frontiers in Future Transportation

TYPE Original Research
PUBLISHED 10 July 2026
DOI 10.3389/ffutr.2026.1765037

Check for updates

OPEN ACCESS

EDITED BY
Meng Xu,
Beijing Jiaotong University, China

REVIEWED BY
Bin Zhou,
The University of Hong Kong, Hong Kong SAR, China
Zhongnan Zhao,
Harbin University of Science and Technology, China

*CORRESPONDENCE
Sachin Salunkhe,
sachinsalunkhe@gazi.edu.tr
Sachin Shinde,
sachin.shinde@dmce.ac.in

RECEIVED 16 December 2025
REVISED 30 May 2026
ACCEPTED 15 June 2026
PUBLISHED 10 July 2026

CITATION
Ramesh Kumar B, Salunkhe S, Shinde S and Cepova L (2026) Dynamic traffic signal scheduling system based on adaptive quad agent Double Deep Q -network algorithm.
Front. Future Transp. 7:1765037.
doi: 10.3389/ffutr.2026.1765037

COPYRIGHT
© 2026 Ramesh Kumar, Salunkhe, Shinde and Cepova. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Dynamic traffic signal scheduling system based on adaptive quad agent Double Deep Q -network algorithm

Bharathi Ramesh Kumar[1], Sachin Salunkhe[2,3]*, Sachin Shinde[4]* and Lenka Cepova[5]

[1]Department of Mathematics, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, Tamilnadu, India, [2]Department of Biosciences, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, India, [3]Department of Mechanical Engineering, Faculty of Engineering, Gazi University, Ankara, Türkiye, [4]Department of Mechanical Engineering, Datta Meghe College of Engineering, Navi Mumbai, Maharashtra, India, [5]Department of Machining, Assembly and Engineering Metrology, Faculty of Mechanical Engineering, VSB-Technical University of Ostrava, Ostrava, Czechia

Introduction: Real-time estimation of vehicle queue lengths at signalized intersections remains a significant challenge, particularly when conventional input–output traffic models fail to capture queues extending beyond detector coverage. Although Deep Q-Networks (DQNs) have demonstrated considerable potential for dynamic traffic signal control, existing approaches often suffer from large state spaces, unstable reward signals, and inefficient utilization of high-quality traffic data. To address these limitations, this study proposes an Adaptive Quad-Agent Double Deep Q-Network (AQDDQN) framework for intelligent traffic signal optimization.

Methods: The proposed AQDDQN framework improves learning stability and Q-value estimation accuracy through a multi-agent reinforcement learning strategy. The model analyzes the relationship between vehicle queue length and reward values to optimize signal control decisions. Historical traffic data are utilized to establish preconditions, time-based prediction errors are computed, and optimal signal phases are selected based on minimum loss across multiple preconditions. The experimental evaluation includes agent-wise behavioral analysis and comparative assessments against Double Deep Q-Network (DDQN), Fixed Point Techniques, and Improved DDQN methods. Performance is evaluated using metrics such as reward values, queue lengths, predicted overflow delays, and queue length–reward relationships.

Results: The proposed adaptive framework demonstrates superior performance compared with existing approaches by improving traffic signal control accuracy, reducing prediction errors, and enhancing overall traffic throughput. Simulation results indicate that the AQDDQN model effectively supports dynamic signal phase adaptation, minimizes congestion, and provides more accurate queue length estimations under complex traffic conditions.

Discussion: The findings confirm the effectiveness and robustness of the proposed AQDDQN framework for real-time intelligent traffic management. By improving learning stability and adaptive decision-making capabilities, the

model offers a practical solution for optimizing traffic operations at signalized intersections and has strong potential for deployment in future smart transportation systems.

KEYWORDS

adaptive traffic management, Deep Q-network, intelligent traffic signal control, overflow rate, traffic optimization, vehicle waiting time

1 Introduction

Urban traffic congestion has become one of the most significant challenges faced by modern transportation systems. Rapid urbanization, population growth, and the continuous increases in vehicle ownership have significantly increased traffic demand in metropolitan areas. Signalized intersections are critical components of urban networks, where inefficient signal control often leads to extreme vehicle waiting times, increased fuel consumption, and higher levels of environmental pollution. Consequently, the development of efficient and adaptive traffic signal control strategies has become essential in intelligent transportation systems. Transportation traffic signal control strategies depend on predefined signal timing plans/limited detector based information. Although these approaches are simple to implement, repeatedly fail to respond effectively to dynamic traffic conditions. In particular, when vehicle queues extend beyond the detection range of traffic sensors, traditional models struggle to accurately estimate traffic states, resulting in inefficient signal timing decisions and increased congestion. Recent advances in deep reinforcement learning have provided promising solutions for adaptive traffic signal control. Mnih et al. (2015) and Van Hasselt et al. (2016) assist traffic controllers to learn optimal signal policies through interaction with the environment, reducing the rule based optimization. DQN approaches can approximate optimal Q-value functions using deep neural networks and dynamically adjust signal phases based on real-time traffic conditions. However, several challenges remain in practical implementation, including large state action spaces that increase computational complexity, sparse and delayed reward signals that delay effective learning, and Q-value overestimation in conventional DQN methods, which may lead to unstable training and suboptimal decision-making policies.

Many researchers have stated various results to address traffic congestion challenges and improve overall traffic signal performance. Review studies in Vandaele et al. (2000) analyze traffic flow models, and highlight the analytical strength of queuing theory in congestion analysis, performance evaluation, and environmental impact assessment. Similarly, the study in Akçelik and Rouphail (1993) reviews delay estimation at signalized intersections, observes that traditional models, including those based on the Highway Capacity Manual, often underestimate interruptions caused by fluctuating demand. The authors propose a peak adjusted framework that significantly improves capacity estimation accuracy. In large-scale traffic signal timing problems, the primary objective is to minimize overall travel time and vehicle waiting time while enhancing coordination among multiple intersections (Guo et al., 2019). The study presented in Jina and Ma (2017) proposes a genetic algorithm integrated with an agent based framework in which each agent represents an individual traffic signal, enabling dynamic improvement of signal timings. The traffic signal optimization problem using a memetic algorithm was investigated in Sabar et al. (2017) focusing on queue waiting time reduction. Dynamic flow networks and accumulation equilibrium with well-posedness conditions were studied in Como and Nilsson (2021). Researchers in Ahmed et al. (2019) developed a queueing model based on differential inclusions with constant routing matrices that allow outflow rates to exceed queue mass. To maintain smooth traffic flow during peak congestion periods, a continuous time mass flow conservation model was introduced in Uppaluru et al. (2022) to distribute green signal time effectively. A virtual arterial corridor model proposed in Li and Ban (2020) enables real-time vehicle communication, thereby reducing traffic signal stops and minimizing vehicle waiting time (Zhu et al., 2022). Traffic congestion modification strategies have also been explored through flow stabilization and experimental traffic analysis. The work presented in Xu et al. (2021) reduced traffic flow fluctuations, while Qu et al. (2021) conducted experimental investigations on complex traffic flow networks. A static equation for order traffic flow system was introduced in Sau et al. (2019) to optimize flow density equilibrium, and dynamically adaptive signal timing strategies (Szoke et al., 2023) proposed to minimize traffic distributions. A priority based algorithm aimed at reducing fuel consumption across traffic networks was introduced in Kim et al. (2023). Arterial intersection complexity and coordination optimization were addressed in Li et al. (2023), while the AM-BAND–PBAND coordination model proposed in Li et al. (2022) maximizes green wave bandwidth and improves travel time, delay, stops, emissions, and fuel consumption, validated through VISSIM simulations. The impact of different traffic signal patterns on delay and safety, considering both pedestrians and vehicles, was analyzed in Montazeri et al. (2022), which proposed a hybrid adaptive signal pattern based on real-time conditions. Decentralized and scalable coordination frameworks were introduced in Augusto Vieira et al. (2024), followed by further optimization methods for complex intersection configurations in Tomar et al. (2022). A comprehensive evaluation of traffic signal characteristics and their impacts on safety and operational efficiency was presented in Leitner et al. (2022). Fuzzy logic based traffic signal control was developed in Taskin and Gumustas (1997), demonstrating improved handling of uncertainty and variability compared with traditional control strategies. Subsequently, deep reinforcement learning integration achieved an 18.46% decrease in queue length, as reported in Meepokgit and Wisayataksin (2024). Queue length estimation using connected vehicles as mobile sensors was proposed in Gao et al. (2019), improving real-time sensing accuracy for adaptive signal control. Flexible routing and coordination for connected and automated vehicles at lane allocation free intersections were introduced in Hao et al. (2023), significantly improving intersection capacity and reducing delays. A dynamic network level signal and speed control framework leveraging connected vehicle data was proposed in Yuan and Zeng (2024), reducing delays, stops, fuel consumption, and emissions. A heterogeneous traffic flow model incorporating human driven vehicles, adaptive travel control was developed in Guan et al. (2023), demonstrating improved traffic stability with higher connected vehicle capacity. Adaptive decentralized traffic signal control using MATSim simulations was investigated in

Thunig et al. (2019), while GPS based delay modeling for signalized intersections was presented in Alkaissi and Hussain (2020), showing performance comparable to actuated systems under varying traffic conditions. Field implementations of arterial control strategies were reviewed in Gomes (2010), identifying RHODES as one of the most effective approaches. An improved greedy algorithm based traffic signal scheduling system was proposed in Sun et al. (2024), achieving reduced delay and improved throughput. Automotive communication technologies and future research directions were reviewed in Abdur Rahima et al. (2020), while real-time optimization methods addressing temporal traffic constraints were discussed in Eom and Kim (2020). Adaptive signal control algorithms aimed at improving timing efficiency and reducing queue lengths were presented in Shelby (2004). Decentralized Nash dealing control strategies demonstrated superior congestion reduction compared with centralized methods in Maha et al. (2021).

Reinforcement learning based adaptive signal controllers integrated with vehicle speed guidance were proposed in Maadi et al. (2022), significantly reducing queue length and stop delays. Vehicle to infrastructure and vehicle to vehicle communication based adaptive signal control was investigated in Recker (2010). A convolutional neural network based multi agent deep Q learning framework for dynamic signal control was introduced in Ramesh Kumar et al. (2024), achieving significant reductions in queue length and waiting time. Signal distribution control algorithms dynamically adjusting signal timing based on vehicle arrivals were analyzed in Solaiappan et al. (2023). Conventional behavior analysis across multiple studies was reviewed in Feng et al. (2021), highlighting limitations related to data quality and ethical considerations.

Comprehensive surveys of traffic control system technique were presented in Majeed and Fahad (2022), comparing traditional, sensor based connected vehicle and AI-enhanced approaches. A deep reinforcement learning traffic signal control strategy integrating temporal traffic pattern learning within an actor critic framework was proposed in Ma et al. (2021), demonstrating superior delay reduction performance. Cooperative multi-agent deep actor-critic models were introduced in Song et al. (2024), while incentive based communication frameworks improving agent coordination were developed (Zhou et al., 2024). Finally, model based reinforcement learning approaches incorporating internal traffic models for optimized control planning were proposed in Shi et al. (2023).

