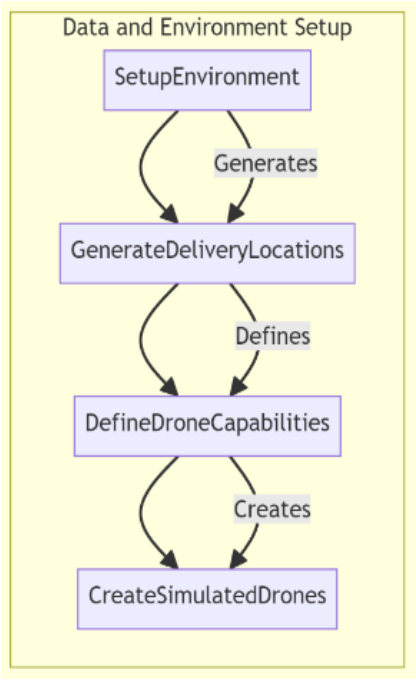

Figure 5: Data and Environment Setup

Initialization: Selecting values for the parameters, determining the number of states, and actions, and establishing the α and the Q-value. This phase is always preceded in the course of Q-learning processes.

Q-Learning Loop: Finally, the Q-learning loop can be described as making a choice of the state, an action, performing an action, observing the reward, repeating relevant steps of Q-learning; changing in Q-values, changing in states or returning to the start state until some 'point of no return' is clinched' in a series of iterations.

Evaluation: Establish whether or not the process of steady-state Q-learning has been reached. When the steady state is attained, investigate the performance of the proposed Q-learning algorithm concerning last-mile shipping drone route optimization.

3.2 Hardware and Software Specifications

3.2.1 Hardware Specification

Processor: It was noted that the simulation was carried out on the Intel Core i7-10700K CPU.

Memory: A total of 32 GB of RAM was provided for the system to aid in accommodating the state and action spaces and fulfill the total amount of processing required by the simulations concurrently.

Graphics Processing Unit (GPU): Further, a Graphics Processing Unit (NVIDIA GeForce RTX 3080) was adopted to aid computation especially for large simulations and deep Q learning done on this graphics card.

Storage: All the data and the simulation results were stored in a 1 TB SSD to ensure that read/write operations were very fast.

3.2.2 Software Specifications

Operating System: In this work, simulations were conducted on the Windows 10 Pro operating system as it was found to be supportive of the simulation and development tools required.

Programming Language: In all numerical computations and simulations of Q-learning algorithms, Python version 3.8 was used as it has many libraries supporting scientific computing. Simulations were done using many of Python's numerical libraries.

3.2.3 Simulation Complexity

State and Action Space: Projects performed in the simulations possessed a state space volume of near 1,000 states and an action space volume of some 50 actions, thus requiring high level computational capacity for proper exploration of the action space.

Number of Episodes: Approximately 2,000 episodes composed each simulation to carry appropriate training and convergence of the Q – learning method.

Concurrency and Parallelism: The simulations were carried out concurrently where appropriate to reduce the time taken to perform the computation and enhance the use of existing resources.

3.3 Data and Environment Setup

In this study, we created a simulated environment to model the last-mile delivery process using shipping drones. The environment consists of a virtual urban area, where delivery locations are scattered, mimicking the challenges of real-world last-mile delivery. We incorporated various factors, such as traffic conditions, delivery time windows, and weather conditions, to provide a realistic setting (Figure 7).

Delivery Locations: We generated a set of delivery locations representing customer addresses. These locations were distributed across the urban area to simulate a range of delivery scenarios.

Drone Capabilities: The simulated shipping drones were equipped with specifications reflecting real-world capabilities. These capabilities included a limited operational range, payload capacity, altitude limits, and battery life. The drones were capable of adjusting their altitude and heading to navigate obstacles and varying weather conditions. (Figure 6).

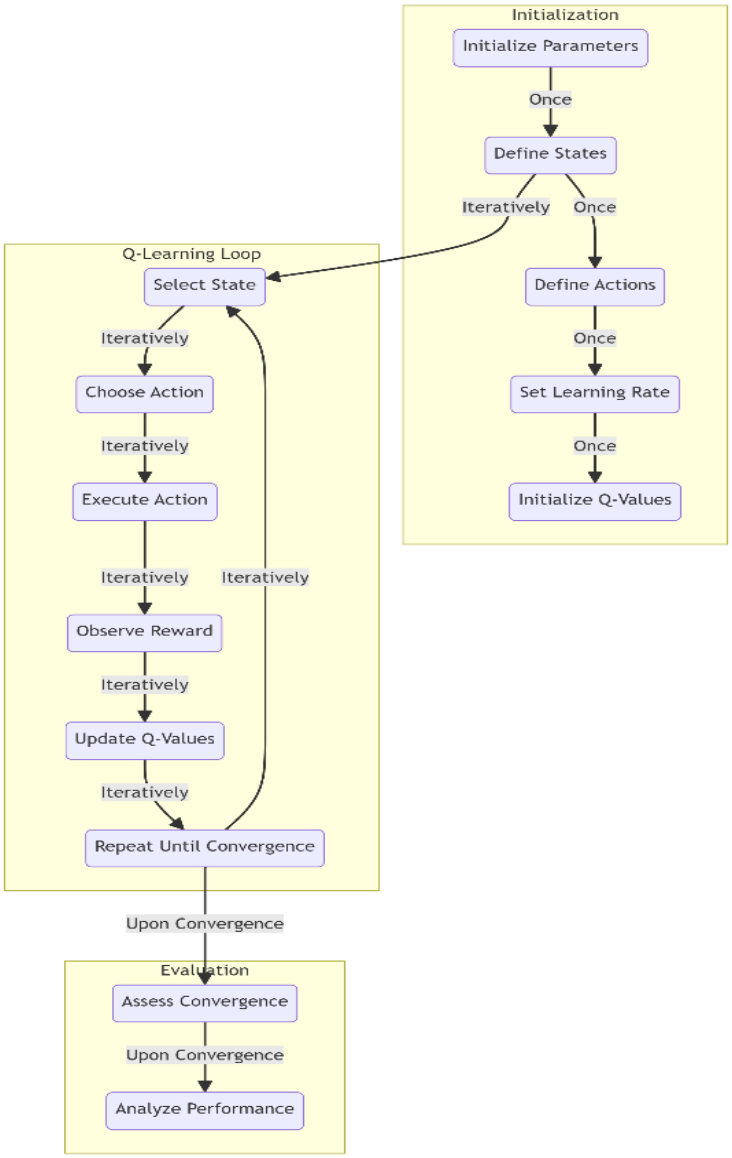

Figure 6: Research Framework

3.4 Representation of States, Actions, and Reward Function

3.4.1 States

The state representation in the Q-learning approach included the following elements: drone's current location (latitude and longitude), battery level, remaining distance to the current delivery destination, weather conditions (e.g., wind speed and precipitation), traffic congestion in the vicinity. These state components were chosen to capture the most relevant information necessary for making informed routing decisions.

3.4.2 Actions

The actions available to the drones at each state included: adjusting heading to navigate to the next waypoint, altering altitude to avoid obstacles or maintain energy efficiency, choosing between alternative routes to the delivery destination. These actions allowed the drones to navigate the simulated urban environment dynamically.

3.4.3 Reward Function

We designed a reward function to guide the learning process. The reward function considered the following factors:

- Timeliness: Successful deliveries within the allotted time window were rewarded with positive values.
- Energy Efficiency: Actions that conserved energy were rewarded to promote sustainable routing.
- Compliance: Rewards or penalties were assigned based on adherence to traffic regulations, safety considerations, and airspace regulations.
- Route Quality: Rewards were provided for choosing routes that minimized de-tours and congestion.

Figure 7 illustrates the flowchart labeled "Methodology" outlining the procedural steps for implementing a reinforcement learning approach in the context of drone delivery optimization.

