Cite this article
Yeon C, Cho A, Kim S, Lee Y and Lee S
Real-time dynamic route generation algorithm of DRT with deep Q-learning.
Proceedings of the Institution of Civil Engineers – Municipal Engineer,
https://doi.org/10.1680/jmuen.24.00082

Research Article
Paper 2400082
Received 30/12/2024;
Accepted 02/04/2025

Municipal Engineer

Real-time dynamic route generation algorithm of DRT with deep Q-learning

Chihyeong Yeon
National Infrastructure and Geospatial Information Research Division, Korea Research Institute for Human Settlements (KRIHS), Sejong, Republic of Korea

Ara Cho
Department of Transportation Engineering, University of Seoul, Seoul, Republic of Korea; Department of Smart Cities, University of Seoul, Seoul, Republic of Korea

Sion Kim
Department of Transportation Engineering, University of Seoul, Seoul, Republic of Korea; Department of Smart Cities, University of Seoul, Seoul, Republic of Korea

Yongryeong Lee
Department of Transportation Engineering, University of Seoul, Seoul, Republic of Korea; Department of Smart Cities, University of Seoul, Seoul, Republic of Korea

Seungjae Lee
Department of Transportation Engineering, University of Seoul, Seoul, Republic of Korea (corresponding author: sjlee@uos.ac.kr)

Demand-responsive transportation (DRT) offers a flexible solution for transportation issues in areas with insufficient public transit. Unlike fixed-route buses, DRT dynamically adjusts routes based on user demand. However, without effective route optimisation, DRT can underperform compared with fixed-route buses. This study proposes a method to reduce passenger waiting and boarding times by applying a deep Q-network (DQN) algorithm to establish dynamic routes for semi-dynamic DRT systems in urban residential areas. A simulation was conducted to compare the performance of fixed-route and dynamic-route systems, analysing how changes in passenger demand affect waiting and travel times. Results indicate that dynamic routes optimised by the DQN algorithm achieved higher boarding and alighting rates across all demand levels compared with fixed routes. In addition, even under high demand, dynamic-route DRTs reduced waiting and travel times, demonstrating superior efficiency. These findings confirm that dynamic-route DRTs enhance service quality and operational performance in residential areas.

Keywords: deep Q-learning/demand-responsive transportation/traffic engineering/transport management/transport planning/UN SDG 9: Industry, innovation and infrastructure

Notation

L	number of alighted passengers
M	number of boarded passengers
N	number of stops
P_i	number of unboarded passengers at stop *i*
$\overline{T}$	average value of total network travel time
t_i	time (in minutes) passengers have spent waiting at stop *i*
t_j	time (in minutes) from the generation of passenger j until boarding

1. Introduction

Suburban and rural areas are known to suffer from inefficiencies in public transportation systems due to declining user populations, increasing operational costs, and insufficient transportation networks (Porru *et al*., 2020). In particular, maintaining a balance between service levels and operational costs has become a major challenge as public transport usage continues to decline and operating expenses rise in low-density areas (Berrada and Poulhès, 2021; Sommer and Deutsch, 2021). In addition, in recent years, new towns have emerged as a result of urban redevelopment and decentralisation, leading to frequent mobility challenges due to a lack of large-scale transportation networks (Liu *et al*., 2023; Tarakci and Turk, 2022; Wu *et al*., 2019). These issues have been further exacerbated by the disruptions caused by the COVID-19 pandemic, which has significantly undermined the efficiency and reliability of existing public transportation systems (Jang *et al*., 2023; Ku *et al*., 2023; Kwak *et al*., 2024). Various strategies have been proposed to revitalise urban environments and enhance the resilience of public transportation systems. However, conventional measures such as reducing bus frequencies or increasing fares have limitations, often causing confusion and inconvenience for users (Dockendorf *et al*., 2000).

As an alternative to address these limitations, demand-responsive transportation (DRT) systems have gained attention (Mulley and Nelson, 2009). Unlike conventional public transport, DRT does

not adhere to fixed routes or schedules but dynamically adjusts travel paths in response to real-time passenger demand. This system is particularly valuable in areas where traditional bus and rail services fail to provide adequate coverage and has been identified as a solution for improving accessibility in new towns, suburban, and rural regions (Coutinho *et al*., 2020; Vansteenwegen *et al*., 2022). In addition, by minimising unnecessary travel time and distance, DRT enhances operational efficiency and reduces costs for public transport providers (Liu *et al*., 2024). Therefore, rather than scaling down existing systems, adopting innovative operational models such as DRT is essential. In response, various research initiatives and pilot programmes have been conducted to mitigate public service operating deficits while ensuring long-term sustainability (Barrett *et al*., 2019).

This study aims to maximise the operational efficiency of DRT systems through route optimisation. Currently, DRT systems rely on fixed routes or have limited flexibility, making them less responsive to real-time demand changes. As a result, inefficiencies such as unnecessary empty vehicle trips, increased waiting time, and reduced service quality arise. Therefore, it is necessary to develop a more efficient DRT operational strategy.

The specific objective of this study is to develop an algorithm that dynamically optimises routes based on deep Q-network (DQN) in response to real-time passenger demand, including waiting passengers. This algorithm aims to minimise passenger waiting and travel times while improving boarding and alighting efficiency. In this context, DQN is an advanced reinforcement learning technique compared with traditional Q-learning, enabling real-time route optimisation in DRT. In addition, this study investigates whether the proposed system can provide a service level comparable with or even capable of replacing traditional fixed-route buses within local community networks. Therefore, to validate the effectiveness of the proposed algorithm, its performance will be analysed in comparison with fixed-route bus systems, evaluating key metrics such as waiting time, service rate, and boarding and alighting performance.

Specifically, Section 2 provides an overview of the concepts and classifications of DRT systems and reviews existing studies on the optimisation of operating systems in DRT and DQN-based decision-making strategies. This review identifies the distinguishing elements of the proposed study compared with previous research. Section 3 describes the experimental design, including the construction of a DRT test network and the process of generating passenger demand data. In addition, it outlines the reinforcement learning environment based on the DQN algorithm, detailing the reward function design and hyperparameter settings. Section 4 presents an empirical comparison between the proposed DQN-based algorithm and fixed-route public transportation systems. The evaluation focuses on boarding and alighting performance, service rate, waiting time, and total service time, providing insights into the operational characteristics of both systems. Section 5 summarises the key findings of the study and discusses the implications of the proposed dynamic-route generation algorithm. In addition, it suggests future research directions to enhance the scalability and practical applicability of the algorithm.

2. Literature review

In this section, we review previous studies related to DRT. First, we explore the emergence, types, routes, and advantages of DRT, highlighting its evolution and operational frameworks. Understanding the key characteristics and benefits of DRT provides valuable insights into improving public transportation systems and addressing diverse passenger demands while enhancing service efficiency and flexibility. Next, we examine the dynamic-route generation algorithm, which is essential for optimising DRT operations. This algorithm plays a crucial role in reducing passenger waiting and travel times by adapting routes based on real-time demand. Finally, we explore decision-making strategies based on DQN. The application of DQN in DRT enables dynamic and efficient routing by learning optimal dispatch strategies through reinforcement learning.