In addition, many existing reinforcement learning frameworks give the intersection control problem as a single agent optimization assignment, where a single learning agent simultaneously determines all signal phase decisions. While this approach simplifies the model design, it often leads to large action spaces and reduced learning stability, particularly in complex multi-phase intersections. Furthermore, most current DRL based approaches rely primarily on data driven learning without incorporating analytical traffic flow models that explicitly describe queue dynamics, delay accumulation, and overflow conditions. As a result, the learned policies efficiency interpretability and fail to accurately represent real world traffic dynamics.

Therefore several research gaps remain in the development of efficient reinforcement learning based traffic signal control system. First, there is limited integration of analytical traffic state models with reinforcement learning frameworks, which restricts the ability of learning algorithms to capture detailed traffic queue and delay dynamics. Second existing approaches often lack structured phase level decomposition, leading to large action spaces and unstable learning performance for multi-phase intersections. Third, current multi-agent reinforcement learning strategies typically rely on conventional max-Q value action selection mechanisms, which may introduce overestimation preconception and reduce learning stability in stochastic traffic environments.

To address these limitations, this study proposes an Adaptive Quad Agent Double Deep Q Network framework for intelligent traffic signal control. The proposed approach decomposes the intersection control problem into four coordinated phase specific agents, allowing each agent to focus on limited traffic states while collectively optimizing global intersection performance. This structured decomposition significantly reduces the action space complexity and improves the stability of the learning process.

In addition, the proposed framework a Dynamic signal Distribution and traffic control (DSTCM) model logically evaluate the queue lengths, delay accumulation, and overflow dynamics. These logically derived traffic metrics are embedded directly into the reinforcement learning state representation and reward formulation, improving both learning interpretability and traffic state representation accuracy. Furthermore, a loss based cooperative phase selection is introduced to coordinate the decision making process among the multiple agents. Unlike conventional approaches that depends on maximizing Q-values, the proposed framework, determines signal phase decisions by minimizing aggregated temporal difference loss across the agents. This process reduce Q-value overestimation and enhances the robustness and stability of the leaning process.

To evaluate the effectiveness of the proposed approach, simulation experiments were conducted low, moderate and heavy traffic demand scenarios. The proposed AQDDQN frameworks was compared with existing methods. The experimental results demonstrate that the proposed framework significantly reduces vehicle waiting time, improves queue management and enhances overall traffic throughput under dynamic traffic conditions.

Though existing MARL based traffic signal control methods employ phase-level decomposition and decentralized learning strategies, most approaches mainly depend on reward maximization and communication based coordination process. In addition many existing frameworks are purely data-driven and do not explicitly incorporate analytical traffic state modeling into the reinforcement learning process. In contrast, the proposed AQDDQN framework integrates three key components: a quad agent DDQN architecture for scalable phase-wise learning a loss-based coordination process that minimizes temporal-difference loss instead of conventional max-Q selection, and the DSTCM analytical traffic model for explicit queue, delay, and overflow estimation. Therefore, the novelty of the proposed framework lies in the combined within a DDQN-based adaptive traffic signal control system.

The key contribution of this work as follows:

- A novel Adaptive Quad Agent Double Deep Q network framework is proposed, where each signal phase is modeled as an independent learning agent reducing state space complexity and improving scalability.
- A loss based coordination mechanism is introduced, where signal phase selection is determined by minimizing temporal difference loss, unlike conventional reward based or communication driven coordination.

- Integrating Dynamic Signal Distribution and Traffic Signal Control Model (DSTCM) with reinforcement learning is developed, enabling explicit modeling of queue length, delay and overflow.
- The proposed approach improves learning stability and reduce Q value overestimation through quad agent decomposition combined with DDQN architecture.

Symbol	Description
k	Signal cycle index
i	Signal phase index $i \in \{1, 2, 3, 4\}$
$V(k, i)$	Queue length waiting time associated with phase 'i' at time step 'k'
$I(k, i)$	Total waiting time interrupted queue length
$V_i(k)$	Waiting time of phase 'i' at cycle 'k'
$Q_i(k)$	Number of vehicle queued in phase 'i' at time step 'k'
$D_i(k)$	Average delay of phase 'i' at cycle 'k'
λ_i	Vehicle arrival rate for phase 'i' assumed to follow a Poisson process
μ_i	Service rate of vehicle in phase 'i'
$G_i(k)$	Green time allocated to phase 'i' during cycle 'k'
$R_i(k)$	Red time duration for phase 'i' during cycle 'k'
$C(k)$	Total signal cycle length at time step 'k'
$\tau(t)$	Vehicle arrival process at time 't'
ν	Time partition parameter within signal cycle
ρ	Overflow rate parameter
$\rho_{i(k)}$	Overflow delay associated with phase 'i' at cycle 'k'
$\rho_{i(s)}$	Accumulated overflow waiting time for phase 'i'
P_i	Service probability of phase 'i'
s_i^{ρ}	Total vehicle delay measure
s_k	Traffic state observed by the agent(s) at time step 'k'
a_k	Control action selected at time step 'k'
r_k	Immediate reward obtained after executing action a_k
$Q(s, a)$	Action value function estimating expected cumulative reward for state action pair (s,a)
γ	Discount factor determining the importance of future rewards
α	Learning rate of the Q – network
e_i	Temporal difference loss computed by phase agent 'i'
$e_i^{'}$	Corresponding loss estimated using the target network for phase agent 'i'
θ	Parameters of Q –network
θ^{-}	Parameters of target Q – network
A	Set of admissible signal control action
DSTCM	Dynamic signal distribution and traffic control model
AQDDQN	Adaptive quad agent Double Deep Q Network

Compared with existing MARL based traffic signal control methods, the proposed AQDDQN framework differs not only through phase-wise decomposition, but primarily through the integration of analytical DSTCM based traffic modeling with a loss driven integrative DDQN coordination process. Distinct conventional MARL approaches that depend on reward-maximization/communication based coordination, the proposed framework performs phase selection by minimizing accumulated temporal-difference loss, thereby improving learning stability, interpretability, and robustness under stochastic traffic conditions.

The remainder of this paper is organized as follows: Section 2 reviews related work in reinforcement learning based traffic signal control. Section 3 presents the proposed AQDDQN framework and analytical traffic modeling approach Parameters in Table 1. Section 4 describes the experimental setup and simulation environment. Section 5 discusses the numerical results and performance analysis. Finally, Section 6 concludes the paper and outlines future research directions.

List of symbols and representation as follows:

2 Studies on the time delay model

Delay is an essential metric at signalized intersections; it is necessary for measuring the traffic parameters. Conversely, the dynamic signal model coordinates the above concerns and analyzes the system performance. The following parameters are important for the signal delay analysis, such as Notations:

d = mean vehicle delay
c = signal cycle time
g = available green duration
x = demand-capacity ratio
q = arrival frequency (veh/sec)
$Q(o)$ = queue backlog at the beginning of the signal cycle
A = arrivals within the signal phase
C = maximum outflow during green interval
ΔC = provided the reserve capacity is zero in the cycle, else

Recent advancements in deep reinforcement learning (DRL) for traffic signal control have focused on improving coordination and scalability in multi-agent environments. Recent studies include spatio-temporal reflection based models, graph based reinforcement learning approaches and game theoretic matching strategies. Similarly, Double Deep Q network models enhances Q-value estimation, and Nash equilibrium-based methods improve cooperation among agents. However, most existing approaches depend on communication mechanisms, with limited focus on integrating analytical traffic flow models/loss based coordination strategies. The proposed AQDDQN framework within recent state of the art developments, additional MARL based traffic signal control approaches from 2023 to 2024, including heterogeneous MARL coordination models, graph attention reinforcement learning architectures and Nash-equilibrium based multi-agent methods have been comparatively analyzed in Table 2.

The primary novelty of the proposed AQDDQN framework does not occur only from the use of phase-level multi-agent

decomposition, since related decomposition strategies have been investigated in earlier MARL studies. Instead, the proposed contribution lies in the integrated design that combines DSTCM-based analytical traffic modeling, loss-based cooperative coordination, and DDQN-based overestimation reduction within a unified traffic signal control framework. This integrated formulation enables improved learning stability, reduced action-space complexity, and enhanced interpretability compared with existing purely data driven MARL approaches.

3 Traffic management system

The proposed traffic signal management framework aims to complete the balance between operational delays and road safety. Traditional traffic management schemes typically include: (i) parking restrictions, (ii) route restrictions, and (iii) right-of-way restrictions. Signal timing strategies are implemented through either traffic-responsive systems or interconnected control systems that require reinforced or wireless communication and are managed within centralized controller databases. According to the Federal Highway Administration (FHWA), the traffic management structure operates at three hierarchical levels. At the local level, intersection controllers collect detector data, adjust traffic parameters, and modify signal phases in real-time.

3.1 Reinforcement learning formulation

The traffic signal control problem is formulated as a reinforcement learning framework defined by a state space, action space and reward function.

State Space (S)

The state Space is defined as: $s_k = \{Q_i(k), V(k,i), D_i(k)/i = 1,2,3,4\}$. This state representation captures both congestion level and temporal dynamics of signal control.

Action Space (A):

The action $a_i \in A$ corresponds to the adaptive adjustment of green time allocation for the four signal phases $a_i = \{\Delta g_1, \Delta g_2, \Delta g_3, \Delta g_4\}$.

Reward Function (R)

TABLE 1 Comparison of various traffic light systems.

(1956) Beckman	Traffic flow signalized at a steady state delay model	Total delay function in the static signal period	$d = \frac{(c-g)}{c(1-(q/s)}\left[\frac{Q_0}{q} + \frac{c-g+1}{2}\right]$
(1961) Webster	Integrated analytical and numerical method for traffic light control	Typical vehicle delay	$d = \frac{c(1-g/c)^2}{2[1-(g/c)x]} + \frac{x^2}{2q(1-x)} - (0.65)\left(\frac{c}{q^2}\right)^{(1/3)} x^{2+5(g/c)}$
(1968) Miller	Standard formulation for generalized traffic flows during a constant signal period	Determined the average number of overflow vehicles	$d = \frac{(1-g/c)}{2(1-q/s)}\left[c(1-g/c) + \frac{2Q_0}{q}\right]$
(1980) Akcelik	Utilized the coordinate transformation technique	Mean delay under uniform arrival assumption in cycle [0, T]	$d = \left\{\begin{array}{ll} \frac{c(1-g/c)^2}{2(1-q/S} & when\ x<1 \\ (c-g)/s & when\ x \geq 1 \end{array}\right\}$
(2020) Birol Roy	Dynamic PCE model	Optimized the delay time at the signalized intersection area	---
(2021) Abdhul Khadhir	Mathematical delay model for mixed and Disordered traffic streams	Total average control delay	$d = \left\{\frac{c(1-\lambda)^2}{2(1-\lambda X)} + \frac{X\sqrt{2(n+1)}}{2q(1-X)}\right\}$
(2022) Borigarla Barhmaiah	Indo-Highway capacity manual (HCM)	Estimation of traffic queue length and total signal delay	$d = 35.47 + 0.5^*C\frac{[1-(g/c)]^2}{\{1-(\min(X,1.0)^*(g/c)\}} + 11.5R_P^2 - 47.38R_P$
(2022) Samrat mukhopadhyay	Queue model analysis in continues –Time Markov chain (CTMC)	Mean delay estimation for signalized intersections under Disordered traffic flow	$d = c\left(\frac{1-(g/c)^2}{2(1-(g/c)x)} + \bar{\omega} - 0.65\left(\frac{c}{\lambda^2}\right)^{(1/3)} x^{2+5(g/c)}\right.$
(2024) Ali Payıdar Akgüngör	Microsimulation model for traffic networks (TRAF-NETSIM)	An analysis was conducted to compare the simulated and proposed delay model results for 48 different scenarios	$d = c\left(\frac{1-(g/c)^2}{2(1-(g/c)x)}\right.$
(2025) Vinaya S. Mattungal	Time-dependent delay models	Reduce the waiting time and residual value	$d = \left\{\frac{c(1-\lambda)^2}{2(1-\lambda X)} + 900T\left[(x-1) + \sqrt{(x-1)^2 + 12\left(\frac{x-x_0}{cT}\right)}\right]\right\}$

TABLE 2 Methodological differences between existing MARL approaches and the proposed framework.

