

V2I-Enhanced State Representation for Adaptive Traffic Signal Control Using Deep Q-Networks

Leila Madi^{1,*}, Ahcene Youcef Benabdallah², and Nouredine Guerssi³

^{1,3} Badji Mokhtar-Annaba University, P.O. Box 12, Annaba 23000, Algeria.

² University 8 Mai 1945, Guelma, Algeria.

* Corresponding author: leila.madi@univ-annaba.dz. Contributing authors: benabdallah.ahcene@gmail.com; guerssi54@yahoo.fr.

Received: 21 July 2026 **Accepted:** 29 August 2026 **Published:** 9 October 2026

DOI: 10.15849/ijasca.164 **License:** Open access under CC BY 4.0. Authors retain copyright.

Publisher: Al-Zaytoonah University of Jordan (ZUJ), Amman, Jordan.

Cite this article: L. Madi, A. Y. Benabdallah, and N. Guerssi, "V2I-enhanced state representation for adaptive traffic signal control using deep Q-networks," *Int. J. Advance Soft Compu. Appl.*, vol. 18, no. 3, pp. 335–356, 356, doi: <https://doi.org/10.15849/ijasca.164>.

ABSTRACT

Traffic congestion remains a major challenge for intelligent transportation systems. Vehicle-to-Infrastructure (V2I) communication provides a promising foundation for adaptive traffic signal control by enabling enhanced real-time traffic detection. This paper proposes an adaptive traffic signal control framework that exploits a V2I-enhanced state representation for isolated intersections using a Deep Q-Network (DQN) agent. The agent learns an optimal control policy directly from V2I-derived state information, including queue lengths, traffic signal phase, and green-light duration, to dynamically adjust signal timing and improve traffic efficiency and stability. The proposed framework was evaluated in a custom stochastic traffic simulation environment implemented in Python/PyTorch, trained for 200 episodes, and assessed over 30 independent stochastic runs under low-, medium-, and high-traffic demand scenarios. Experimental results demonstrate that the proposed approach significantly outperforms both heuristic V2I and fixed-time baseline controllers, reducing the mean queue length by 29.8% under medium traffic demand and 58.3% under high traffic demand, while substantially reducing queue variability across stochastic traffic realizations. Furthermore, a comparative evaluation involving DQN, Double DQN, PPO, and A2C revealed no statistically significant performance differences among the reinforcement learning algorithms, indicating that the observed performance gains primarily originate from the proposed V2I-enhanced state representation rather than from a specific reinforcement learning algorithm. These findings demonstrate the potential of V2I-enabled reinforcement learning as a robust framework for adaptive traffic signal control at isolated intersections.

Keywords: Traffic signal control; Congestion; V2I; Deep Q-Network (DQN).

1 Introduction

Urban road traffic is experiencing sustained growth driven by the expansion of the global vehicle fleet, rapid urbanization, and the concentration of economic activities. This trend leads to an intensification of traffic flow across infrastructures with limited capacity, resulting in recurrent congestion, frequent delays, bottlenecks, and increased travel time variability. Traffic congestion has therefore become one of the major challenges facing modern Intelligent Transportation Systems (ITS), with significant economic, environmental, and road safety consequences [8,13].

Traffic Signal Control (TSC) represents a fundamental component of urban traffic management, particularly at signalized intersections where conflicting traffic flows converge. Conventional approaches, including fixed-time and actuated traffic signal control, have been widely deployed because of their simplicity and operational reliability. However, their performance remains highly dependent on predefined signal timing plans or limited local traffic observations, making them less effective under highly dynamic and stochastic traffic conditions [10,20,21]. Consequently, the effectiveness of adaptive traffic signal control increasingly depends on the availability of accurate, timely, and fine-grained traffic information [13,21].

Recent advances in Reinforcement Learning (RL), particularly Deep Reinforcement Learning (DRL), have demonstrated considerable potential for adaptive traffic signal optimization by enabling controllers to learn efficient signal control policies directly through interaction with the traffic environment [6,20,21]. Deep Q-Networks (DQN) further extend

these capabilities by efficiently approximating optimal control policies in high-dimensional state spaces. Nevertheless, most existing RL-based traffic signal control approaches primarily rely on locally observed traffic states and do not explicitly exploit the richer contextual information provided by Vehicle-to-Infrastructure (V2I) communication [6,20,21]. Conversely, recent V2I-enabled traffic management approaches generally depend on explicit traffic prediction models, heuristic optimization strategies, or model-based control schemes rather than model-free reinforcement learning [13,14].

The existing literature therefore reveals a clear research gap. On the one hand, reinforcement learning-based approaches have demonstrated strong adaptive control capabilities but generally neglect the integration of V2I-derived traffic information into the learning state. On the other hand, V2I-enabled traffic management methods effectively exploit real-time traffic information but remain largely dependent on predictive models or heuristic optimization techniques. Consequently, the potential benefits of combining a V2I-enhanced traffic state representation with model-free deep reinforcement learning for adaptive traffic signal control at isolated intersections remain insufficiently investigated [6,13,14,20,21].

To address this gap, this paper proposes an adaptive traffic signal control framework integrating Vehicle-to-Infrastructure (V2I) communication with a Deep Q-Network (DQN) agent. The proposed framework exploits a V2I-enhanced state representation composed of queue lengths, traffic signal phase, and green-light duration, enabling the DQN agent to learn an adaptive signal control policy directly through continuous interaction with the traffic environment without relying on explicit traffic prediction models.

The proposed framework is evaluated using a custom stochastic traffic simulation environment developed in Python/PyTorch under low-, medium-, and high-traffic demand scenarios. Its performance is compared with two representative baseline controllers: a conventional fixed-time controller and a proactive V2I-based heuristic controller. Experimental results demonstrate that the proposed framework significantly reduces both the mean and variability of queue lengths while outperforming the baseline strategies across different traffic demand levels.

The main contributions of this work are summarized as follows:

- A V2I-enhanced state representation integrating queue lengths, traffic signal phase, and green-light duration for adaptive traffic signal control.
- A model-free Deep Q-Network framework capable of learning adaptive signal control policies directly from V2I-derived traffic information without relying on explicit traffic prediction models.
- A comprehensive experimental evaluation under low-, medium-, and high-traffic demand scenarios, including statistical validation over 30 independent evaluation runs, demonstrating significant improvements over both fixed-time and proactive V2I heuristic controllers.

2 Related work

Existing research on urban traffic optimization can be broadly categorized into three main directions: swarm intelligence and reinforcement learning-based optimization approaches, reinforcement learning-based intelligent traffic signal control methods, and V2X/V2I-enabled approaches for congestion detection and management.

First, Swarm Intelligence and Reinforcement Learning for Optimization: Several studies combine swarm optimization techniques with reinforcement learning to enhance convergence quality in complex optimization problems. Works [1–3] propose hybrid variants integrating

Particle Swarm Optimization (PSO) with reinforcement learning mechanisms in order to avoid local minima and accelerate convergence. These approaches have been applied to general optimization tasks and trajectory planning for autonomous vehicles. The efficacy of these systems is often predicated on precise vehicle detection and tracking, which have been extensively explored through ground-level image processing [7] and spatiotemporal analysis of satellite video [9].

Although these studies demonstrate the effectiveness of hybrid optimization strategies in complex search spaces, their primary objective is optimization or traffic perception rather than adaptive traffic signal control. Consequently, they neither exploit V2I communication nor address real-time signal control at isolated urban intersections.

Second, Reinforcement Learning for Intelligent Traffic Signal Control: Deep reinforcement learning has recently emerged as a promising paradigm for adaptive traffic signal control. This evolution follows long-standing mathematical models, such as those based on the modified Webster function [10], which provided the initial groundwork for traffic light optimization. Ault and Sharon [20] introduced standardized reinforcement learning benchmarks for traffic signal control, highlighting the potential of RL-based methods while also emphasizing challenges related to training stability and policy generalization across varying traffic scenarios. Recent overviews of emerging techniques further highlight the shift towards predictive analytics for urban mobility management [8].

While some studies employ PPO or DQN for network-wide control [6], others focus on evaluating the relative efficiency of learning-based methods against conventional control. For instance, Mahato [21] provided a comparative analysis between Multi-Agent Reinforcement Learning (MARL) and fixed-time strategies, highlighting the superior adaptability of autonomous agents in dynamic environments.

