



Reward Scheme Analysis for DQN-Based Adaptive Traffic Signal Control in Road Repair Zones

Yoanda Alim Syahbana¹, Muhammad Wahyudi², Fikri Muhaffizh Imani³, Ardianto Wibowo⁴, Fatur Rizqy Al-Farisz⁵

¹Computer Engineering Technology, Information Technology Department, Politeknik Caltex Riau, Pekanbaru, Indonesia

²Applied Master of Computer Engineering, Information Technology Department, Politeknik Caltex Riau, Pekanbaru, Indonesia

³Information System, Information Technology Department, Politeknik Caltex Riau, Pekanbaru, Indonesia

⁴Département Mathematical and Electrical Engineering, IMT Atlantique Brest, Brest, France

^{4,5}Informatics Engineering, Information Technology Department, Politeknik Caltex Riau, Pekanbaru, Indonesia

*yoanda@pcr.ac.id

Abstract

Traffic congestion caused by temporary road repairs often forces bidirectional traffic to alternate through a single lane, leading to increased delays and imbalanced traffic flow. This study investigates the impact of reward scheme design on the performance of a Deep Q-Network (DQN)-based adaptive traffic signal control system in such constrained environments. Using the Simulation of Urban Mobility (SUMO), a traffic scenario involving 1,656 vehicles over 1,800 seconds was modeled to evaluate six reward scheme configurations combining Traffic Flow (TF), Waiting Time (WT), and Average Speed (AS): TF-TF, TF-WT, WT-TF, WT-WT, AS-TF, and AS-WT. The DQN agent, implemented with a two-layer neural network and trained for 50 epochs, dynamically adjusted signal timing to balance traffic from opposing directions. Experimental results indicate that the AS-WT configuration achieved the most balanced performance, producing the best fairness index (1.04) while maintaining stable traffic flow. In contrast, schemes with misaligned or redundant metrics showed significantly poorer performance. These findings highlight the importance of reward design in reinforcement learning-based traffic signal control and suggest that carefully selected reward schemes can improve fairness and efficiency in temporary road repair zones.

Keywords: adaptive traffic signals; deep Q-network; reinforcement learning; reward schemes; SUMO simulation

How to Cite: Y. A. Syahbana, M. Wahyudi, F. M. Imani, A. Wibowo, and F. R. Al-Farisz, "Reward Scheme Analysis for DQN-Based Adaptive Traffic Signal Control in Road Repair Zones", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 10, no. 3, pp. 609 - 616, Jun. 2026.

Permalink/DOI: <https://doi.org/10.29207/resti.v10i3.7381>

Received: December 15, 2025

Accepted: May 23, 2026

Available Online: June 8, 2026

This is an open-access article under the [CC BY 4.0 License](#)

Published by [Ikatan Ahli Informatika Indonesia](#)

1. Introduction

Traffic congestion represents a pervasive and costly challenge in urban environments, significantly impacting socioeconomic activities and environmental quality. Globally, it leads to increased journey times, higher fuel consumption, and excessive operational costs [1]. Furthermore, the environmental quality is critical, as heightened emissions from idling and stop-and-go vehicles contribute significantly to urban air pollution and climate change [2].

A particularly disruptive scenario arises from temporary road blockages, such as alternating lane closures caused by road repair and maintenance works. This temporary road blockages forces opposing traffic flows to share a single open lane, creating a bottleneck.

As illustrated in Figure 1, this alternating flow arrangement demands an adaptive coordination to prevent severe congestion and vehicle delays.

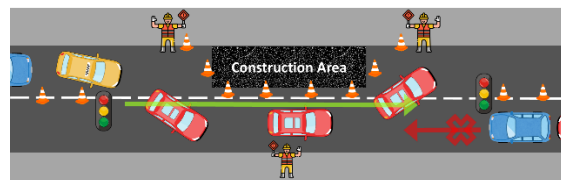


Figure 1. Illustration of research case study: Road Repair Zones

The deployment of temporary traffic lights is a common solution in alternating flow arrangement [3], [4]. However, conventional Fixed-time Traffic Signal (FTS) systems prove inadequate in this dynamic, asymmetric traffic flow. Since FTS utilizes

predetermined cycles, it fails to adapt to the unpredictable fluctuations in traffic volume between opposing directions. This inflexibility results in inefficiencies, where one direction may receive excessive green time while the other endures prolonged waiting periods, extended congestion that leads to driver frustration and reduced compliance [5]. This emphasizes the critical need for optimization systems.

Recent literature has increasingly highlighted Reinforcement Learning (RL) as a promising method for intelligent traffic signal control [6-10]. Unlike supervised learning, RL agents learn optimal control policies through direct interaction with the environment, receiving immediate feedback via rewards or punishment [11]. This approach is invaluable in complex, novel scenarios like road repair zones, where historical data is often non-existent or unrepresentative of the current constraints. The advent of Deep Q-Networks (DQN) [12], which integrates Q-learning with deep neural networks to manage high-dimensional state representations, has proven effective in complex simulations, such as those conducted using the Simulation of Urban Mobility (SUMO) [13], [14].

