Multi-Agent Reinforcement Learning for Traffic Signal Optimization in Tier-2 Cities

Sura Hamed Mousa[1], Mahmood Anees Ahmed[2] and Subrata Sahana[3]
[1] Al-Turath University, Baghdad, Iraq, 10013 Baghdad, Iraq
[2] Medical Technical College, Al-Farahidi University, 10065 Baghdad, Iraq
[3] Sharda School of Computing Science & Engineering, Sharda University, 201310 Greater Noida, India
Sura.hami@turath.edu.iq, mahmood.anees@life-rdh.org, subrata.sahana@sharda.ac.in

Keywords: Multi-Agent Reinforcement Learning (MARL), Traffic Signal Optimization, Tier-2 Cities, Adaptive Traffic Control, Intelligent Transportation Systems (ITS), Deep Reinforcement Learning (DRL).

Abstract: Traffic jamming in cities is also one major issue of concern with the major concern being the situation in Tier-2 cities where urbanization has been taking an overgrowth without the development of the infrastructure. The older fixed time and actuated traffic signal systems are not usually able to adjust to irregular and mixed traffic, which causes long delays, more fuel use as well as higher emissions. In order to solve these problems, this paper suggests a Multi-Agent Reinforcement Learning (MARL)-based adaptive traffic signal control scheme. According to the suggested model, every intersection will be described as a self-governed agent, which monitors the state of local traffic, chooses the best possible signal phases, and modifies its policy according to the information provided by the environment. The system was tested with the realistic simulation environment of a Tier-2 city with mixed vehicle vehicles, pedestrian movement and noisy sensor data. Findings indicate that the MARL framework can reduce vehicle delays as well as pedestrian wait times by a significant factor than fixed-time, actuated, and single-agent DRL models in enhancing throughput and decreasing emissions. In addition, ablation experiments ratified the significance of multi-objective reward design in the attainment of a balanced optimization. This study identifies the opportunity of using MARL as a scalable and cost-effective solution to enhance the management of traffic in resource-limited urban settings, and this research paper prepares the way to deploy it in real-world Tier-2 city networks in the future.

1 INTRODUCTION

The fast rate of urbanization in emerging economies has posed a great challenge to the control of transportation systems in an urban area. The traditional types of traffic signal control systems (fixed-time or actuated control) in Tier2 cities have an irregular infrastructure and traffic pattern that do not adjust to dynamic traffic flow. The effects of such practices are usually prolonged delays, use of more fuel and pollution. The necessity of smart and dynamic traffic signal control systems is now paramount especially in towns which do not have developed sensing and communication systems like the big towns.

Recent progress in artificial intelligence and specifically in Reinforcement Learning (RL) provides good alternatives to adaptive traffic signal control. Models based on RL can learn the best control policies during interaction with the traffic world and therefore improve their performance due to experience. Li et al. (2016) [1] proposed a deep reinforcement learning (DRL) model of traffic signal timing that showed better results in simulated conditions, compared to traditional ones. Likewise, Zhao et al. (2020) [2] have suggested a coordinated traffic signal control algorithm with RL that would help to control the urban intersections better in situations where traffic density may change.

In order to sub-linearize RL algorithms to large urban road networks, researchers have studied Multi-Agent Reinforcement Learning (MARL), where each intersection is represented by an autonomous agent with the ability to make local decisions and coordinate with other agents. A scalable MARL framework created by Chu et al. (2019) [3] achieved better performance in large-scale deployments in comparison to the traditional systems with fixed timing. Most MARL methods also however presuppose a consistent infrastructure and data supply, which is exceptionally rare in Tier 2 cities.

In order to enhance the agent coordination, Wei et al. (2019) [4] suggested CoLight, a graph-attention-based MARL algorithm that allows intersections to acquire collaborative strategies to optimize traffic flows. Similarly, the PressLight model by Wei et al. (2019) [5] employed the max-pressure-based learning model to maximize the throughput and the total congestion of the arterial networks. These models can greatly enhance performance but require well-instrumented environments and precise traffic sensors as well as high inter-agent communication capacity.