Other studies, such as [5], apply an advanced variant of Double Q-Learning with Prioritized Experience Replay (DQL-PER) to the vehicle routing problem, demonstrating superior convergence speed and reduced estimation bias compared with conventional Q-learning and DQN approaches. A different perspective is presented in [4], where phase selection is governed by an urgency function represented as an explainable tree structure and optimized through genetic programming, emphasizing interpretability rather than purely value-based learning.

Despite their promising results, most reinforcement learning-based approaches either target large-scale traffic networks or rely exclusively on locally observed traffic states. As a result, they do not explicitly exploit the richer contextual information provided by V2I communication to improve decision making at isolated intersections.

Third, V2X/V2I-Based Approaches: V2X technologies enable rich information exchange between vehicles and infrastructure, opening new opportunities for more informed traffic control strategies. Recent literature emphasizes the fundamental role of Intelligent Transportation Systems (ITS) as a cornerstone for the development of sustainable smart cities [18]. In this context, Vehicular Ad Hoc Network (VANET) technology has been extensively investigated for its capacity to mitigate urban congestion through decentralized monitoring strategies [17,18]. A key challenge in these networks is maintaining communication reliability; to this end, the implementation of load-aware and priority-adaptive methods [15] ensures that critical information flows remain efficient, even under conditions of high network density. While some research focuses on the stability of distributed congestion control protocols [19], others leverage VANET frameworks for precise state estimation and cooperative detection. For instance, Zhang and Zhang [18] demonstrate the effectiveness of a V2V-based cooperative approach for congestion detection, which operates autonomously without reliance on centralized infrastructure.

Furthermore, the advancement of cooperative V2X perception is supported by the development of large-scale real-world datasets such as V2X-Real [11] and V2V4Real [12]. These contributions provide valuable multimodal benchmarks for training and evaluating cooperative perception systems.

Regarding Traffic Signal Control (TSC), recent studies have progressively shifted toward predictive and anticipatory strategies. Qiao et al. [13] proposed a V2I-based predictive control strategy optimized through genetic algorithms within a receding-horizon framework, whereas Huang et al. [14] introduced an anticipatory control scheme incorporating a metamodel to bridge the gap between theoretical traffic models and real-world dynamics.

Although these studies demonstrate the potential of V2I-enabled traffic management, many existing approaches remain dependent on explicit traffic prediction models, heuristic optimization, or carefully calibrated metamodels. Such assumptions may limit their adaptability to highly dynamic traffic environments.

In contrast, the proposed framework integrates V2I-derived traffic information into a model-free Deep Q-Network, enabling the control policy to be learned directly through continuous interaction with the traffic environment without relying on explicit traffic prediction models.

Overall, the reviewed literature reveals three major research gaps: (i) optimization-oriented approaches that are not specifically designed for adaptive traffic signal control, (ii) reinforcement learning methods that generally rely on locally observed traffic states without exploiting V2I-enhanced information, and (iii) V2I-based approaches that remain largely dependent on predictive or heuristic models. To address these limitations, the proposed framework combines a V2I-enhanced state representation with model-free deep reinforcement learning for adaptive traffic signal control at isolated intersections.

To further position the proposed approach with respect to the existing literature, Table 1 presents a comparative analysis of representative reinforcement learning–based and V2I-enabled traffic signal control studies, highlighting their methodologies, evaluation environments, principal contributions, and the research gaps addressed by the proposed framework.

Table 1. Critical comparison of representative reinforcement learning–based and V2I-enabled adaptive traffic signal control approaches.

Reference	Methodology	Evaluation Environment	Performance Metrics	Main Contribution	Research Gap
Zhang et al. [10]	Modified Webster traffic flow model	Analytical traffic flow model	Average delay, cycle efficiency	Establishes a deterministic baseline for traffic signal timing	Unable to adapt to stochastic traffic conditions or real-time traffic fluctuations
Liao et al. [4]	Genetic Programming with explainable urgency trees	Synthetic traffic scenarios	Average delay, policy interpretability	Introduces interpretable traffic signal decision rules	Limited flexibility due to predefined tree structures; lacks deep representation learning
Skoropad et al. [6]	Deep Reinforcement Learning (DQN/PPO)	Urban traffic datasets	Traffic throughput, prediction accuracy	Demonstrates the effectiveness of deep RL for urban traffic optimization	Focuses mainly on large-scale optimization without exploiting V2I-derived traffic information
Qiao et al. [13]	V2I predictive traffic signal control using Genetic Algorithms	V2I receding-horizon simulation	Queue length, throughput	Performs anticipatory traffic signal optimization using	Relies on explicit prediction models and heuristic optimization

				short-term traffic prediction	
Huang et al. [14]	Model-based anticipatory traffic signal control	Traffic simulation with metamodeling	Travel time, delay	Bridges analytical traffic models and realistic traffic dynamics	Requires continuous model calibration and accurate traffic prediction
Ault & Sharon [20]	Reinforcement Learning benchmark for Traffic Signal Control	Standard RL benchmark environments	Queue length, cumulative delay	Demonstrates the superiority of RL over fixed-time control	Uses conventional traffic measurements without V2I-enhanced information
Mahato [21]	Multi-Agent Reinforcement Learning (MARL)	Benchmark traffic scenarios	Queue length, delay, convergence	Shows the adaptability of MARL under dynamic traffic conditions	Does not integrate V2I communication into the learning process
Proposed Framework	Model-free Deep Q-Network with V2I-enhanced state representation	Custom Python/PyTorch stochastic traffic simulator (low-, medium-, and high-demand scenarios)	Mean queue length, queue variability, cumulative reward, training loss, statistical significance (30 independent runs)	Integrates V2I-derived traffic information into a model-free RL framework, achieving statistically validated improvements over Fixed-Time and heuristic V2I controllers	Currently validated for a single isolated intersection under ideal V2I assumptions (100% penetration, no latency or packet loss); future work will consider realistic communication conditions and multi-intersection coordination.

3 Proposed approach

Building upon the limitations identified in the existing literature, the present work seeks to address the gap between V2I-enabled traffic state awareness and adaptive learning-based traffic signal control. By integrating communication-driven perception with data-driven decision-making, this study aims to establish a unified framework for intelligent intersection management.

Accordingly, we propose a Deep Q-Network (DQN)-based adaptive signal control framework that explicitly incorporates V2I communication data into the traffic state representation. This integration enables the learning of an efficient signal timing control policy through continuous interaction with the environment. Unlike purely predictive or heuristic approaches, the proposed method learns control decisions directly from traffic dynamics rather than relying on predefined threshold rules. The objective is not solely congestion detection or prediction, but the learning of a robust control policy capable of dynamically adjusting green signal duration to minimize both average queue length and queue variability under varying traffic demand levels. The overall architecture of the proposed DQN-V2I traffic signal control framework is presented in (Figure 1).

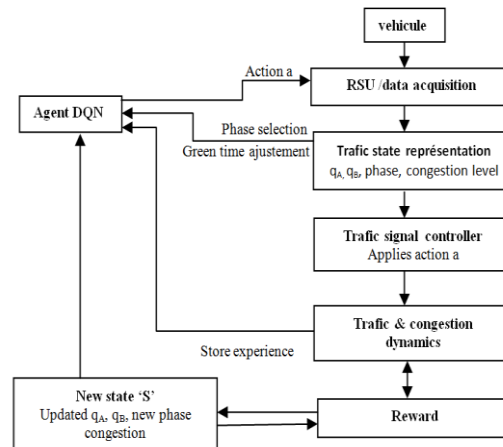


Figure 1. DQN-V2I Traffic Signal Control

In the following section, we describe the formalization of the intersection environment, the state and action spaces, the reward formulation, and the DQN architecture, providing a complete methodological foundation for the proposed control strategy.