While previous studies have successfully demonstrated RL's potential in general intersection optimization, often focusing on metrics like reduced waiting time [3], there remains a significant gap in addressing the issue of fairness and equity in highly asymmetric and constrained environments, such as alternating lane closures. Research into reward scheme design for adaptive signals has shown that misalignment between state representation and the reward function can lead to biased outcomes, where the agent prioritizes maximizing flow for one direction at the expense of severe queuing in the opposite direction [15], [16]. Thus, ensuring balanced performance metrics that minimize disparity in waiting times and queue lengths between opposing flows is important for achieving fair treatment of all road users [17], [18].

This research is motivated by the limitations of existing RL approaches in handling these directional imbalances. The purpose of this study is to develop and evaluate a DQN-based adaptive traffic signal control system for temporary road blockages, with a core focus on a systematic reward scheme analysis. Specifically, this work investigates how different combinations of key traffic metrics influence overall system performance and fairness. The evaluated metrics include Traffic Flow (TF), Waiting Time (WT), and Average Speed (AS), which are used as state variables and reward signals.

The key research questions guiding this investigation are 1) Which combination of traffic metrics, when configured as the RL agent's state and reward scheme, achieves the best fairness index in balancing opposing traffic directions under alternating lane closures?, 2) How is the identified optimal scheme compared against conventional fixed-timing baselines and sub-optimal

RL schemes in terms of reducing overall waiting times and improving average vehicle speeds?

By systematically comparing six distinct schemes (TF-TF, TF-WT, WT-TF, WT-WT, AS-TF, and AS-WT), this research aims to identify the optimal design principles for RL-based traffic control systems operating in asymmetric traffic flow from temporary road blockages. The findings are expected to contribute to robust design guidelines for implementing fair and efficient Intelligent Transportation Systems (ITS) in temporary road blockages.

This study contributes to the growing body of research on reinforcement learning-based traffic signal control by focusing on the role of reward function design under non-ideal traffic conditions, specifically in road repair zones that require alternating traffic flow. The main novelty of this work lies in the systematic analysis of multiple reward scheme configurations within a DQN framework to evaluate their impact on both traffic efficiency and fairness between opposing traffic directions. Unlike many existing studies that emphasize throughput or delay reduction alone, this research investigates how combinations of traffic flow, waiting time, and average speed metrics influence the learning behavior of the agent and the resulting traffic balance. From a theoretical perspective, the study provides insights into how reward structure affects policy learning in RL-based control systems. From a practical standpoint, the findings offer guidance for designing adaptive traffic signal strategies in ITS, particularly in temporary and constrained environments such as road repair zones where maintaining fairness and operational stability is critical.

This paper is organized as follows: Section 2 details the materials and methods, including the SUMO simulation environment and the DQN implementation. Section 3 presents and analyzes the experimental results, highlighting the performance of each reward scheme, and discusses the implications of the findings. Finally, Section 4 concludes the paper with a summary and suggestions for future work.

2. Methods

This study employs a simulation-based experimental framework to evaluate DQN performance for adaptive traffic signal control in road repair scenarios. The research methodology integrates the SUMO traffic simulator with a custom DQN implementation, creating a closed-loop system where an RL agent learns to optimize traffic light timing based on synthetically generated traffic patterns. The following sections detail the simulation environment, traffic data generation, DQN architecture, and experimental design for comparative reward scheme analysis.

2.1 Simulation Environment and Traffic Data

The traffic simulation environment was developed using Eclipse Simulation of Urban Mobility (SUMO) version 1.18.0, a widely adopted traffic simulation

platform known for its realistic modeling of vehicular dynamics and comprehensive Application Programming Interface (API) [10], [13], [19-21]. The simulation scenario represents a straight two-lane road segment with opposing traffic directions, incorporating a temporary blockage zone where road repairs restrict traffic to alternating single-lane operation. This configuration models realistic construction sites where one lane is closed while the other remains operational for bidirectional traffic flow.

Synthetically generated traffic patterns were created using SUMO's python script to ensure controlled and reproducible experimental conditions. The script generated 1,656 vehicle trips over a 1,800-second simulation period, resulting in a constant arrival rate of approximately 0.92 vehicles per second. Vehicle origins and destinations were randomly assigned while enforcing a minimum trip distance of 500 meters to avoid unrealistically short journeys.

All generated vehicles were classified as passenger cars, with trip data exported in XML format containing detailed attributes including departure time, speed profiles, travel duration, waiting time, and route length. This programmatically generated traffic demand provided consistent baseline conditions for fair comparison across all reward scheme evaluations.

2.2 Reinforcement Learning (RL) Framework

RL is a machine learning technique that designs an agent to learn to make decisions in sequential time-dependent situations, without a supervisor, where each decision made by the agent affects the subsequent data it will receive [22], [23]. RL has six main elements that support its algorithm: Environment, Agent, State, Reward, Value, and Policy. First, the Environment is the setting that directly interacts with the RL agent. Since the agent usually has no prior knowledge of the environment's conditions, it is often treated as a black box. The Agent is the entity that performs the main actions within this environment. The State or Configuration represents all possible situations that the agent might encounter. Rewards serve as feedback from the environment, guiding the agent in its learning process; it is typically expressed as positive rewards indicating success and negative rewards indicating failure. The Value of each state or configuration provides a basis for the agent to decide which actions to take. Lastly, the Policy is the agent's learning strategy, consisting of two options: exploitation, where the agent leverages known information, and exploration, where the agent seeks new information. Together, these elements form the foundation of RL and enable the agent to learn effectively through trial and error.