2.1 Overview of DRT: emergence, types, and benefits

DRT is a system that can operate flexibly according to demand on a fixed route and without a set schedule, and it can be said to be a form that combines the economic feasibility of buses and the flexibility of taxis (Nelson *et al*., 2004). DRT began in the early twentieth century in the USA in the form of Jitney, which picks up and drops off passengers in vans for a certain fee (Khattak and Yim, 2004). At that time, Jitney was a popular means of transportation that transported short-distance passengers quickly and inexpensively, but it gradually disappeared for political and economic reasons. Since then, the concept of Dial-a-Ride, a reservation system through phone, has emerged, expanding the possibility of DRT (Deka *et al*., 2023). In this process, the need for advanced scheduling and dispatching technologies has emerged, and DRT is currently playing an important role as a private or semi-public transportation system by establishing itself as a type of shared mobility service (Martí *et al*., 2023).

For the efficient operation of such DRT, Sungatek *et al*. (2023) emphasised the need to distinguish service provision methods by route type and termination type and established the type system of DRT as shown in Table 1.

This DRT addresses the time and space limitations of existing public transportation by offering high flexibility. Its effectiveness is particularly notable in areas with limited public transit, where it enhances community accessibility and supports resident mobility, contributing to local community development and improved quality of life (Logan, 2007). In addition, DRT provides a more affordable

Table 1. Classification and characteristics of DRT service types (Sungatek *et al*., 2023)

Type		Classification		
Main category	Subcategory	Boarding/alighting stops	Operation route	Operation timetable
Fixed	Shuttle	Fixed (only origin and destination exist)	Fixed	Fixed, flexible
	Route	Fixed (fixed origin and destination, multiple passenger designation)	Fixed	Fixed, flexible
Flexible fixed	Out-of-route	Semi-fixed (fixed origin and destination, some on-demand allowed)	Fixed (main route or some detour allowed)	Fixed, flexible
	Adapted route	Semi-fixed (fixed origin and destination, some on-demand allowed)	Flexible (some detour with optimised routing after)	Fixed, flexible
Semi-dynamic		Flexible (only origin and destination are fixed)	Flexible	Fixed, flexible
Feeder service		Semi-fixed (origin or destination is public transit stops)	Flexible (designated area only)	Flexible
Dynamic	Route-based	Flexible	Semi-fixed (some detour allowed)	Flexible
	Area-based	Flexible	Flexible (designated area service)	Flexible
	Unconstrained	Flexible	Flexible	Flexible

alternative to personal vehicle travel, which can help reduce greenhouse gas emissions (Bürstlein *et al*., 2021).

The primary goals of DRT are to minimise operating costs through flexible service provision and to maximise user satisfaction by optimising waiting and travel time (Campisi *et al*., 2021). Accordingly, this paper proposes a dynamic-route generation algorithm for DRT and assesses its service quality by comparing it with that of a fixed-route bus system. We evaluate the service quality of the proposed DRT system using service rate and total service time. Service rate measures the proportion of total demand successfully served, while total service time includes both waiting time until boarding and travel time until alighting, providing a comprehensive assessment of service efficiency.

2.2 Optimisation of operating systems in DRT

Various attempts have been made worldwide to optimise public transportation schedules and operations. In this context, DRT systems are gaining attention for their ability to dynamically adapt to real-time demand, making the optimisation of vehicle dispatch and operational planning a critical challenge.

Accordingly, Zhou *et al*. (2024) developed an optimisation model for reservation-based DRT operations that considers multiple vehicle sizes and travel distances, utilising the ant colony optimisation algorithm to enhance dispatch planning. A case study on the Shanghai highway route demonstrated that the proposed model effectively optimised the number of vehicles, reducing total empty travel distance by 23.34%, thereby contributing to cost reduction and improved resource utilisation.

Unlike reservation-based approaches, real-time DRT systems must account for both scheduled passengers and immediate requests. Huang *et al*. (2020) proposed an optimisation method to address the dynamic vehicle dispatch problem, considering both on-demand and scheduled requests. Their study introduced a constrained spatio-temporal value function to estimate vehicle value and applied a randomised best-fit and online planning algorithm for dispatch optimisation. Experimental results using real-world ride-hailing service data showed that the proposed approach outperformed existing dispatch algorithms in operational efficiency and profitability, particularly in complex environments with multiple vehicles. Shen *et al*. (2021) introduced a path-based dynamic vehicle dispatch strategy to effectively handle real-time reservation requests in DRT systems. To optimise dispatch by incorporating future demand patterns, they employed an approximate dynamic programming approach, demonstrating that this method significantly improved vehicle operational efficiency and revenue generation. A case study using real-world DRT data from Qingdao, China, showed that their approach increased total operational revenue by 6.03% to 23.30% compared with conventional myopic dispatch methods.

Beyond these traditional optimisation techniques, recent research has increasingly focused on leveraging reinforcement learning for real-time DRT operations. Zhang and Li (2024) proposed a graph prompt learning (GPL) approach to optimise vehicle dispatch in DRT systems. Their study addressed the generalisation limitations of pre-trained models by incorporating structural and task-specific tokens, enabling adaptation to varying scales and demand changes. Experimental validation using data from Wangjing, Beijing, demonstrated that their approach significantly reduced operational costs and computation time while maintaining stability in large-scale dynamic demand scenarios.

The key constraints considered in the aforementioned studies are summarised in Table 2.

According to Table 2, existing studies have primarily focused on optimisation based on vehicle capacity constraints, time constraints, and passenger destination considerations. This review has revealed that these studies have been conducted mainly based on reserved or predicted demand while lacking mechanisms to incorporate real-time waiting passengers into the decision-making process. Therefore, to enhance passenger satisfaction and improve the effectiveness of DRT operations, an approach that dynamically integrates real-time waiting passengers and their destination information will be essential in advancing future DRT systems.

2.3 Decision-making strategy based on DQN

Recent advancements in artificial intelligence and algorithms have greatly helped solve complex scheduling and location optimisation challenges in transportation systems (Bencekri *et al*., 2024; Choi *et al*., 2024a). Among these advancements, DQN has emerged as a reinforcement learning algorithm that integrates Q-learning with deep neural networks, enabling efficient learning in high-dimensional and continuous state spaces (Lai *et al*., 2024a; Ye *et al*., 2024).

Q-learning is a model-free reinforcement learning method that predicts the expected total reward for selecting a specific action in a given state based on the Q-function (Watkins and Dayan, 1992). The agent utilises a Q-table to compare the Q-values of all possible actions in the current state and selects the optimal action while adjusting the balance between exploration and exploitation during the learning process to derive the optimal policy (Sutton, 2018). Here, the policy refers to the strategy an agent employs to select the optimal action in a given state. However, as the state and action spaces expand, the Q-table grows exponentially, leading to high memory requirements and increased computational complexity (Jang *et al*., 2019; Shoufeng *et al*., 2008).

To address this limitation, DQN approximates the Q-value function using neural networks, allowing for effective learning even in large state-action spaces (Mnih *et al*., 2015). Unlike conventional Q-learning, which updates only the states it has encountered, DQN leverages neural networks to estimate Q-values for unvisited states, enhancing generalisation and decision-making efficiency. DQN incorporates three key components: a neural network, a target network, and experience replay memory. It employs convolutional neural networks to process the current state as input and approximate Q-values for various actions to determine the optimal choice. To improve training stability, a target network is introduced, which updates periodically, thereby reducing variability in target values and ensuring stable learning. In addition, experience replay memory is utilised to randomly sample past experiences, mitigating data correlations and enhancing generalisation (Moreno-Malo *et al*., 2024).