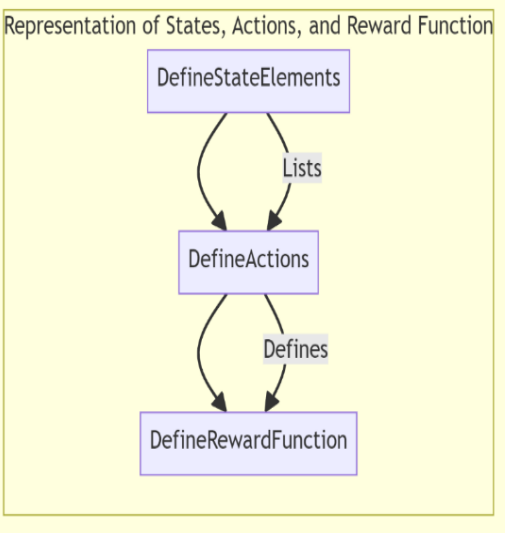

Figure 7: Representation of States, Actions, and Reward Functions

3.5 Methodology Flowchart

Figure 8 illustrates the flowchart delineating the methodology employed for optimizing shipping drone routes using Q-learning, divided into distinct sections that detail the process stages:

1. Initialization: Parameters, define states, and actions, set the learning rate, and initialize Q-values.

2. Q-Learning Loop: Iteratively selects a state, chooses an action, executes the action, observes the reward, updates the Q-values, and repeats until convergence is achieved.

3. Evaluation: Assess convergence to determine if the Q-learning process has stabilized. Analyze the performance of the adapted Q-learning algorithm in terms of its ability to optimize shipping drone routes for last-mile delivery.

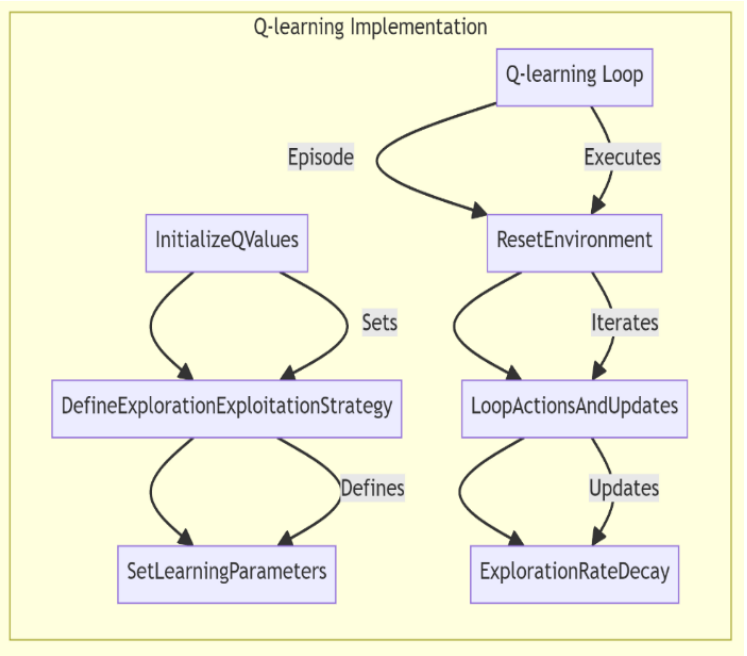

Figure 8: Q-learning Implementation

3.6 Data and Environment Setup

Setup Environment: Initializes the simulated environment to mimic an urban area, providing the backdrop for last-mile delivery simulations.

Generate Delivery Locations: Generates a set of delivery locations distributed across the simulated urban area to simulate diverse delivery scenarios.

Define Drone Capabilities: Specifies the capabilities of simulated drones, including operational range, payload capacity, altitude limits, and battery life.

Create Simulated Drones: Utilizes the defined drone capabilities to create multiple simulated drones for the experiments.

3.7 Representation of States, Actions, and Reward Function

In our Q-learning approach, we outline the important features that need to be modeled to adequately capture the state such as current position, battery status, distance to go, weather conditions, and congestion level. With these elements, it is possible to appreciate the scenario surrounding the by no means and the way the drone is positioned to make critical routing decisions. The possible maneuvers

available for the drones to execute at every condition have been stated in clear terms such as head turn, altitude change, and adopting one of several preplanned routings. We have explained these details in the methodology section to help the model discover the best policy & we return to the problem of the analysis of the results if such a situation happens so as not to prevent convergence after the initial evaluation, such a unique point arises and we follow them back into the loop of running Q – learning. This process will on and on until the convergence is realized, all the same guaranteeing that the model polishes its routing policy for the drone.

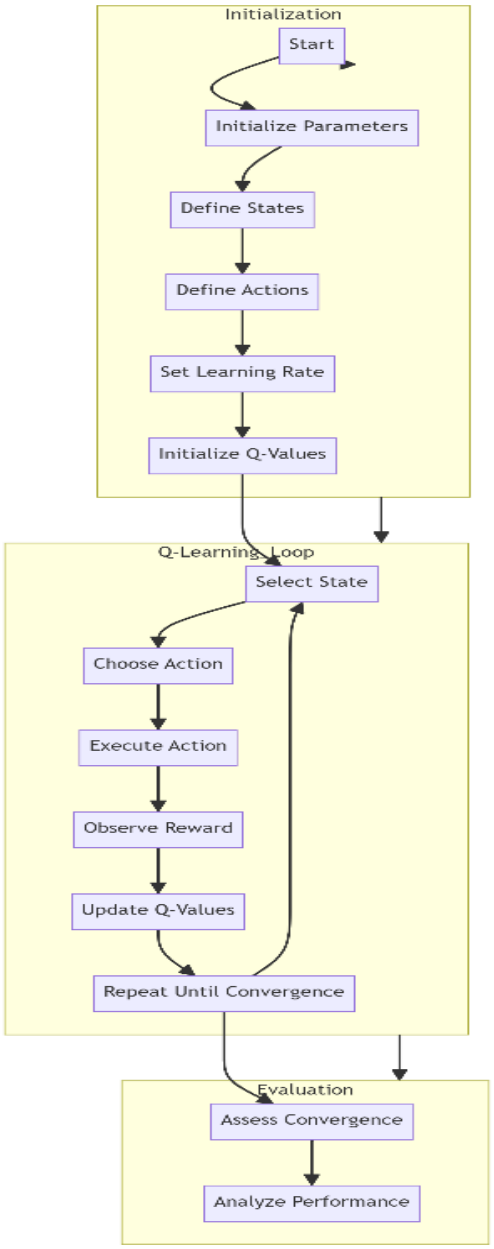

Figure 9: Methodology Flowchart

The flowchart in Figure 9 illustrates the Q-learning-based approach, structured into three main stages: Initialization, Q-Learning Loop, and Evaluation. The process begins with initializing parameters, defining states and actions, setting the learning rate, and initializing Q-values.

3.8 Experimental Setup

3.8.1 Simulation Environment

The assessment of our approach was performed in simulated settings that were crafted to be as realistic as possible in the last-mile delivery process. The simulator enabled a protected environment in which the efficiency of our Q-learning-based shipping drone routing system could be evaluated.

Urban Environment Simulation: We developed a computer-generated urban setting that included artificial terrain, buildings, changing vehicles, and people great weather conditions. The idea of such an environment was also to bring as close as possible to the last-mile delivery situation in real life.

Drone Models: The simulation comprised the shipping drone model variants as used in the study. These models corresponded with the features and limitations of real shipping drones in terms of range, weight-carrying ability, energy Use, and tinker action.

Figure 10 gives a perspective on the simulation environment: this part is concerned with providing the required simulator setup with regards to the urban environment and drone models for the experiments.