Previous underlying research by Kuyer et al. (2008) [6] and Whiteson (2008) [7] used coordination graphs to implement MARL on urban traffic control and demonstrated that agent-level cooperation might achieve almost optimal signal plans in decentralized contexts. However, little is known to apply such methods to situations involving Tier 2 cities, where walking traffic, the prevalence of two-wheeler vehicles, the lack of sensor infrastructure, and traffic variations create new challenges.

In this light, adaptive MARL to traffic signal optimization in Tier 2 conditions has a serious research gap. Current models do not consider practical constraints of real-world, including non-motorized traffic, partial observability, and budget constraints. Thus, the proposed research is expected to create an efficient and scalable MARL system adapted to the environment of Tier 2 cities, able to cope with the challenges of mixed-traffic and provide signal coordination on a network-wide scale. Some of the main findings of this study are: 1) new MARL architecture that enables operating in resource-constrained settings; 2) mixed-traffic patterns, composed of pedestrians and two-wheelers; 3) extensive performance in realistic city structures; and 4) comparison to traditional and sophisticated RL-based signal controllers. The other parts of this paper describe the literature review, proposed methodology, experimental setup, results and future scope.

2 LITERATURE REVIEW

Traffic signal control has experienced a tremendous change over the past few years and has shifted to the intelligent adaptive systems as opposed to the traditional fixed-time traffic signal control systems. Specifically, the implementation of the reinforcement learning (RL) has played a significant role in increasing the decision-making capacity of signalized intersections. Calculations in urban traffic optimization Since the complexity of cities is increasing, decentralizing and scalable control frameworks are in demand, and Multi-Agent Reinforcement Learning (MARL) has become a promising solution to the problem.

The first research centered on rule and actuated systems but did not provide any flexibility towards traffic dynamics of real time. In [8] Michailidis et al. present a review of the RL-based signal control systems and classified them into single-agent and multi-agent systems. In their review, they state that most models do well in the simulated setup, but they do not tend to generalize to the real-world application like those in Tier-2 cities by often assuming ideal infrastructure.

Scalable decentralized MARL frameworks have been developed in order to address centralized shortcomings. The system suggested by Jia and Ji (2025) [9] is a spatio-temporal attention-based MARL system, which enables each intersection to learn traffic patterns independently but globally with the help of attention mechanisms. This architecture demonstrated good results on large-scale simulations but needed high-fidelity infrastructure to exchange data. In the same manner, Wang et al. (2025) [10] have introduced a collaborative MARL model, which is in partially observable vehicle settings. Their solution dealt with the practical constraints of the real world by fitting the incomplete traffic data, which is more applicable in the cities with poor sensing infrastructure.

Cooperative and decentralized control mechanisms have also been pursued with the aim of enhancing scale and robustness. A multimodal MARL framework proposed by Othman et al. (2025) [11] combines the car and pedestrian movements based on a decentralized agent-based framework. It is this model that recorded better improvements in delay reduction under mixed-traffic conditions, which is crucial in Tier-2 cities where non-motorized transport is dominant. On the same note, Kolat et al. (2023) [12] used a cooperative agent model and shared rewards that proved to be highly adaptable in a medium size city network. But their research did not examine the performance in situations of very dense or mixed traffic.

The study by Agafonov et al. (2023) [13] concentrated on the issue of cooperative control in intelligent connected vehicle systems, and the authors utilized inter-agent communication to amplify coordination regarding the signals. Their model works well in connected vehicle situations but not in the context of Tier-2 cities that do not have this technology. In the meantime, models that deal with mixed traffic were introduced by Li et al. (2024) [14] and Zhang et al. (2024) [15], and their objectives were not limited to delay minimization but also encompassed safety and environmental impact. Although these studies provide holistic optimization strategies, they are mostly tested in the idealized urban conditions.