4 Material and Methods

This section delineates the proposed V2I-enabled adaptive traffic signal control framework, designed to optimize intersection efficiency by minimizing congestion and queue accumulation via real-time connectivity. To evaluate this framework, a custom discrete-time stochastic traffic simulation environment was developed in Python utilizing the PyTorch library and deployed on the Google Colab cloud infrastructure. The environment models an isolated, two-phase signalized intersection under stochastic demand. Unlike general-purpose traffic simulators such as SUMO, the proposed custom simulation environment provides direct control over traffic state transitions and reinforcement learning interactions, enabling a controlled evaluation of the proposed V2I-enhanced framework. It was adopted as a proof-of-concept platform, while validation using SUMO is considered future work. Each discrete time step (t) corresponds to a synchronous operational update cycle where vehicle arrivals, departures, queue evolutions, and phase transitions are computed simultaneously. Vehicle arrivals at each approach are modeled as independent Poisson streams, where the probability of observing (k) vehicle arrivals during a given time step is defined by the Poisson probability mass function:

$$P(K = k) = \frac{\lambda^k e^{-\lambda}}{k!} \quad (1)$$

where λ denotes the arrival rate parameter controlling the expected traffic demand at a given approach. Vehicle arrivals at each approach were modeled as independent Poisson processes to reproduce the stochastic nature of traffic demand while providing a controlled and reproducible simulation environment. Three arrival-rate configurations were defined to represent low-, medium-, and high-demand scenarios, enabling the proposed controller to be evaluated under progressively increasing traffic loads while maintaining identical experimental conditions across all control strategies. This scenario-based traffic modeling is commonly adopted in studies on isolated intersection traffic signal control and provides a systematic framework for assessing controller robustness under different congestion levels [10,13]. The corresponding arrival rates are summarized in Table 2.

Table 2. Traffic demand scenarios defined by the Poisson arrival rates at the two intersection approaches.

Scenario	Approach A Arrival Rate (λ_A)	Approach B Arrival Rate (λ_B)
Low demand	0.3	0.4
Medium demand	0.6	0.9
High demand	0.9	1.2

The maximum holding capacity of each queue is bounded by $Q_{max} = 30$ vehicles to represent physical storage constraints at the intersection approaches, such that:

$$q_A, q_B \in [0, Q_{max}]$$

4.1 Reinforcement Learning Formulation

State Representation

The system state is constructed from traffic information assumed to be available through V2I communication. The state vector at time step is defined as:

$$S_t = [q_A, q_B, g_t, p_t]$$

where:

- q_A and q_B denote the queue lengths of approaches A and B.
- g_t represents the current green time duration.
- p_t indicates the active signal phase.

This representation provides the agent with real-time traffic awareness necessary for adaptive decision-making. These four state variables were deliberately selected to provide a compact yet informative representation of the traffic environment. Queue lengths directly quantify the congestion level on each approach, while the current signal phase reflects the controller's operational status. The green-light duration enables the agent to adapt signal timing dynamically according to evolving traffic conditions. Together, these variables capture the essential information required for effective decision-making while avoiding an unnecessarily large state space that could increase learning complexity and slow convergence.

Action Space

The action space is discrete and consists of incremental adjustments to the current green time:

$$A = \{-1, 0, +1\}$$

Each action modifies the green duration within predefined limits:

$$g_{t+1} = \text{clip}(g_t + a_t, g_{min}, g_{max}) \quad (2)$$

Where g_{min} and g_{max} define operational bounds.

The structural signal constraints strictly enforce that the final allocated green time remains bounded within a predefined operational range:

$$g \in [g_{min}, g_{max}], \text{ where } g_{min} = 1 \text{ and } g_{max} = 6$$

Reward Function

The reward is defined as the variation in total queue length between two consecutive time steps:

$$r_t = Q_{t-1} - Q_t \quad (3)$$

Where $Q_t = q_A + q_B$.

The variation in total queue length was selected as the reward because it provides an immediate and directly observable measure of congestion evolution from the V2I-derived state information. Unlike metrics such as average delay, waiting time, or throughput, which generally require longer observation windows, queue-length variation provides rapid feedback to the learning agent, thereby facilitating stable policy optimization. Moreover, this simple reward formulation was intentionally adopted to isolate and evaluate the contribution of the proposed V2I-enhanced state representation without introducing additional optimization objectives.

To improve training stability and prevent excessively large temporal-difference updates caused by abrupt stochastic queue fluctuations under high traffic demand, a symmetric reward clipping function is applied. This clipping bounds the reward within the interval $([-5, 5])$, thereby reducing reward variance and stabilizing neural network optimization during reinforcement learning.

$$r = \max(-5, \min(5, r_{raw})) \quad (4)$$

4.2 DQN Architecture

The decision agent is powered by a Deep Q-Network (DQN) which approximates the optimal action-value function $Q^*(s, a)$. The network architecture is a Fully Connected Multi-Layer Perceptron (MLP) consisting of an input layer (dim=4), two consecutive hidden layers with 64 neurons each utilizing Rectified Linear Unit (ReLU) activations ($f(x) = \max(0, x)$), and a linear output layer mapping to the 3 discrete actions (dim=3) as shown in (Figure2).

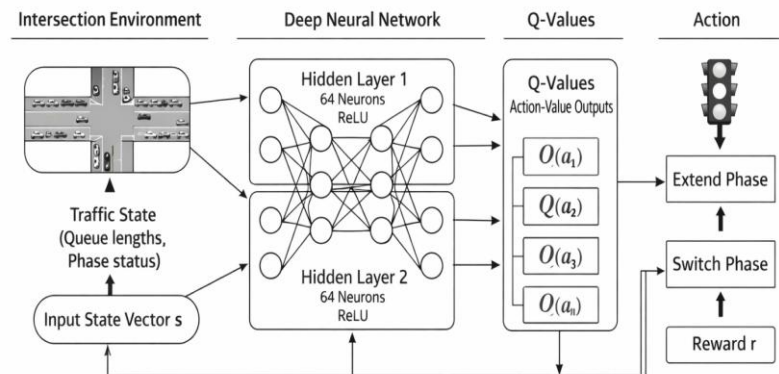


Figure 2. Architecture of the Proposed Deep Q-Network for Adaptive Traffic Signal Control

The agent is trained over $M = 200$ episodes, with each episode running for $T = 1000$ discrete time steps. To balance exploration and exploitation, an exponential ϵ -greedy strategy is deployed, initialized at $\epsilon_0 = 0.3$ and decayed at the end of each episode according to:

$$\epsilon_{ep} = \max(0.05, \epsilon_{ep-1} * 0.99) \quad (5)$$

Once the memory contains at least 32 transitions, mini-batches of size $b = 32$ are randomly sampled from D to update the network. For each sampled transition $i = (s_i, a_i, r_i, s'_i)$ in the mini-batch, a target value is first computed as:

$$y_i = r_i + \gamma \max_{a'} Q(s'_i, a'; \theta) \quad (6)$$

where s_i and s'_i denote, respectively, the state and the subsequent state recorded at the time step at which the transition was originally collected. The network parameters θ are then updated by minimizing the temporal-difference (TD) error through the Mean Squared Error (MSE) loss function:

$$L(\theta) = 1/b \sum_{i=1}^b (Q(s_i, a_i; \theta) - y_i)^2 \quad (7)$$

It should be noted that, for the standard DQN agent, both the current action-value estimate $Q(s_i, a_i; \theta)$ and the target y_i are computed using the same online parameters θ , consistent with the single-network configuration reported in Table 3. A separate target network, updated every five episodes, is employed only for the Double DQN baseline introduced for comparison purposes (Section 5.4), in order to mitigate the overestimation bias inherent to value-based bootstrapping

To guarantee complete algorithmic replicability, global system random seeds (random, numpy, and torch) were strictly fixed to 0. For performance validation, multi-run evaluations across $N_{eval} = 30$ independent stochastic simulation runs. All hyperparameters used in this custom simulation framework are consolidated in Table 3.

Table 3. Implementation Hyperparameters and Simulation Settings.