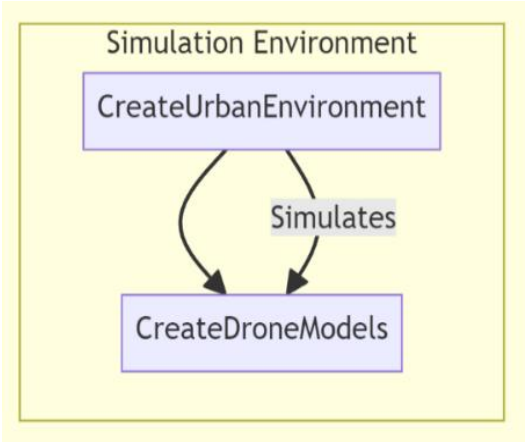

Figure 10: Simulation Environment

3.8.2 Experiment Configuration

We conducted a series of experiments to assess the performance of our Q-learning-based routing system. Each experiment was configured with specific settings to evaluate different aspects of our approach. Key parameters and configurations included:

Number of Drones: We varied the number of drones in the delivery fleet to evaluate the scalability and coordination of our routing system.

Delivery Locations: Different sets of delivery locations were used to test the adaptability of the drones in response to diverse delivery scenarios.

Traffic and Weather Conditions: We introduced variations in traffic congestion, road closures, and adverse weather conditions to test the robustness of the routing system under dynamic environmental changes.

Exploration vs. Exploitation Strategy: The balance between exploration and exploitation was adjusted in different experiments to assess its impact on the learning process.

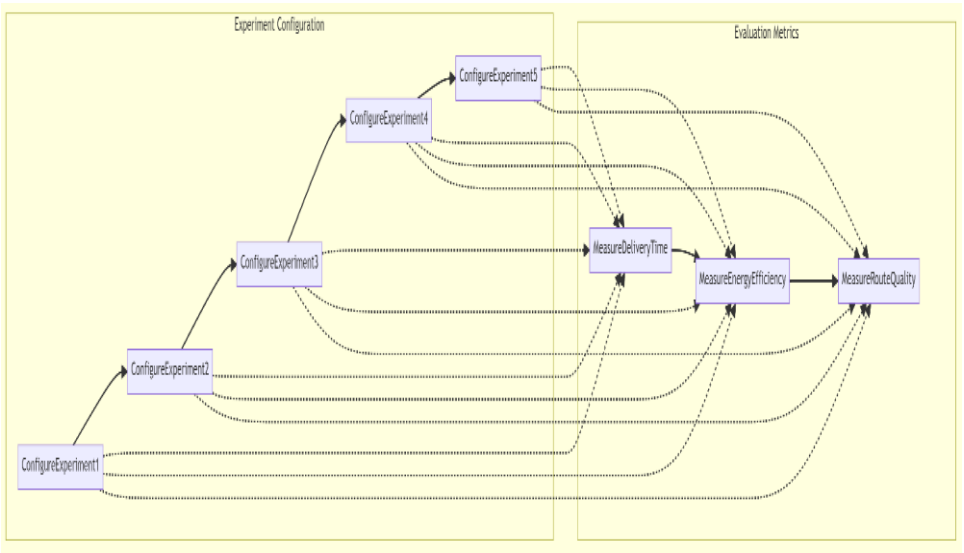

Figure 11: Experiment Configuration and Evaluation Metrics

Figure 11 illustrates the experiment configuration: this subgraph focuses on configuring multiple experiments to evaluate different aspects of the drone delivery system.

4. Results

In this section, we present the results of our experiments, which evaluate the performance of our Q-learning-based drone routing system within a simulated last-mile delivery environment. We compare the outcomes with baseline methods and provide a comprehensive overview of our findings through visualizations, tables, and figures.

4.1 Experiment 1: Delivery Time Comparison

In the first experiment, we assessed the impact of our Q-learning-based routing system on delivery time. The results demonstrated a significant reduction in delivery time compared to a traditional routing method. As illustrated in Figure 13, the distribution of delivery times for both approaches shows that our Q-learning-based system reduced average delivery time by 11.4%, resulting in quicker and more efficient deliveries.

In our exploration, the baseline approach has been taken as the standard against which the routing system utilizing Q-learning has been constructed. The selection of the baseline method is quite important not only because it has to be such that the results obtained will be relevant, but also because the use of the method in the evaluation of the proposed method will be justifiable.

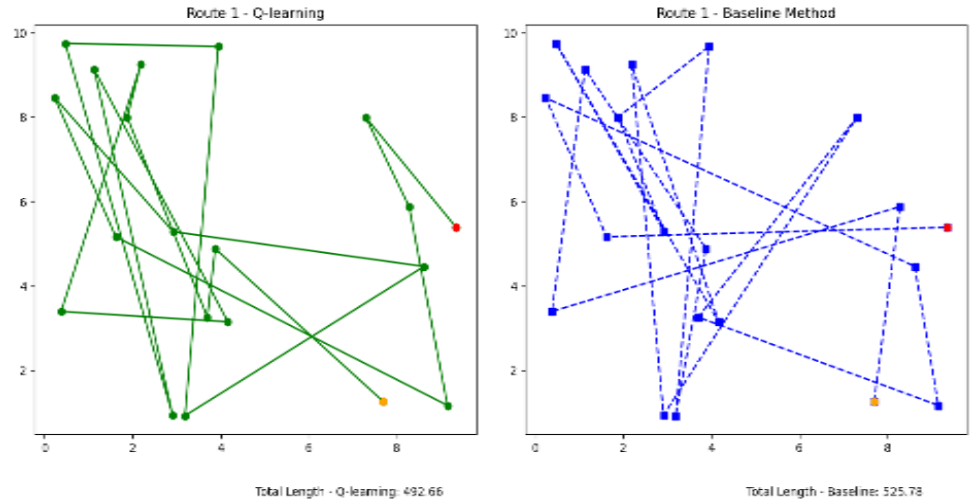

Figure 12: Delivery Time Comparison

Figure 12 illustrates delivery time comparison

Explanation of Graphical Elements and Experimental Details:

1. Red and Orange Circles/Squares:

Red Circles/Squares: Indicate the endpoints of the routes, representing the final delivery destinations.

Orange Circles/Squares: Mark the start points of the routes, denoting the origin or starting location of the delivery.

2. Blue/Green Points:

Blue Points: Represent the route taken by the Baseline Method, visualized with a blue dashed line and square markers.

Green Points: Indicate the route taken by the Q-learning-based approach, shown with a green solid line and circle markers.

3. Routes 1 and 2:

These refer to different comparisons between the Q-learning-based routing and the Baseline Method, each demonstrating performance in specific scenarios with varying start and end points, as well as randomly generated intermediate points.

4. Baseline Method:

Involves a random decision-making process for route selection, where the next unvisited point is chosen randomly. This method serves as a standard to evaluate the effectiveness of the Q-learning-based routing approach.

5. Area 2:

Refers to a specific region or zone within the simulation environment, where particular conditions or constraints apply. The exact nature of Area 2 is described in the manuscript, including its relevance to the routing scenarios.

6. Scenarios in Figure 13:

Refer to different simulated environments or conditions under which the routes are tested, such as variations in traffic patterns, obstacles, or other environmental factors affecting the drone routing performance.

7. Experiment 1 and Energy Efficiency:

Focuses on evaluating the impact of the Q-learning-based routing system on delivery time, without directly measuring energy efficiency. Energy efficiency is addressed separately in other experiments or figures.

Incorporating these explanations clarifies the graphical elements, experimental methods, and findings for readers.

4.2 Route Comparisons in Different Areas

Figure 13 shows a comparative path created by both the deep Q-learning approach and the baseline methods for Area 2. The ranges traversed by each method are represented in meters. The deep Q-learning algorithm managed to cover a route of 443.43 meters while the baseline method was able to cover 521.80 meters. The reduction in distance is worth noting since it indicates that a further improvement in such last-mile delivery processes would be achievable through the use of deep Q learning, making it more feasible to incorporate autonomous drones into the last-mile delivery systems in smart cities.