The learning process updates the Value estimation of states that use value iteration, which is calculated using Equation 1,

$$V(S_t) \leftarrow V(S_t) + \alpha_{RL}[V(S_{t+1}) - V(S_t)] \quad (1)$$

where $V(S_t)$ denotes the value of the state at $t \in \{1, 2, \dots, T\}$ and α_{RL} denotes the learning rate, which ranges from 0 to 1. This equation indicates that the new value of the State t is equal to its current value plus the difference between the Value of the next state and the current state's value, all multiplied by a learning rate α_{RL} (assuming the reward for the intermediate state is 0).

In recent years, there have been significant advancements in RL, especially in addressing various challenges [24]. In the field of Intelligent Transportation Systems (ITS), RL can be applied to optimize several aspects of transportation, including road traffic, vehicle management, and public transit. For example, RL can be utilized for traffic signal control by learning traffic patterns and adjusting light durations to minimize delays and congestion [25]. Additionally, RL can be applied to vehicle routing and scheduling to reduce travel time and operational costs [26]. In public transportation, RL can be used to optimize routes and schedules while estimating passenger numbers and arrival times [27]. Overall, RL holds substantial potential for optimizing transportation and improving overall efficiency. Implementing RL in ITS can help reduce congestion, enhance mobility, and lower vehicle operational costs.

DQN is a RL technique that integrates Q-Learning with Deep Neural Networks (DNN) [12], [28]. DQN algorithm was initially introduced by Mnih in 2015 [12] and has since served as a foundational study and reference point for numerous advancements in deep machine learning. The algorithm has been applied to train an adaptive traffic light model, enabling it to learn how to adjust the durations of green, yellow, and red lights based on data related to vehicle waiting times [29].

The adaptive traffic signal control implements a DQN, a value-based reinforcement learning algorithm that combines Q-learning with deep neural network function approximation [12], [22]. The DQN agent interacts with the SUMO environment through the Traffic Control Interface (TraCI) API, forming a standard RL cycle where the agent observes traffic state (S_t), selects control actions (A_t), executes these actions in the simulation, and receives corresponding rewards (R_t) that guide learning optimization.

DQN architecture employs a feedforward neural network with two hidden layers, each containing 256 neurons with Rectified Linear Unit (ReLU) activation functions. The input layer receives state vectors of dimension 2 (representing one traffic metric per lane), while the output layer produces Q-value estimates for two possible control actions. Network training utilizes Mean Squared Error (MSE) loss between predicted and target Q-values, optimized via the Adam algorithm with a learning rate ($\alpha_{optimizer}$) of 0.1. Experience replay with a memory capacity of 100,000 transitions stabilizes training by decorrelating sequential

experiences, while mini-batch gradient descent with a batch size of 1,024 ensures efficient parameter updates. This implementation was adapted from the Dynamic Traffic Light Management System by Maunish Dave [30], which was modified according to the needs of this study. The hyperparameters used in the DQN model were selected to ensure stable training behavior and efficient convergence within the simulated traffic environment. The network architecture, replay memory configuration, discount factor, and training settings

were adapted from prior reinforcement learning-based traffic signal control studies, particularly the implementation proposed by [30], with adjustments made to accommodate the asymmetric traffic conditions and reward scheme evaluation in this study. The selected parameters provided stable learning performance during preliminary experiments while maintaining computational efficiency for repeated SUMO simulations. Flow chart of the DQN is shown in Figure 2 .