Accordingly, DQN has emerged as a versatile tool, effectively applied across numerous industries (Jang *et al*., 2019). Recently, its application has expanded to demand-responsive public transit systems, where it is predominantly used for operating dynamic-route buses, enabling real-time optimisation of routes and schedules. Wu *et al*. (2021) effectively addressed the dynamic dispatch problem within the demand-responsive public transit system by utilising the DQN algorithm. In this system, the DQN algorithm analysed passenger departure and destination data to determine the most suitable vehicle type (small, medium, or large buses) and enhanced service efficiency by minimising unnecessary stops. Vehicle-passenger matching was dynamically conducted by incorporating various cost factors, including estimated passenger waiting and travel times. Consequently, the system's overall operational efficiency was significantly improved, leading to an average reduction in passenger waiting time by over 23%, optimised travel time, and a notable decrease in operating costs. Similarly, Ai *et al*. (2022) applied DQN to model bus dispatch optimisation as a sequential decision-making process. Actions like 'departure' and 'no departure' were determined every minute, dynamically adjusting dispatch intervals based on real-time passenger demand and traffic conditions. Key

Table 2. Comparison of research approaches in DRT system operations

Study	Research focus	Constraints	Passenger routing strategy
Zhou *et al*. (2024)	Minimising vehicle count and travel distance by way of ACO	Pickup and drop-off time constraints, vehicle capacity	Fixed pickup and drop-off points based on reservations
Huang *et al*. (2020)	Balancing scheduled and real-time requests using a spatio-temporal value function	Dynamic time windows, vehicle availability	Immediate routing based on real-time requests
Shen *et al*. (2021)	Predicting future demand and optimising dispatch with approximate dynamic programming	Future demand response time, vehicle capacity restrictions	Dynamic routing integrating predicted demand patterns
Zhang and Li (2024)	Real-time adaptive route generation using GPL	Real-time scheduling to minimise idle time and optimise vehicle load	Reinforcement learning-based adaptive routing for efficiency

state variables, including load factor, waiting passengers, and capacity utilisation, enabled real-time operational adjustments. The reward function minimised passenger waiting time, reduced empty seats, and addressed unmet demand. Reinforcement learning utilised experience replay and a target network to enable stable training, demonstrating that the DQN-based approach could surpass traditional fixed dispatch methods in both operational efficiency and service quality.

In addition to its applications in transportation, the DQN algorithm has proven effective in addressing optimisation challenges across various engineering and technical domains. Cui (2024) proposed an optimisation algorithm combining DQN with a dynamic feedback mechanism to address the limitations of existing multitask scheduling algorithms, such as slow convergence and high computational costs. The algorithm leverages deep reinforcement learning to efficiently navigate high-dimensional state spaces while dynamically adjusting learning rates and exploration strategies to adapt to real-time environmental changes. It demonstrated that the DQN-based approach outperforms traditional algorithms, including ant colony, particle swarm, and adaptive genetic algorithms, by achieving faster convergence and a shorter execution time. This highlights its capability to ensure both efficiency and adaptability in multitask scheduling. Yang *et al*. (2020) used the DQN algorithm for comparative purposes to derive the optimal performance of the scheduling Q-learning algorithm for the robot-based unmanned warehouse delivery system of the 'goods to people' model. Starting from the problem of basic multi-robot route planning, a simulation was planned to solve the two disadvantages of the robot route planning problem, slow convergence and excessive randomness, and the reward function according to the arrival time was calculated. As a result, it was shown that the DQN algorithm converges faster than the existing Q-learning algorithm and learns solutions to path planning problems faster.

Consequently, DQN has been shown to efficiently explore optimal solutions even in complex environments and to operate effectively in various decision-making processes. In contrast, Q-learning methods rely on pre-learned data, making them less capable of responding immediately to dynamically changing environments. As a result, in public transportation systems where external factors cause real-time variations, the increased computational burden makes it difficult to ensure optimal performance, and large-scale service applications may also face limitations (Lai *et al*., 2024b; Rahmani *et al*., 2022). On the other hand, DQN can dynamically adjust its actions based on real-time data, suggesting that it can serve as an effective decision-making tool in evolving public transportation systems.

3. Methodology

This study aims to develop a real-time, dynamic path generation algorithm tailored for DRT services within a defined residential area. To achieve this, DQN is utilised, as discussed in Section 2, due to its capability to learn stably even in complex state–action spaces. This characteristic makes it particularly effective in addressing stop selection problems where multiple variables influence decision making.

Accordingly, this study employs the DQN algorithm to optimise dynamic-route configurations by learning from real-time passenger demand. Specifically, the algorithm determines routes across all stops based on real-time demand and dynamically adjusts paths by considering the passenger waiting demand from all stops to the destination, which has not been explicitly considered in previous studies. This approach ensures adaptive responsiveness to changing demand conditions.

The study framework is depicted in Figure 1, illustrating the methodological flow and procedural steps of the algorithm's development and implementation process.

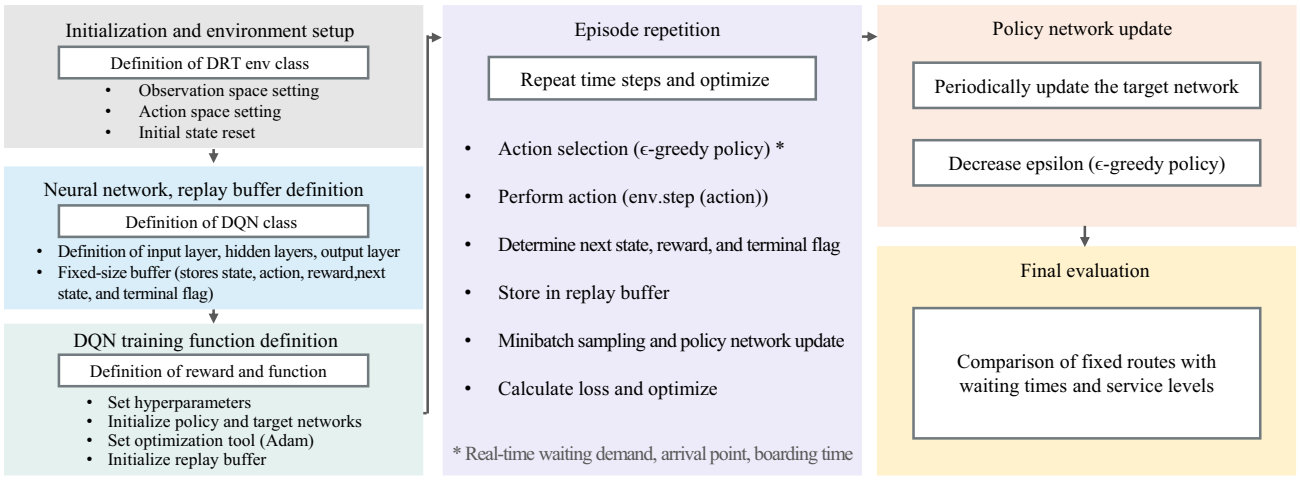

Figure 1. Framework of study

3.1 Construction of the DRT test network

In this study, Sejong City, South Korea, where DRT services are currently in operation within a defined residential area, was selected as the test network for evaluating the DRT operation planning methodology. Sejong City, a newly developed urban area aimed at decentralising the capital region's population, is still in the early stages of infrastructure development and lacks a fully established public transportation network. Analysis of transportation card data indicates that existing bus routes only meet about 10% of the overall transit demand, resulting in low operational efficiency, a gap that the introduction of DRT aims to address.