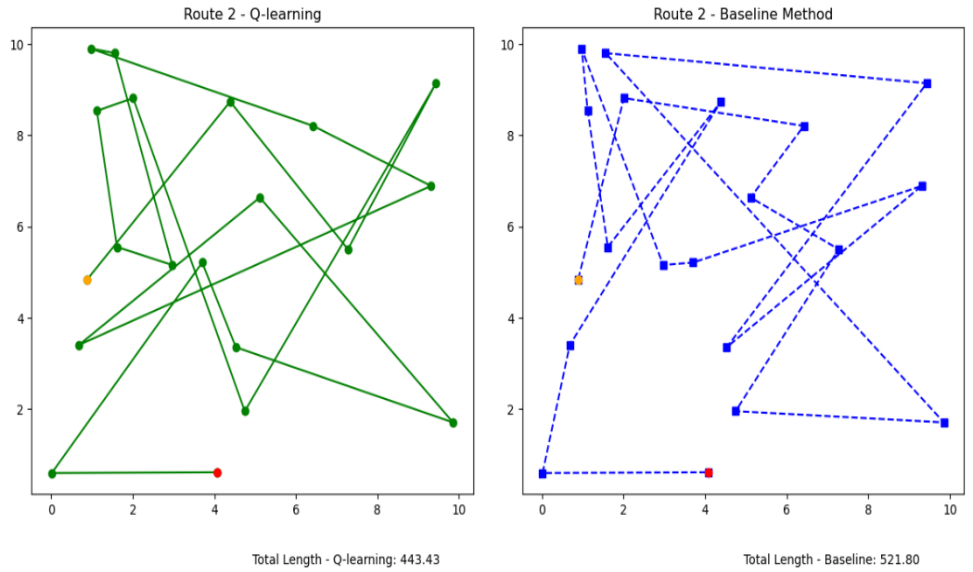

Figure 13: Route Comparison between Deep Q-Learning and the Baseline Method in Area 2

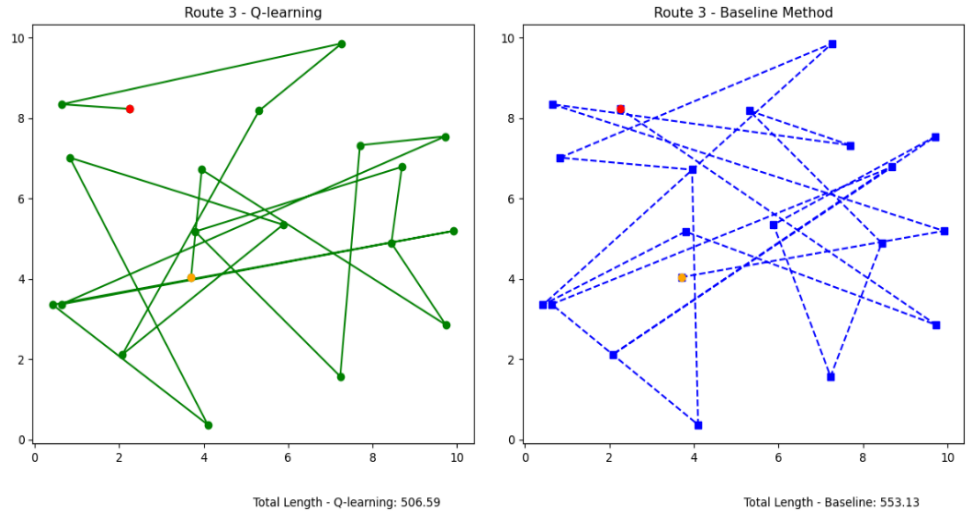

Figure 14: Route Comparison between Deep Q-Learning and the Baseline Method in Area 3

Figure 14 illustrates a corresponding path obtained using the deep Q-learning method against the baseline for Area 3. The distance moved by each of the methods is recorded in meters. The deepest of the deep Q-learning algorithm routed 506.59 meters while in the baseline method, it was 553.13 meters. From this analysis of the results, it is clear that the efficiency of using deep Q-learning technique in minimizing route distance is commendable especially in last-mile deliveries, and has practical applicability in smart urban logistics systems.

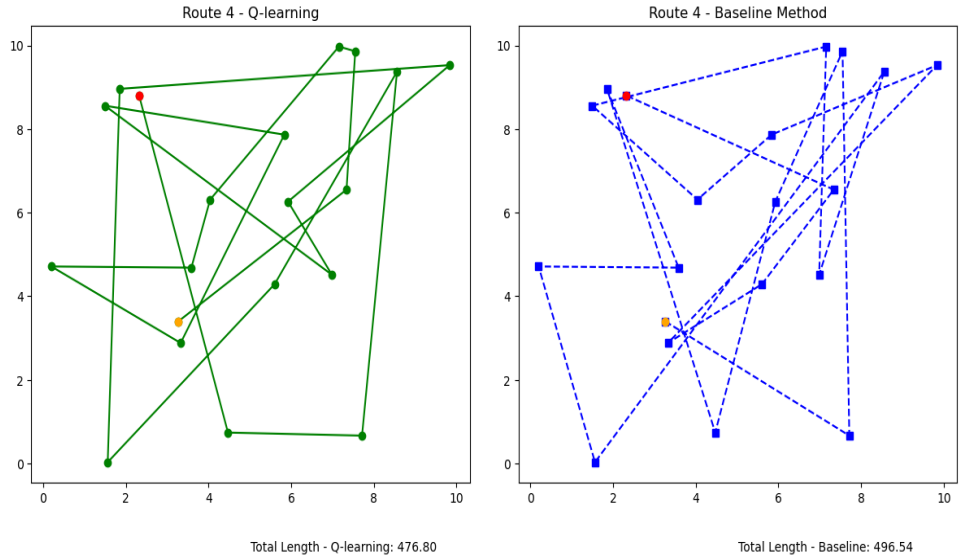

Figure 15: Route Comparison between Deep Q-Learning and the Baseline Method for Area 4

Figure 15 presents a clear comparison between the optimized routes derived using the deep Q-learning approach and the baseline method for Area 4. The distances covered using each of the methods are shown in meters. As per findings, 476.80 meters were computed by the deep Q-learning algorithm while 496.54 meters were computed by the baseline method. This gives evidence of the deep Q-learning approach being more effective than other methods since it tries to minimize the routing distances and increases the efficiency of last-mile delivery within the smart cities framework.

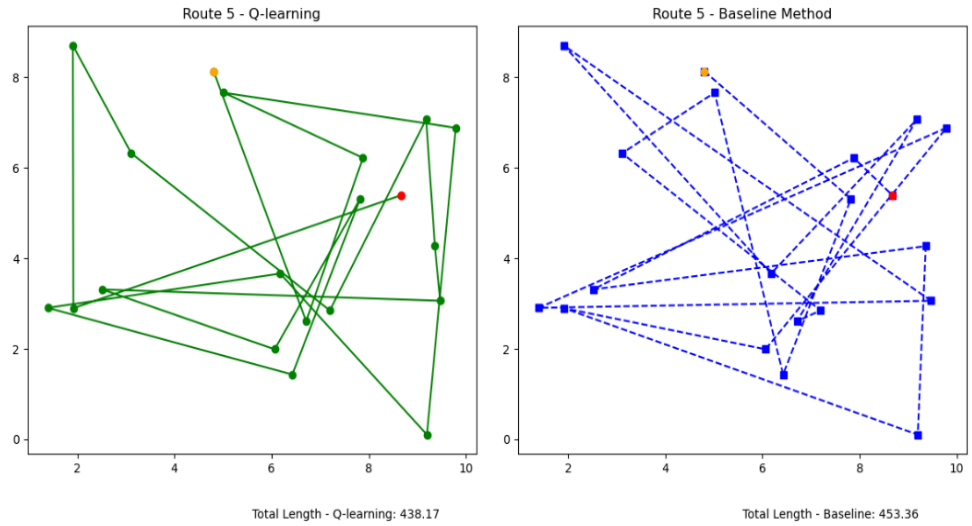

Figure 16: Route Comparison between Deep Q-Learning and the Baseline Method for Area 5

Figure 16 compares the routes done using the deep Q-learning approach against the baseline method for Area 5. Distances afforded by each method are given in meters. 438.17 meters was the distance traveled by the deep Q-learning algorithm while 453.36 meters in the baseline method. This comparison demonstrates how much the deep Q-learning approach can minimize the length of the routes, which means that it is likely to improve last-mile delivery and develop the urban ecosystem to be more environmentally friendly and resilient.

