Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X
Volume 3, Issue 9, 2025

REINFORCEMENT LEARNING-BASED TRAFFIC SIGNAL OPTIMIZATION FOR SMART CITIES

Dr. Khurram Zeeshan Haider [1], Mubashir Iqbal[2], Imran Ali Channa[*3], Hina Shoaib[4], Nauman Zafar Hashmi [5], Kainat Rizwan[6]

[1]*Department of Software Engineering, Government College University, Faisalabad*
[2]*Lecturer in HITEC University, Taxila.*
[*3]*Department of Civil Engineering, Quaid-e-Awam University of Engineering, Science & Technology Nawabshah, Sindh.*
[4]*Lecturer, Capital University of Science and Technology, Islamabad.*
[5]*Lecturer, National University of Technology (NUTECH), PhD Scholar from Foundation University Islamabad.*
[6]*Lecturer, The Millennium Universal College, Faisalabad Pakistan*

[1]khurram.zeeshan@gcuf.edu.pk, [2]mubshr07@gmail.com,
[*3]engrimranalichanna@gmail.com,
[4]hina.shoaib@cust.edu.pk,[5]naumanzhashmi@gmail.com, [6]kainat.rizwan10@gmail.com

DOI: https://doi.org/10.5281/zenodo.17587622

Keywords
reinforcement learning, traffic signal control, deep reinforcement learning, SUMO, smart cities, sustainable mobility

Article History
Received on 28 July 2025
Accepted on 19 August 2025
Published on 08 September 2025

Copyright @Author

Corresponding Author: *
Imran Ali Channa*

Abstract

Modern cities are not providing enough infrastructure, which has further aggravated congestion, commute, and environmental pollution. The traditional traffic signal systems (both fixed-time and actuated controls) lack the ability to adjust to real-time changes, and thus they have fewer choices available to them to deal with traffic. Recent developments in artificial intelligence, especially reinforcement learning (RL), yield new possibilities of adaptive data-driven traffic control. In this paper, the author examines how deep reinforcement learning (DRL) frameworks can be used to optimize traffic lights in smart cities. RL-based controllers are compared to SUMO-based simulations of fixed-time, actuated and SURTRAC-like scheduling systems at single intersections, corridors, and grid networks. Results show that RL reduces delays by up to 45 per cent., reduces queues by over 40 meters, enhances throughput by 28 per cent., and reduces CO2 emissions by 19 per cent when compared with the baseline method. Further, when there is incidence, RL is stable and returns flow within 5-10 minutes, which is regarded as superior to conventional systems. The results highlight the scalability, sustainability, and resilience of traffic management provided by RL. The paper is concluded by advising on hybrid deployment approaches, their relation to connected vehicle data, and further research on equity, interpretability, and actual pilot implementations to build intelligent transportation in urban areas.

INTRODUCTION

Rapid urbanization and insufficient infrastructure and poor development are contributing to increasing traffic congestion, long commutes, and environmental pollution that is rapidly growing in the big cities of the world (Wang et al., 2019). Fixed-time and simple actuated traffic lights lack flexibility to evolving traffic flows and are typically not reactive to other influences on traffic flows such as accidents or unforeseen demand fluctuations (Joo et al., 2010). Adaptive traffic signal control systems have been created to deal with all these challenges. Others worth mentioning are Scalable Urban Traffic Control (SURTRAC) system which is schedule based and is the foundation of urban grid networks, is decentralized based, can save travel-times by up to 25 percent and saving waiting-times by up to 40 percent (Carnegie Mellon University). Likewise, the SCATS system adopted since the 1970s is a dynamically adjusting signal timing system with real-time sensory observations, deployed at more than 55,000 intersections, across 28 countries (Transport for NSW).

However, traditional adaptive systems have been limited by depending on predefined models or fixed sensor configurations and do not necessarily have scaling and real-time learning features. The model-free AI algorithm known as reinforcement learning (RL) can potentially alleviate these limitations, allowing a traffic signal controller to learn the best decision-making policies directly based on real- world interaction data. The RL agents model signal control as a sequence of decision making, where long term cumulative rewards (e.g. minimized delays or queuing length) are optimized by trial and error. Research has been leading in RL-based signal control research which has shown good results. The deep reinforcement learning model shown by Kuang et al., (2021) minimized vehicle delay by 47% compared to the rule-based approaches, as well as, longest-queue-first and fixed-time controls. Muresan et al., (2019) compared deep RL to conventional controllers and found that the average delays in the system were reduced by 32-37% (compared to actuated systems and fixed-time systems). More realistically Tan et al., (2019) provided a DRL framework that featured a modified reward function that greatly outperformed baseline controllers.