Parameter Category	Hyperparameter / System Property	Value	Formal Notation
Traffic Simulation	Simulation platform	Python 3.x, PyTorch (CPU)	—
Traffic Simulation	Traffic generation model	Stochastic Poisson arrival process	—
Traffic Simulation	Time steps per episode	1000	T
Traffic Simulation	Independent stochastic evaluation runs	30	N_{eval}
Traffic Simulation	Traffic demand (Approaches A/B)	Low: (0.3, 0.4) Medium: (0.6, 0.9) High: (0.9, 1.2)	(λ_A, λ_B)
Traffic Simulation	Maximum physical queue capacity	30 vehicles	$Q_{max} = 30$
Traffic Simulation	Green time bound constraints	1–6 steps	$g \in [g_{min}, g_{max}]$
Traffic Simulation			
DQN Architecture	State space dimension	4	$S_t = [q_A, q_B, g_t, p_t]$
DQN Architecture	Action space dimension	3	$A = \{-1, 0, +1\}$
DQN Architecture	Network topology	Fully connected MLP (4→64→64→3, ReLU)	—
RL Training	Optimization algorithm	Adam	—
RL Training	Learning rate	0.001	$\alpha = 10^{-3}$
RL Training	Discount factor	0.9	$\gamma = 0.9$
RL Training	Replay buffer capacity	1000 transitions	—
RL Training	Training batch size	32	—
RL Training	Exploration strategy	ϵ -greedy	—
RL Training	Initial / minimum ϵ	0.30 → 0.05	$\epsilon_0 = 0.3, \epsilon_{min} = 0.05$
RL Training	ϵ decay rate	0.99 (per episode)	$\epsilon \leftarrow \max(\epsilon_{min}, 0.99 \cdot \epsilon)$
RL Training	Target network update	Every 5 episodes (Double DQN only)	—

RL Training	Gradient clipping	Max norm = 1.0 (Double DQN only)	$ \nabla \leq 1.0$
RL Training	Training episodes	200	—

Table 3 summarizes the hyperparameters and simulation settings used in the experimental study. To ensure a fair comparison, all reinforcement learning methods were trained and evaluated under identical experimental conditions, including the same traffic demand scenarios, training duration, and evaluation protocol. Common parameters, such as the discount factor, number of training episodes, and state-action space dimensions, were kept consistent across all methods whenever applicable. Algorithm-specific hyperparameters were configured according to the requirements of each reinforcement learning model.

4.3 Integration With V2I Communication

The DQN controller's state representation is structurally designed to reflect data obtainable through Vehicle-to-Infrastructure (V2I) communication. In a connected traffic environment, vehicles periodically transmit Basic Safety Messages (BSMs) containing position estimates, instantaneous velocity, and movement status to a Roadside Unit (RSU) at the intersection. In the Google Colab implementation, this architecture is emulated through direct access to environment state variables: q_A , q_B , and the current signal phase pt serve as analytical abstractions of V2I-acquired traffic information. To evaluate the fundamental feasibility of the control loop, the simulation assumes an ideal connectivity scenario with 100% vehicle penetration, zero communication latency, and lossless packet transmission, thereby providing high-fidelity, noise-free state observations. This idealized setup establishes a performance reference for assessing the optimization capability of the proposed framework before introducing realistic constraints such as transmission delays, packet loss, or partial V2I adoption.

4.4 Baseline Methods

To rigorously evaluate the proposed V2I+DQN framework, five baseline methods were selected, including two conventional traffic signal controllers (Fixed-Time and V2I Proactive Heuristic) and three reinforcement learning algorithms (Double DQN, A2C, and PPO). The Fixed-Time controller serves as a conventional benchmark widely adopted in traffic signal control studies, whereas the V2I Proactive Heuristic isolates the contribution of V2I-derived traffic information without reinforcement learning [13,21]. Double DQN, A2C, and PPO were selected because they represent well-established value-based and policy-based reinforcement learning approaches that are widely used as benchmark algorithms in adaptive traffic signal control research [20,22–25]. To ensure a fair comparison, all learning-based methods were trained and evaluated under identical traffic demand scenarios, training episodes, and evaluation protocols, while algorithm-specific hyperparameters were configured according to the requirements of each method. This experimental design allows the performance gains associated with the proposed V2I-enhanced state representation to be distinguished from those attributable to the underlying reinforcement learning algorithm, thereby providing a rigorous assessment of the proposed framework.

4.4.1 Fixed-Time Controller

The Fixed-Time controller serves as the conventional baseline. The green signal duration is maintained at a constant value of three-time steps ($BASE_GREEN = 3$), regardless of traffic

conditions or queue dynamics. Since no traffic observations or V2I information are considered, this controller provides a reference for evaluating the benefits of adaptive traffic signal control.

4.4.2 V2I Proactive Heuristic Controller

The proposed V2I Proactive Heuristic controller implements a rule-based adaptive signal control strategy using traffic observations derived from V2I communication. At each time step t , the expected queue growth for each approach is estimated as:

$$\hat{q}_A(t) = q_A(t) + \lambda_A, \hat{q}_B(t) = q_B(t) + \lambda_B \quad (8)$$

where $q_A(t)$ and $q_B(t)$ denote the current queue lengths, while λ_A and λ_B represent the expected vehicle arrival rates. The green signal duration is then adjusted according to a threshold-based policy. When the predicted congestion of the active approach exceeds the maximum queue capacity Q_{max} , the green duration is increased to improve traffic discharge:

$$g_{t+1} = \min((g_t + 1, g_{max})) \quad (9)$$

Otherwise, the signal duration is progressively restored toward the nominal value g_{base} . This rule-based baseline provides adaptive signal control using V2I-derived traffic information without reinforcement learning, thereby isolating the contribution of the learning component.

4.4.3 Double Deep Q-Network (DDQN)

Double Deep Q-Network (DDQN) extends the standard DQN by employing a separate target network for target value estimation, thereby reducing Q-value overestimation and improving training stability. In the implemented framework, the target network is synchronized with the online network every five training episodes.

4.4.4 Advantage Actor-Critic (A2C)

Advantage Actor-Critic (A2C) is a policy-gradient reinforcement learning algorithm based on an actor-critic architecture, where the actor learns the control policy and the critic estimates state values to guide policy updates through the advantage function.

4.4.5 Proximal Policy Optimization (PPO)

Proximal Policy Optimization (PPO) is a policy-gradient algorithm that improves training stability through a clipped surrogate objective, preventing excessively large policy updates during optimization.

5 Results

The experimental validation covers three stochastic demand configurations (low: $\lambda_A=0.3$, $\lambda_B=0.4$; medium: $\lambda_A=0.6$, $\lambda_B=0.9$; high: $\lambda_A=0.9$, $\lambda_B=1.2$), with 30 independent runs of 1,000 steps per configuration. The primary metric is mean total queue length (μ), with stability quantified by standard deviation (σ) and 95% confidence intervals. Statistical significance across five pairwise comparisons per scenario (V2I+DQN vs. Fixed-Time, V2I Proactive Heuristic, Double DQN, A2C, and PPO) is assessed via the non-parametric Mann-Whitney U test, with Bonferroni correction ($\alpha^* = 0.01$) controlling the family-wise error rate. Results are consolidated in Table 4.

Table 4. Statistical Performance Metrics Across 30 Stochastic Runs (Mean $\pm$ 95% Confidence Interval) for different demand levels.

Scenario	Method	Mean (μ)	Std (σ)	95% CI ($\pm$)	95% CI Range
----------	--------	----------------	------------------	------------------	--------------

Low	Baseline	0.363	0.023	0.008	[0.355, 0.371]
	V2I	0.362	0.016	0.006	[0.356, 0.368]
	V2I+ DQN	0.388	0.026	0.009	[0.379, 0.398]
Medium	Baseline	1.089	0.068	0.024	[1.065, 1.114]
	V2I	1.077	0.078	0.028	[1.049, 1.105]
	V2I+ DQN	0.758	0.029	0.010	[0.748, 0.768]
High	Baseline	2.590	0.317	0.113	[2.476, 2.703]
	V2I	2.689	0.337	0.121	[2.568, 2.809]
	V2I+ DQN	1.079	0.026	0.009	[1.070, 1.089]

5.1 Performance Analysis Under Low Traffic Demand

Under low-demand traffic conditions, the intersection operates well below its saturation level, resulting in similar performance across all evaluated control strategies. The Fixed-Time controller records a mean queue length of $\mu = 0.363$ ($\sigma = 0.023$, 95% CI ± 0.008), while the V2I Proactive Heuristic achieves $\mu = 0.362$ ($\sigma = 0.016$, 95% CI ± 0.006). The proposed V2I+DQN framework yields $\mu = 0.388$ ($\sigma = 0.026$, 95% CI ± 0.009), corresponding to a statistically significant but operationally negligible increase of only 0.025 vehicles compared with the Fixed-Time controller (Mann–Whitney U test, $p < 0.05$, Bonferroni corrected). This behavior is consistent with the sparse-reward nature of reinforcement learning in lightly loaded traffic environments. Because queues remain short and traffic conditions change only marginally over time, the reward signal provides limited information for distinguishing between alternative actions. As a result, adaptive learning offers little practical advantage over conventional control strategies when traffic demand is low. Despite this slight performance degradation, the narrow confidence intervals (95% CI $\leq \pm 0.009$) indicate highly consistent results across the 30 independent stochastic runs, confirming the robustness of all evaluated controllers under uncongested conditions. Importantly, this limitation disappears under medium- and high-demand scenarios, where richer traffic dynamics allow the DQN agent to exploit V2I-derived information more effectively. Figure 3 illustrates the distribution of the average queue lengths obtained by the three evaluated traffic signal control strategies under the low-demand scenario.