To optimise the DRT operation plan through reinforcement learning, each stop provides real-time passenger waiting information, which includes destination details due to the demand-responsive nature of the DRT system. The test network used in this study for establishing the DRT operational framework comprises eight designated stops within a road network extracted from OpenStreetMap data. All stop and depot locations are precisely defined based on geographic coordinates, allowing for an accurate and coordinated operational setup (Figure 2).

The distance from each departure point to the destination was computed based on the actual road length, ensuring route accuracy. The algorithm developed in this study does not include predefined knowledge of stop locations within the test network; instead, it operates with data on the travel time required between each pair of stops. That is, the value shown in Table 3 represents the actual time required for the DRT service to travel from one stop to another.

Passengers arrive at each of the eight stops per minute and enter their destinations by way of an application or through the infrastructure

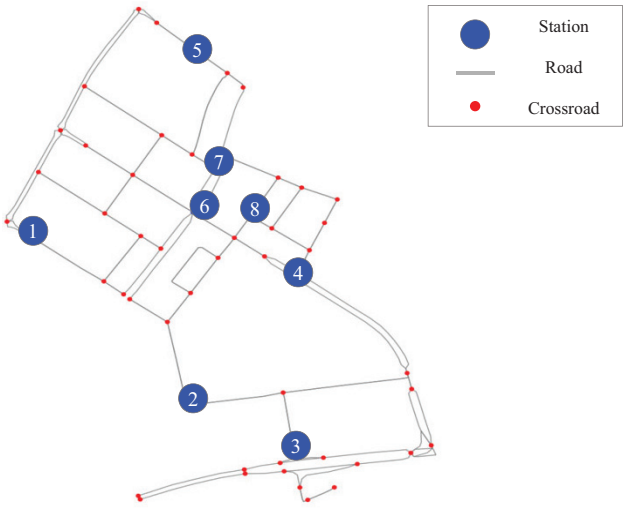

Figure 2. DRT test network structure

Table 3. Travel time matrix between stops within the test network

		Departure: min							
		1	2	3	4	5	6	7	8
Arrival	1	—	10	10	9	7	5	7	7
	2	7	—	2	3	8	3	7	5
	3	9	2	—	3	10	5	9	6
	4	7	4	4	—	8	3	7	3
	5	5	9	9	8	—	4	3	6
	6	4	5	5	4	5	—	4	2
	7	5	6	6	5	3	1	—	3
	8	6	5	5	3	7	2	6	—

provided at the bus stops. Consequently, the algorithm must process inputs that include the current bus location, the number of passengers on board, and the number of passengers waiting at each stop, allowing for efficient route and dispatch optimisation.

3.2 Demand scenario generation and decision-making process

In this study, the DQN algorithm for managing DRT operations dynamically accounts for the real-time number of passengers waiting at each stop and the current load of passengers on the bus. Consequently, data are generated on a minute-by-minute basis to reflect the number of arrivals at each stop, incorporating these details into the algorithm's observation space to inform subsequent operational decisions, as illustrated in Figure 3. This observation space includes not only real-time wait counts at each stop and passenger arrival locations but also factors in variations based on weekday against weekend patterns, daily changes, and seasonal influences, enabling the algorithm to respond adaptively to different temporal demand characteristics.

This approach enables the DQN algorithm to adaptively learn the variability in passenger demand across different conditions, thereby forming a robust foundation for effective decision making in future dynamic route planning for DRT systems. In this study, the DQN algorithm was trained by modifying the frequency of call requests generated per episode, applying this to both fixed-route and dynamic-route bus types to observe and analyse the resulting performance.

By training the DQN algorithm across a range of demand scenarios and repeatedly learning from diverse datasets for each episode, as shown in Table 4, this methodology prepares the algorithm to respond flexibly across various operational contexts. This adaptability is crucial for developing a DRT system capable of addressing unpredictable changes in demand rather than relying on fixed or predictable patterns. Within this framework, the dynamic-route generation algorithm determines the subsequent stop based on the current bus location, onboard passenger count, the number of waiting passengers at each stop, and the respective destination of

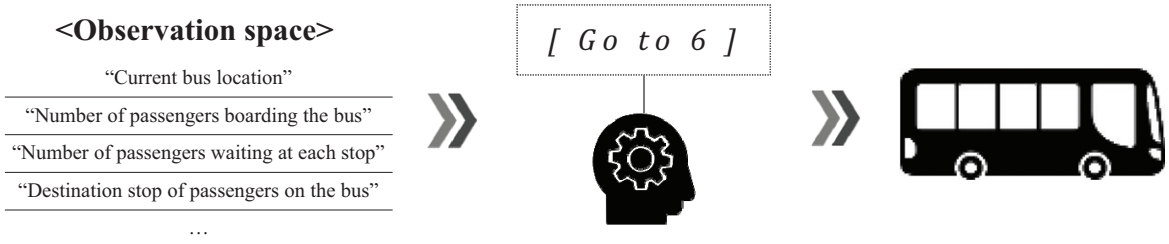

Figure 3. Decision-making process of the DQN algorithm

Table 4. Scenario types and number of generated passengers (example)

	Fixed-route bus								Dynamic-route DRT							
Generated passengers	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
6 passengers/h	—	1	1	—	3	—	1	—	1	—	1	—	2	—	2	1
8 passengers/h	1	—	2	3	—	1	—	1	—	1	1	2	1	1	2	—
…	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—
22 passengers/h	5	5	—	4	—	3	2	3	3	—	5	1	3	2	3	5
24 passengers/h	—	3	3	3	4	3	4	1	1	1	3	7	8	1	1	2

each passenger on board, ensuring real-time responsiveness and route optimisation.

3.3 DQN-based reinforcement learning framework for DRT operations

3.3.1 Reinforcement learning environment configuration

In this study, an environmental class was developed to configure the DQN reinforcement learning environment. This class establishes the foundational structure of reinforcement learning, functioning as a crucial module that defines temporal progression, passenger count at each stop, DRT departure and arrival processes, and boarding procedures, thereby generating corresponding outcomes.

To utilise the test network structure outlined in Section 3.1, the network comprises eight stops, with each assigned to a separate destination list. To ensure analytical accuracy, passengers generated at a stop were restricted from returning to the same location, preventing excessive passenger generation that could skew result interpretation. For determining the DRT's subsequent action (i.e. next stop), which is central to this study's goal of dynamic-route operation planning, the neural network continuously learns from DRT stop operations and outcomes over a repeated 1-h episode.

The status data includes a real-time list of waiting passengers at each stop, the current bus location, onboard passenger count, and the destinations reserved by passengers. Notably, the model does not precisely predict the actual passenger count on each route, as additional passengers may arrive at stops during travel time. This approach is designed to balance model training time with real-world operational demands, enhancing the model's practical applicability for DRT route optimisation.

Once the departure and operation plans are established, each DRT vehicle will operate across all eight designated stops as directed by the algorithm, utilising OpenStreetMap coordinates for routing within the network. Passengers will get off at their designated stops after boarding. The DRT vehicles will proceed through an operational loop, mirroring real-world processes: performing an arrival matrix check, facilitating passenger drop-off, onboarding new passengers at each stop, and then departing for the next stop. Throughout each cycle, detailed data will be recorded on passenger boarding, alighting, and overall DRT operation, capturing a comprehensive dataset for analysis.