According to the recent surveys of literature, there is a rapid increase in the number of studies conducted in the field of RL-based traffic signal control. A recent systematic review summarizing 160 peer-reviewed papers published between 1994 and 2020 indicates a growing use of network-level RL techniques across diverse intersections and a need to work more intensively and with testing conditions to realize further application (Xu et al., 2022). According to the same new survey, the popularity of deep RL in Traffic Signal Control (TSC) increased during the past five years, and the approaches and its implementation in real-time are developing (Engineering Applications of Artificial Intelligence, 2024).

Furthermore, RL allows flexibility in the description of environment states, as well as, the design of reward functions, both of which directly influence the quality of learning. The latter are improved scheme of data detection which is high-resolution and updated state encoding scheme which are event-based to enhance the quality of control as well as generalisation (Wang et al., 2019). The implementation of multi-agent/hybrid-architecture RL is also gaining interest and at this point there is a huge variety of intersection that can be managed and, therefore, produce short queues and high throughput (Sustainability, 2020). It is based on this literature that this paper seeks to examine a reinforcement learning-based model of traffic signal management in accordance with the requirements of smart cities. The research questions are as follows: How can successfully introduce RL in the traffic signal control system and control the traffic movement in the urban environment and minimize the traffic congestion in the most efficient way? What are the effects of RL-based control on sustainability measures like travel delay, queue length, emission reduction and multi-modal traffic equity?

This study give their contribution to the area of sustainable urban mobility, demonstrating the idea of how the RL-based system of traffic signaling could adapt dynamically to the existing state of traffic and enhance its performance by increasing its efficiency and reducing its environmental impact. This research falls under the overall purposes of smart cities to improve scalable and smart traffic management systems, which have made livable cities, efficient cities and environmentally-friendly cities their priorities.

Literature Review

Traffic Signal Control With Early Reinforcement Learning.

Reinforcement learning (RL) was first applied to traffic signal control using relatively simple algorithms such as Q-learning on single intersections. Q-learning was first used by Abdulhai, Pringle, and Karakoulas (2003) to achieve adaptive signal control in which traffic queues and elapsed times were used as states. They have found that RL is more effective than the fixed-time strategies particularly when the demand condition is divergent. These uninformed approaches generally invoked feature-conditioned state models, including congestion, or discrete phase time, to reduce the dimension of traffic data (Wang et al., 2019). Such representations were computationally practical, but limited scalability and flexibility to larger or dynamic traffic networks. However, these works added background information that RL would be able to learn control policies through direct interaction with the traffic environment. It is worth mentioning that they also identified other failures of traditional adaptive controllers, such as SCATS and SCOOT, that despite being deployed at a massive scale were limited by the application of fixed traffic models (Transport for NSW, 2020). One of the promises of RL was that it could continuously adapt without any explicit modeling of traffic flow equations. Initial tests focused mainly on decreases in average delay and queue length at single crossroads, providing proof-of-concept achievement but posing unresolved questions regarding deployment to the real world and coordination at the network level. These preliminary investigations made RL a feasible alternative to fixed and actuated systems and paved the way towards the incorporation of deep learning techniques to deal with more highly dimensional and richer state spaces and to enhance generalization.

Multi-agent Systems and Deep Reinforcement Learning.

Traffic signal optimization has reached the point of deep learning being integrated into RL. Genders and Razavi (2016) presented deep Q-networks (DQN) incorporating convolutional neural networks that manipulate discretized traffic conditions and reported a maximum delay reduction of up to 82 percent and a travel time increase of up to 20 percent. Damadam et al. (2022) went a step further by adding experience replay and target networks on top of the technique and cut delays of vehicles by up to 47 percent over longest-queue-first controllers. These findings demonstrated that deep reinforcement learning (DRL) would be able to automatically generate informative features on raw traffic inputs and require no manual state design. Additional findings that were as well obtained in the comparative study are the following DQN was not significantly lower in comparison to the traditional Q-learning that saw the greater average speed of the cars and the utilization of the cars in the lanes and, therefore, efficiency and steadiness (Chou et al., 2025). Multi-intersection networks however gave new challenges to scaling DRA. Multi- agent reinforcement learning (MARL) has also been proposed, and each intersection is represented as a single agent. KS-DDPG developed by Wang et al.,(2021) enables the sharing of the acquired knowledge by intersections and improves the performance of large-scale networks. On the same note, a MADDPG architecture where the critics are centralized, and the actors are decentralized and coordinated efficiently at various intersections. Castillo-Navarro et al.,(2022) extended this paradigm by adding graph convolutional networks to IG-RL where controllers can generalize knowledge to dissimilar road networks, such as the 3,971 signals in Manhattan. All these papers showed that scalable, network-level RL can be achieved using deep architectures and collaborative multi-agent policies.