Method	Learning architecture	Coordination strategy	Traffic modeling	Overestimation handling	Action space complexity	Main limitation
Spatio-temproal MARL (2025)	Hierarchical	Attention-based coordination	Data-driven only	Not explicitly addressed	High	High computational complexity
Heterogeneous MARL (2024)	Multi-agent RL	Reward sharing mechanism	Limited analytical modeling	Partial	High	Requires extensive real world training data
GAT-DDPG (2025)	Graph attention actor-critic	Graph-based communication	Network-level representation	Moderate	Very high	Expensive training and scalability issues
Dueling DDQN (2024)	Single-agent DDQN	Centralized control	Queue-based state only	DDQN based reduction	Moderate	Limited scalability for multi-phase intersections
Nash-based MARL (2024)	Game theoretic MARL	Nash equilibrium optimization	Implicit traffic representation	Not explicitly handled	High	Complex convergence behavior
Existing multi-agent DDQN	Distributed DDQN	Max-Q coordination	Purely data-driven	Partial	High	Q-value instability under stochastic traffic
Proposed AQDDQN	Quad-agent DDQN	Loss-based cooperative coordination	DSTCM analytical queue delay overflow modeling	Explicit DDQN overestimation reduction	Reduced via phase decomposition	Moderate computational complexity with improved interpretability and stability

The immediate reward r_k is designed to minimize congestion and delay while maintaining stable signal operations: $r_k = -\sum_{i=1}^{4}(Q_i(k) + D_i(k) + \gamma)\,\rho_i(t)$.This reward structure directly links queue evolution and delay reduction to the learning objective.

Learning Objective:

The AQDDQN agent learns an optimal policy π^* that maximizes the expected cumulative discounted reward $\pi^* = \arg\max_{\pi} E[\sum_{t=0}^{T}\delta^t r_t]$

The Double Deep Q-learning framework with Quad Agent decomposition is employed to reduce overestimation and improve convergence stability. The above RL formulation is applied in Algorithm 1 and AQDDQN architecture shown in Figure 1.

3.2 Double Deep Q-Network (DDQN)

The Double Deep Q-Network (DDQN) extends the Deep Q-network (DQN) architecture by addressing the overestimation bias present in standard Q-learning. It employs two separate neural networks to decouple action selection and action evaluation during Q-value update process.

The agent network enables the DQN structure to estimate the accurate Q-value and reward value in an urban traffic environment. However, greedy approaches may quickly get optimization while increasing the possibility of overestimation preconception and unsteady convergence under stochastic traffic conditions (Mnih et al., 2015; Sun et al., 2024). The maximum target Q-value was selected by $\overset{max}{a'}\,Q(s_{t+1}, a'; \theta^-)$, and the result can be significant to flow deviation and predict the overestimation action values. Consider two distinct networks for the action selection process and evaluation. This model efficiently minimizes the overestimation error values and improves the stability level.

The evaluation network $Q(s, a; \theta)$ selects actions, while the target network $Q(s, a; \theta^-)$ evaluates them. This separation improves stability and reduces overestimation errors.

The action selection process for state s_{t+1} is defined as:

$$a' = \arg\overset{sup}{a} \in A\, Q(s_{t+1}, a; \theta) \tag{3.1}$$

The target Q-value is computed as:

$$y_t = r_t + \delta * Q(s_{t+1}.a'; \theta^-) \tag{3.2}$$

The decoupled mechanism effectively reduces Q-value overestimation and improves learning stability:

$$Q(s_{t+1}, a^*; \theta^-) \leq \max_{a \in A} Q(s_{t+1}, a; \theta^-) \tag{3.3}$$

Using Equation 3.2, we obtain Equation 3.4 within the DQN framework.

$$y_t = r_t + \delta^* Q(s_{t+1}, \arg\overset{Inf}{a} \in A\, Q((s_{t+1}, a; \theta), \theta^-) \tag{3.4}$$

The loss function is defined as the mean squared error between predicted and target Q-values:

$$L(\theta) = E_{(s,a)}\left[(y_t - Q(s, a; \theta)^2\right] \tag{3.5}$$

The network parameters are updated using gradient descent:

$$\nabla L(\theta) = E\left[y_t - Q(s, a, \theta)\nabla Q(s, a, \theta)\right] \tag{3.6}$$

Initially, parameters are updated $\theta' = \theta + \nabla\theta$ using fixed the step size σ, after which the target network parameters $\theta^- = \theta$ are periodically synchronized with the evaluation network. In

Figure 2, the DDQN process integrates reinforcement learning and gradient descent to refine the action value function and learn an optimal control policy.

3.3 Quad Double Deep Q-Network (QDDQN)

The Quad-DQN architecture decomposes the Q-function into two components: the state value function e and the advantage function e', which are subsequently recombined to obtain the final Q-value estimate. This decomposition enables improved differential between state importance and action contribution.

To address unpredictability and overestimation issues observed in conventional DQN methods under stochastic traffic conditions, a Quad-Agent DQN framework is proposed. A preconditioned learning $q_{e(s')}$ mechanism is incorporated and the temporal difference loss is computed using structured target values. Phase selection $q_{e'(s')}$ is performed by minimizing the TD loss rather than directly maximizing Q-values, thereby improving convergence stability and estimation accuracy.

In the proposed AQDDQN framework, the temporal-difference (TD) loss is used as an indicator of learning reliability for each phase-specific agent. A lower TD-loss implies that the predicted Q-values are nearer to the target Q values, indicating more stable and accurate learning of traffic conditions. Therefore, instead of selecting actions only through maximum Q values, the proposed method selects the phase with minimum TD-loss. This strategy reduces the effect of Q-values overestimation under stochastic traffic conditions and improves learning stability, convergence robustness, and signal coordination performance.

The loss values of e and e' is evaluated relative to the learning rate to ensure stable updates, as in Figure 3. The action selection follows the Double Q-learning principle, where arg max operation $q_{e(s')}$ is applied to the evaluation network and the resulting action is evaluated using the target network. This strategy further reduces overestimation preconception and enhances training robustness.

The formulation of Equations 3.7, 3.7b are obtained form, while Equation 3.7b from Van Hasselt et al. (2016), and are presented below:

The Classical Q learning update is:

$$Q^*(s,a) = E\left[r_t + \delta \max_{a'} Q^*(s',a') \middle| s,a\right] \tag{3.7}$$

The target Q-value becomes:

$$y_t = r_t + \delta \max_{a'} Q(s',a';\theta) \tag{3.7a}$$

The Expanded Double DQN target is:

$$y_t = r_t + \delta Q\left(s', \arg\max_{a'} Q(s',a';\theta), \theta^-\right) \tag{3.7b}$$

Form the above equation the QDDQN update is:

$$Q(s,a;\theta) = r_t + \delta Q\left(s', \arg\max_{a'} Q(s',a';\theta), \theta\right) \tag{3.8a}$$

But, proposed system controls four signal phases, thus

$$Q_i(s_i,a_i;\theta_i) = r_{t,i} + \delta Q_i\left(s', \arg\max_{a_i'} Q_i(s',a_i';\theta_i), \theta_i\right) \tag{3.8b}$$

Define action selection form the evaluation network is $e^* = \arg\max_e Q(s,e;\theta)$, then the corresponding value is evaluated using target network:

$$D(q(e)) = Q(s', e^*;\theta) \tag{3.9a}$$

So the complete Double Q learning target becomes:

$$y_t = r_t + \delta Q\left(s', \arg\max_{a'} Q(s,e;\theta), \theta\right) \tag{3.9b}$$

Define the loss function; it should represent the mean squared temporal difference error, The target value:

$$T_e = r_t + \delta Q(s', e^*;\theta) \tag{3.10}$$

Therefore, a Transformed Q-value, often representing a temporal difference error/loss adjusted Q-value is $D(q(e)) = y_e - Q(s,e;\theta)$, which represents the TD error (i.e., difference between the target value and predicted Q value)

Thus, the loss function becomes:

$$L(\theta) = \left[r_t + \delta Q\left(s', \arg\max_e Q(s,e;\theta);\theta\right) - Q(s,e;\theta)\right]^2 \tag{3.11}$$

$q(e)$- Q value associated with phase (e')

Performance Metrics:

$$\text{Total reward per episode: } R = \sum_{i=1}^{T} r_t \tag{3.12}$$

$$\text{Average Travel time: } T_{avg} = \frac{1}{N}\sum_{i=1}^{N} T_i \tag{3.13}$$

$$\text{Average return value: } R_{avg} = \frac{1}{N}\sum_{i=1}^{N} R_i \tag{3.14}$$

$$\text{Average number of steps: } S_{avg} = \frac{1}{N}\sum_{i=1}^{N} S_i \tag{3.15}$$

$$\text{Average reward per step: } PR = \frac{1}{N}\sum_{i=1}^{N} \frac{R_i}{S_i} \tag{3.16}$$

The single-intersection formulation enables controlled evaluation of the proposed AQDDQN framework before extending it to large-scale network-level traffic systems.

4 Dynamic signal distribution under variable demand and time conditions

4.1 Definition of DSTCM and integration AQDDQN

Definition of DSTCM:

The Dynamic Signal Distribution and Traffic Control Model (DSTCM) is an analytical traffic state estimation model that computes phase-wise queue evolution, delay accumulation and overflow dynamics under stochastic vehicle arrivals. For each phase $i \in \{1,2,3,4\}$ and cycle k

Vehicle arrivals follow a Poisson process: $A_i(k) \sim P(\lambda_i, G_i(k))$

Queue update as follows: $Q_i(k+1) = \max[Q_i(k) + A_i(k) - \mu_i G_i(k), 0]$

Over Estimation:

Overflow occurs if $Q_i(k) > Q_i^{\max}$

Overflow delay term: $\rho_i(k) = \max[Q_i(k) - Q_i^{\max}, 0]$

Delay Accumulation:

Average delay: $D_i(k) = \frac{Q_i(k)}{\mu_i}$

Total phase delay: $I_i(k) = Q_i(k) * D_i(k)$

Total Intersection Delay: $D_{total}(k) = \sum_{i=1}^{4} I_i(k)$

4.2 Integration with AQDDQN

DSTCM does not act as an independent controller: Instead it provides structured analytical inputs to the reinforcement learning framework:

State Construction: For each agent 'i': $s_i(k) = \{Q_i(k), D_i(k), \rho_i(k), G_i(k), C(k)\}$

Global state: $s(k) = \cup_{i=1}^{4} s_i(k)$. Thus, DSTCM outputs directly define the RL state space.