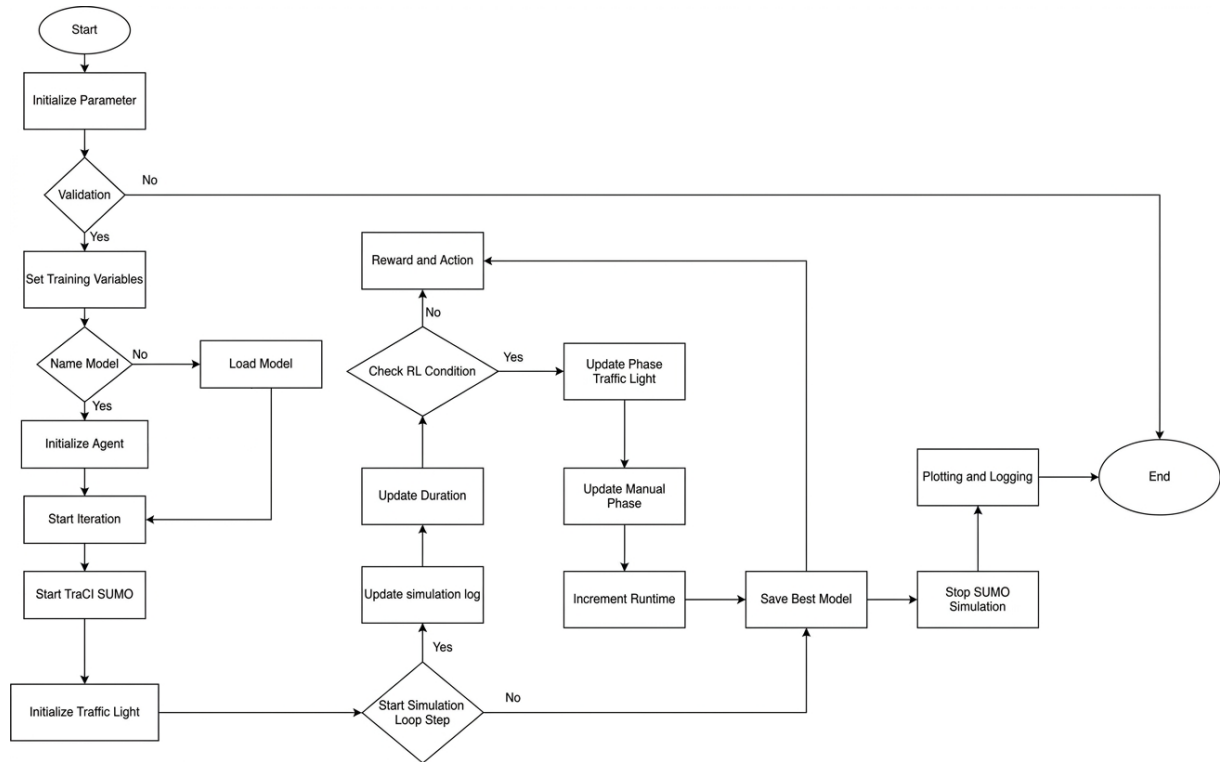


Figure 2. DQN flowchart

2.3. State Representation and Action Space

The state observation provided by the DQN agent consists of normalized traffic metrics captured from each controlled lane. Depending on the reward scheme under evaluation, these metrics include either traffic flow (vehicle count), waiting time (seconds), or average speed (meters/second). Each lane's selected metric is sampled when vehicles are positioned beyond 10 meters from the intersection, ensuring that only vehicles actively influencing traffic dynamics are considered. This filtering prevents distortion from vehicles that have already cleared the intersection or are too distant to affect immediate control decisions.

The action space comprises two discrete control options representing extreme allocation strategies: Action 0 allocates the entire green phase duration to the first traffic direction while halting the opposing direction, whereas Action 1 reverses this allocation. This binary action design establishes a challenging control problem where the agent must learn to alternate efficiently between directions rather than implementing proportional allocation. Each action determines not

only which direction receives right-of-way but also the duration of the subsequent green phase, with maximum phase duration constrained to 30 seconds to prevent excessive waiting for stopped vehicles. Figure 3 illustrates RL framework, DQN state representation and action space.

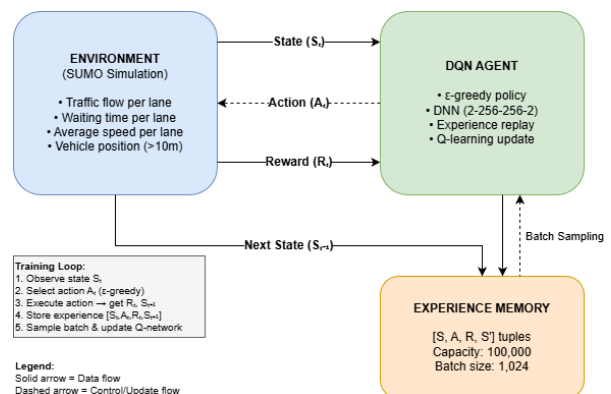


Figure 3. DQN Agent-Environment Interaction Diagram

2.4 Reward Scheme Design and Evaluation

The experimental investigation evaluates six reward schemes that vary the relationship between state observation metrics and reward calculation objectives as shown in Table 1. These schemes combine the three fundamental traffic metrics: Traffic Flow (TF), Waiting Time (WT), and Average Speed (AS) in different configurations. The naming convention follows "X-Y" format, where X denotes the state observation metric and Y represents the reward calculation metric. For instance, the TF-WT scheme uses traffic flow for state representation while optimizing waiting time reduction, whereas the TF-TF scheme uses the same metric for both purposes.

Table 1. Reward Scheme Configurations

Reward Scheme	State Observation	Reward Calculation	Design Rationale
TF-WT [30]	Traffic Flow	Waiting Time	Decouple congestion measurement from delay optimization
TF-TF	Traffic Flow	Traffic Flow	Direct congestion minimization
WT-WT	Waiting Time	Waiting Time	Direct delay minimization
WT-TF	Waiting Time	Traffic Flow	Delay-informed congestion optimization
AS-WT	Average Speed	Waiting Time	Mobility-informed delay optimization

Each reward scheme was evaluated based on fairness in waiting time distribution between opposing traffic directions, quantified as the ratio between maximum and minimum directional waiting times. A fairness value approaching 1.0 indicates fair treatment, while larger values signify increasing imbalance. Additional performance metrics included total system waiting time, average vehicle speed, and traffic throughput. Comparative evaluation included baseline performance from conventional fixed-time traffic control with equal phase durations, establishing

2.5 Experimental Procedure and Implementation

The experimental implementation was developed in Python 3.8, utilizing PyTorch for neural network operations and the SUMO-TraCI interface for real-time simulation control. Training consisted of a 50 epoch per reward scheme, with each epoch comprising a complete 1,800-second simulation.

The DQN agent employed an ϵ -greedy exploration strategy with ϵ initialized at 0.3 and decayed linearly to 0.05 over training, balancing exploration of novel strategies with exploitation of learned knowledge. A discount factor (γ) of 0.8 emphasized medium-term consequences of control decisions. Code implementation followed modular design principles with separate components for simulation management, DQN agent operations, traffic light control logic, and data logging.