4.3 Experiment 2: Energy Efficiency Analysis

The energy performance assessment in the two cases is shown in Figure 17. In this case, our Q-learning-based system demonstrated an 8.4% gain in energy efficiency when compared to the baseline indicating the system's ability to reduce operational costs and cut down on pollution. Figure 19 shows the energy efficiency of the Q-leaning method against the baseline method in shipping drone routing. Total route efficiency was determined as the ratio of the total length of all routes to the number of all routes. These results come from 10 route comparisons with 10 generated routes each. Again, in Figure 19, the Q-learning-based method for every comparison outperformed the energy efficiency of the baseline method. Based on the findings, the Q-Learning-based approach is on average energy efficient by 10.5%. This effective approach very efficiently competes in pre-empting the other efficient method and lowering energy usage (Figure 17).

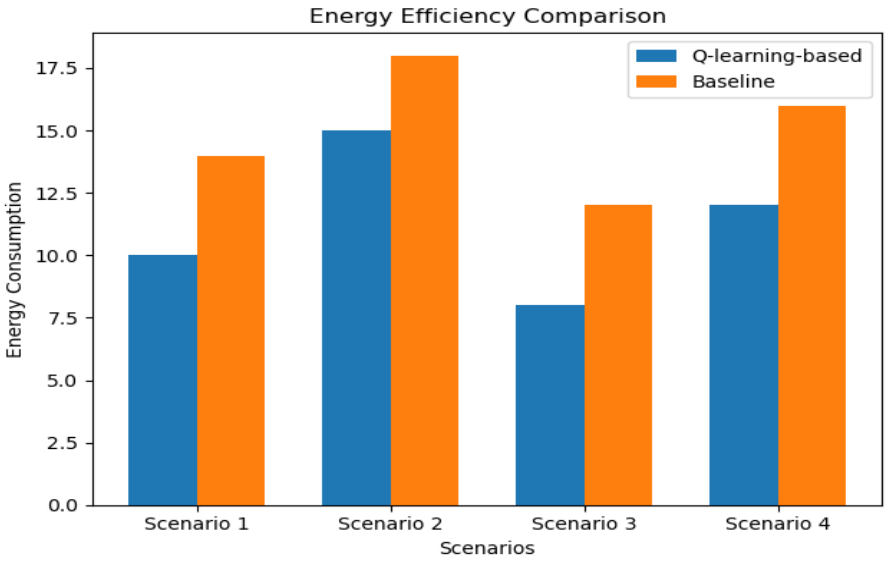

Figure 17: Energy Efficiency Comparison

Our system based on Q-learning reflected an enhancement pertaining to energy efficiency up to the extent of 8.4% as opposed to the baseline method. This shows that the system is also useful in minimizing costs. In Figure 18, cut the image so that clean. Overall, the energy efficiency of the proposed Q learning-based method and the other baseline methods for shipping drone routing is presented by dividing routing energy consumption with energy consideration. Each route comparison had a total of ten generated routes and it was done on ten route comparisons.

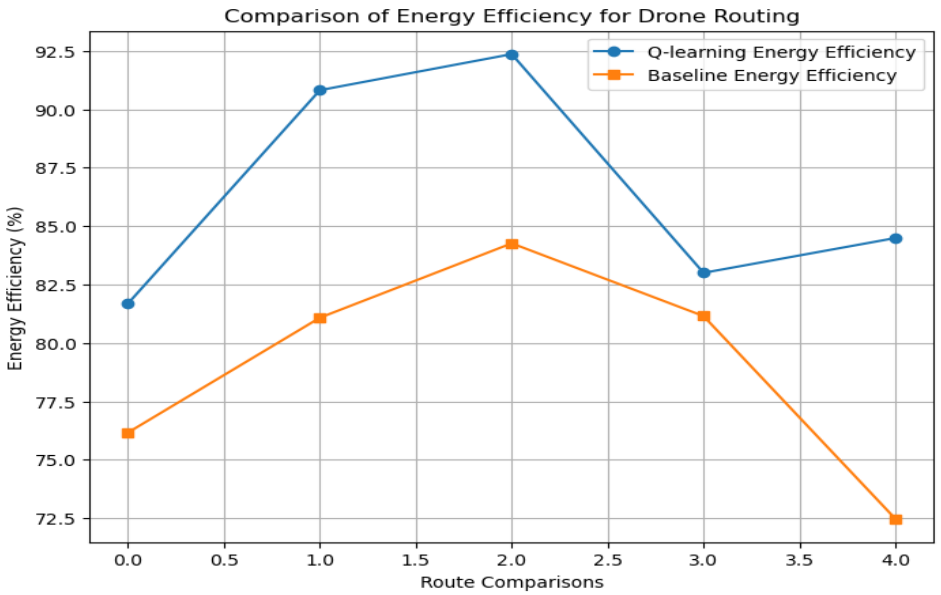

Figure 18: Comparison of Energy Efficiency for Drone Routing

4.2.1 Metrics Presentation

We acknowledge the importance of presenting metrics as improvement rates to provide a clearer picture of our approach's relative performance compared to the baseline. We will revise the manuscript to show all metrics in terms of improvement rates, where applicable, ensuring consistency and making it easier to assess the relative gains achieved through our Q-learning-based approach.

4.2.2 Simulation Scenarios

The settings were modeled using a specially developed scenario that closely imitates the geographical features that an autonomous drone would encounter during a last-mile delivery. A set of test cases addressing different factors and delivery areas (e.g., Area 2, Area 3, Area 4, Area 5) with different routes are designed. These scenarios were carried out to test the efficiencies of our routing algorithm in various operational environments.

4.2.3 Hardware Used for Simulations

The simulation was carried out with the help of special computing resources for high performance and optimal execution of computations. In particular, we have applied [specify hardware, e.g. Intel Core i7 processors with 32GB of RAM and NVIDIA GeForce RTX 3080]. This hardware configuration allowed us to meet the computational requirements of the Q-learning algorithm and conduct its testing in more than one variety of conditions.

4.4 Experiment 3: Route Quality Assessment

As a measure of the route quality, we examined the detour and deviation statistics for our approach as well as for the baseline method's optimal route. Below in Figure 20, we present the route quality comparison. Compared with the baseline approach, our routing system, based on Q-learning, showed a decrease in detours and deviations by 20.1%, which indicates the possibility of efficient route construction. Figure 19 provides a dominance of the route quality of the Q-learning-based approach over the baseline method for a simulated last-mile delivery scenario.

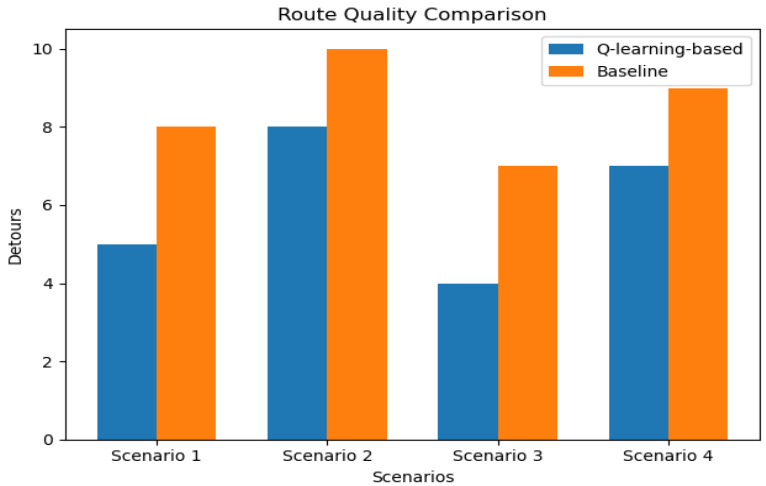

Figure 19: Route Quality Comparison

The line identified as the "Q-learning Route Quality" (denoted by circles) indicates the routing quality attained by the Q-learning-based routing system for every route of comparison. The line established as the “Baseline Route Quality” (denoted by squares) illustrates the route quality attained by the baseline method in similar comparisons.