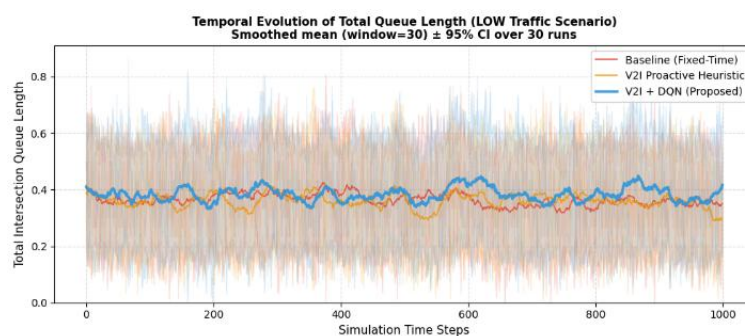


Figure 3. Temporal evolution of total queue length across control methodologies: Performance evaluation under low-density traffic conditions.

5.2 Performance Analysis Under Medium Traffic Demand

The medium-demand scenario represents the first traffic condition in which meaningful performance differences among the evaluated control strategies become evident. The Fixed-Time controller deteriorates to $\mu = 1.089$ ($\sigma = 0.068$, 95% CI ± 0.024), while the V2I Proactive Heuristic achieves a comparable $\mu = 1.077$ ($\sigma = 0.078$, 95% CI ± 0.028), indicating that simple threshold-based adaptations are insufficient to cope with sustained stochastic traffic demand. In contrast, the proposed V2I+DQN framework achieves $\mu = 0.758$ ($\sigma = 0.029$, 95% CI ± 0.010), corresponding to statistically significant reductions of 30.4% and 29.6% compared with the Fixed-Time and V2I Proactive Heuristic controllers, respectively ($p < 0.001$). These improvements arise because medium traffic demand generates richer traffic dynamics and more informative reward signals than low-demand conditions, enabling the DQN agent to learn a more effective control policy. By exploiting V2I-derived traffic information, the agent continuously adapts green signal durations to the current traffic state instead of relying on fixed timing plans or threshold-based rules. Beyond reducing the average queue length, the proposed framework also improves traffic stability. The standard deviation decreases by 57.4% relative to the Fixed-Time controller and by 62.8% relative to the V2I Proactive Heuristic, indicating a more robust policy under stochastic traffic conditions. Figure 4 illustrates the temporal evolution of the total queue length under the medium-demand scenario. The smoother trajectory of the proposed controller confirms its ability to maintain stable traffic flow throughout the 1,000 simulation time steps.

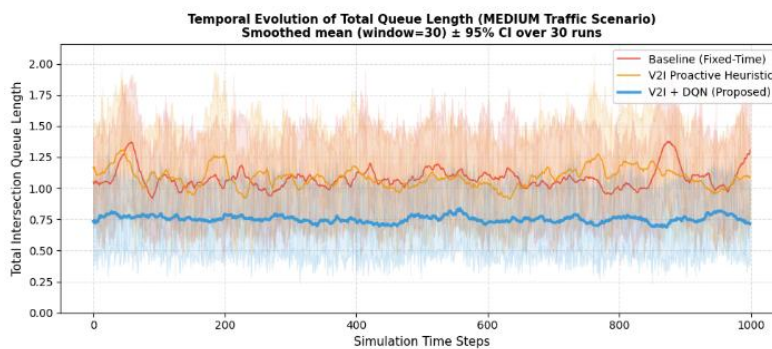


Figure 4. Temporal evolution of total queue length across control methodologies: Performance evaluation under medium -density traffic conditions.

5.3 High-Demand Analysis: The Robustness of RL

The high-demand scenario represents the most challenging operating condition, where the differences between the evaluated control strategies become the most pronounced. Under these near-saturated traffic conditions, the Fixed-Time controller reaches a mean queue length of $\mu = 2.590$ ($\sigma = 0.317$, 95% CI ± 0.113), while the V2I Proactive Heuristic records $\mu = 2.689$ ($\sigma = 0.337$, 95% CI ± 0.121), indicating that static and threshold-based control strategies are unable to effectively accommodate persistent queue growth under heavy stochastic traffic demand. In contrast, the proposed V2I+DQN framework maintains a mean queue length of $\mu = 1.079$ ($\sigma = 0.026$, 95% CI ± 0.009), corresponding to reductions of 58.3% and 59.9% compared with the Fixed-Time and V2I Proactive Heuristic controllers, respectively ($p < 0.001$). This improvement is attributed to the richer learning opportunities available under high traffic demand. As congestion becomes more frequent, the DQN agent is exposed to a wider range of traffic states, allowing it to progressively learn a more effective signal control policy. Combined with the real-time traffic information provided by the V2I communication framework, the learned policy adapts green signal durations more efficiently, thereby limiting queue growth under saturated traffic conditions.

The proposed framework also demonstrates excellent robustness. The low standard deviation and the narrow 95% confidence interval ([1.070, 1.089]) indicate highly consistent performance across the 30 independent stochastic simulation runs. Figure 5 illustrates the temporal evolution of the total queue length, showing that the proposed controller maintains a stable traffic flow throughout the simulation, whereas the Fixed-Time and V2I Proactive Heuristic controllers exhibit larger oscillations and sustained queue accumulation.

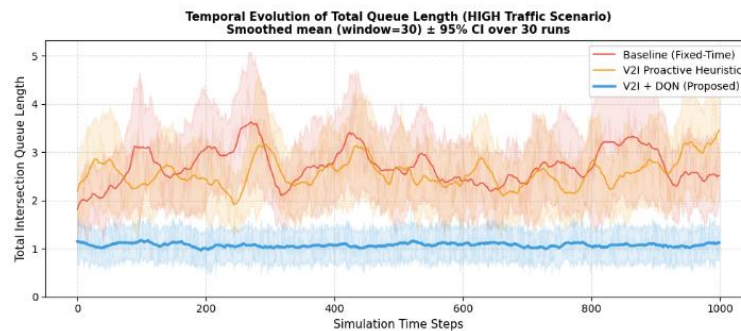


Figure 5. Temporal evolution of total queue length across control methodologies: Performance evaluation under high-density traffic conditions.

5.4 Comparative Performance Of RL Variants

To evaluate the robustness of the proposed V2I integration framework independently of the specific reinforcement learning algorithm employed, four RL variants spanning both value-based and policy gradient paradigms — DQN, Double DQN, A2C, and PPO — are evaluated under identical experimental conditions and hyperparameter configurations. Table 5 presents the complete pairwise statistical test results between V2I+DQN and all competing methods.

Table 5. Statistical Significance Tests — V2I+DQN vs. All Methods (Mann-Whitney U, Bonferroni Corrected)

Scenario	Comparison	p-value	P (Bonferroni)	Significant ($\alpha=0.05$)	Interpretation
Medium	DQN vs Baseline	< 0.001	< 0.001	YES	DQN significantly better
Medium	DQN vs Heuristic	< 0.001	< 0.001	YES	DQN significantly better
Medium	DQN vs Double DQN	0.399	1.000	NO	No statistically significant difference
Medium	DQN vs A2C	0.734	1.000	NO	Confirms generalizability No statistically significant difference
Medium	DQN vs PPO	0.325	1.000	NO	No statistically significant difference
High	DQN vs Baseline	< 0.001	< 0.001	YES	DQN significantly better
High	DQN vs Heuristic	< 0.001	< 0.001	YES	DQN significantly better
High	DQN vs Double DQN	0.056	0.282	NO	No statistically significant difference
High	DQN vs A2C	0.072	0.362	NO	No statistically significant difference
High	DQN vs PPO	0.460	1.000	NO	No statistically significant difference

Under medium traffic conditions, the mean queue lengths of all RL variants converge within a narrow range of [0.752, 0.777] vehicles, with no statistically significant pairwise difference detected relative to V2I+DQN ($p > 0.28$ in all cases, Bonferroni corrected). Under high traffic

conditions, convergence is similarly observed, with all RL variants recording mean queue lengths in the range [1.066, 1.094], none of which differs significantly from V2I+DQN ($p > 0.05$ in all cases). These observations are discussed and interpreted in Section 4.