3. Results and Discussions

This section presents the experimental findings from implementing six reward schemes for DQN-based adaptive traffic signal control in road repair scenarios. We analyze performance based on fairness metrics, waiting time distributions, and comparative effectiveness of different reward configurations.

3.1 Experimental Results Overview

The performance evaluation of six reward schemes was conducted through 30-minute simulation periods (1,800 seconds) with synthetically generated traffic patterns comprising 1,656 vehicles. Table 2 summarizes the waiting time distributions and fairness values obtained for each scheme, comparing traffic flow from right-to-left (RTL) and left-to-right (LTR) directions.

Table 2. Performance Comparison of Reward Schemes

Reward Scheme	Waiting Time RTL (s)	Waiting Time LTR (s)	Fairness Value	Relative Performance
Timing-based (baseline)	15.25	16.53	1.28	Reference
TF-WT [30]	25.00	26.23	1.23	Moderate improvement
TF-TF	0.00	2657.53	2657.53	Severe imbalance
WT-WT	15.37	16.53	1.16	Good improvement
WT-TF	0.00	2657.53	2657.53	Severe imbalance
AS-WT	18.04	17.00	1.04	Optimal performance
AS-TF	2675.31	0.00	2675.31	Severe imbalance

The fairness metric, calculated as the ratio between maximum and minimum directional waiting times as shown in Equation 2

$$fairness = \frac{\max(wt_{RTL}, wt_{LTR})}{\min(wt_{RTL}, wt_{LTR})} \quad (2)$$

where wt_{RTL} denotes directional waiting times of RTL and wt_{LTR} denotes directional waiting times of LTR. It serves as the primary indicator of fair traffic management.

3.2 Performance Analysis by Reward Scheme Category

The experimental results show a performance hierarchy among the evaluated reward schemes. The AS-WT scheme achieved optimal performance with the best fairness value of 1.04 and balanced waiting times (18.04s RTL, 17.00s LTR), leveraging average speed for comprehensive traffic state representation while using waiting time to directly optimize delay reduction. This approach provides congestion and user experience awareness. The WT-WT scheme also performed strongly (fairness: 1.16), though its slightly inferior results suggest that incorporating contextual speed information enhances decision-making beyond pure waiting-time optimization. The TF-WT scheme demonstrated moderate improvement (fairness: 1.23) but suffered from fundamental limitations, as traffic flow metrics fail to distinguish between vehicles with

short versus extended waiting times, potentially prioritizing high-volume directions regardless of individual delay durations.

Critically, schemes TF-TF, WT-TF, and AS-TF exhibited extreme performance degradation with fairness values exceeding 2,657. The schemes consistently favored one traffic direction entirely while neglecting the other. This failure occurs because the agent learns that maximizing the metric for one direction while minimizing it for the other results in the highest cumulative reward, leading to a "rich-get-richer" effect that worsens traffic imbalances instead of resolving them.

3.3 Discussion of Implementation Insights

The choice of state representation significantly influences agent performance. Average speed proved most effective as it encapsulates both congestion levels, through reduced speeds, and traffic flow quality. In contrast, traffic flow alone provides quantitative but not qualitative information, for example a lane may have many vehicles moving efficiently or few vehicles experiencing severe congestion. Waiting time as a state representation creates a direct but potentially myopic optimization target, as it focuses exclusively on current delays without considering broader traffic dynamics.

Our experimental results support several reward design principles for RL-based traffic control. These include the Complementarity Principle, Fairness Explicitness, and the Multi-metric Advantage, where state and reward metrics should provide complementary rather than identical information; the need for Fairness Explicitness, as efficiency-focused rewards naturally favor already-advantaged directions; and the Multi-metric Advantage, suggesting that real-world implementations would likely benefit from combining fairness, efficiency, and safety considerations.

Several implementation-level factors influenced the experimental outcomes, including the Binary Action Space that amplified performance differences, the Deterministic Traffic Patterns that simplified real-world variability, and the Single Intersection Focus that excluded coordination complexities present in larger networks.

Our findings align with and extend previous research in adaptive traffic control. While Dave [30] demonstrated the feasibility of TF-WT for basic traffic management, our results show that AS-WT achieves superior fairness. Koch et al. [4] emphasized the importance of multi-objective optimization, a perspective supported by our discovery that single-metric approaches often produce imbalanced outcomes. The severe imbalance observed in TF-TF, WT-TF, and AS-TF represents a novel contribution, highlighting a previously undocumented pitfall in reward scheme design.

The superiority of average speed as a state metric corroborates findings from Kumar et al. [5], who noted that speed-based metrics better capture traffic quality

than volume-based measures, though our research extends this insight by demonstrating that speed metrics are most effective when paired with delay-focused rewards rather than volume-focused ones.

The experimental results offer several practical implications for traffic management in road repair zones. Traffic authorities implementing RL-based adaptive control should prioritize AS-WT or WT-WT schemes over TF-based approaches for better fairness outcomes. Systems should track fairness metrics alongside traditional efficiency measures to detect and correct emerging imbalances. In addition, reward schemes might require adaptive adjustment based on time of day, traffic patterns, or specific repair zone characteristics. Combining multiple reward signals or implementing scheme switching based on traffic conditions could further optimize performance.