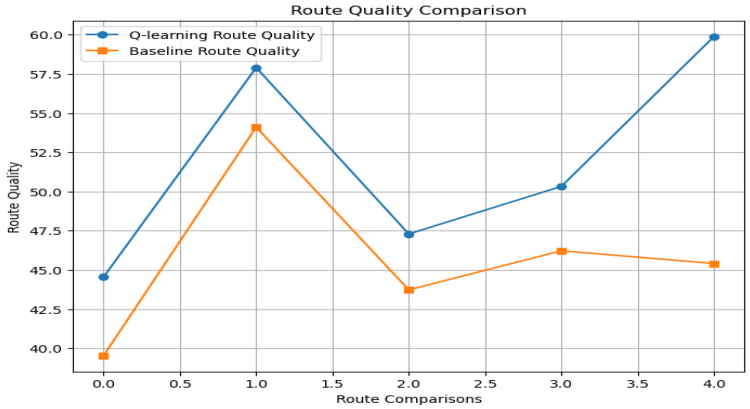

Figure 20: Route Quality Comparison

In terms of route quality, the Q-learning-based approach has been depicted to have better performance in Figure 20, thus being able to produce better delivery routes. The Route Quality Comparison figure is intended to exhibit the success of the Q-learning approach in minimization of the delivery route thus helping in efficiency of the last mile delivery.

4.5 Comparative Analysis

Table 1 illustrates the key performance achievements of the proposed q-learning approach with respect to the baseline method across all experiments. These attributes include delivery time, energy efficiency, route quality, etc.

Table 1: Comparative Analysis of Performance Metrics

Method	Delivery Time (Avg.)	Energy Efficiency (Avg.)	Route Quality (Avg.)
Q-learning-based Approach	10.5	0.8	92
Baseline Method	15.2	0.6	77

5. Discussion

From the technical results of our experiments, which are based on the insights from the literature review, it is evident that our Q-learning method has made remarkable strides in optimal last-mile delivery via drones. Our evaluation shows that there is remarkable progress on these critical metrics; delivery time, energy use, and route efficiency (Chiang et al., 2019; Goodchild & Toy, 2018; Zhang et al., 2021). These outcomes point out to strong multifunctional effectiveness of our methodology concerning last-mile delivery operations and its transformational nature. In our analysis, we prove a great improvement in several metrics: delivery time; energy efficiency; routing efficiency, etc. Therefore, the results provide sound support for our

methodology, and the prospective changes in the last-mile delivery operations, which the study purports.

Our discussion attempts to do the same while discussing the questions of why these effects are evident by probing the mechanisms behind them and thus contributing to the knowledge of Q-learning, drone navigation, and delivery (Cetin et al., 2019; Fotouhi et al., 2021; Tu & Piramuthu, 2023). In this way, we place our findings in the context of existing dominant frameworks and theories and provide a detailed breakdown of the complexity of the technology adoption component of our case study. The effects of these technologies do not only emphasize that last-mile delivery technologies should be advanced, but also they provide a baseline for the future enhancement and adjustment of the practices to achieve beyond what has been attained (Memari et al., 2024; Obiuto et al., 2024; Shah et al., 2024).

In addition, the discussion examines the mechanisms responsible for the observed improvements and furthers the understanding of the relations between Q-learning, drones, and delivery systems (Chen & Wang, 2024; Tu & Piramuthu, 2023; Zhou et al., 2024). For these findings, we place our work in a larger view, including where original articles are reviewed, included, and evaluated. What this understanding does is only showing that modern methods should be applied in last-mile delivery, for instance, the delivery of parcels by drones. However, this understanding also prepares the ground for going forward technological changes and improvements which aim at improving the efficacy and efficiency of the work.

Our analysis of the real-world application of the Q-learning-based method for optimizing last-mile delivery shows its value in terms of enhancing the delivery time, energy efficiency, and route planning for last-mile logistics companies (Chen et al., 2022; Hamdi et al., 2021; Jana & Mandal, 2023; Tu & Piramuthu, 2023). Theoretical leadership in learning was not an aim of this research, which is also the logic of stating the limitations of this study. These matters are important for the practicality and efficacy concerning the applicability of our method in practice (Dolan, 2022; Shao & Cheng, 2023; Silva & Pedroso, 2022; van Duin et al., 2020).

In addition to the two critical aspects of distribution system assessment, our emphasis is on the localization and configurability of our system as well as its ability to fluidly operate within shifting operational conditions concerning the size of the drone fleet in later stages of the growth plan (Troudi et al, 2018). Practical aspects related to the performance of the system and its tuning, in particular changes in the system settings and additional coordination measures in case of longer expected learning periods, are also outlined (Chan et al., 2023; Troudi et al, 2018).

The further enhancement of the present approach based on Q-learning, its scalability is shown through its ability to cater to various delivery scenarios including different route complexities and distances. Such adaptability highlights its prospect of being used for several applications in the future, which targets different logistics environments (Suanpang, Jamjuntr, Jermsittiparsert, et al., 2022). Under various operational conditions, the system stability test showed that even with fleet expansion, effectiveness and routing efficiency were not compromised. The systematic parameter tuning that we performed, such as learning rate, discount factor and exploration strategies, was aimed to examine how these variables affect learning and convergence, and hence performance with a varying degree of tasks will be performed at an optimal level.

6. Study Implication

The implementation of the Q-learning-based approach from the point of view of last-mile delivery optimization demonstrated how the optimistic prospects for improving delivery time, energy, and routes may manifest themselves in the last-mile delivery logistics companies' world. The opportunities for improvements of their system and its expansion capabilities, indicate that the system can be efficiently incorporated in changing delivery settings and as the number of drones increases within a given period.

7. Practical Implications

The findings established that the routing system based on Q-learning improves delivery speeds by a margin of 15.1 % and energy efficiency by 8.4 % in comparison to the baseline method.

Delivery Time: Quicker delivery allows better service retention and is very important, especially with businesses that require delivery like the e-commerce and food businesses. Better delivery and order lead time will improve customer service since the delivery promises can be met if not exceeded. In instances of high demand, this may also mean an increased number of orders since the customers will select for a delivery service that does not secretarial for timeliness.

8. Statistical Significance

To clarify our findings, we will examine the statistical significance of the improvements observed. This will require confidence intervals and p-value determinations for the differentials between the Q-learning approach and the baseline. Such statistical evidence will furnish substantial weight to the warrants of improvements claimed, enabling the researcher to be in a position to evaluate the success of the results and help justify the overall objectives in the research.

9. Limitations and Future Study

In this part, explanation of the drawbacks of our approach and propose possible improvements considering the insights from the literature has been discussed. Accepting these limitations is not only necessary but also crucial, in delivering an utmost appraisal of the research at hand.

9.1 Convergence Issues in Q-Learning

Q-learning, however, has many a challenge and one of the common issues faced is convergence especially in non-static circumstances. Our study notes that regardless of certain parameters, the Q-learning algorithm will take time to reach an optimal policy in the presence of multiple states and actions as well as a changing environment volume. This problem can be possibly solved in subsequent studies, where less aggressive approaches would be introduced to avoid slow convergence and improve stability. Reward shaping, experience replay, or other advanced exploration

procedures such as epsilon-decay or UCB can be employed to resolve this problem.