3.3.2 Reward function configuration for DQN operational algorithm

In reinforcement learning algorithms, the performance of the learning model is highly influenced by the configuration of the reward function. In this study, the reward function is designed to appropriately reward the algorithm based on passenger waiting time, boarding, and alighting activities. The reward structure consists of three components, each optimised through multiple iterative trials to enable the bus to effectively meet demand in a balanced manner.

First, the algorithm is assigned a stop-specific penalty reward for unboarded passengers, thereby consistently penalising passenger waiting time to minimise it as much as possible. Accordingly, the expression for the stop-specific penalty (negative reward) for unboarded passengers is defined as follows:

$$R_{unboarded} = \quad -\sum_{i=1}^{N} P_i \times t_i$$

where N is the number of stops, P_i is the number of unboarded passengers at stop i, and t_i is the time (in minutes) passengers have spent waiting at stop i.

Equation 1. Stop-specific penalty for unboarded passengers

The algorithm is designed to incentivise boarding by using the average stop access time across all stops, allowing as many passengers as possible to board. However, an additional incentive, scaled by a factor of 10, is applied based on a 10-min service standard. This standard reflects the maximum acceptable waiting time suggested by Arhin *et al*. (2019). Consequently, the reward function based on the average network travel time upon passenger boarding can be expressed as follows:

$$R_{board} = \sum_{j=1}^{M} [10 \times \overline{T} \times 1(t_j \leq 10) + \overline{T} \times 1(t_j > 10)]$$

where M is the number of boarded passengers, $\overline{T}$ is the average value of total network travel time, t_j is the time (in minutes) from the generation of passenger j until boarding, and $1(\cdot)$ is an indicator function that returns 1 if the condition is true and 0 otherwise.

Equation 2. Average network travel time reward for passenger boarding

Also, the algorithm was induced to increase the service rate as much as possible by rewarding the average value of the entire stop access time in the same way as boarding when a passenger gets off. The expression is as follows:

$$R_{alight} = \sum_{k=1}^{L} \overline{T}$$

where L is the number of alighted passengers, $\overline{T}$ is the average value of total network travel time.

Equation 3. Average network travel time reward for passenger alighting

The alighting reward function does not incorporate additional rewards or penalties for factors such as detours or direct routes. Implementing such a mechanism caused the algorithm to prioritise faster routes to board more passengers rather than facilitating efficient alighting. This bias was determined to be misaligned with the fundamental objective of public transportation services. Therefore, a reward function proportional to the number of passengers alighting within the episode was applied, ensuring that the system effectively encourages passenger alighting during the episode.

3.3.3 Hyperparameter settings

Hyperparameters refer to variables that significantly influence the learning process of a deep-learning network, such as the learning rate, batch size, and DQN-specific parameters, including the total memory size for storing experiences, the minimum memory size required before learning begins, and the discount rate (γ). The agent explores the environment and learns an optimal policy over a specified number of episodes, which is determined based on the total episodes of interaction with the environment for training. A sufficient number of episodes allows the agent to experience a variety of scenarios, enhancing learning stability and performance; however, it also increases training time. In this study, 10 000 episodes were conducted for each demand scenario.

The discount rate (γ), a value between 0 and 1, determines the extent to which the agent considers future rewards. For instance, a γ value of 0.99 encourages the agent to consider future rewards as nearly as important as immediate rewards, thus promoting a long-term perspective in decision making. In this study, the γ value was set to 0.99 to ensure sufficient consideration of future rewards.

The ε-start parameter represents the initial exploration probability within the ε-greedy policy, which balances exploration and exploitation in the learning process. At the start of training, this parameter sets a high likelihood that the agent will take random actions, enabling extensive exploration of the environment by testing various behaviours. As training progresses, the ε-end gradually reduces the probability of exploration, encouraging the agent to rely more on its learned policy in the later stages.

The ε-decay parameter determines the rate at which ε decreases with each episode, thus defining the transition from exploration to exploitation over time. Early in the learning process, a higher exploration rate is maintained, while in later stages, the learned policy is increasingly applied. In this study, ε-decay was set to 0.9995, allowing for a gradual reduction in exploration probability across episodes. By the final 1,000 episodes, the exploration rate stabilises at 0.01, ensuring that most actions are based on the agent's learned policy.

This gradual transition allows the later episodes to serve as effective evaluation metrics for dynamic route optimisation. As illustrated in Figure 4, the increasing reward values indicate the agent's improved learning performance over episodes, while Figure 5 shows the decreasing number of moves as the agent refines its decision-making

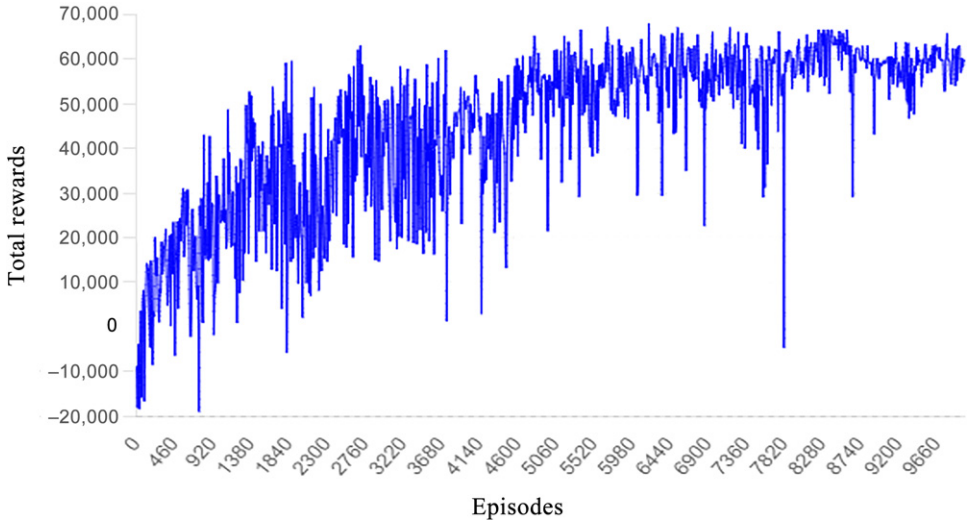

Figure 4. Graph of algorithm reward values acquired through episode training

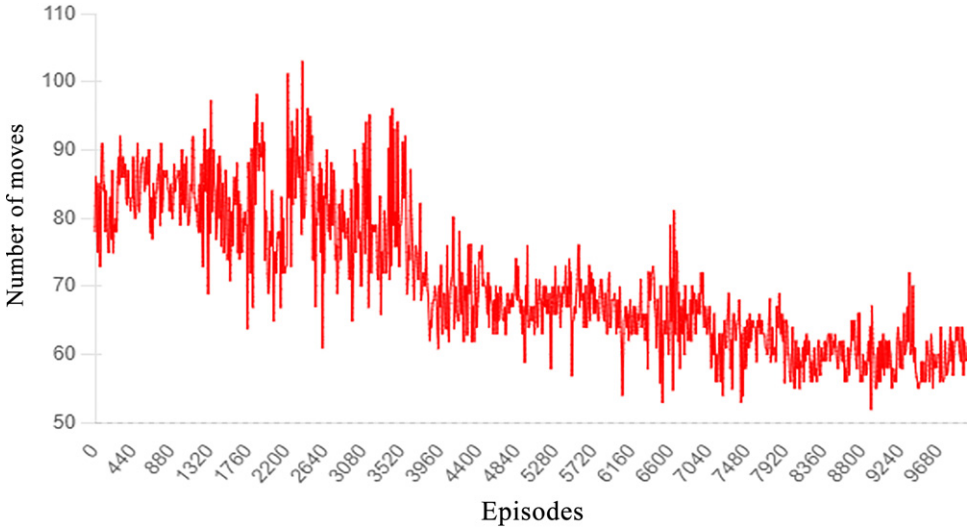

Figure 5. Graph of algorithm action lengths through episode training

process. The decrease in ε correlates with improved reward acquisition, and the algorithm reduces unnecessary movements across the network, reflecting an optimised operational performance.