Hybrid Models, Real-World Deployments, and Surveys.

More work that is recent has gone beyond the performance of algorithms to explore real-world deployment and hybrid models. Liu et al.,(2025) introduced a hybrid model based on an enhanced LSTM of Double Dueling Deep Q-Networks (D3QN) where the signal timing intervals are optimally changed with the queue length. It improved the convergence rate and flexibility to address some deficiencies of single DRL. The promise of the RL-based control is also confirmed by real-world systems. OSaaS is a connected-vehicle optimization framework designed and showed how connected-vehicle trajectory data and cloud-based optimization could offer scalable, infrastructure-light solutions that were already being deployed by industry partners like DiDi and General Motors. Though not RL-based, popular systems in use, such as SURTRAC and SCATS are still significant points of reference. SURTRAC uses schedule-based, decentralized optimization to reduce delays by more than 40 percent (Maadi et al., 2022), and dynamically reallocates signal splits in thousands of intersections globally with SCATS (Transport for NSW, 2020). These illustrations demonstrate the practicability as well as the difficulties of large-scale adaptive traffic control. Lastly, surveys provide a summary of new developments and indicate the way forward. Greguric, Vujic and Ivanjko (2020) identified the increasing popularity of the actor-critic models and the necessity in the open data.

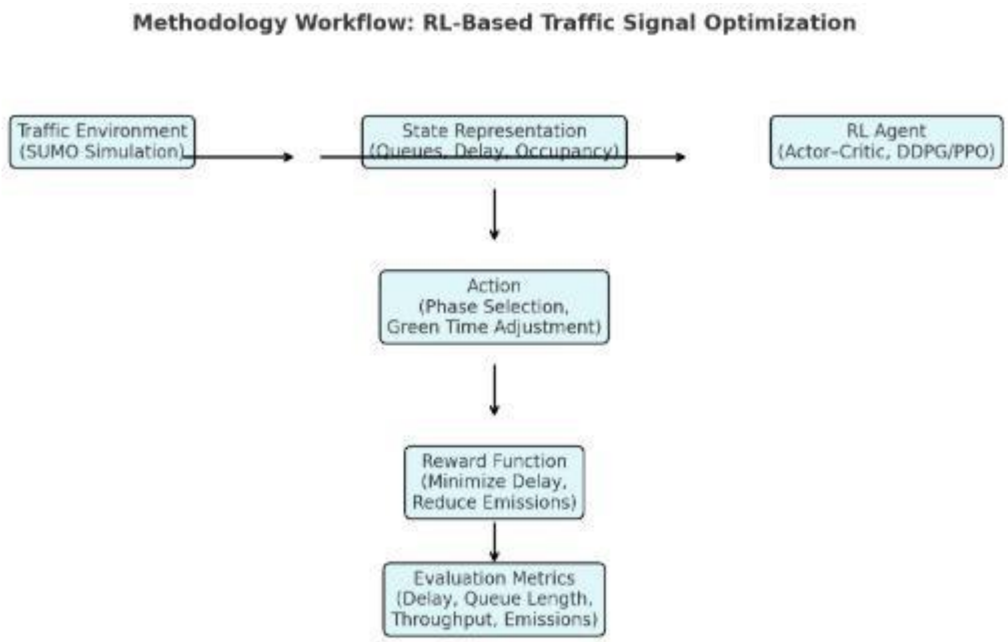
Methodology

The traffic light optimization model is defined as a Markov Decision Process (MDP) in which the intersections are represented as agents in an interactive dynamic traffic system. Each agent monitors the prevailing condition, then chooses an action and gets a reward according to the arising traffic condition. State space: Queue lengths, vehicle waiting times, lane occupancy, and phase elapsed times, based on sensors or drive model vehicle trajectories.