Immediate reward: $r(k) = -\sum_{i=1}^{4} [\alpha Q_i(k) + \beta D_i(k) + \lambda \rho_i(k)]$

Where, α, β, γ are represents the weighting coefficients.

To analyze the contribution of the DSTCM segment, comparative experiments were conducted using reinforcement learning both with and without analytical traffic state integration, the results indicate that DSTCM improves queue estimation accuracy, overflow prediction capacity, and reward stability by providing structured traffic state representations to the AQDDQN framework. This analytical integration significantly improves learning interpretability and stabilizes policy updates under stochastic traffic demand conditions. In comparative simulations, the AQDDQN framework without DSTCM integration demonstrated higher queue instabilities, less stable reward evolution, and reduced overflow estimate accuracy under moderate and heavy traffic conditions. In contrast, the inclusion of DSTCM improved queue estimation consistency and produced more stable adaptive signal control behavior. These observation confirm the contribution of analytical traffic state modeling toward improving reinforcement learning stability and traffic performance.

4.3 Derivation of total vehicle delays for traffic signal cycle length in dynamic signal distribution

A dynamic traffic model with Poisson-distributed arrivals is analyzed. A traffic overflow formula and deterministic factors guide real-time adjustments. The use of the TRAF-NETSIM [2024] simulation enhances efficiency and reduces overflow.

The total waiting time of the interrupted queue length can be expressed as the sum of components obtained at each cycle phase is:

$$I(k,i) = \sum_{i=1}^{4} V(k,i) \quad k < i \tag{4.1}$$

The relationship between signals phases are currently active phase corresponds to the green signal and the remaining phases correspond to red signals.

$$E(I_t(k,1)) = \rho_q(0)\nu_1 + q\left[\frac{\nu_1(\nu_1+1)}{2}\right] \tag{4.2}$$

The exit flow rate in each direction can be estimated and defined as:

$$E(S_i) = p^* T_i, T_i = \frac{V_{ki}}{V_k}, p = \frac{\mu}{T_c} \tag{4.3}$$

Similarly,

$$E\left(\rho_q(\nu_1+\nu_2)\right) = \nu_2 q_{s_1} + \nu_1 q_{s_1} + \rho_q(0), \tag{4.4}$$

$$E\left(\rho_q(\nu_1+\nu_2+\nu_3)\right) = \nu_3 q_{s_3} + \nu_2 q_{s_2} + \nu_1 q_{s_1} + \rho_q(0) \tag{4.5}$$

$$E\left(\rho_q(\nu)\right) = \nu_4 q_{s_4} + \nu_3 q_{s_3} + \nu_2 q_{s_2} + \nu_1 q_{s_1} + \rho_q(0) \tag{4.6}$$

Using these results in (4.3), we get

$$V(k,1) = \nu_1 \rho_q(0) + \left(\frac{\nu_1(\nu_1+1)}{2} + \frac{\nu_1 M_{f_1}}{2P_1}\right. \tag{4.7}$$

$$V(k,2) = \nu_2\left(\rho_q(0) + P_1\nu_1\right) + (0.5)\rho_{s_2} + \frac{(\nu_2(\nu_2+1))}{2} \tag{4.8}$$

$$V(k,3) = \nu_3\left(\rho_q(0) + P_1\nu_1 + P_2\nu_2\right) + (0.5)\rho_{s_4} + \frac{(\nu_3(\nu_3+1))}{2} \tag{4.9}$$

$$V(k,4) = \nu_4\left(\rho_q(0) + P_1\nu_1 + P_2\nu_2 + P_3\nu_3\right) + (0.5)\rho_{s_4} + \frac{(\nu_4(\nu_4+1))}{2} \tag{4.10}$$

Therefore, the total waiting time of the interruption queue length is

$$I(k,i) = \sum_{i=1}^{4} V(k,i) \tag{4.11}$$

The average change in queue length is.

$$(\gamma(\tau)) = \sum_{i=1}^{4} P_i \nu_i \tag{4.12}$$

The total vehicle delays for the traffic signal cycle length are

$$I^*(k,i) = (I(k,i)) + (0.5)\rho_{i(s)} \quad i = 1,2,3,4 \tag{4.13}$$

where ρ_{is} is defined as the total vehicle waiting time accumulated across all signal phases during the interrupted cycle.

Figure 4, illustrates the performance-based optimization framework in which traffic outflows are regulated through coordinated signal operations and iterative validation. Real-time traffic data collected from Radio Frequency Identification (RFID) sensors are used to identify the current traffic state and extract key variables, which are provided as inputs to the DSTCM. It predicts signal timings inputs and traffic conditions within a specified prediction horizon and functions as the traffic state modeling and delay estimation component of the adaptive framework. The phase-wise queue lengths, interruption delays, and overflow measures generated by DSTCM, used to construct the state space, reward function, and loss evaluation. Consequently, the proposed framework integrates DSTCM with reinforcement learning to achieve adaptive, data-driven and real-time traffic signal optimization.

5 Adaptive quad agent- dynamic signal distribution control optimization (AQDSD)

The proposed AQDSD framework aims to optimize urban traffic signal operations using a multi-agent reinforcement learning architecture. Each agent independently controls a signal phase and together adjusts with signal durations based on real-time traffic data collected through Radio Frequency Identification (RFID) sensors. The system continuously collects traffic information and evaluates traffic flow performance, signal cycle length and reward values for each phase is illustrated in Figure 1. Unlike conventional Deep Q-Network (DQN) approaches, the proposed framework first performs a precondition matching process, followed by computation of the prediction error (loss value) for the current traffic state. The computed error is integrated into the Double Deep Q Network (DDQN) learning process, phase selection is then performed by comparing the loss value is selected for activation. Finally, Quad-Agent Q network reduces prediction error by comparing the learned error values e and e', thereby improving the learning accuracy and stability of the system.

Figure 1, workflow of the Adaptive Quad Agent traffic signal control framework. Initially, the cycle length, signal phases and system processing start time are utilized to zero. Real-time traffic data are collected from the each agent and corresponding traffic state variables are updated, accordingly, based on updated state information, and the signal duration is adjusted. Evaluates, each agent performance using Quad agent network by comparing the corresponding loss values associated with updated phase durations. During this process, memory management is maintained by looping existing values when the memory capacity is exceeded. After that, update the agent information and signal timing are coordinated with neighboring agents to preserve intersection level synchronization. The vehicle waiting time, phase duration and cycle length are continuously recalculated using the updated traffic data. Subsequently, signal phase selection and timing adjustments are optimized through reinforcement learning techniques to reduce vehicle delays and improve traffic flow efficiency. Finally, the overall system performance is analyzed using the Quad Agent framework, where key traffic performance indicators, including queue length, cycle duration, and vehicle delay, are evaluated to measure improvements in traffic flow conditions.

In Table 3, the phase wise traffic states derived from the DSTCM presented for each decision step. In particular, the traffic variables, including queue length, delay and overflow are supplied to the corresponding agents. These state variables, together with the defined reward structure and learning parameters, are utilized in the Agent DQN update and phase selection process, as illustrated in Figure 1.

5.1 Adaptive quadruple agent system

5.1.1 Quad agent system architecture and coordination mechanism

The proposed framework adopts a centralized training with decentralized execution architecture for traffic signal control at a four phase intersection. The control problem is decomposed into four phase specific reinforcement learning agents $i \in \{1, 2, 3, 4\}$. Each agent is responsible for optimizing the green-time allocation of its corresponding signal phase.

Phase Level Agent Representation:

Each agent observes a localized traffic data form DSTCM model:

$$s_i(k) = \{Q_i(k), V_i(k).D_i(k), \rho_i(k)\} \tag{5.1}$$

Using this state representation, each agent estimates its phase specific action value function using a Double DQN:

$$Q_i(s_i, a_i; \theta_i) \tag{5.2}$$

This decomposition reduces the dimensionality of the joint action space and enables phase wise learning aligned with physical traffic dynamics. Here, agents do not communicate directly each other. Instead of coordination is achieved through a centralized loss aggregation module located at the intersection controller level.

At each decision approach (one signal cycle)

Temporal Difference Loss Computation:

$$e_i(k) = \left(y_i(k) - Q_i(s_i, a_i; \theta_i)\right)^2 \tag{5.3}$$

where $y_i(k)$ represents the target Q-value

Global Phase Selection:

The computed loss values are transmitted to the centralized coordination layer. The optimal signal phase is selected by minimizing the aggregated loss:

$$a^*(k) = \arg \min_{i \in \{1,2,3,4\}} e_i(k) \tag{5.4}$$

Thus, the phase corresponding to the minimum temporal difference error is activated for the next signal interval.

5.1.2 Action space complexity reduction

In a conventional single agent formulation, the action space grows exponentially with the number of signal phases. If each phase 'm' discrete actions, then joint action space is

$$\left|A_{joint}\right| = m^4 \tag{5.5}$$

In compare, the quad agent framework decomposes the problem into four independent subspaces:

$$\left|A_{decomposed}\right| = 4m \tag{5.6}$$

This reduces the complexity from exponential growth $O(m^4) = O(m)$

5.1.3 Q-learning update rule

The classical Q-learning update rule is expressed as

$$Q(s, a) \leftarrow Q(s, a) + \alpha\delta \tag{5.7}$$

Where the temporal difference error is defined as

$$\delta = r + \gamma \max_{a'} Q(s', a') - Q(s, a) \tag{5.8}$$

5.1.4 Variance reduction in Q-estimation

Let the single agent Q-update rule be defined as

$$Q(s,a) \leftarrow Q(s,a) + \alpha\delta \tag{5.9}$$

Where the temporal difference error is defined as

$$\delta = r + \gamma \max_{a'} Q(s',a') - Q(s,a) \tag{5.10}$$

here, $Q(s,a)$ represents the action value function.

In high dimensional joint action spaces, the variance of TD-error increases with the size of action space.

Therefore, the variance of the TD error can be expressed as

$$Var\left(\delta^{joint}\right) \propto \left|A^{joint}\right| \tag{5.11}$$

where, A_i represents the action space corresponding to each phase specific agent.

Since

$$|A_i| < \left|A_{joint}\right| \tag{5.12}$$

It follows that

$$Var\left(\delta^{decpmposition}\right) < Var\left(\delta^{joint}\right) \tag{5.13}$$

Thus, quad agent decomposition significantly reduces the variance of the TD error. This improvement leads to enhanced learning stability, smoother gradient updates, and improved convergence reliability of the reinforcement learning algorithms.