4. Results

As outlined in Section 3, our DQN algorithm calculates the real-time number of waiting passengers at each stop, the current bus location, the number of passengers on board, and the destinations of the passengers. The algorithm updates the status across all stops and selects an action (next stop) at each decision point. Each episode represents a 1-h operation, where the algorithm continuously processes passenger demand within the network for an hour to learn an optimal operational strategy.

Upon selecting an action, the algorithm determines the next stop based on the chosen action value, accounting for passengers waiting at the stop, passengers alighting from the bus, and new passengers boarding. The waiting time for passengers and the travel time for those alighting are continuously recorded at each stop. Consequently, waiting time for passengers who have not boarded serves as a penalty in the reward function, while rewards for boarding and alighting events contribute positively to the total reward. The total sum of rewards obtained during the episode over

the hour becomes the final outcome for that episode. The algorithm adjusts its actions in subsequent episodes based on the reward from the previous episode, with learning repeated over a total of 10 000 episodes. Ultimately, as the DRT system adaptively selects stops, it optimises service time for passengers and dynamically adjusts routes to maximise passenger load, thereby achieving efficient and flexible operations.

Call demand within each 1-h episode was generated using a uniform Monte Carlo method in both time and space, rather than relying on historical driving records or demand forecasts. This approach provided an equal probability of real-time call demand occurring across all times and locations. To regulate real-time call demand within each episode, the number of calls was set to range from a minimum of 6 to a maximum of 22 passengers per episode. This constraint was applied because, at demand levels exceeding 24 passengers, the outcomes of fixed-route and dynamic-route operations tended to converge with similar results.

As explained in Section 1, some regions are transitioning from fixed-route buses to DRT systems to enhance the flexibility of public transportation. Currently, DRT still operates in a manner similar to fixed-route buses; however, by integrating real-time dynamic-route generation algorithms, it is gradually evolving into a more adaptive system. Therefore, this study analyses how dynamic-route DRT and fixed-route buses respond to increasing demand within the same network.

While fixed-route buses operate by following predefined routes and circulating through all stops, dynamic-route DRT determines its next stop dynamically based on real-time waiting demand. Through repeated learning, the model acquires sufficient state conditions, serving as an indicator for validating the feasibility of real-time route optimisation in actual DRT operations. Based on this analysis, the experiment results were categorised by demand levels to evaluate the operational performance of fixed-route buses and DRT, identifying the quantitative differences that DRT can offer over conventional fixed-route bus systems.

4.1 Comparison of service rate

First, a comparison was conducted between the number of boarding and alighting services provided by fixed-route buses and dynamic-route DRT based on real-time ride requests within the same episode. In the case of dynamic-route DRT, rewards are granted for boarding and alighting services, while penalties are continuously applied for unmet demand based on waiting time. Consequently, the dynamic-route DRT actively moves between stops to pick up passengers, serving as an indicator for evaluating the proposed algorithm's ability to respond to demand changes.

Table 5 showed that for boarding, fixed-route buses could accommodate all demand up to 12 passengers. However, at 18 passengers or more, the dynamic-route DRT was able to serve a larger number of passengers compared with the fixed route. For alighting, fixed routes struggled to meet growing demand as the number of passengers increased, whereas the dynamic-route DRT demonstrated higher performance by adapting to the increased demand.

While the dynamic-route DRT exhibited slightly lower performance in boarding compared with the fixed-route buses, this can be attributed to its reward function, which prioritises minimising passenger waiting time. This approach may have been efficient in reducing delays but might have slightly constrained the system's overall boarding capacity.

The boarding service rate of a fixed-route bus and a dynamic-route DRT was compared in response to real-time call demand. The service rate is defined as the proportion of demand that successfully completes their journeys, from boarding to alighting, within the episode. Since the dynamic-route DRT receives rewards for boarding and alighting events, it is designed to provide these services to as many passengers as possible while moving between stops. Therefore, the object is to maximise the service rate for boarding and alighting in relation to total demand.

As a result, Table 6 shows that the proportion of successful boarding and alighting services within each episode was higher for the dynamic route compared with the fixed route, except at a demand level of 14

Table 5. Comparison of boarding and alighting performance between fixed and dynamic-route systems

(Demand)										
		Call demand								
		6	8	10	12	14	16	18	20	22
Boarding	Fixed route (A)	6	8	10	12	12	13	12	13	14
	Dynamic route (B)	6	8	9	11	10	12	14	15	14
	Difference (B – A)	0	0	−1	−1	−2	−1	2	2	0
Alighting	Fixed route (A)	5	6	7	8	9	10	9	9	10
	Dynamic route (B)	6	8	9	10	8	11	14	14	13
	Difference (B – A)	1	2	2	2	−1	1	5	5	3

Table 6. Comparison of service rate between fixed- and dynamic-route systems

(%)	Call demand								
	6	8	10	12	14	16	18	20	22
Fixed route (A)	91.7	87.5	85.0	83.3	75.0	71.9	58.3	55.0	54.5
Dynamic route (B)	100.0	100.0	90.0	87.5	64.3	71.9	77.8	72.5	61.4
Difference (B – A)	8.3	12.5	5.0	4.2	–10.7	0.0	19.4	17.5	6.8

passengers. The lower performance at this specific demand level was attributed to overfitting during the training process. Overall, the results demonstrate that the dynamic-route DRT achieved a relatively higher service rate than the fixed route, highlighting its greater effectiveness in responding to ride requests within the same time frame.

4.2 Comparison of total service time

Dynamic-route DRT receives an hourly penalty for unboarded demand, prompting it to move between stops while quickly boarding as many passengers as possible. This contributes to minimising the waiting time for each passenger until they receive boarding service. Analysis of Table 7 showed that the waiting time until boarding tends to increase for both fixed-route buses and DRT as demand increases. However, except for certain demand levels, DRT was found to have a shorter boarding waiting time.

In addition, DRT is rewarded based on the number of passengers alighting, encouraging it to drop off as many passengers as possible while moving between stops. As a result, while the service time for fixed-route buses increased with rising demand, DRT maintained consistently shorter travel times even as demand increased (Table 8).

Overall, the total service time is calculated as a concept that includes both the waiting time until boarding and the travel time until alighting. The analysis of Tables 7–9 confirms that the reduction in travel time significantly contributes to the decrease in total service time. Except in cases of algorithm overfitting, DRT consistently exhibits shorter service times across all demand levels. This demonstrates the potential to reduce service time compared with fixed-route systems, thereby enhancing passenger satisfaction.