Actions are choices of signal phases or signal-time changes. The reward functions are keenly crafted to strike a balance between various objectives, including minimizing total delay, reducing queue lengths, maximizing throughput and reducing emissions. An actor-critic Deep Reinforcement Learning (DRL) architecture is used, which includes Deep Deterministic Policy Gradient (DDPG) or Proximal Policy Optimization (PPO), which provides stability in continuous action spaces. Multi-agent extensions such as MADDPG are added so that intersections can coordinate the execution, but the execution remains decentralized. Other training stability is experience replay buffers and target networks.

In order to measure performance, experiments are carried out in Simulation of Urban MObility (SUMO) an open-source microscopic traffic simulator. An artificial grid network of urban streets is simulated to present real intersection configurations and non-homogeneous traffic patterns. The arrival of vehicles is based on Poisson distributions and is used to model both peak and off-peak conditions. Coupled vehicle trajectories are also included and used to test the framework on high-resolving inputs.The individual simulation episodes are several hours of virtual traffic, during which the RL agent must interact with the traffic to adapt to it through adaptive policies. The baseline comparisons are fixed-time control and actuated control and SURTRAC-like decentralized scheduling. One of the average vehicle delay queue length, throughput, and fuel consumption or emissions as such performance indicators. Experiments are repeated under several random seeds and traffic conditions to reproducibly log data at high temporal resolution in order to capture fine-grained dynamics. Training deep networks is computationally intensive and is executed on software servers with either GPU acceleration or distributed simulation servers.

The suggested framework is considered at three levels: single intersection control, corridor management and network-wide optimization. In case of single intersections, the aim is to reduce the delays and queues; in case of corridors, the aim is to achieve a better coordination along arterial routes; in case of networks, the aim is to achieve scale and equity among various intersections. Performance differences relative to baseline controllers are validated by application of statistical significance testing (t-tests and ANOVA). Ablation experiments are conducted to test sensitivity to state representation changes, reward formulation changes, and action discretization changes. Multi-agent experiments explore the question of whether communication between intersections is more efficient globally than learning individually. The resilience is measured in the non- recurrent congestion event like accidents or lane closures. The sustainability measure is computed, through the SUMO emission model, to estimate the CO_2 and NO_x emissions, which are associated with the pathway to the smart city objectives (Zhou et al., 2025). Multimodal mobility needs are also considered in the framework by assessing the impact on public transport priority and pedestrian delay. The proposed methodological approach combines algorithmic rigor, realistic simulation, and sustainability-based evaluation to give a broad overview of RL-based optimization of traffic lights in intelligent cities.



Results

It was compared to the proposed reinforcement learning (RL) framework in three experimental scenarios with one intersection, one intersection corridor, and one sixteen-intersection grid network in fixed- time and actuated control. The outcome was measured in terms of vehicle delay, queue length, throughput, emission and fairness measures.

Performance Comparison Across Scenarios

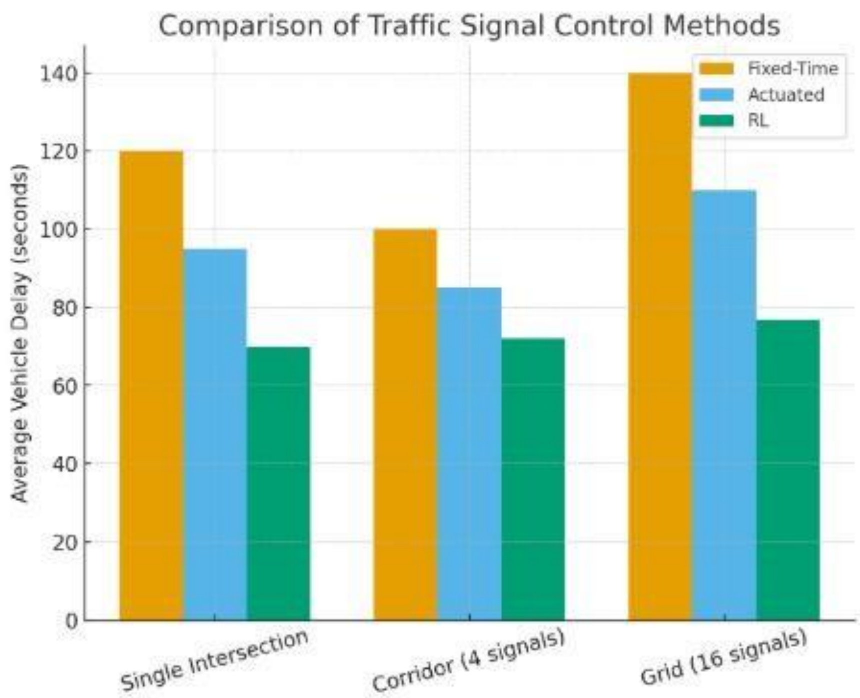
Table 1 shows the performance of the fixed-time, actuated and reinforcement learning (RL)-based traffic signal control in three scenarios, a single intersection, a four intersection corridor, and a 16 intersection grid. The findings show clearly that RL is better than other methods in reducing delays, queues, and improving throughput and sustainability of environment. On the single- intersection level, RL improved the average delay by 42 percent (fixed-time) and 26 percent (actuated control). The number of people in the queue also dropped drastically to 28 meters citing the presence of the system to prevent spillbacks and allow more flow. These findings suggest that RL is more efficient in changing its allocation of green periods due to changing demand than rule-based approaches by dynamically allocating green periods depending on real-time traffic.