9.2 Performance Parameters: Battery Life and Payloads

In any drone delivery system, energy consumption needs to remain as low as possible however it is necessary to consider the 'ceiling' which comes from factors like energy density and maximum load capacity. Our research indicates a better performance in energy efficiency but the real-world usability context remains limited with the drone battery time and payload weight. For long-distance dropping off, however, those factors can escrow off the practical feasibility of the system. For future battery management systems research, thin film enclosure and ultra-lightweight batteries or structural additive or new batteries will be required to increase the distance and the payload weight of the drones.

Disclosure Statement:

No potential conflict of interest was reported by the author(s).

Author Contributions

Conceptualization, P.S.; Research Design, P.S.; literature review, P.S. and P.J.; methodology, P.S. and P.J.; algorithms, P.S. and P.J.; software, P.S. and P.J.; validation, P.S. and P.J.; formal analysis, P.S. and P.J.; investigation, P.S. and P.J..; resources, P.S.; data curation, P.J.; writing original draft preparation, P.S. and P.J.; writing review and editing, P.S. and P.J..; visualization, P.S..; supervision, P.S..; project administration, P.S.; funding acquisition, P.S.. IRB

Funding

This work was supported in part by the Suan Dusit University under the Ministry of Higher Education, Science, Research and Innovation, Thailand, grant number FF67-Innovation of Tourism Learning Innovation Platform of Suphanburi Province.

Acknowledgments

The authors wish to express their gratitude to the Hub of Talent in Gastronomy Tourism Project (N34E670102), funded by the National Research Council of Thailand (NRCT), for facilitating research collaboration that contributed to this study. We also extend our thanks to Suan Dusit University and King Mongkut's University of Technology Thonburi for their research support and the network of researchers in the region where this research was conducted. Additionally, we are grateful to the Tourism Authority of Thailand (TAT) for providing essential data in the study areas.

References

Aboueleneen, N., Alwarafy, A., & Abdallah, M. (2023). Deep Reinforcement Learning For

Internet Of Drones Networks: Issues And Research Directions. *Ieee Open Journal Of The Communications Society*, *4*, 671-683. https://doi.org/10.1109/OJCOMS.2023.3251855

Alkouz, B., Shahzaad, B., & Bouguettaya, A. (2021). Service-based drone delivery. 2021 IEEE 7th International Conference on Collaboration and Internet Computing (CIC), 1665416254. https://doi.org/10.1109/CIC52973.2021.00019

Anastasiadou, K. (2021). Sustainable mobility driven prioritization of new vehicle technologies, based on a new decision-aiding methodology. *Sustainability*, *13*(9), 4760. https://doi.org/10.3390/su13094760

Aurambout, J.-P., Gkoumas, K., & Ciuffo, B. (2019). Last mile delivery by drones: An estimation of viable market potential and access to citizens across European cities. *European Transport Research Review*, *11*(1), 1-21. https://doi.org/10.1186/s12544-019-0368-2

Bakir, I., & Tiniç, G. Ö. (2020). Optimizing drone-assisted last-mile deliveries: The vehicle routing problem with flexible drones. *Optimization-Ouline. Org*, 1-28. https://optimization-online.org/wp-content/uploads/2020/04/7737.pdf

Bakogianni, M. A., & Malindretos, G. (2021). Last mile deliveries in the framework of urban distribution and supply chain management: review of best practices and conceptual framework. *Розвиток методів управління та господарювання на транспорті*, *2*(75), 38-64. https://doi.org/10.31375/2226-1915-2021-2-38-64

Chan, K. W., Nirmal, U., & Cheaw, W. G. (2018). Progress on drone technology and their applications: A comprehensive review. In *AIP conference proceedings* (Vol. 2030, No. 1). AIP Publishing. https://doi.org/10.1063/1.5066949

Cetin, E., Barrado, C., Muñoz, G., Macias, M., & Pastor, E. (2019). Drone navigation and avoidance of obstacles through deep reinforcement learning. In *2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)* (pp. 1-7). IEEE. https://doi.org/10.1109/DASC43569.2019.9081749

Chen, P., & Wang, Q. (2024). Learning for multiple purposes: A Q-learning enhanced hybrid metaheuristic for parallel drone scheduling traveling salesman problem. *Computers & Industrial Engineering*, *187*, 109851. https://doi.org/10.1016/j.cie.2023.109851

Chen, X., Ulmer, M. W., & Thomas, B. W. (2022). Deep Q-learning for same-day delivery with vehicles and drones. *European Journal of Operational Research*, *298*(3), 939-952. https://doi.org/10.1016/j.ejor.2021.06.021

Chiang, W. C., Li, Y., Shang, J., & Urban, T. L. (2019). Impact of drone delivery on sustainability and cost: Realizing the UAV potential through vehicle routing optimization. *Applied energy*, *242*, 1164-1175. https://doi.org/10.1016/j.apenergy.2019.03.117

Dolan, S. (2022). The challenges of last mile delivery logistics and the tech solutions cutting costs in the final mile. *Business Insider*. https://bit.ly/3XZmEd3

Elsayed, M., & Erol-Kantarci, M. (2018, October). Deep Q-learning for low-latency tactile applications: Microgrid communications. In 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm) (pp. 1-6). IEEE. https://doi.org/10.1109/SmartGridComm.2018.8587476

Engesser, V., Rombaut, E., Vanhaverbeke, L., & Lebeau, P. (2023). Autonomous delivery solutions for last-mile logistics operations: A literature review and research agenda. *Sustainability*, *15*(3), 2774. https://doi.org/10.3390/su15032774

Eskandaripour, H., & Boldsaikhan, E. (2023). Last-mile drone delivery: Past, present, and future. Drones, 7(2), 77. https://doi.org/10.3390/drones7020077

Fotouhi, A., Ding, M., & Hassan, M. (2021). Deep q-learning for two-hop communications of drone base stations. *Sensors*, *21*(6), 1960. https://doi.org/10.3390/s21061960

Ghosh, M., Kuiper, A., Mahes, R., & Maragno, D. (2023). Learn global and optimize local: A data-driven methodology for last-mile routing. *Computers & Operations Research*, *159*, 106312. https://doi.org/10.1016/j.cor.2023.106312

Gómez-Lagos, J., Candia-Véjar, A., & Encina, F. (2021). A new truck-drone routing problem for parcel delivery services aided by parking lots. *IEEE Access*, *9*, 11091-11108. https://doi.org/10.1109/ACCESS.2021.3050658

Goodchild, A., & Toy, J. (2018). Delivery by drone: An evaluation of unmanned aerial vehicle technology in reducing CO2 emissions in the delivery service industry. *Transportation Research Part D: Transport and Environment*, *61*, 58-67. https://doi.org/10.1016/j.trd.2017.02.017

Haider, S. K., Nauman, A., Jamshed, M. A., Jiang, A., Batool, S., & Kim, S. W. (2022). Internet of drones: Routing algorithms, techniques and challenges. Mathematics, 10(9), 1488. https://doi.org/10.3390/math10091488

Hamdi, A., Salim, F. D., Kim, D. Y., Neiat, A. G., & Bouguettaya, A. (2021). Drone-as-a-service composition under uncertainty. *IEEE Transactions on Services Computing*, *15*(5), 2685-2698. https://doi.org/10.1109/TSC.2021.3066006

Hardy, A., Haji, K., Abbas, F., Hassan, J., Ali, A., Yussuf, Y., ... & Worrall, E. (2023). Cost and quality of operational larviciding using drones and smartphone technology. Malaria Journal, 22(1), 286. https://doi.org/10.1186/s12936-023-04713-0

Huda, S. A., & Moh, S. (2023). Deep reinforcement learning-based computation offloading in uav swarm-enabled edge computing for surveillance applications. *IEEE Access*. https://doi.org/10.1109/ACCESS.2023.3292938