5. Conclusion

This study proposed a dynamic-routing approach for DRT using the DQN algorithm to optimise DRT operations within a service area. The optimal stop sequence was determined based on passenger waiting status, destination requests, the current bus location, and the number of passengers. Operating the DRT on a dynamic route using this approach resulted in a higher level of service to passengers compared with fixed-route bus operations.

Table 7. Waiting time until boarding based on passenger demand

(min)	Call demand								
	6	8	10	12	14	16	18	20	22
Fixed route (A)	8.81	8.33	9.60	9.41	10.35	11.00	15.38	17.40	18.65
Dynamic route (B)	3.17	7.88	9.80	8.25	15.35	12.25	11.94	14.22	16.20
Difference (B – A)	–5.64	–0.45	0.20	–1.16	5.00	1.25	–3.44	–3.18	–2.45

Table 8. Travel time until alighting based on passenger demand

(min)	Call demand								
	6	8	10	12	14	16	18	20	22
Fixed route (A)	16.70	18.08	18.50	21.58	18.83	18.30	20.00	22.61	22.62
Dynamic route (B)	14.17	12.63	15.44	9.72	14.00	10.66	17.50	12.24	13.81
Difference (B – A)	–2.53	–5.45	–3.06	–11.86	–4.83	–7.64	–2.50	–10.37	–8.81

Table 9. Comparison of total service time between fixed and dynamic route systems

(min)	Call demand								
	6	8	10	12	14	16	18	20	22
Fixed Route (A)	25.51	26.41	28.10	30.99	29.18	29.30	35.38	40.01	41.27
Dynamic Route (B)	17.34	20.51	25.24	17.97	29.35	22.91	29.44	26.46	30.01
Difference (B-A)	−8.17	−5.90	−2.86	−13.02	0.17	−6.39	−5.94	−13.55	−11.26

The analysis revealed that, in demand simulations involving 8 or fewer passengers and 22 passengers, there was little difference between the fixed and dynamic-route configurations in terms of boarding and alighting numbers. However, in most other demand simulations, the dynamic-route configuration demonstrated superior performance indicators.

Regarding total service time, the dynamic route consistently exhibited better indicators, except in the 14-passenger scenario, where overfitting occurred. Particularly, the findings indicate that the total service time could be reduced by a minimum of V 2.86 min and a maximum of 13.55 min, demonstrating the potential of implementing dynamic-route DRT to enhance service time efficiency compared with fixed-route buses. This study further validated that, by planning and operating the optimal stop sequence based on passenger waiting time and destination requests at departure, dynamic-route DRT could accommodate and disembark more passengers within a single 1-h episode compared with fixed-route buses. Moreover, reinforcement learning trained on 10,000 operational scenarios demonstrated that even in low-demand situations within the test network representing residential areas, dynamic routes outperformed fixed routes in service time and service rate during actual operations. However, during the analysis, instances of suboptimal performance caused by overfitting were observed, highlighting the need for further research to address this issue. Future studies should focus on refining the learning process and improving environmental configurations to prevent overfitting and ensure robust model performance. In addition, for DRT systems to be widely adopted and fully realise their potential, integrating with other multimodal transportation options, such as mobility hubs and micro-mobility services, should be considered (Choi *et al*., 2024b; Ku *et al*., 2022). Such integration could further enhance the operational efficiency and connectivity of DRT systems, building upon the demonstrated advantages of dynamic routing in this study and enabling first- and last-mile solutions within broader urban mobility networks.

6. Discussion

This study confirmed that a DQN-based dynamic-route generation algorithm can improve the operational efficiency of DRT, yet the overfitting problem encountered during the training process remains a major limitation that needs to be addressed. In particular, during the 14-passenger demand simulation, repeated learning experiences at a specific stop led to a phenomenon where the DRT did not move to that location despite passengers waiting. This issue may stem from insufficient information about that stop or the algorithm persistently learning incorrect actions. Potential solutions include redesigning the reward function, adjusting state variables, balancing the training dataset, or increasing the number of episodes. For instance, incorporating time-of-day information into the learning process could guide the algorithm towards selecting more appropriate routes under congested conditions. If travel time to certain stops increases during specific periods, the algorithm may learn to avoid those stops. This approach would naturally account for congestion, enabling more realistic route optimisation.

Moreover, because the algorithm in this study was designed to mitigate the imbalance between demand and supply, it particularly focuses on responding to real-time passenger demand. However, factors such as network complexity, traffic congestion, and signal systems may influence route selection in real urban environments—constraints that were not fully accounted for in this study. Therefore, future research should develop optimisation strategies reflecting real-time traffic flow, congestion, and signal cycles, thus enhancing both the practical applicability and generalisability of the algorithm.

This study demonstrated the feasibility of applying DQN to transition from fixed-route buses to a dynamic-route DRT, confirming its potential to efficiently manage real-time demand changes in urban settings. If subsequent research incorporates additional elements such as congestion and time-of-day variations, DRT systems could further enhance the operational efficiency and passenger satisfaction in real-world traffic environments.

Funding

This work was supported by the government of the Republic of Korea (MSIT) and the National Research Foundation of Korea (Grant Nos. NRF-2024K2A9A2A06014158 and FY2024) for Seungjae Lee. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (Grant No. RS-2024-00452221) for Sion Kim.

Author contributions

C. Yeon and A. Cho have equally contributed to this work. C. Yeon and A. Cho were primarily responsible for data analysis, algorithm development, and simulation design. S. Kim and Y. Lee assisted with data validation, drafting the manuscript, and creating visual representations. S. Lee supervised the study, providing critical guidance and ensuring its overall quality. All authors contributed to the research and provided feedback on the manuscript.

REFERENCES

Ai G, Zuo X, Chen G and Wu B (2022) Deep reinforcement learning based dynamic optimization of bus timetable. *Applied Soft Computing* **131**: 109752.

Arhin SA, Gatiba A, Anderson M, Manandhar B and Ribbisso M (2019) Acceptable wait time models at transit bus stops. *Engineering, Technology & Applied Science Research* **9(4)**: 4574–4580.

Barrett, S, Santha, N and Khanna, A (2019), *On-demand public transport: Key learnings from global pilots*. LEK Special Report.

Bencekri M, Van Fan Y, Lee D, Choi M and Lee S (2024) Optimizing shared bike systems for economic gain: integrating land use and retail. *Journal of Transport Geography* **118**: 103920.

Berrada J and Poulhès A (2021) Economic and socioeconomic assessment of replacing conventional public transit with demand responsive transit services in low-to-medium density areas. *Transportation Research Part A: Policy and Practice* **150**: 317–334.

Bürstlein J, López D and Farooq B (2021) Exploring first-mile on-demand transit solutions for North American suburbia: a case study of Markham, Canada. *Transportation Research Part A: Policy and Practice* **153**: 261–283.

Campisi T, Canale A, Ticali D and Tesoriere G (2021) Innovative solutions for sustainable mobility in areas of weak demand. Some factors influencing the implementation of the DRT system in Enna (Italy). *AIP Conference Proceedings*, vol. 2343. AIP Publishing.

Choi M, Van Fan Y, Lee D, Kim S and Lee S (2024a) Location and capacity optimization of EV charging stations using genetic algorithms and fuzzy analytic hierarchy process. *Clean Technologies and Environmental Policy*. Springer, pp. 1–14.

Choi M, Kang G and Lee S (2024b) Autonomous driving parking robot systems for urban environmental benefit evaluation. *Journal of Cleaner Production* **469**: 143215.