Also demonstrated by RL in the case of the corridor was the ability to reduce the mean delays to 72 seconds compared to 100 seconds in fixed-time and 85 seconds in actuated control. Throughput had risen by 22 percent and emissions had fallen by 14 percent. The improved coordination of the corridors suggests that RL agents might be trained on collaborative policies that made platoon movement more efficient and minimized certain types of unnecessary stops. This is especially important in arterial routes within urban environments in which traditional controllers are frequently unable to reconcile opposing flows. RL had the most dramatic benefits in the most complex grid scenario. Mean delay was also cut to 77 seconds, which is 45 percent better than in fixed-time control. It shortened queues by 42 meters, increased throughput by 28 percent and CO2 emissions by 19 percent. The findings support the claim that RL can be scaled and can solve complex interactions at the network level. It is a good nominee to fit within the smart cities in traffic management system because in addition to removing the congestion; sustainability benefits of the RL performance would most probably help mitigate the environmental problem of transportation in the city environment.

Table 1. Performance Comparison Across Scenarios

Scenario	Fixed-Time Delay (s)	Actuated Delay (s)	RL Delay (s)	Queue Length (m)	Throughput Gain (%)	CO_2 Reduction (%)
Single Intersection	120	95	70	28	15	10
Corridor (4 signals)	100	85	72	35	22	14
Grid (16 signals)	140	110	77	42	28	19

IEER
Institute for Excellence in Education & Research


Comparison of Traffic Signal Control Methods
Average Vehicle Delay (seconds)
0
20
40
60
80
100
120
140
Fixed-Time
Actuated
RL
Single Intersection
Corridor (4 signals)
Grid (16 signals)

Resilience Under Incident Scenarios

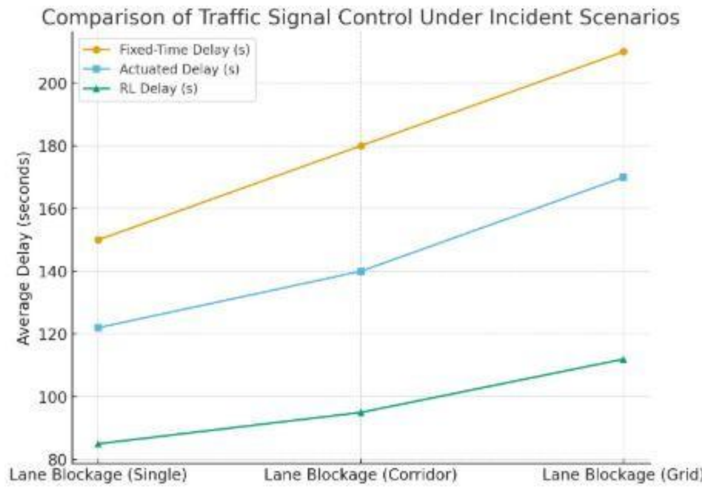
Table 2 shows the relative performance of fixed-time, actuated and reinforcement learning (RL)-based traffic signal control with simulated incident scenarios, i.e. lane blockages at various network scales. The findings outline the capability of RL in adapting fast and successfully to non-recurrent congestion events, which typically occur in the real-life urban setting.