5.1.5 Loss based coordination mechanism

Instead of selecting actions based on maximum Q-values

$$a^{\star} = \arg\max_{a} (s,a) \tag{5.14}$$

The proposed framework selects actions by minimizing prediction error:

$$a^{\star} = \arg\max_{i} e_i(k) \tag{5.15}$$

The proposed coordination strategy differs from conventional max-Q action selection frequently used in reinforcement learning methods. In standard DDQN based traffic control, the action with the highest estimated Q values is selected, which may lead to unstable decisions due to temporary Q value overestimation under stochastic traffic conditions. In compare, the proposed AQDDQN framework selects the signal phase with minimum temporal difference (TD) loss. Since TD-loss measures the difference between predicted and target Q-values, lower loss indicates more reliable and stable learning. Consequently, the proposed loss based coordination process improves convergence stability, reduces overestimation preconception, and enables more consistent adaptive traffic signal control.

5.1.6 System learning objective

The overall objective of the proposed framework is to maximize the cumulative discounted reward:

$$\pi^{\star} = \arg\max_{\pi} E\left[\sum_{t=0}^{T} \gamma^t r_t\right] \tag{5.16}$$

The reward function integrate consequences associated with queue length, delay and overflow derived from the DSTCM analytical model.

5.1.7 Comparative analysis of single agent DDQN, proposed quad agent AQDDQN and transformer based centralized control models

In single agent DDQN architecture, the control structure is fully centralized, where a single Q-network manages the joint action space of signal phases. This results in exponential growth of the action space $O(m^4)$, increased temporal difference variance, and higher sensitivity to overestimation preconception under heavy traffic demand. Although the implementation is relatively simple, convergence may be shower due to the high dimensional joint decision space.

In compare, the proposed model decomposes the intersection into four coordinated phase level agents, thereby reducing the effective action space complexity to linear order $O(m)$. Each agent operates on structured state variables derived from the DSTCM, which improves interpretability and stabilizes the learning dynamics. Instead of using conventional max-Q selection, global coordination is achieved through aggregated temporal difference loss minimization, which reduces overestimation preconception and enhances convergence robustness.

The proposed decomposition improves scalability, reduces variance, and achieves significant delay reductions under stochastic traffic conditions while maintaining moderate computational cost. Unlike transformer based centralized models, which depends on high dimensional embedding, large training datasets, and complex attention mechanisms, the proposed framework maintains stability and interpretability. Its analytically integrated structure avoids over parameterization, and provides a balanced trade-off between efficiency, robustness, and real-time deplorability for isolated intersection traffic signal control.

Figure 5 illustrates the operational workflow of the reinforcement learning framework based on the AQDDQN approach. Initially, the algorithm monitors the predefined limit. After initialization, the initial state value of the Q network is evaluated. The agents selects and executes actions within environment. As a result of action, the system observes a new state and corresponding reward. The transition tuple consisting of state, action, reward and next state is represented as (s_t, a_t, r_t, s_{t+1}). These observed tuples are stored in a memory replay buffer for future training. Periodically, the parameters of Q network are updated to maintain learning stability. This mechanism helps reduce the correlation between sequential updates and improves convergence stability. The training objective is to minimize the difference between the predicted Q-value and the target Q-value. The learning process updates the Q-network parameters iteratively until the loss function converges.

The proposed model evaluates traffic performance using a combined measure of queue length and vehicle waiting time to direct optimal traffic signal decisions. To predict traffic flow variations, two consecutive signal cycles denotes as λ_i and λ_{i+1} are analyzed. The change between these cycles determines the direction and scale of traffic flow.

```
Input:
Traffic Environment, State Space(S), Action Space (A),
Learning rate (α), Discount factor (γ), exploration
rate (ϵ)
Output:
Optimal Phase wise traffic signal control policy
Initialize four phase specific agents i ∈ {1,2,3,4}
Initialize the Q-network Q_i(s,a_i;θ_i) and target Q network
Q_i(s,a_i;θ_i^-) for each agent
Initialize an experience replay memory D_i for each agent
For each training episode:
    Observe the initial traffic state s_0
    For each decision step k:
      For each phase agent 'i' do
      Select an action a_i, using a ϵ greedy policy
      Execute the selected and observe reward r_i and the
      next state s_{k+1}
      Store the transition (s_k,a_i,r_i,s_{k+1}) in the
      replay memory D_i
      Sample a mini-batch from D_i
      Updated the Q-network by minimizing the temporal
      difference loss:
      L(θ_i) = E[(y_i − Q_i(s_i,a_i: θ_i)^2]
      End for
    Select the signal phase corresponding to the minimum
    loss among all phase agents.
    Periodically update the target networks θ_i → θ_i^-
    Update the current state s_k → s_{k+1}
    End the training loop and return the learned optimal
    traffic signal control policy.
End for
```

Algorithm 1. Adaptive Quad Agent Double Deep Q Network.

5.2 Convergence and learning stability analysis

The learning stability of the proposed framework is analyzed from an empirical deep reinforcement learning perspective. Since the proposed implementation employs a fixed learning rate of $\alpha = 0.001$, the classical Robbins-Monro stochastic approximation conditions required for strict asymptotic convergence guarantees are not satisfied. Therefore, the proposed framework does not claim formal theoretical convergence to a global optimum. Instead, the convergence behavior observed in this study should be interpreted as empirical learning stability established through simulation experiments under varying traffic demand scenarios. The use of a fixed learning rate is a common practical choice in deep reinforcement learning implementations to maintain stable and computationally efficient training behavior.

Instead, the stability of the learning process is attained through several established mechanisms commonly adopted in Deep Reinforcement Learning architectures.

For each phase specific agent 'i', the Q-network parameters are updated using gradient descent:

$$\theta_i^{t+1} = \theta_i^t - \alpha_i \nabla_{\theta_i} L(\theta_i) \tag{5.17}$$

where, $L(\theta_i)$ denotes the temporal difference loss function with the DDQN update process.

The loss function is defined as:

$$L(\theta_i) = E\left[\left(y_i - Q(s_i, a_i; \theta_i)\right)^2\right] \tag{5.18}$$

The proposed framework improves empirical learning stability through the following mechanisms:

1. Experience replay memory reduces temporal correlation among training samples.
2. Target network synchronization stabilize the Q-value update process and mitigates oscillator behavior during training.
3. The DDQN structure reduces Q-value overestimation by decoupling action section and target evaluation.
4. The quad agent decomposition reduces the effective action space dimensionality and lowers the variance of temporal difference updates.

Consequently, the proposed framework proves stable learning behavior and consistent policy improvement across simulation episodes below in low, moderate, and heavy traffic demand scenarios. The convergence characteristics reported in this study should therefore be interpreted as empirical convergence behavior observed during simulation experiments rather than as a formal stochastic approximation proof.

6 Numerical example and simulation results

To validate the effectiveness of proposed work, a comprehensive simulation study was conducted using a four way urban signalized intersection model. The simulation environment represents a signalized junction located near Sholinganallur, with stochastic vehicle arrivals modelled as a Poisson process under varying demand conditions. Although, the Sholinganallur station involves a multiple intersections junction area, the present study focuses on a single isolated four phase intersection as a main validation environment of the proposed work. This experimental setup permits the detailed evaluation of reinforcement learning stability, queue development features, overflow estimation and phase wise coordination performance of the proposed model without introducing additional network level suspicions. The isolated intersection configuration is commonly adopted in preliminary traffic signal control studies to validate algorithmic effectiveness before extending the framework to large scale coordinated urban traffic networks.

Three traffic demand levels were considered to evaluate the robustness of the proposed approach: low traffic (0.10 veh/s),

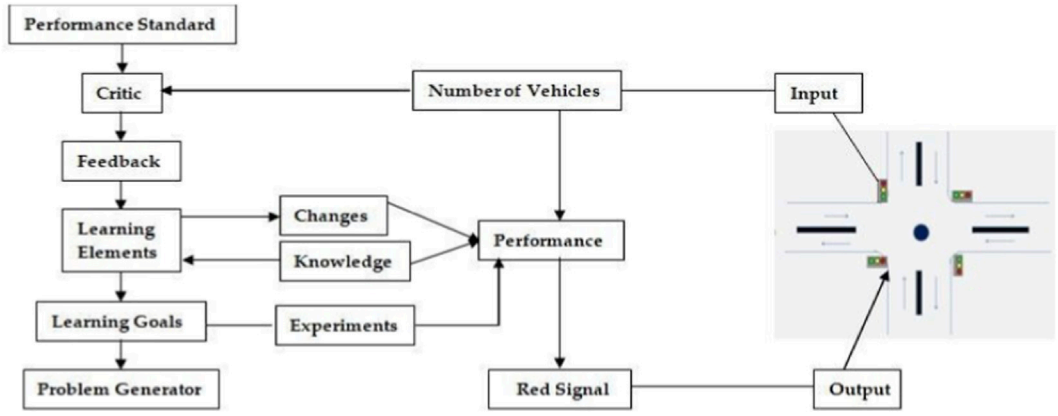

FIGURE 1
Adaptive Quad-Agent-based architecture for traffic control System.

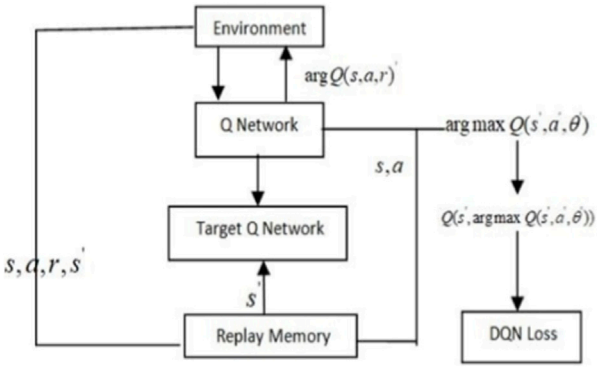

FIGURE 2
DDQN algorithm.

moderate traffic (0.30 veh/s), and heavy traffic (0.60 veh/s). Each experiment was executed for 3,600 s, and a 30 independent simulation runs were performed to make sure statistical reliability. The described performance values correspond to the mean results obtained from the 30 independent simulations runs. The related standard deviations were additionally analyzed to evaluate the robustness and statistical consistency of the proposed framework under stochastic traffic conditions. The observed variations remained within acceptable limits across all traffic demand scenarios, indicating stable learning and reliable traffic control performance.

The stated performance metrics correspond to the mean values obtained through the 30 independent simulation runs, while the associated standard deviations were analyzed to evaluate the stability and strength of the proposed framework under stochastic traffic conditions.

The performance of the proposed framework was compared with three baseline methods: Fixed time control, Double Deep Q Network, and Improved DDQN. The evaluation metrics include vehicle waiting time, queue length evolution, reward convergence, and overflow delay estimation.

The hyper parameter settings used for training the proposed reinforcement learning model are summarized in Table 3.