Coutinho FM, van Oort N, Christoforou Z *et al.* (2020) Impacts of replacing a fixed public transport line by a demand responsive transport system: case study of a rural area in Amsterdam. *Research in Transportation Economics* **83**: 100910.

Cui M (2024) DQN and dynamic feedback for multitask scheduling optimization in engineering management. *International Journal of Low-Carbon Technologies* **19**: 2279–2286.

Deka U, Varshini V and Dilip DM (2023) The journey of demand responsive transportation: towards sustainable services. *Frontiers in Built Environment* **8**: 942651.

Dockendorf J, Levinson HS, Fichter D, *et al.* (2000) Bus transportation—a look forward. *A1E01: Committee on Bus Transit Systems, Transportation in the New Millennium, State of the Art and Future Directions, Transportation Research Board, Washington, DC*.

Huang T, Fang B, Bei X and Fang F (2020) Dynamic trip-vehicle dispatch with scheduled and on-demand requests, *Uncertainty in Artificial Intelligence*, PMLR, pp. 250–260.

Jang B, Kim M, Harerimana G and Kim JW (2019) Q-learning algorithms: a comprehensive classification and applications. *IEEE Access* **7**: 133653–133667.

Jang Y, Ku D, Choi M *et al.* (2023) Relational analysis between economic activities and mobility during coronavirus disease. *Proceedings of the Institution of Civil Engineers – Municipal Engineer* **176(3)**: 139–149.

Khattak AJ and Yim Y (2004) Traveler response to innovative personalized demand-responsive transit in the San Francisco Bay Area. *Journal of Urban Planning and Development* **130(1)**: 42–55.

Ku D, Choi M, Lee D and Lee S (2022) The effect of a smart mobility hub based on concepts of metabolism and retrofitting. *Journal of Cleaner Production* **379**: 134709.

Ku D, Kim D and Lee S (2023) Mobility challenges for senior passengers during the covid-19 pandemic. *Proceedings of the Institution of Civil Engineers – Municipal Engineer* **176(3)**: 161–172.

Kwak J, Jeong I, Lee D, Ku D and Lee S (2024) Lessons learned from spatiotemporal effect of COVID on the population density in the CBD. *Proceedings of the Institution of Civil Engineers – Municipal Engineer* **177(4)**: 169–179.

Lai L, Dong Y, Andriotis CP, Wang A and Lei X (2024a) Synergetic-informed deep reinforcement learning for sustainable management of transportation networks with large action spaces. *Automation in Construction* **160**: 105302.

Lai X, Yang Z, Xie J and Liu Y (2024b) Reinforcement learning in transportation research: frontiers and future directions. *Multimodal Transportation* **3(4)**: 100164.

Liu Y, Chen X, Xiao H and Duan J (2023) Study on the smart transformation strategy of old neighbourhoods based on urban renewal. *Proceedings of the Institution of Civil Engineers – Smart Infrastructure and Construction* **177(1)**: 15–24, https://doi.org/10.1680/jsmic.23.00013.

Liu T, You H, Gkiotsalitis K and Cats O (2024) Human-machine collaborative decision-making approach to scheduling customized buses with flexible departure times. *Transportation Research Part A: Policy and Practice* **187**: 104184.

Logan P (2007) Best practice demand-responsive transport (DRT) policy. *Road & Transport Research: A Journal of Australian and New Zealand Research and Practice* **16(2)**: 50–59.

Martí P, Jordán J, González Arrieta MA and Julian V (2023) A survey on demand-responsive transportation for rural and interurban mobility. *International Journal of Interactive Multimedia and Artificial Intelligence* **8(3)**: 43–54.

Mnih V, Kavukcuoglu K, Silver D *et al.* (2015) Human-level control through deep reinforcement learning. *Nature* **518(7540)**: 529–533.

Moreno-Malo J, Posadas-Yagüe J-L, Cano JC *et al.* (2024) Improving traffic light systems using deep Q-networks. *Expert Systems with Applications* **252**: 124178.

Mulley C and Nelson JD (2009) Flexible transport services: a new market opportunity for public transport. *Research in Transportation Economics* **25(1)**: 39–45.

Nelson JD, Ambrosino G and Romanazzo M (2004) *Demand Responsive Transport Services: Towards the Flexible Mobility Agency.* Newcastle University.

Porru S, Misso FE, Pani FE and Repetto C (2020) Smart mobility and public transport: opportunities and challenges in rural and urban areas. *Journal of Traffic and Transportation Engineering* **7(1)**: 88–97.

Rahmani AM, Naqvi RA, Yousefpoor E *et al.* (2022) A Q-learning and fuzzy logic-based hierarchical routing scheme in the intelligent transportation system for smart cities. *Mathematics* **10(22)**: 4192.

Shen S, Ouyang Y, Ren S and Zhao L (2021) Path-based dynamic vehicle dispatch strategy for demand responsive transit systems. *Transportation Research Record* **2675(10)**: 948–959.

Shoufeng L, Ximin L and Shiqiang D (2008) Q-learning for adaptive traffic signal control based on delay minimization strategy. *2008 IEEE International Conference on Networking, Sensing and Control*, IEEE, pp. 687–691.

Sommer C and Deutsch V (2021) Nahverkehrsplanung und Netzgestaltung des ÖPNV. *Stadtverkehrsplanung Band 3: Entwurf, Bemessung und Betrieb*, Springer, pp. 255–285.

Sungatek C, Gurjung K, Junsik P, *et al.* (2023) Impact Analysis of Implementing the Demand Responsive Transit System in Metropolitan Areas.

Sutton RS (2018) *Reinforcement Learning: An Introduction*. A Bradford Book.

Tarakci S and Turk SS (2022) Public value capture capacity in the urban renewal project process: Fikirtepe case. *Proceedings of the Institution of Civil Engineers – Municipal Engineer* **175(1)**: 42–56.

Vansteenwegen P, Melis L, Aktaş D *et al.* (2022) A survey on demand-responsive public bus systems. *Transportation Research Part C: Emerging Technologies* **137**: 103573.

Watkins CJCH and Dayan P (1992) Q-learning. *Machine Learning* **8(3–4)**: 279–292.

Wu S, Zhuang Y, Chen J *et al.* (2019) Rethinking bus-to-metro accessibility in new town development: case studies in Shanghai. *Cities* **94**: 211–224.

Wu M, Yu C, Ma W, Wang L and Ma X (2021) Reinforcement learning based demand-responsive public transit dispatching. *CICTP 2021*, pp. 387–398.

Yang Y, Juntao L and Lingling P (2020) Multi-robot path planning based on a deep reinforcement learning DQN algorithm. *CAAI Transactions on Intelligence Technology* **5(3)**: 177–183.

Ye R, Zhang D, Chen R, *et al.* (2024) Research on emergency control strategy for transient instability of power system based on deep reinforcement learning and graph attention network. *2024 7th International Conference on Energy, Electrical and Power Engineering (CEEPE)*, IEEE, pp. 1040–1048.

Zhang K and Li M (2024) *Graph Prompt Learning Method for the Demand-Responsive Transport Routing Problem, IEEE Transactions on Big Data*. IEEE.

Zhou X, Zhang Y and Guo H (2024) Scheduling method of demand-responsive transit based on reservation considering vehicle size and mileage. *Applied Sciences* **14(19)**: 8836.