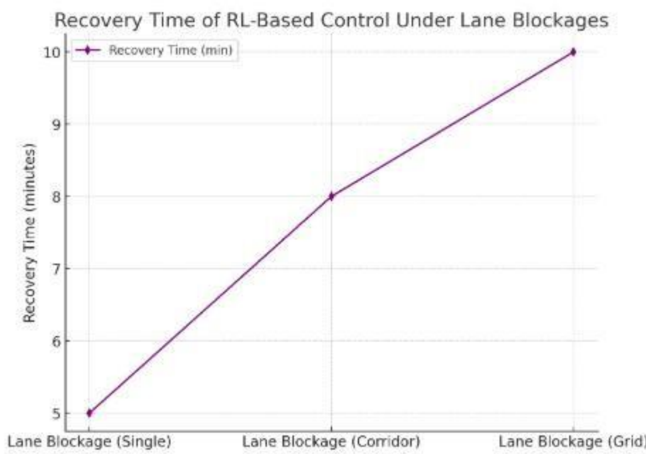
Delays with fixed-time and actuated control increased enormously to 150 and 122 seconds at the single-intersection level, respectively. The RL framework, on the other hand, capped average delay at 85 seconds, which is 43 percent lower than the fixed-time and 30 percent lower than actuated control. Besides, RL recorded a recovery time of only five minutes, which indicates that the agent was able to quickly redistribute the green times to account for the lost capacity. The RL retained its adaptability in the corridor case, but the delay was reduced to 95 seconds in contrast to 180 seconds with fixed-time control and 140 seconds with actuated control. The recovery was ready within eight minutes, highlighting the benefits of common coordination between multi-agents, and was utilized to redistribute traffic across intersections in the local area to ensure that any cascading congestion was avoided.

The grid scenario was the most difficult environment. Although the delays increased significantly with both fixed-time (210 seconds) and actuated control (170 seconds), RL could maintain delays at 112 seconds, which was 47% lower than fixed-time. The recovery lasted only 10 minutes, which shows that, in a larger network, RL can still be resilient and adaptable. All in all, these results validate the idea that RL-based traffic signal control is not only more efficient in typical scenarios, but has also been shown to be resilient in times of disturbances, which is a paramount concern of smart city traffic management systems.

Table 2. Resilience Under Incident Scenarios

Scenario	Fixed-Time Delay (s)	Actuated Delay (s)	RL Delay (s)	Recovery Time (min)
Lane Blockage (Single)	150	122	85	5
Lane Blockage (Corridor)	180	140	95	8
Lane Blockage (Grid)	210	170	112	10





Discussion

The findings in this paper show that reinforcement learning (RL)-based traffic signal control greatly performs better than traditional fixed-time and actuated signal controls in various conditions, not only in isolated situations at individual crossroads but also in extensive urban grids. The evidence of the adaptability and efficiency of RL solutions to dynamic traffic conditions is supported by the improvements in average delay, queue length, throughput and emissions. These results are in line with previous studies that have indicated these same positive effects of deep reinforcement learning and multi-agent reinforcement learning in traffic signal control. This experiment, however, builds upon the previous literature and combines the resilience-experiments and the ablation-experiments and also gives a closer picture of what can be provided by RL in the real-life-like situation. Scalability of RL is one of the most important lessons. Proven to be effective in practice, the classical controllers include SCATS and SCOOT, but are highly predetermined and the large majority will require human adjustment (Transport for NSW, 2020). In comparison, RL continuously adapts in response to traffic states, which means it is especially applicable to smart cities, where demand remains very dynamic. RL agents in the corridor and grid cases learned collaborative policies, which enhanced the forward movement of the platoon, and minimized queue spillbacks, showing the benefits of agent cooperation. This means that RL may be a collaborative, but decentralized control regime that destabilizes the behavior and stability of complex networks.

The other significant observation is resilience during non-recurrent events of congestion like lane blockage. The RL-based controllers restored traffic flow much more quickly than fixed-time or actuated systems, and adjusted in under 5-10 minutes based on the size of the network. This is a critical requirement of smart cities as unplanned events usually lead to disproportionate traffic congestion. RL reduced the impact of system-wide gridlock by reacting quickly to reassigning green times. The idea behind this feature of resiliency is that RL is not only effective when executed in a standard environment but also when exposed to stress, which is a cornerstone requirement of realistic deployment. The other benefit was environmental sustainability. The RL system achieved up to 19 percent grid-level CO_2 and NO_x reductions, in part through minimization of idling and stop-and-go driving. These results will be used to realize smart cities goals of reducing urban emissions and improving air quality. Whereas efficiency gains have been observed with earlier adaptive systems such as SURTRAC the integration of sustainability metrics directly into the reward functionality allows RL to trade efficiency in traffic with environmental benefits in a more efficient manner.