5.5 Comparative Performance Evaluation And Statistical Validation

The overall results demonstrate that the effectiveness of the proposed V2I+DQN framework increases with traffic demand. Under low-demand conditions, all controllers achieve comparable performance because congestion remains limited, leaving little opportunity for adaptive learning to improve signal control. As traffic demand increases, however, the proposed framework consistently outperforms both the Fixed-Time and V2I Proactive Heuristic controllers, with the performance gap becoming progressively larger under medium and high traffic conditions..

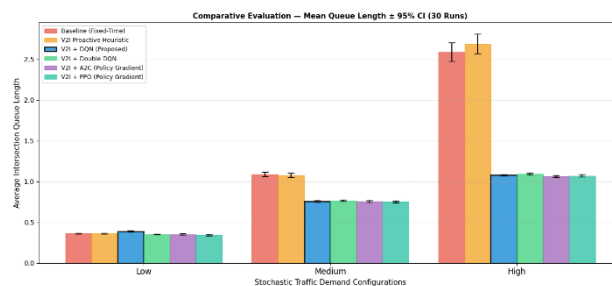


Figure 6. Average Queue Length Comparison Across Low, Medium, and High Traffic Demand Scenarios.

Figure 6 illustrates this overall trend, showing that while all methods achieve similar average queue lengths under low traffic demand, the proposed V2I+DQN controller increasingly outperforms the conventional strategies as traffic saturation grows. This behavior confirms that the benefits of reinforcement learning become more significant when continuous adaptation to changing traffic conditions is required. Across the 30 independent stochastic runs, the proposed controller also demonstrates superior robustness. Compared with the Fixed-Time controller, the mean queue length is reduced by 29.8% under medium demand and 58.3% under high demand, while the standard deviation decreases by up to 92%, indicating improvements in both traffic efficiency and operational stability. To further assess the consistency of these results, Figure 7 presents the distribution of the mean queue lengths over the 30 stochastic runs for each traffic demand scenario. Under low traffic demand, all controllers exhibit compact distributions, reflecting the limited influence of the control strategy in uncongested conditions. In contrast, under medium and high traffic demand, the V2I+DQN controller exhibits both lower median queue lengths and a noticeably smaller interquartile range than the conventional controllers. The limited dispersion of the results confirms that the learned policy remains stable across different stochastic traffic realizations, further demonstrating the robustness of the proposed framework.

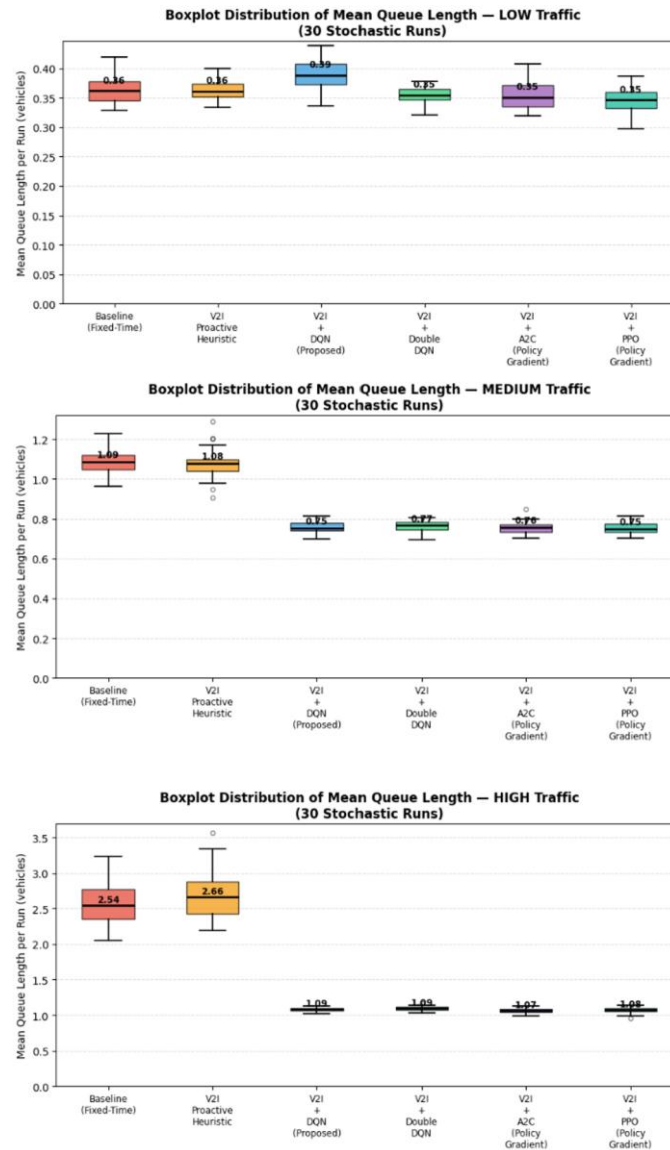


Figure 7. Boxplot distributions of mean queue length over 30 stochastic runs under low-, medium-, and high-traffic demand, respectively. Boxes represent the interquartile range (IQR), horizontal lines indicate the median, and whiskers extend to $1.5 \times \text{IQR}$.

Under low traffic demand, all controllers exhibit compact and largely overlapping distributions, with median queue lengths ranging from 0.35 to 0.39 vehicles and no observable outliers. This confirms that, under uncongested conditions, the choice of the control strategy has only a limited influence on traffic performance, as congestion rarely develops.

As traffic demand increases, the differences between the evaluated methods become progressively more pronounced. Under medium traffic demand, the reinforcement learning-based controllers achieve substantially lower median queue lengths (approximately 0.75–0.77 vehicles) than the Fixed-Time and V2I Proactive Heuristic controllers (approximately 1.08–1.09 vehicles). In addition, their narrower interquartile ranges indicate greater consistency across stochastic realizations, whereas the heuristic controller exhibits several outliers, reflecting occasional sensitivity to traffic fluctuations and a less stable adaptive behavior.

Under high traffic demand, the superiority of the proposed V2I+DQN framework becomes most evident. While the Fixed-Time and V2I Proactive Heuristic controllers display wide interquartile ranges and pronounced variability, the proposed controller maintains a median queue length of approximately 1.09 vehicles with a remarkably compact distribution. This limited dispersion across the 30 independent stochastic runs demonstrates that the observed performance improvements are not the result of a few favorable simulations but rather reflect the consistent robustness of the learned control policy under saturated traffic conditions.

5.6 Computational Cost Analysis

To assess the practical feasibility of the proposed V2I-enabled traffic signal control framework, the computational cost of the evaluated reinforcement learning algorithms was analyzed by measuring both the offline training time and the online inference time. The results obtained under the three traffic demand scenarios are summarized in Table 6.

Table 6. Computational cost of the evaluated reinforcement learning algorithms under different traffic demand scenarios.

Scenario	DQN Training (s)	DQN Inference (ms/decision)	Double DQN Training (s)	Double DQN Inference (ms/decision)	A2C Training (s)	A2C Inference (ms/decision)	PPO Training (s)	PPO Inference (ms/decision)
Low	382.63	0.119	462.72	0.106	145.34	0.113	132.50	0.133
Medium	381.03	0.097	458.35	0.097	142.93	0.132	130.82	0.119
High	387.45	0.111	499.74	0.140	168.87	0.127	152.34	0.119

As shown in Table 6, the offline training time varies across the evaluated algorithms. Double DQN requires the longest training time, ranging from 458.35 s to 499.74 s, due to the additional computations associated with maintaining separate online and target networks. In contrast, A2C and PPO exhibit considerably shorter training times, while DQN requires an intermediate training time of approximately 381–387 s. Since model training is performed offline, these differences do not affect the operational performance of the proposed traffic signal controller.

More importantly, all evaluated algorithms exhibit extremely low inference latency, with an average decision-making time ranging from 0.097 ms to 0.140 ms per control action across all traffic demand scenarios. Such latency is negligible compared with the control intervals typically employed in adaptive traffic signal systems, demonstrating that the proposed framework satisfies the computational requirements of real-time operation.