The demonstrated effectiveness of AS-WT in achieving both efficiency and fairness suggests promising potential for real-world deployment, particularly in constrained environments like road repair zones where equitable access is both practical and ethical. Recent studies on reinforcement learning-based traffic signal control have primarily focused on improving global traffic efficiency metrics such as queue length, throughput, and average waiting time using DQN and other deep reinforcement learning variants. Recent approaches have also explored enhanced learning architectures, multi-objective optimization, and distributed reinforcement learning to improve scalability and robustness in urban traffic environments [31], [32]. In parallel, fairness-aware traffic signal control has gained increasing attention, as several studies have shown that efficiency-oriented reward functions may produce significant imbalance between traffic flows [33], [34]. However, most existing studies focus on generalized urban intersections and symmetric traffic conditions, while studies specifically addressing temporary road repair zones with alternating lane closures remain relatively limited. In contrast to previous work that mainly prioritizes congestion reduction, this study systematically evaluates how different reward scheme combinations influence fairness and adaptive decision-making behavior in constrained and asymmetric traffic scenarios. The findings indicate that combining complementary reward metrics can produce more balanced control policies, reinforcing the growing view that reward engineering plays a critical role in reinforcement learning-based Intelligent Transportation Systems operating under non-ideal traffic conditions.

Although the simulation results demonstrate the effectiveness of the proposed reward schemes, several challenges remain for real-world deployment. Practical implementation of reinforcement learning-based adaptive traffic signal control requires reliable traffic sensing, continuous real-time data acquisition, and stable communication between roadside infrastructure and control systems. Traffic conditions in actual road

repair zones are also more dynamic than simulated environments, as they may be influenced by weather conditions, driver behavior variability, emergency vehicles, and unexpected traffic disturbances. In addition, real-world deployment must consider computational latency, sensor inaccuracies, and integration with existing Intelligent Transportation System (ITS) infrastructure. These factors may affect the stability and adaptability of the learned policy, particularly in highly asymmetric traffic conditions.

4. Conclusions

This study investigated the impact of reward scheme design on a DQN-based adaptive traffic signal control in temporary road repair zones characterized by alternating lane closures. Six reward configurations combining Traffic Flow (TF), Waiting Time (WT), and Average Speed (AS) were evaluated through systematic SUMO simulations. The results show that the AS-WT configuration achieved the most balanced performance, producing the best fairness index (1.04) while maintaining stable traffic flow. Other configurations demonstrated weaker performance, with several schemes leading to severe directional imbalance.

The findings highlight the central role of reward design in reinforcement learning-based traffic control. In asymmetric traffic environments, reward structures that combine complementary metrics enable more balanced decision-making. Average speed reflects global traffic conditions, while waiting time directly captures localized delay reduction. From a theoretical perspective, these results support the view that effective adaptive decision-making in RL systems depends on aligning state representation with optimization objectives, allowing agents to learn policies that balance competing system goals.

The main contribution of this study lies in demonstrating how carefully designed reward schemes can improve fairness in adaptive traffic signal control under non-ideal and constrained traffic conditions. The results provide practical guidance for ITS, suggesting that fairness-aware reward functions should be prioritized when managing temporary bottlenecks such as road repair zones.

This study has several limitations, including a simplified binary action space, synthetic traffic demand, and a single-intersection scenario without network coordination. Future research should investigate proportional green-time allocation, multi-agent reinforcement learning for larger networks, hybrid reward designs incorporating additional objectives such as emissions reduction, and validation through real-world traffic data. Future research should also validate the proposed reward schemes using real-world traffic datasets and field observations from actual road repair zones to evaluate model robustness under dynamic traffic conditions.

Overall, the findings reinforce the potential of RL to support adaptive traffic management in constrained environments and contribute to the development of more robust and fairness-aware control strategies for next-generation ITS.

Acknowledgements

We acknowledge the support received from the Politeknik Caltex Riau under Internal Research Grants Program 2023. We send special thanks to Yokota Laboratory, Gifu University and In-Luck Company for case study of temporary road blockages in Japan.