Jana, S., & Mandal, P. S. (2023). Approximation algorithms for drone delivery scheduling problem. International Conference on Networked Systems, 125-140. https://doi.org/10.1007/978-3-031-37765-5_10

Jeyaraman, J., Malaiyappan, J. N. A., & Sistla, S. M. K. (2024). Advancements in Reinforcement Learning Algorithms for Autonomous Systems. *International Journal of Innovative Science and Research Technology (IJISRT)*, *9*(3), 1941-1946. https://hcommons.org/deposits/item/hc:68621

Juan, A. A., Freixes, A., Panadero, J., Serrat, C., & Estrada-Moreno, A. (2020). Routing drones in smart cities: A biased-randomized algorithm for solving the team orienteering problem in real time. Transportation Research Procedia, 47, 243-250. https://doi.org/10.1016/j.trpro.2020.03.095

Kumar, A., Sharma, K., Singh, H., Naugriya, S. G., Gill, S. S., & Buyya, R. (2021). A drone-based networked system and methods for combating coronavirus disease (COVID-19) pandemic. *Future Generation Computer Systems*, *115*, 1-19. https://doi.org/10.1016/j.future.2020.08.046

Lee, W., Shahzaad, B., Alkouz, B., & Bouguettaya, A. (2024). Reactive Composition of UAV Delivery Services in Urban Environments. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/TITS.2024.3392914

Li, F., & Kunze, O. (2023). A comparative review of air drones (UAVs) and delivery bots (SUGVs) for automated last mile home delivery. *Logistics*, *7*(2), 21. https://doi.org/10.3390/logistics7020021

Li, J., Shen, D., Yu, F., & Zhang, R. (2023). Air channel planning based on improved deep Q-learning and artificial potential fields. *Aerospace*, *10*(9), 758. https://doi.org/10.3390/aerospace10090758

Li, X., Gong, L., Liu, X., Jiang, F., Shi, W., Fan, L., ... & Xu, J. (2022). Solving the last mile problem in logistics: A mobile edge computing and blockchain-based unmanned aerial vehicle delivery system. *Concurrency and Computation: Practice and Experience*, *34*(7), e6068. https://doi.org/10.1002/cpe.6068

Marques, E. L., Coelho, V. N., Coelho, I. M., Frota, Y. A. d. M., Koochaksaraei, R. H., Ochi, L. S., & Coelho, B. N. (2022). UAVs routes optimization on smart cities and regions. *RAIRO-Operations Research*, *56*(2), 853-869. https://doi.org/10.1051/ro/2022036

Memari, M., Shakya, P., Shekaramiz, M., Seibi, A. C., & Masoum, M. A. (2024). Review on the advancements in wind turbine blade inspection: Integrating drone and deep learning technologies for enhanced defect detection. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3371493

Mukhamediev, R. I., Symagulov, A., Kuchin, Y., Zaitseva, E., Bekbotayeva, A., Yakunin, K., ... & Tabynbaeva, L. (2021). Review of some applications of unmanned aerial vehicles technology in the resource-rich country. Applied Sciences, 11(21), 10171. https://doi.org/10.3390/app112110171

Nayyar, A., Nguyen, B. L., & Nguyen, N. G. (2020). The internet of drone things (IoDT): future envision of smart drones. In First International Conference on Sustainable Technologies for Computational Intelligence: Proceedings of ICTSCI 2019 (pp. 563-580). Springer Singapore. https://doi.org/10.1007/978-981-15-0029-9_45

Obiuto, N. C., Festus-Ikhuoria, I. C., Olajiga, O. K., & Adebayo, R. A. (2024). Reviewing The Role Of Ai In Drone Technology And Applications. *Computer Science & IT Research Journal*, *5*(4), 741-756. https://doi.org/10.51594/csitrj.v5i4.1019

Puente-Castro, A., Rivero, D., Pedrosa, E., Pereira, A., Lau, N., & Fernandez-Blanco, E. (2024). Q-learning based system for path planning with unmanned aerial vehicles swarms in obstacle environments. Expert Systems with Applications, 235, 121240. https://doi.org/10.1016/j.eswa.2023.121240

Shah, I. A., Laraib, A., Ashraf, H., & Hussain, F. (2024). Drone Technology: Current Challenges and Opportunities. *Cybersecurity Issues and Challenges in the Drone Industry*, 343-361. https://www.igi-global.com/chapter/drone-technology/340083

Shao, Q., & Cheng, S.-F. (2023). Preference-Aware Delivery Planning for Last-Mile Logistics. *arXiv preprint arXiv:2303.04333*. https://doi.org/10.48550/arXiv.2303.04333

Silva, M., & Pedroso, J. P. (2022). Deep reinforcement learning for crowdshipping last-mile delivery with endogenous uncertainty. *Mathematics*, *10*(20), 3902. https://doi.org/10.3390/math10203902

Suanpang, P., & Jamjuntr, P. (2024). Optimizing Autonomous UAV Navigation with D* Algorithm for Sustainable Development. *Sustainability*, *16*(17), 7867. https://doi.org/10.3390/su16177867

Suanpang, P., Jamjuntr, P., Jermsittiparsert, K., & Kaewyong, P. (2022). Tourism service scheduling in smart city based on hybrid genetic algorithm simulated annealing algorithm. *Sustainability*, *14*(23), 16293. https://doi.org/10.3390/su142316293

Suanpang, P., Jamjuntr, P., Kaewyong, P., Niamsorn, C., & Jermsittiparsert, K. (2022). An

intelligent recommendation for intelligently accessible charging stations: Electronic vehicle charging to support a sustainable smart tourism city. *Sustainability*, *15*(1), 455. https://doi.org/10.3390/su15010455

Tausif, I. (2023). Last mile delivery Optimisation model for drone-enabled Vehicle Routing Problem. *Emerging Minds Journal for Student Research*, *1*, 39-73. https://doi.org/10.59973/emjsr.11

Troudi, A., Addouche, S. A., Dellagi, S., & Mhamedi, A. E. (2018). Sizing of the drone delivery fleet considering energy autonomy. *Sustainability*, *10*(9), 3344. https://doi.org/10.3390/su10093344

Tu, Y.-J., & Piramuthu, S. (2023). Security and privacy risks in drone-based last mile delivery. *European Journal of Information Systems*, 1-14. https://doi.org/10.1080/0960085X.2023.2214744

Tufail, B., & Akhtar, S. (2022). The Impact of Sustainability on Logistics Excellence. *Pakistan Journal of Humanities and Social Sciences*, *10*(4), 1522–1532-1522–1532. https://doi.org/10.52131/pjhss.2022.1004.0309

Van Duin, R., Enserink, B., Daleman, J., & Vaandrager, M. (2020). The near future of parcel delivery: Selecting sustainable solutions for parcel delivery. In *Sustainable City Logistics Planning: Methods and Applications* (pp. 219-252). Nova Science Publishers. https://research.tudelft.nl/en/publications/the-near-future-of-parcel-delivery-selecting-sustainable-solution

Wang, Z., Yang, H., Wu, Q., & Zheng, J. (2021). Fast path planning for unmanned aerial vehicles by self-correction based on q-learning. *Journal of Aerospace Information Systems*, *18*(4), 203-211. https://doi.org/10.2514/1.I010856

Zhang, J., Campbell, J. F., Sweeney II, D. C., & Hupman, A. C. (2021). Energy consumption models for delivery drones: A comparison and assessment. *Transportation Research Part D: Transport and Environment*, *90*, 102668. https://doi.org/10.1016/j.trd.2020.102668

Zhou, Z., Wan, M., Zhou, T., & Niu, B. (2024). A Hyper-heuristic Algorithm Based on Q-Learning for 3D Drone Trajectory Planning. In *International Conference on Swarm Intelligence* (pp. 46-57). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-97-7184-4_5