Furthermore, considering that no statistically significant performance differences were observed among DQN, Double DQN, PPO, and A2C, the slight differences in computational cost do not translate into meaningful improvements in traffic control performance. These results reinforce the conclusion that the effectiveness of the proposed approach primarily stems from the V2I-enhanced state representation rather than from the choice of a specific reinforcement learning algorithm.

6 Discussion

The experimental findings presented in Section 5 provide further insight into the contribution of the proposed V2I-enabled reinforcement learning framework and allow its positioning with respect to the current state of the art in adaptive traffic signal control.

Positioning relative to existing approaches: As summarized in Table 1 (Related Work), existing intelligent traffic management studies mainly investigate reinforcement learning-based traffic signal control, predictive optimization methods, and V2I-enabled traffic management frameworks separately [6,14,15,21,22]. Reinforcement learning approaches generally rely on traffic states obtained from local sensors or simulation environments, whereas V2I-oriented studies commonly employ explicit predictive models or heuristic optimization strategies. In contrast, the proposed framework explicitly integrates V2I-derived observations into the reinforcement learning process by constructing the agent state from queue lengths, signal phase, and current green-time duration. Because these variables are continuously updated through V2I communication, the agent benefits from a richer and more timely representation of intersection conditions, enabling adaptive decisions before congestion becomes critical. This distinction explains the observed performance gap between the V2I Proactive Heuristic controller and the proposed V2I+DQN framework: although both methods exploit the same V2I-derived traffic information, only the learning-based controller is able to continuously improve its control policy through interaction with the environment, leading to queue-length reductions ranging from 29.8% to 58.3% under medium and high traffic demand.

Explaining the observed performance: The experimental results further indicate that the effectiveness of reinforcement learning strongly depends on traffic demand. Under low-demand conditions, queue lengths remain close to zero for most of the simulation, producing sparse reward signals with only minor variations between successive states. Consequently, the DQN agent receives limited informative feedback for distinguishing between alternative actions, reducing the practical benefit of adaptive learning. Since the intersection operates well below its saturation level, even conventional Fixed-Time and heuristic controllers maintain satisfactory traffic flow.

As traffic demand increases, queue dynamics become considerably more complex and informative. Larger and more frequent queue fluctuations generate richer reward signals, allowing the DQN agent to progressively learn more effective control policies. Under these conditions, adaptive optimization becomes increasingly beneficial, explaining the significant improvements observed under medium and high traffic demand. These findings suggest that reinforcement learning is particularly advantageous when congestion management requires continuous adaptation to rapidly evolving traffic conditions.

Robustness across reinforcement learning variants: Another important contribution of this study is the observation that the performance improvements appear to originate primarily from the V2I-enriched traffic-state representation rather than from the specific reinforcement learning algorithm employed. After Bonferroni correction, no statistically significant differences were observed between DQN and the evaluated reinforcement learning variants (Double DQN, A2C, and PPO) under either medium- or high-demand conditions ($p > 0.05$). This result indicates that the proposed framework remains robust across both value-based and policy-gradient methods. Consequently, the experimental evidence suggests that enhanced traffic-state observability provided by V2I communication constitutes the dominant factor contributing to the overall performance improvements [12–14].

Practical implications: From an operational perspective, these findings suggest that transportation agencies seeking to deploy adaptive traffic signal control may not necessarily need to adopt a specific reinforcement learning algorithm. Instead, the choice between value-based and policy-gradient approaches may be guided by implementation constraints such as computational latency, hardware resources, interpretability, or deployment requirements, provided that reliable V2I-derived traffic information is available. This flexibility facilitates

practical deployment while preserving the performance benefits associated with enhanced traffic-state observability.

Nevertheless, the idealized assumptions adopted in this study—namely full V2I penetration, zero communication latency, and lossless packet transmission—represent an upper-bound performance scenario. Under realistic deployment conditions, communication delays, packet losses, and incomplete market penetration may reduce the observed performance gains, although improvements over conventional non-adaptive traffic signal control are still expected [16–20].

Soft-computing perspective. From the perspective of computational intelligence, the proposed framework can be regarded as a soft-computing approach because it relies on reinforcement learning to learn an adaptive traffic signal control policy directly from uncertain and stochastic traffic observations rather than from explicit analytical traffic models. Unlike conventional optimization techniques based on deterministic formulations, the proposed controller progressively improves its decision policy through continuous interaction with the environment, which is a fundamental characteristic of computational-intelligence methods. Although the present study does not explicitly incorporate fuzzy logic, evolutionary optimization, or swarm-intelligence techniques, it complements these paradigms by demonstrating that enriching the state representation through V2I communication substantially improves the effectiveness and robustness of learning-based adaptive traffic signal control. Consequently, the contribution of this work lies not only in the proposed V2I-enhanced state representation but also in illustrating how connected-vehicle information can strengthen soft-computing approaches for intelligent transportation systems.

Limitations and future work: Several limitations should be acknowledged. First, the proposed framework was evaluated only for an isolated two-phase intersection, without considering network-level interactions such as spillback propagation, corridor coordination, or multi-intersection synchronization [6,22]. Second, vehicle arrivals were modeled using independent Poisson processes to provide a controlled stochastic environment suitable for evaluating the learning framework. Although this assumption is widely adopted in traffic signal control research, it does not fully represent the complexity of real urban traffic, including turning movements, heterogeneous driver behavior, pedestrian crossings, and multimodal interactions. Finally, the V2I communication model assumes ideal connectivity and therefore neglects realistic wireless communication impairments such as latency, packet loss, and partial penetration of connected vehicles.

These simplifying assumptions were intentionally adopted to isolate and evaluate the contribution of the proposed V2I-enabled reinforcement learning framework under controlled experimental conditions. Future work will therefore focus on extending the proposed approach to multi-agent reinforcement learning for coordinated multi-intersection traffic control, while incorporating realistic communication constraints and validating the framework within high-fidelity microscopic traffic simulators such as SUMO.

7 Conclusion

This study proposed a Vehicle-to-Infrastructure (V2I)-enabled Deep Q-Network (DQN) framework for adaptive traffic signal control at an isolated intersection. By integrating real-time traffic information provided through V2I communication with reinforcement learning, the proposed controller dynamically adjusts green signal durations according to the current traffic state. Experimental evaluation under three stochastic traffic demand scenarios demonstrated that the proposed framework significantly outperforms conventional Fixed-Time and V2I Proactive Heuristic controllers, achieving mean queue-length

reductions of 29.8% and 58.3% under medium- and high-demand conditions, respectively ($p < 0.001$). In addition to improving traffic efficiency, the proposed controller substantially reduced performance variability, with standard deviations decreasing by up to 92% under congested traffic conditions.

An important finding of this study is that the observed performance improvements appear to be driven primarily by the V2I-enriched traffic-state representation rather than by the particular reinforcement learning algorithm employed. The comparable performance achieved by DQN, Double DQN, A2C, and PPO suggests that providing accurate and timely traffic-state information through V2I communication may have a greater influence on adaptive traffic signal control than the selection of a specific reinforcement learning architecture, which is consistent with recent studies highlighting the importance of connected-vehicle information for intelligent traffic management [13,20,21].

Several limitations should nevertheless be acknowledged. The proposed framework was evaluated for a single isolated intersection under idealized communication assumptions, including full V2I penetration, zero communication latency, and lossless data transmission. Moreover, the traffic model does not explicitly represent network-level interactions, heterogeneous traffic participants, or realistic wireless communication impairments. Consequently, the reported results should be interpreted as a controlled proof of concept rather than as a direct representation of real-world deployment.

Future research will focus on extending the proposed framework toward Multi-Agent Reinforcement Learning (MARL) for coordinated multi-intersection traffic signal control, while incorporating realistic V2I communication constraints, heterogeneous traffic conditions, and validation within high-fidelity microscopic traffic simulators. These research directions are consistent with current trends in intelligent transportation systems and connected vehicular networks [6,15,16,21,22]. Overall, this work provides further evidence that enhancing traffic-state observability through V2I communication constitutes a robust foundation for next-generation adaptive traffic signal control and represents a promising direction for future intelligent transportation systems.

DECLARATIONS

Funding. “The authors received no external funding for this research.”

Conflicts of interest. The authors declare that they have no competing interests.