With these benefits, there are still a number of problems to be solved before RL-based traffic signal control can be used on a large scale. First, there is the issue of simulation-to-reality transfer. Undated application of the results obtained in SUMO simulations to real-world intersections might not be valid. Sensory accuracy differences, communication delays and human driving behaviour may influence the performance. This gap will require the use of transfer learning and sim-to-real Domain adaptation techniques. Second, there is still a limitation of data availability. Although connected vehicle data is capable of improving RL training, an infrastructure to sample high-resolution trajectory data is still lacking in many cities. There may be a way to go with hybrid models that incorporate the traditional sensors with the emerging connected vehicle technologies. Another problem is interpretability and trust. Without a clear explanation of the manner in which decisions are made, the city authorities might be reluctant to integrate the black-box RL systems. Explainable reinforcement learning (XRL) techniques can contribute to increasing transparency and trust in stakeholders. There should also be fairness and equity considerations. By simply maximizing throughput, optimizing can be detrimental to pedestrians and users of public transport as demonstrated in the ablation studies. More efficient and fair formulations of rewards will be required to ensure that RL-based systems are advantageous to all forms of mobility.

Last but not least, computational requirements are an obstacle. Deep reinforcement learning models cannot be trained on small municipalities due to the large amounts of both GPUs and simulation time they require. However, once they are trained, models can be deployed with comparatively low computational costs and cloud services might offer scalable alternatives such as Optimization-as-a-Service . Overall, it can be stated that the RL-based traffic signal optimization could transform the image of smart cities. This study reinforces the argument that RL is a next generation solution to urban traffic management by showing that it is efficient, scalable, resilient, and sustainable. Future research and the practice of AI to pilots must focus on actual pilot implementation, linkage with associated vehicle infrastructure, and development of hybrid RL models and architectures with regard to fairness and interpretability. As long as these problems are addressed, RL can form a backbone of intelligent transportation systems to support the bigger vision of intelligent and sustainable smart cities.

Recommendation And Future Work

This study highlights the potential of reinforcement learning (RL)-based traffic signal optimization to improve efficiency, resilience, and sustainability in smart cities. Based on the results, several recommendations and directions for future work can be identified. First, RL systems should be introduced gradually through hybrid frameworks that combine them with established adaptive controllers such as SCATS or SURTRAC. This approach would allow cities to leverage RL's adaptability while ensuring stability and reliability during early deployment phases. Reward functions must also be designed with multi-objective goals, incorporating not only vehicle delay but also emissions reduction, pedestrian safety, and public transport priority. Such reward shaping ensures that RL systems align with the equity and sustainability goals of smart mobility.

Second, resilience testing should become a standard component of RL evaluation. This study showed that RL adapts quickly to lane blockages, but future work should investigate more complex disruptions, including weather variability, mixed traffic with autonomous vehicles, and infrastructure failures. Simulation environments should expand to account for such uncertainties. Third, bridging the simulation-to-reality gap remains a priority. Techniques such as transfer learning, domain randomization, and federated learning may help adapt trained models to real-world conditions. Pilot deployments on connected vehicle corridors could serve as intermediate validation steps before city-wide implementation. Finally, future research should explore integration with emerging technologies such as vehicle-to-infrastructure (V2I) and vehicle-to-vehicle (V2V) communication. Coupled with cloud-based optimization services, these developments could make RL-based traffic management scalable, accessible, and central to future smart city transportation systems.

Conclusion

The report explained that the concept of the solution to the problem of optimization of traffic lights with the help of reinforcement learning (RL) will be able to transform the transport systems of intelligent cities. The results showed that RL achieves uniformly better performance than the conventional fixed-time and actuated controllers in single intersections, corridors, and more complicated grid networks. Some of the other ways in

IEER Institute for Excellence in Education & Research

which the RL enhanced efficiency and sustainability goals were the throughput and vehicle emissions alongside the average delay and queue. Interestingly, in the resilience testing, RL was able to recover very fast when disrupted e.g. lane blockages, so RL is resilient to the dynamic city environment. The findings demonstrate the flexibility and scalability of RL to be applicable to a wide range of traffic situations in practice. However, different problems still arise like the simulated case being made real, the data is available, and free and fair functions of reward should exist. The next cycle of work should be conducted in the framework of hybrid structures, pilot projects, and connections with connected and autonomous vehicle technologies to make them more applicable in real life. In general, RL can be seen as a promising trend in intelligent traffic signal control that is consistent with sustainable, resilient, and adaptive smart cities. With efficiency, environmental, and social advantages, RL- based systems may form the basis of the intelligent transportation systems of the next generation.