6.1 Implementation details and training protocol

The experimental implementation details of the proposed framework with explicitly defined. Each phase specific agents employs a fully connected network consisting of three hidden layers with 128, 64 and 32 neurons respectively. Rectified Linear Unit (ReLU) activation functions are used in all hidden layers, while the output layer corresponds to discrete green time adjustment actions. The model is trained using in Adam optimizer with a learning rate of 0.001. A replay buffer of size 10,000 is maintained, and mini-batches of size 128 are sampled during training. The discount factor is set of 0.99, and the target network is updated every 200 iterations to make sure learning stability. A greedy exploration strategy is adopted, where ∈ decays from 1.0 to 0.01 over the training

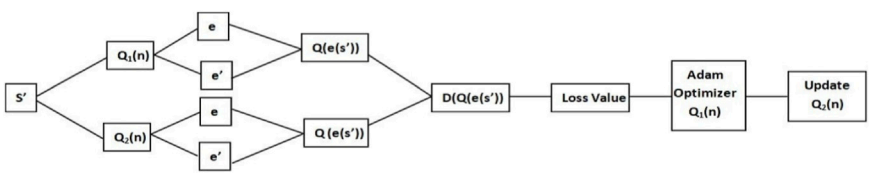

FIGURE 3
Quad−DDQN.

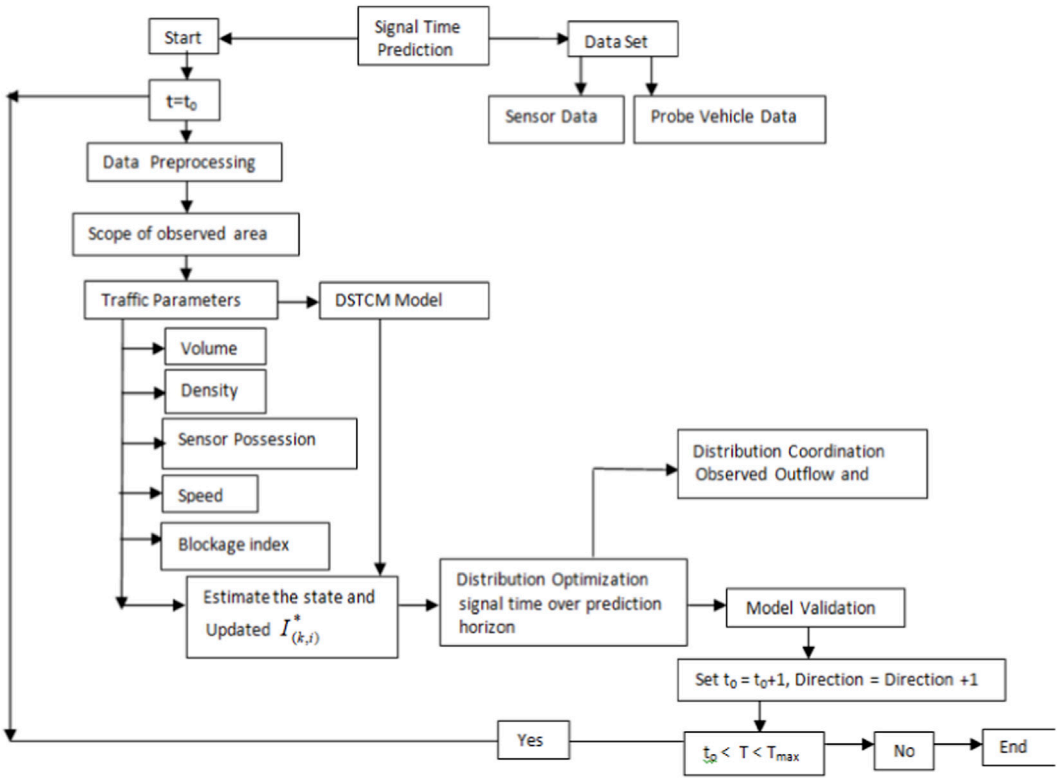

FIGURE 4
Proposed methodology.

TABLE 3 Hyper parameter setup.

Hyper-parameter	Value
Initial examination parameter	0.99
Minimum examination parameter	0.01
Reward damping parameter	0.6
Learning rate	0.001
Update sampling interval	200
Mean predictive weight	5
Inflation parameter	2
Selection chance	0.3
Training memory size	10,000
Maximum episode length	500
Iteration limit	200
Batch size	128
Performer achievement rate	0.001
Reminiscence defense dimension	64
Mini set quantity	0.99

process. Each experiment is conducted over 500 episodes, with a maximum of 500 steps per episode. The simulation and learning framework are implemented using a custom MATLAB-based simulation depends on integrating the DSTCM analytical traffic model with the AQDDQN reinforcement learning architecture.

The complete hyper-parameter setting, training configuration, state representation reward formulation, and simulation environment specifications are clearly provided to improve experimental reproducibility and simplify future comparative studies.

6.1.1 Overflow delay analysis

Figure 6 shows the comparison of predicted overflow delay across different analytical models. The proposed DSTCM model captures stochastic traffic variations more effectively than traditional model such as HCM and CTMC. Though conventional approaches tend to minimize delay under high congestion, DSTCM identifies peak congestion intervals more precisely, particularly during high demand periods.

6.1.2 Reward and queue dynamics

To evaluate the empirical convergence characteristics of the proposed AQDDQN framework, training reward evolution and

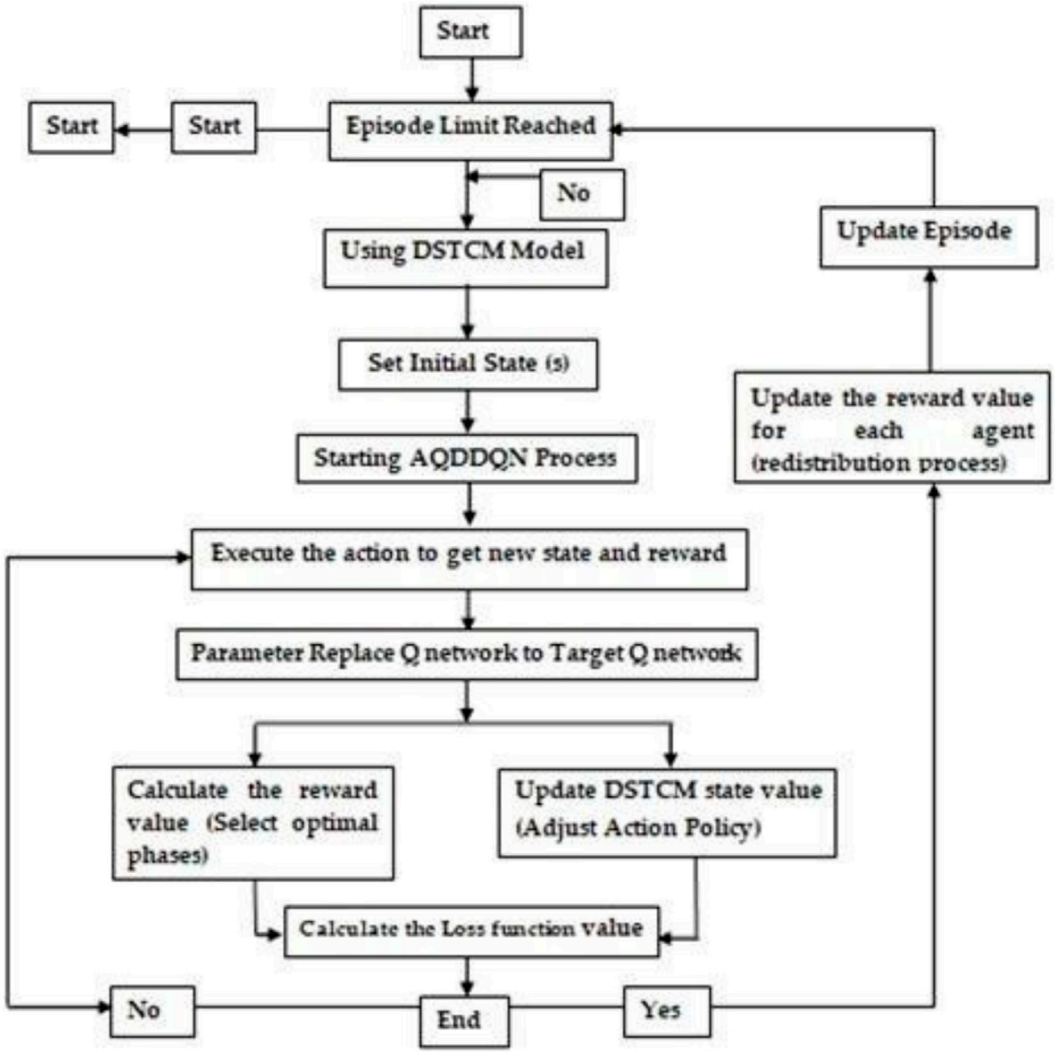

FIGURE 5
Methodology procedure.

queue stabilization behaviors are analyzed through Figures 7–9.

Figure 7 shows the total reward obtained per signal cycle for each phase. The reward values exhibit stable convergence behavior after initial exploration, indicating effective learning of optimal signal policies. Phase wise analysis exposes that the proposed framework dynamically adjusts green time allocation to minimize delay and queue accumulation.

Figure 8 discusses the relation between queue length and reward values over the 80 signal control cycles. While the reward trajectory does not complete asymptotic convergence within the training period. The proposed modes validate progressive policy stabilization and consistent developments in traffic performance. The observed reward variations are estimated under dynamic multi-traffic environments, and results should be construed as empirical learning developments rather than full convergence strategy.

Figure 9 illustrates the relationship between queue length and reward value during the simulation period. The queue length increases with vehicle arrival rates, while the reward values fluctuate between −40 and 40 reflecting the controller's adaptive response to traffic conditions. The observed tendency indicates measured convergence toward an optimal signal control policy.

To evaluate the empirical convergence features of the proposed AQDDQN framework, training reward evolution and queue stabilization performances are analyzed through Figures 7–9. The figures priminiarly discuss the cycle wise traffic performance matrices and adaptive policy stabilization during simulation. While these results indicate progressive improvement in learning behavior and traffic control stability, they should be interpreted as empirical learning developments rather than strict asymptotic convergence guarantees. Further extensions of this work may incorporate episode level cumulative reward learning curves for more comprehensive convergence visualization.