References

- [1] H. B. Faheem, A. M. El Shorbagy, and M. E. Gabr, "Impact of Traffic Congestion on Transportation System: Challenges and Remediations - A Review," *Mansoura Eng. J.*, vol. 49, no. 2, pp. 1–15, Jan. 2024, doi: [10.58491/2735-4202.3191](https://doi.org/10.58491/2735-4202.3191).
- [2] T. Afrin and N. Yodo, "A Survey of Road Traffic Congestion Measures Towards a Sustainable and Resilient Transportation System," *Sustainability*, vol. 12, no. 11, p. 4660, Jun. 2020, doi: [10.3390/su12114660](https://doi.org/10.3390/su12114660).
- [3] B. Wang, Z. He, J. Sheng, and Y. Chen, "Deep Reinforcement Learning for Traffic Light Timing Optimization," *Processes*, vol. 10, no. 11, p. 2458, Nov. 2022, doi: [10.3390/pr10112458](https://doi.org/10.3390/pr10112458).
- [4] L. Koch, T. Brinkmann, M. Wegener, K. Badalian, and J. Andert, "Adaptive Traffic Light Control with Deep Reinforcement Learning: An Evaluation of Traffic Flow and Energy Consumption," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 12, pp. 15066–15076, Dec. 2023, doi: [10.1109/TITS.2023.3305548](https://doi.org/10.1109/TITS.2023.3305548).
- [5] R. Kumar, N. V. K. Sharma, and V. K. Chaurasiya, "Adaptive Traffic Light Control Using Deep Reinforcement Learning Technique," *Multimed. Tools Appl.*, vol. 83, no. 5, pp. 13851–13872, Feb. 2024, doi: [10.1007/s11042-023-16112-3](https://doi.org/10.1007/s11042-023-16112-3).
- [6] F. Rasheed, K.-L. A. Yau, R. M. Noor, C. Wu, and Y.-C. Low, "Deep Reinforcement Learning for Traffic Signal Control: A Review," *IEEE Access*, vol. 8, pp. 208016–208044, 2020, doi: [10.1109/ACCESS.2020.3034141](https://doi.org/10.1109/ACCESS.2020.3034141).
- [7] W. Jia and M. Ji, "Multi-agent Deep Reinforcement Learning for Large-scale Traffic Signal Control with Spatio-temporal Attention Mechanism," *Appl. Sci.*, vol. 15, no. 15, p. 8605, Aug. 2025, doi: [10.3390/app15158605](https://doi.org/10.3390/app15158605).
- [8] M. Elharoun, S. M. El-Badawy, E. A.-E. Shwaly, and U. E. Shaddah, "Adaptive Traffic Signal Control Using Deep Reinforcement Learning: A Multi-objective Approach for Single and Multi-intersection Scenarios," *IATSS Res.*, vol. 49, no. 4, pp. 481–492, Dec. 2025, doi: [10.1016/j.iatssr.2025.10.004](https://doi.org/10.1016/j.iatssr.2025.10.004).
- [9] A. Saadi, N. Abghour, and Z. Chiba, "A Survey of Reinforcement and Deep Reinforcement Learning for Coordination in Intelligent Traffic Light Control," *J. Big Data*, vol. 12, no. 1, p. 84, 2025, doi: [10.1186/s40537-025-01104-x](https://doi.org/10.1186/s40537-025-01104-x).
- [10] P. Michailidis, I. Michailidis, C. R. Lazaridis, and E. Kosmatopoulos, "Traffic Signal Control via Reinforcement Learning: A Review on Applications and Innovations," *Infrastructures*, vol. 10, no. 5, p. 114, May 2025, doi: [10.3390/infrastructures10050114](https://doi.org/10.3390/infrastructures10050114).
- [11] K. L. Tan, S. Poddar, A. Sharma, and S. Sarkar, "Deep Reinforcement Learning for Adaptive Traffic Signal Control," in *Proc. IEEE 22nd Int. Conf. Intell. Transp. Syst. (ITSC)*, Auckland, NZ, 2019, pp. 353–358, doi: <https://doi.org/10.48550/arXiv.1911.06294>.
- [12] V. Mnih et al., "Human-level Control Through Deep Reinforcement Learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015, doi: [10.1038/nature14236](https://doi.org/10.1038/nature14236).
- [13] P. A. Lopez et al., "Microscopic Traffic Simulation Using SUMO," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Maui, HI, USA, 2018, pp. 2575–2582, doi: [10.1109/ITSC.2018.8569938](https://doi.org/10.1109/ITSC.2018.8569938).
- [14] Y. H. Chow, K. J. A. Ooi, M. A. S. Bhuiyan, M. B. I. Reaz, and C. W. Yuen, "Computation and Optimization of Traffic