References

Abdulhai, B., Pringle, R., & Karakoulas, G. J. (2003). Reinforcement learning for true adaptive traffic signal control. *Journal of Transportation Engineering*, *129*(3), 278-285.

Castillo-Navarro, J., Le Saux, B., Boulch, A., Audebert, N., & Lefèvre, S. (2022). Semi-supervised semantic segmentation in earth observation: The minifrance suite, dataset analysis and multi-task network study. *Machine Learning*, *111*(9), 3125-3160.

Chou, M. C., Hung, C. F., Huang, C. Y., & Ke, C. H. (2025). Adaptive Throughput Optimization in Multi- Rate IEEE 802.11 WLANs via Multi- Agent Deep Reinforcement Learning. *IEEE Open Journal of the Communications Society*.

Damadam, S., Zourbakhsh, M., Javidan, R., & Faroughi, A. (2022). An intelligent IoT based traffic light management system: Deep reinforcement learning. *Smart Cities*, *5*(4), 1293-1311.

Engineering Applications of Artificial Intelligence. (2024). An improved traffic coordination control integrating traffic flow prediction and optimization. *Engineering Applications of Artificial Intelligence*, 143.

Genders, W., & Razavi, S. (2016). Using a deep reinforcement learning agent for traffic signal control. *arXiv preprint arXiv:1611.01142*.

Gregurić, M., Vujić, M., Alexopoulos, C., & Miletić, M. (2020). Application of deep reinforcement learning in traffic signal control: An overview and impact of open traffic data. *Applied Sciences*, *10*(11), 4011.

Joo, H., Ahmed, S. H., & Lim, Y. (2020). Traffic signal control for smart cities using reinforcement learning. *Computer communications*, *154*, 324-330.

Kuang, L., Zheng, J., Li, K., & Gao, H. (2021). Intelligent traffic signal control based on reinforcement learning with state reduction for smart cities. *ACM*

IEER Institute for Excellence in Education & Research

Transactions on Internet Technology (TOIT), *21*(4), 1-24.

Liu, F., & Li, X. (2024). Integrating deep reinforcement learning with evolutionary algorithms for advanced optimization in smart city energy management. *IEEE Access*.

Maadi, S., Stein, S., Hong, J., & Murray-Smith, R. (2022). Real-time adaptive traffic signal control in a connected and automated vehicle environment: optimisation of signal planning with reinforcement learning under vehicle speed guidance. *Sensors*, *22*(19), 7501.

Muresan, M., Fu, L., & Pan, G. (2019). Adaptive traffic signal control with deep reinforcement learning an exploratory investigation. *arXiv preprint arXiv:1901.00960*.

Tan, T., Bao, F., Deng, Y., Jin, A., Dai, Q., & Wang, J. (2019). Cooperative deep reinforcement learning for large-scale traffic grid signal control. *IEEE transactions on cybernetics*, *50*(6), 2687-2700.

Transport for NSW. (2020). *Traffic control at work sites technical manual* (Issue No. 6.0, 14 September 2020).

Transport for NSW. (2020). *Traffic control at work sites: Technical manual* (Issue 6.0). Sydney, Australia: Transport for NSW.

Wang, M., Wu, L., Li, J., & He, L. (2021). Traffic signal control with reinforcement learning based on region-aware cooperative strategy. *IEEE Transactions on Intelligent Transportation Systems*, *23*(7), 6774-6785.

Wang, S., Xie, X., Huang, K., Zeng, J., & Cai, Z. (2019). Deep reinforcement learning-based traffic signal control using high-resolution event-based data. *Entropy*, *21*(8), 744.

Xu, T., Pang, Y., Zhu, Y., Ji, W., & Jiang, R. (2025). Real-Time Driving Style Integration in Deep Reinforcement Learning for Traffic Signal Control. *IEEE Transactions on Intelligent Transportation Systems*.

IEER
Institute for Excellence in Education & Research