6.1.3 Comparative performance analysis

Percentage calculations:

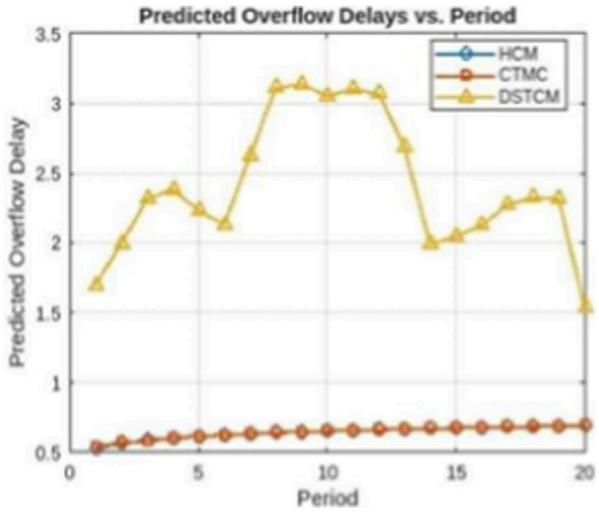

FIGURE 6
Comparison of predicted overflow delay vs. period.

$$\text{Formula: } Improvement\,(\%) = \frac{Baseline - AQDDQN}{Baseline} \times 100$$

Figure 10 presents the comparative performance analysis of DDQN, Improved DDQN and the proposed AQDDQN framework under low, moderate, and heavy traffic conditions. The result express in Table 4, proposed AQDDQN framework consistently achieves lower vehicle waiting times compared with the existing approaches. Under low traffic conditions, the AQDDQN model reduces the waiting time from 121 units (Improved DDQN) to 91 units, corresponding to a reduction of approximately 36.36%. Similarly, compared with DDQN, the proposed framework achieves a significant reduction in waiting time. Under moderate traffic conditions, the AQDDQN framework continues to provide improved delay reduction and traffic flow stability. In heavy traffic conditions the proposed framework maintains superior performance by effectively minimizing queue accumulation and vehicle waiting time through its phase-wise decomposition and loss based coordination process. Overall, the obtained results confirm that the proposed AQDDQN framework improves traffic signal control efficiency and provides more stable performance under varying traffic demand conditions.

The Improved performance observed in the proposed framework is strongly associated with the introduced loss-based coordination strategy. By selecting signal phases according to minimum temporal difference loss rather than maximum Q-values, the proposed controller achieves more stable phase's transitions and reduces the impact of Q-values overestimation during training. This behavior is reflected in the smoother reward convergence and lower queue accumulation observed under moderate and heavy traffic demand scenarios. The results therefore provide empirical evidence that the proposed loss-base coordination process contributes to improved traffic signal timing decisions and enhanced learning stability.

6.1.4 Single agent vs. multi agent analysis

Figure 11 compares the performance of single agent reinforcement learning with the proposed quad agent multi agent



FIGURE 7
Total reward per cycle for each phase.

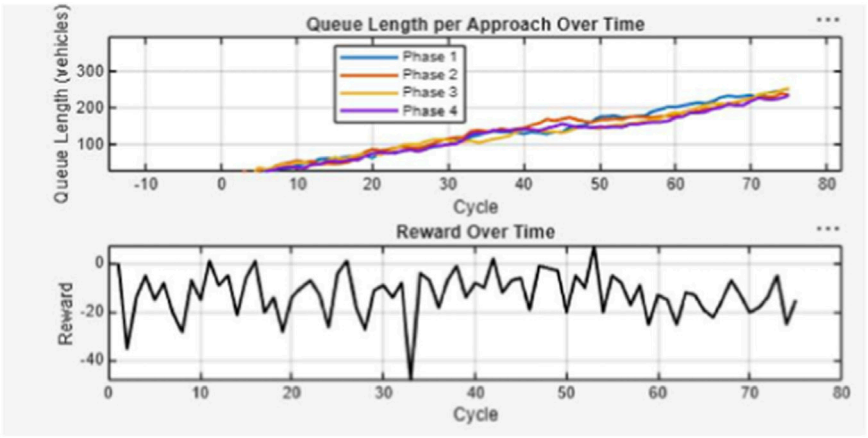

FIGURE 8
Queue length approach reward over time.

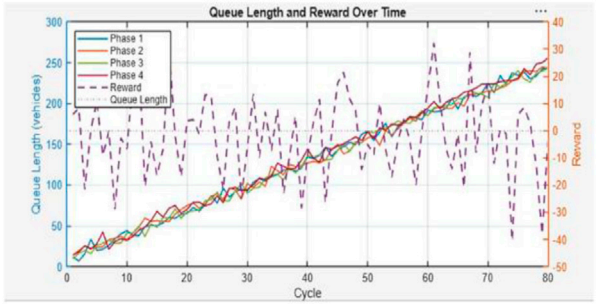

FIGURE 9
Queue length vs. reward value.

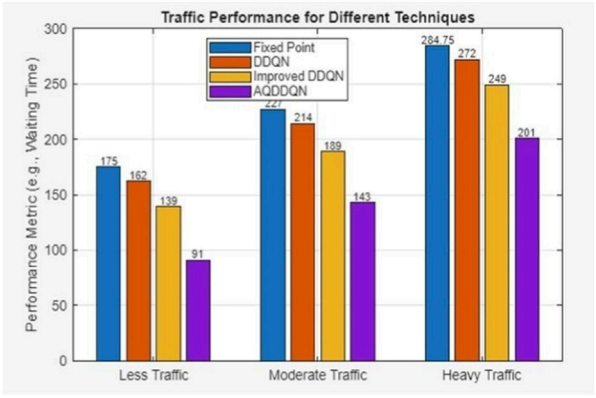

FIGURE 10
Traffic performance for different techniques.

framework. The results demonstrate that multi agent architecture significantly improves traffic control performance in terms of queue length reduction, convergence speed and signal coordination. The quad agent structure decomposes the intersection control problem into phase specific decision processes, supporting localized learning and reducing variance in temporal difference updates. This leads to improved stability and faster convergence compared with centralized single agent approaches.

7 Conclusion

This study presented an Adaptive Quad Agent Double Deep Q Network framework for intelligent traffic signal control at urban intersections. The proposed approach integrates a decentralized multi-agent reinforcement learning architecture with the DSTCM to effectively capture stochastic traffic dynamics and optimize signal timing decisions. By decomposing the intersection control problem into four coordinated agents, the proposed model enables each agent to focus on localized traffic states while contributing to global intersection optimization. The experimental results demonstrate that the proposed framework significantly improves traffic performance compared with conventional signal control methods, including fixed time control, standard DDQN and improved DDQN approaches. In particular, the model achieves substantial reductions in vehicle waiting time, queue length, and traffic congestion, especially under moderate and heavy traffic conditions. The decentralized quad agent architecture improves scalability and coordination among signal phases, allowing the system to adapt efficiently to dynamic traffic patterns.

TABLE 4 **Performance improvements (%).**

Traffic level	Vs. DDQN	Improved DDQN
Low	43.83%	36.46%
Moderate	33.18%	24.34%
Heavy	26.10%	19.28%
Mean waiting time and SD	136 ± 5.2	121 ± 4.7

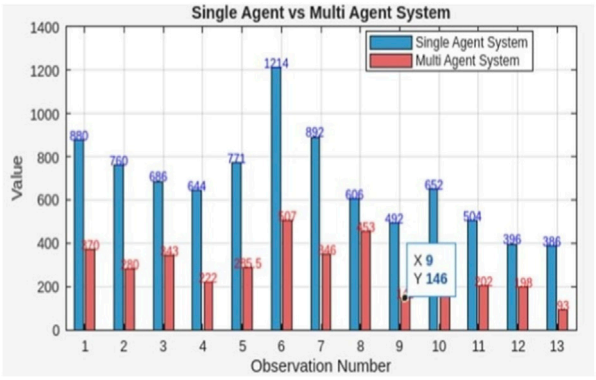

FIGURE 11
Single agent vs. Multi agent.

Though existing MARL based traffic signal control approaches that primarily depend on reward maximization and communication driven coordination, the proposed framework integrates analytical traffic state estimation with loss based cooperative DDQN learning, thereby providing improved interpretability and stable phase wise optimization under stochastic traffic environments.

Furthermore, the incorporation of precondition based learning enhances the stability of policy updates by restricting phase transitions to contextually valid actions. This mechanism reduces the exploration of suboptimal control decisions and accelerates the convergence of the reinforcement learning process. Additionally, the integration of the Double Deep Q Network framework mitigates Q-value overestimation, resulting in more reliable policy learning and smoother signal control actions. These features collectively contribute to improved robustness and reduced performance variance across multiple simulation runs.

Simulation experiments conducted under different traffic demand scenarios confirm that the propose model consistently outperforms of existing approaches. The simulation results shows that the proposed AQDDQN framework attains significant reductions in average vehicle waiting time under low, moderate, and heavy traffic conditions compared with conventional Fixed-Time, DDQN, and improved DDQN approaches. The obtained results consistently express improved queue management, reduced congestion, and improved traffic flow stability across varying traffic demand scenarios. These finding highlight the ability of the proposed model to effectively manage traffic congestion and improve intersection throughput in high-density traffic environments.

The proposed loss-based cooperative coordination process further improves policy stability by prioritizing reliable Q-value estimation instead of conventional max-Q action selection.

Though the proposed AQDDQN framework demonstrates effective performance under isolated intersection conditions, the present study is limited to a single four-phase intersection and does not fully capture large-scale urban traffic complexities. In real traffic networks, neighboring intersections interact through spillback propagation, queue interference, and coordinated signal operations, which may influence traffic stability and queue evolution. Therefore, the reported performance improvements should be considered as preliminary validation under controlled stochastic conditions. Nevertheless, the proposed quad-agent architecture provides a scalable foundation for future multi-intersection traffic control using cooperative multi-agent reinforcement learning and inter-intersection coordination process.

The present study has certain limitations. The simulation framework considers only a single isolated signalized intersection, whereas real urban traffic environments such as Sholinganallur junction involve multiple interconnected intersections with complex traffic interactions and network level dependences. As a result, large-scale traffic dynamics, inter-intersection dependencies, and network-level flow propagation are not fully captured.

Moreover, the model validates stable empirical learning performance; a formal theoretical convergence make sure under stochastic approximation is not established due to the implementing of a fixed learning rate. Though the proposed AQDDQN framework demonstrates stable empirical convergence behavior across multiple simulation scenarios, the use of a fixed learning rate does not satisfy the classical Robbins-Monro conditions required for formal stochastic approximation convergence proofs. This implementation alternative follows common practice in deep reinforcement learning to ensure stable and efficient network training.