- Network Topologies Using Eclipse SUMO," *Ann. Emerg. Technol. Comput.*, vol. 6, no. 4, pp. 31–37, Oct. 2022, doi: [10.33166/AETIC.2022.04.004](https://doi.org/10.33166/AETIC.2022.04.004).
- [15] L. Kuyer, S. Whiteson, B. Bakker, and N. Vlassis, "Multiagent Reinforcement Learning for Urban Traffic Control Using Coordination Graphs," in *Proc. Eur. Conf. Mach. Learn. Knowl. Discov. Databases (ECML PKDD)*, Antwerp, Belgium, 2008, pp. 656–671, doi: [10.1007/978-3-540-87479-9_61](https://doi.org/10.1007/978-3-540-87479-9_61).
- [16] K.-L. A. Yau, J. Qadir, H. L. Khoo, M. H. Ling, and P. Komisarczuk, "A Survey on Reinforcement Learning Models and Algorithms for Traffic Signal Control," *ACM Comput. Surv.*, vol. 50, no. 3, Art. no. 34, May 2018, doi: [10.1145/3068287](https://doi.org/10.1145/3068287).
- [17] H. Luo and S. Jin, "Reinforcement Learning for Traffic Signal Control in Hybrid Action Space," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 6, pp. 5225–5241, Jun. 2024, doi: [10.1109/TITS.2023.3349248](https://doi.org/10.1109/TITS.2023.3349248).
- [18] C. Li, X. Ma, L. Xia, Q. Zhao, and J. Yang, "Fairness Control of Traffic Light via Deep Reinforcement Learning," in *Proc. IEEE 16th Int. Conf. Autom. Sci. Eng. (CASE)*, Hong Kong, 2020, pp. 752–757, doi: [10.1109/CASE48305.2020.9217032](https://doi.org/10.1109/CASE48305.2020.9217032).
- [19] Eclipse Foundation, "SUMO User Documentation," *SUMO - Simulation of Urban Mobility*, 2025. [Online]. Available: <https://sumo.dlr.de/docs/index.html>. [Accessed: Mar. 20, 2025].
- [20] I. Essamlali, H. Nhaila, and M. E. Khaili, "Impact of Urban Block Shape on Traffic and Air Quality: A SUMO-based Comparative Study of Rectangular, Radial, and Triangular Forms," *Transp. Res. Interdiscip. Perspect.*, vol. 31, p. 101413, 2025, doi: [10.1016/j.trip.2024.101413](https://doi.org/10.1016/j.trip.2024.101413).
- [21] A. Sevim, Q. Guo, and E. E. Ozguven, "A Simulation-based Framework for Leveraging Shared Autonomous Vehicles to Enhance Disaster Evacuations in Rural Regions with a Focus on Vulnerable Populations," *J. Infrastruct. Preserv. Resilience*, vol. 6, no. 1, p. 10, 2025, doi: [10.1186/s43065-024-00093-y](https://doi.org/10.1186/s43065-024-00093-y).
- [22] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*, 2nd ed. Cambridge, MA, USA: MIT Press, 2018.
- [23] C. Szepesvári, "Algorithms for Reinforcement Learning," *Synth. Lect. Artif. Intell. Mach. Learn.*, vol. 4, no. 1, pp. 1–103, 2010, doi: [10.1007/978-3-031-01551-9](https://doi.org/10.1007/978-3-031-01551-9).
- [24] D. Silver et al., "A General Reinforcement Learning Algorithm That Masters Chess, Shogi, and Go Through Self-play," *Science*, vol. 362, no. 6419, pp. 1140–1144, Dec. 2018, doi: [10.1126/science.aar6404](https://doi.org/10.1126/science.aar6404).
- [25] M. Koziarski and B. Cyganek, "Impact of Low Resolution on Image Recognition with Deep Neural Networks: An Experimental Study," *Int. J. Appl. Math. Comput. Sci.*, vol. 28, no. 4, pp. 735–744, 2018, doi: [10.2478/amcs-2018-0056](https://doi.org/10.2478/amcs-2018-0056).
- [26] M. Nazari, A. Oroojlooy, L. Snyder, and M. Takac, "Reinforcement Learning for Solving the Vehicle Routing Problem," in *Proc. 32nd Int. Conf. Neural Inf. Process. Syst. (NeurIPS)*, Montréal, Canada, 2018, pp. 9861–9871, doi: <https://doi.org/10.48550/arXiv.1802.04240>.
- [27] S. Yoo, J. B. Lee, and H. Han, "A Reinforcement Learning Approach for Bus Network Design and Frequency Setting Optimisation," *Public Transp.*, vol. 15, no. 2, pp. 503–534, Jun. 2023, doi: [10.1007/s12469-022-00319-y](https://doi.org/10.1007/s12469-022-00319-y).
- [28] M. Roderick, J. MacGlashan, and S. Tellex, "Implementing the Deep Q-Network," *arXiv*, 2017, doi: <https://doi.org/10.48550/arXiv.1711.07478>.
- [29] S. Park, E. Han, S. Park, H. Jeong, and I. Yun, "Deep Q-network-based Traffic Signal Control Models," *PLoS ONE*, vol. 16, no. 9, p. e0256405, Sep. 2021, doi: [10.1371/journal.pone.0256405](https://doi.org/10.1371/journal.pone.0256405).
- [30] M. Dave, "Dynamic Traffic Light Management System," GitHub, 2021. [Online]. Available: <https://github.com/Maunish-dave/Dynamic-Traffic-light-management-system>. [Accessed: Mar. 20, 2025].
- [31] C. Cai and M. Wei, "Adaptive urban traffic signal control based on enhanced deep reinforcement learning," *Scientific Reports*, vol. 14, no. 1, Art. no. 14116, Jun. 2024, doi: [10.1038/s41598-024-64885-w](https://doi.org/10.1038/s41598-024-64885-w).
- [32] K. Cao, L. Wang, S. Zhang, L. Duan, G. Jiang, S. Sfarra, H. Zhang, and H. Jung, "Optimization Control of Adaptive Traffic Signal with Deep Reinforcement Learning," *Electronics*, vol. 13, no. 1, p. 198, Jan. 2024, doi: [10.3390/electronics13010198](https://doi.org/10.3390/electronics13010198).
- [33] Y. Zheng, J. Luo, H. Gao, Y. Zhou, and K. Li, "Pri-DDQN: learning adaptive traffic signal control strategy through a hybrid agent," *Complex & Intelligent Systems*, vol. 11, Art. no. 47, 2025, doi: [10.1007/s40747-024-01651-5](https://doi.org/10.1007/s40747-024-01651-5).
- [34] J. Bao, C. Wu, Y. Lin, L. Zhong, X. Chen, and R. Yin, "A scalable approach to optimize traffic signal control with federated reinforcement learning," *Scientific Reports*, vol. 13, Art. no. 19184, Nov. 2023, doi: [10.1038/s41598-023-46074-3](https://doi.org/10.1038/s41598-023-46074-3).