Ethics approval. This study did not involve human participants, animals, or human biological material. Therefore, ethics approval was not required.

Informed consent. Not applicable.

Data availability. The datasets used in this study were generated through simulation. All relevant simulation settings, experimental configurations, and reinforcement learning hyperparameters necessary to reproduce the reported results are provided in the manuscript.

Author contributions. Conceptualization: L. Madi; Methodology: L. Madi; Software: L. Madi; Validation: L. Madi ; Formal analysis: L. Madi ; Investigation: L. Madi; Writing – original draft: L. Madi; Writing – review & editing: L. Madi; Supervision: A. Y. Benabdallah, and N. Guerssi.

Use of artificial intelligence. Artificial intelligence tools were used solely to improve the language, grammar, readability, and formatting of the manuscript. All scientific ideas, methodology, experimental design, implementation, data analysis, interpretation of the results, and conclusions were developed and verified by the authors.

Acknowledgements. The authors would like to thank the anonymous reviewers for their valuable comments and suggestions, which helped improve the quality of this manuscript.

REFERENCES

- [1] Li, W.; Liang, P.; Sun, B.; Sun, Y.; Huang, Y. Reinforcement Learning-Based Particle Swarm Optimization with Neighborhood Differential Mutation Strategy. *Swarm Evol. Comput.* 2023, 78, 101274. doi:10.1016/j.swevo.2023.101274.
- [2] Lin, S.; Wang, J.; Huang, B.; Kong, X.; Yang, H. Bio Particle Swarm Optimization and Reinforcement Learning Algorithm for Path Planning of Automated Guided Vehicles in Dynamic Industrial Environments. *Sci. Rep.* 2025, 15, 463. doi:10.1038/s41598-024-84821-2.

- [3] Ozsoydan, F.B. Reinforcement Learning Enhanced Swarm Intelligence and Trajectory-Based Algorithms for Parallel Machine Scheduling Problems. *Comput. Ind. Eng.* 2025, 202, 110948. doi:10.1016/j.cie.2025.110948.
- [4] Liao, X.-C.; Mei, Y.; Zhang, M. Learning Traffic Signal Control via Genetic Programming. *arXiv* 2024, arXiv:2403.17328. doi:10.48550/arXiv.2403.17328.
- [5] Patil, M.; Tambolkar, P.; Midlam-Mohler, S. Optimizing Traffic Routes with Enhanced Double Q-Learning. *IET Intell. Transp. Syst.* 2025, 19, e70002. doi:10.1049/itr2.70002.
- [6] Skoropad, V.N.; Deđanski, S.; Pantović, V.; Injac, Z.; Vujičić, S.; Jovanović-Milenković, M.; Jevtić, B.; Lukić-Vujadinović, V.; Vidojević, D.; Bodolo, I. Dynamic Traffic Flow Optimization Using Reinforcement Learning and Predictive Analytics: A Sustainable Approach to Improving Urban Mobility in the City of Belgrade. *Sustainability* 2025, 17, 3383. doi:10.3390/su17083383.
- [7] Anandhalli, M.; Baligar, V.P.; Baligar, P.; DeepSir, P.; Iti, M. Vehicle Detection and Tracking for Traffic Management. *IAES Int. J. Artif. Intell.* 2021, 10, 66–73. doi:10.11591/ijai.v10.i1.pp66-73.
- [8] Mystakidis, A.; Koukaras, P.; Tjortjis, C. Advances in Traffic Congestion Prediction: An Overview of Emerging Techniques and Methods. *Smart Cities* 2025, 8, 25. doi:10.3390/smartcities8010025.
- [9] Li, M.; Fan, D.; Dong, Y.; Li, D. Satellite Video Moving Vehicle Detection and Tracking Based on Spatiotemporal Characteristics. *Sensors* 2023, 23, 5771. doi:10.3390/s23125771.
- [10] Zhang, H.; Yuan, H.; Chen, Y.; Yu, W.; Wang, C.; Wang, J.; Gao, Y. Traffic Light Optimization Based on Modified Webster Function. *J. Adv. Transp.* 2021, 2021, 3328202. doi:10.1155/2021/3328202.
- [11] Xiang, H.; Zheng, Z.; Xia, X.; Xu, R.; Gao, L.; Zhou, Z.; Han, X.; Ji, X.; Li, M.; Meng, Z.; et al. V2X-Real: A Large-Scale Dataset for Vehicle-to-Everything Cooperative Perception. In *Computer Vision – ECCV 2024*; Springer: Cham, Switzerland, 2024; pp. 455–470. doi:10.1007/978-3-031-72943-0_26.
- [12] Xu, R.; Xia, X.; Li, J.; Li, H.; Zhang, S.; Tu, Z.; Meng, Z.; Xiang, H.; Dong, X.; Song, R.; et al. V2V4Real: A Real-World Large-Scale Dataset for Vehicle-to-Vehicle Cooperative Perception. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, BC, Canada, 17–24 June 2023; pp. 13712–13722. doi:10.1109/CVPR52729.2023.01318.
- [13] Qiao, Y.; Meng, T.; Qin, H.; Hu, Z.; Zhong, Z. Vehicle-to-Infrastructure-Based Traffic Signal Optimization for Isolated Intersection. *Sustainability* 2023, 15, 6631. doi:10.3390/su15086631.
- [14] Huang, W.; Hu, Y.; Zhang, X. Enhancing Model-Based Anticipatory Traffic Signal Control with Metamodeling and Adaptive Optimization. *Mathematics* 2022, 10, 2640. doi:10.3390/math10152640.
- [15] Tadesse, E.M.; Demilew, S.A.; Mengesha, A.Z.; Geto, A.D.; Endris, N. Load-Aware and Priority Adaptive Traffic Congestion Control Method in Vehicular Ad Hoc Network. *Int. J. Wirel. Commun. Mob. Comput.* 2024, 11, 39–51. doi:10.11648/j.wcmc.20241102.13.
- [16] Touluni, H.; Miyara, M.; Filali, Y.; Tékouabou, S.C.K. Preventing Urban Traffic Congestion Using VANET Technology in Urban Area. *E3S Web Conf.* 2023, 418, 02005. doi:10.1051/e3sconf/202341802005.
- [17] Dongre, M.M.; Bawane, N.G. Effective Road Model for Congestion Control in VANETs. *Int. J. Wirel. Mob. Netw.* 2016, 8, 11–21.
- [18] Zhang, E.-Z.; Zhang, X. Road Traffic Congestion Detecting by VANETs. In *Proceedings of the 2nd International Conference on Electrical and Electronic Engineering (EEE 2019)*; Atlantis Press, 2019; pp. 242–248. doi:10.2991/eee-19.2019.39.

- [19] Taherkhani, N.; Pierre, S. Improving Dynamic and Distributed Congestion Control in Vehicular Ad Hoc Networks. *Ad Hoc Netw.* 2015, 33, 112–125.
- [20] Ault, J.; Sharon, G. Reinforcement Learning Benchmarks for Traffic Signal Control. In *Proceedings of the Thirty-Fifth Conference on Neural Information Processing Systems (NeurIPS 2021), Datasets and Benchmarks Track*, 2021.
- [21] Mahato, S. Smart Traffic Signals: Comparing MARL and Fixed-Time Strategies. *arXiv* 2025, arXiv:2505.14544. doi:10.48550/arXiv.2505.14544.
- [22] Van Hasselt, H.; Guez, A.; Silver, D. Deep Reinforcement Learning with Double Q-Learning. *Proc. AAAI Conf. Artif. Intell.* 2016, 30, 2094–2100. doi.org/10.1609/aaai.v30i1.10295.
- [23] Mnih, V.; Badia, A.P.; Mirza, M.; Graves, A.; Harley, T.; Lillicrap, T.P.; Silver, D.; Kavukcuoglu, K. Asynchronous Methods for Deep Reinforcement Learning. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*; New York, NY, USA, 19–24 June 2016; pp. 1928–1937.
- [24] Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal Policy Optimization Algorithms. *arXiv* 2017, arXiv:1707.06347.
- [25] Chu, T.; Wang, J.; Codeca, L.; Li, Z. Multi-Agent Deep Reinforcement Learning for Large-Scale Traffic Signal Control. *IEEE Trans. Intell. Transp. Syst.* 2020, 21, 1086–1095. doi:10.1109/TITS.2019.2901791.