Future will address these limitations by extending the framework to multi-intersection and network level traffic control using cooperative multi agent coordination strategies and real world traffic datasets. Additionally, convergence analyses incorporating gradually reduce the learning rate schemes will be discovered to reinforce the theoretical base of the proposed approach.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

BR: Visualization, Investigation, Data curation, Formal Analysis, Resources, Validation, Project administration, Methodology, Writing – review and editing, Software, Writing – original draft, Conceptualization, Funding acquisition, Supervision. SuS: Writing – original draft, Formal Analysis, Project administration, Validation, Data curation, Visualization, Resources, Supervision, Conceptualization, Writing – review and editing, Investigation, Software, Funding acquisition, Methodology. SiS: Investigation, Conceptualization, Writing – review and editing, Resources, Funding acquisition, Supervision, Writing – original draft, Software, Project administration, Formal Analysis, Data curation, Visualization, Methodology, Validation. LC: Writing – original draft, Formal Analysis, Project administration, Visualization, Funding acquisition, Supervision, Methodology, Data curation, Software, Writing – review and editing, Conceptualization, Resources, Investigation, Validation.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This article was co-funded by the European Union under the REFRESH–Research Excellence for Region Sustainability and High-tech Industries project number CZ.10.03.01/00/22_003/0000048 via the Operational Programme Just Transition and has been done in connection with project Students Grant Competition SP2024/087 Specific Research of Sustainable Manufacturing Technologies "financed by the Ministry of Education, Youth and Sports and Faculty of Mechanical Engineering VŠB-TUO.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

Abdur Rahima, Md., Arafatur Rahmanc, Md., Rahmana, M. M., Taufiq Asyharid, A., Zakirul Alam Bhuiyane, Md., and Ramasamy, D. (2020). Evolution of IoT-enabled connectivity and applications in automotive industry. *Veh. Commun.* 27, 100285. doi:10.1016/j.vehcom.2020.100285

Ahmed, A., Ali Naqvi, S. A., Watling, D., and Dong, N. (2019). Real-Time dynamic traffic control based on traffic-state estimation. *Transp. Res. Rec.* 2673 (5), 584–595. doi:10.1177/0361198119838842al

Akçelik, R., and Rouphail, N. M. (1993). Estimation of delays at traffic signals for variable demand conditions. *Transp. Res. Part B Methodol.* 27 (2), 109–131. doi:10.1016/0191-2615(93)90003-s

Alkaissi, Z., and Hussain, A. (2020). Delay time analysis and modelling of signalized intersections using global positioning system (GPS) receivers. *Mater. Sci. Eng.* 671, 012110. doi:10.1088/1757-899X/671/1/012110

Augusto Vieira, M., Galv, G., Vieira, M., Louro, P., Vestias, M., and Vieira, P. (2024). Enhancing urban intersection efficiency: visible light communication and learning-based control for traffic signal optimization and vehicle management. *Symmetry* 16, 240. doi:10.3390/sym16020240

Como, G., and Nilsson, G. (2021). On the well-posedness of deterministic queuing networks with feedback control. *Transp. Res. Part B Methodol.* 150, 323–335. doi:10.1016/j.trb.2021.06.010

Eom, M., and Kim, B.-I. (2020). The traffic signal control problem for intersections. *Eur. Transp. Res. Rev.* 12 (50), 1–20. doi:10.1186/s12544-020-00440-8

Feng, Y., Duives, D., Daamen, W., and Hoogendoorn, S. (2021). Data collection methods for studying pedestrian behaviour: a systematic review. *Build. Environ.* 187, 107329. doi:10.1016/j.buildenv.2020.107329

Gao, K., Han, F., Dong, P., Xiong, N., and Du, R. (2019). Connected vehicle as a mobile sensor for real time queue length at signalized intersections. *Sensors* 19, 2059. doi:10.3390/s19092059

Gomes, G. (2010). *Effectiveness Adaptive Traffic Control for Arterial Signal Management Modeling Results*.

Guan, S., Ma, C., and Wang, J. (2023). Traffic flow state analysis considering driver response time and V2V communication delay in heterogeneous traffic environment. *Sustainability* 15, 8459. doi:10.3390/su1511845

Guo, J., Kong, Ye, Zong, zhi Li, Huang, W., Cao, J., and Wei, Y. (2019). A model and genetic algorithm for area-wide intersection signal optimization under user equilibrium traffic. *Math. Comput. Simul.* 155, 92–104. doi:10.1016/j.matcom.2017.12.003

Hao, R., Zhang, Y., Ma, W. Y. , C., Sun, T., and van Arem, B. (2023). Managing connected and automated vehicles with flexible routing at "lane-allocation-free" intersections. *Transp. Res. Part C Emerg. Technol.* 152, 104152. doi:10.1016/j.trc.2023.104152

Jina, J., and Ma, X. (2017). A group-based traffic signal control with adaptive learning ability. *Eng. Appl. Artif. Intell.*, 65, 282–293. doi:10.1016/j.engappai.2017.07.022

Kim, M., Schrader, M., Yoon, H.-S., and Bittle, J. A. (2023). Optimal traffic signal control using priority metric based on real-time measured traffic information. *Sustainability* 15, 7637. doi:10.3390/su15097637

Leitner, D., Meleby, P., and Miao, L. (2022). Recent advances in traffic signal performance evaluation. *J. Traffic Transp. Eng. Engl. Ed.* 9 (4), 507–531. doi:10.1016/j.jtte.2022.06.002

Li, W., and Ban, X. (2020). Connected vehicle-based traffic signal coordination. *Engineering* 6 (12), 1463–1472. doi:10.1016/j.eng.2020.10.009

Li, M., Luo, D., Liu, B., Zhang, X., Liu, Z., and Li, M. (2022). Arterial coordination control optimization based on AM–BAND–PBAND model. *Sustainability* 14, 10065. doi:10.3390/su141610065

Li, H., Li, S., and Zhang, X. (2023). Coordination optimization of real-time signal priority of self-driving buses at arterial intersections considering private vehicles. *Appl. Sci.* 13, 10803. doi:10.3390/app131910803

Ma, D., Zhou, B., Song, X., and Dai, H. (2021). A deep reinforcement learning approach to traffic signal control with temporal traffic pattern mining. *IEEE Trans. Intelligent Transp. Syst.* 23 (8), 11789–11800. doi:10.1109/tits.2021.3107258

Maadi, S., Stein, S., Hong, J., and Murray-Smith, R. (2022). Real-Time adaptive traffic signal control in a connected and automated vehicle environment: optimization of signal planning with reinforcement learning under vehicle speed guidance. *Sensors* 22, 7501. doi:10.3390/s22197501

Maha, E., Hossam, M., and Rakha, H. A. (2021). Adaptive traffic signal control: game-theoretic decentralized vs. centralized perimeter control. *Sensors* 21, 274. doi:10.3390/s21010274

Majeed, M. F., and Fahad, A. A. (2022). Traffic control system techniques. *Int. J. Nonlinear Anal. Appl.* 13 (2), 829–835. doi:10.22075/ijnaa.2022.27438.3600

Meepokgit, T., and Wisayataksin, S. (2024). Traffic signal control with state-optimizing deep reinforcement learning and fuzzy logic. *Appl. Sci.* 14, 7908. doi:10.3390/app14177908

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., et al. (2015). Human level control through deep reinforcement learning. *Nature* 518, 529–533. doi:10.1038/nature14236

Montazeri, F., Errico, F., and Pellecuer, L. (2022). Comparison of the performance of hybrid traffic signal patterns and conventional alternatives when accounting for both pedestrians and vehicles. *Sustainability* 14, 13667. doi:10.3390/su142013667

Qu, D., Jia, Y., Wang, T., Bin, L., and Han, L. (2021). Research on coordinated control of vehicle's speed in new mixed traffic flow. *J. Intelligent Transp. Syst.*, 26 (6), 704–716. doi:10.1080/15472450.2021.1973897

Ramesh Kumar, B., Kumaran, N., Udaya Prakash, J., Salunkhe, S., Raja, V., Shanmugam, R., et al. (2024). A dynamic traffic light control algorithm to mitigate traffic congestion in metropolitan areas. *Sensors* 24, 3987. doi:10.3390/s24123987

Recker, W. (2010). *Optimization of Control Parameters for Adaptive Traffic Actuated Signal Control*. Irvine, CA: Institute of Transportation Studies University of California.

Sabar, N. R., MinhKieu, L., Chung, E., Tsubota, T., and Almeida, P. E. M. (2017). A memetical algorithm for real world multi-intersection traffic signal optimization problems. *Eng. Appl. Artif. Intell.* 63, 45–53. doi:10.1016/j.engappai.2017.04.021

Sau, J., Monteil, J., Billot, R., and El Faouzi, N. E. (2019). Steady-state analysis of second-order traffic models and application to traffic flow control. *Transp. B Transp. Dyn.* 7, 1–23. doi:10.1080/21680566.2019.1630690

Shelby, S. G. (2004). *Single-Intersection Evaluation of Real-Time Adaptive Traffic Signal Control Algorithms*.

Shi, X., Feng, S., and Lu, Y. (2023). Traffic optimization using model based reinforcement learning. *Transp. Res. Part C Emerg. Technol.* 135, 103568wa.

Solaiappan, S., Kumar, B. R., Anbazhagan, N., Song, Y., Joshi, G. P., and Cho, W. (2023). Vehicular traffic flow analysis and minimize the vehicle queue waiting time using signal distribution control algorithm. *Sensors* 23, 6819. doi:10.3390/s23156819

Song, X. B., Zhou, B., and Ma, D. (2024). Cooperative traffic signal control through a counterfactual multi-agent deep actor critic approach. *Transp. Res. Part C Emerg. Technol.* 160, 104528. doi:10.1016/j.trc.2024.104528

Sun, G., Qi, R., Liu, Y., and Xu, F. (2024). A dynamic traffic signal scheduling system based on improved greedy algorithm. *Plos One* 19 (3), e0298417. doi:10.1371/journal.pone.0298417

Szoke, L., Aradi, S., and Bécsi, T. (2023). Traffic signal control with successor feature-based deep reinforcement learning agent. *Electronics* 12, 1442. doi:10.3390/electronics12061442

Taskin, H., and Gumustas, R. (1997). "Simulation of traffic flow system and control using fuzzy logic," in *Proceedings of the 12th IEEE, International Symposium on Intelligent Control* (Istanbul, Turkey), 16–18.

Thunig, T., Kühnel, N., and Nagel, K. (2019). Adaptive traffic signal control for real-world scenarios in agent-based transport simulations. *Transp. Res. Procedia* 37, 481–488. doi:10.1016/j.trpro.2018.12.215

Tomar, I., Sreedevi, I., and Pandey, N. (2022). State-of-Art review of traffic light synchronization for intelligent vehicles: current status, challenges, and emerging trends. *Electronics* 11, 465. doi:10.3390/electronics11030465

Uppaluru, H., Liu, X., Emadi, H., and Rastgoftar, H. (2022). *A Continuous-Time Optimal Control Approach to Congestion Control, European Control Conference (ECC)*. London, United Kingdom, 1572–1577. doi:10.23919/ECC55457.2022.9838036

Van Hasselt, H., Guez, A., and Silver, D. (2016). Deep reinforcement learning with Double Q-Learning. *Proceeding AAAI Conf. Artif. Intell.* 30 (1), 2094–2100. doi:10.1609/aaai.v30i1.10295

Vandaele, N., Van Woensel, T., and Verbruggen, A. (2000). *A queueing based traffic flow model*. Transportation Research Part D: Transport and Environment, 5 (2), 121–135. doi:10.1016/S1361-9209(99)00028-0

Xu, Y., Jia Zhang, D., Kin, X. Z., Lai, K., and Su, B. (2021). Research on the traffic flow control of urban occasional congestion based on catastrophe theory. *J. Adv. Transp.* 2021, 1341729. doi:10.1155/2021/1341729

Yuan, Z., and Zeng, X. (2024). Dynamic network-level traffic speed and signal control in connected vehicle environment. *Sensors* 24, 6597. doi:10.3390/s24206597

Zhou, B., Zhou, Q., Hu, S., Ma, D., Jin, S., and Lee, D. H. (2024). Cooperative traffic signal control using distributed agent-based deep reinforcement learning with incentive communication. *IEEE Trans. Intelligent Transp. Syst.* 25 (8), 10147–10160. doi:10.1109/tits.2024.3352730

Zhu, Y., Wu, Q., and Xiao, Na (2022). Research on highway traffic flow prediction model and decision-making method. *Sci. Rep.* 12, 1–11. doi:10.1038/s41598-022-24469-y