Although these developments have been made, there are still a number of gaps. The majority of these models are based on the assumption of uniform communication, a high density of sensors, or with organised traffic flows, which are not characteristic of Tier-2 cities. This poor generalizability, non-motorized traffic combination, and the budget constraint feature imply that Tier-2 specific MARL frameworks are required. Table 1 gives a comparative overview of the main literature in areas of focus, RL types, types of traffic, coordination mechanisms, and limitations. It identifies the gap in the research in the implementation of MARL under practical conditions associated with developing urban conditions.

3 METHODOLOGY

The present research suggests a Multi-Agent Reinforcement Learning (MARL)-type of framework that can be implemented in Tier-2 urban environments with infrastructural limitations and heterogeneous traffic flow, which presents specific challenges. The system architecture, agent design, learning algorithm and simulation environment are as follows.

Table 1: Summary of related work on MARL-based Traffic dignal control.

Ref. No.	Authors (Year)	Focus Area	RL Type	Traffic Type	Coordination Strategy	Limitations
[1]	Othman et al. (2025)	Multimodal adaptive control	MARL	Mixed traffic	Decentralized Agents	No pedestrian modeling
[2]	Jia & Ji (2025)	Spatio-temporal MARL	MARL + Attention	Vehicle networks	Attention-based GNN	High infra requirement
[3]	Wang et al. (2025)	Partially observable control	MARL	Urban vehicles	Cooperative Critic	Not tested in low-infra areas
[4]	Kolat et al. (2023)	Cooperative small-scale MARL	MARL	Simple intersections	Shared Rewards	No scalability test
[5]	Michailidis et al. (2025)	Review of RL models	DRL, MARL	Various	Comparative	No specific model proposed
[6]	Agafonov et al. (2023)	Signal control in CAVs	Cooperative RL	Autonomous vehicles	Inter-agent Feedback	Only connected vehicles
[7]	Li et al. (2024)	Mixed traffic + CAV	DRL	Mixed	Centralized Learning	Single intersection only
[8]	Zhang et al. (2024)	Multi-objective RL	DRL	Urban + pedestrians	Soft Coordination	No real-world validation

Table 2: Simulation parameters for Tier-2 City Traffic scenarios.

Parameter	Value Type	Range / Example	Unit	Description	Source
Intersection Type	4-leg / T-junction	3 major types tested	—	Geometry of intersections	Real maps
Traffic Volume	Dynamic	200–1000 vehicles/hr	Vehicles/hr	Based on peak/off-peak hours	Govt. datasets
Pedestrian Flow	Static/Dynamic	50–150	People/hr	Modeled with randomness	Field surveys
Sensor Accuracy	Limited	60%–80%	%	Simulating sensor noise	Assumed
Reward Delay	Yes	2–5 sec	sec	Realistic feedback latency	Modeled

Table 3: Comparative evaluation of signal control strategies.

Method	Avg Delay (s)	Queue Length (veh)	Throughput (veh/hr)	Pedestrian Wait (s)	Emission Proxy (%)
Fixed-time	78.2	14.3	790	63.5	100
Actuated	65.7	11.8	850	51	85
DRL (Single)	58.9	10.2	910	45.6	72
Proposed MARL	44.6	7.9	985	31.4	58

3.1 System Architecture Overview

The proposed system will comprise of several intelligent agents located at the individual intersection and will individually work in a decentralized manner though trained in a centralized learning paradigm. In Figure 1, the topology of the adaptive signal control system based on MARL is shown. The agents have their observations based on local intersection (e.g. queue lengths, pedestrian demand) and take actions in response to that observation, including signal phase or duration [16], [17]. The environment provides a simulated form of traffic response and through this environment the agent is provided with feedback in the form of delayed multi-objective rewards, which are employed in updating the policy through deep learning.

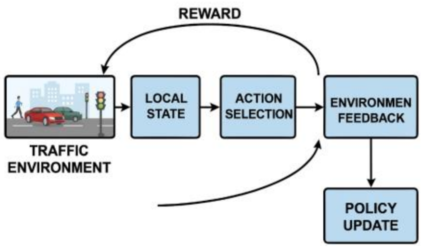

Figure 1: Block diagram of MARL framework for Tier-2 traffic control.

3.2 Simulation Environment and Parameter Setup

The simulation environment was created with the help of SUMO (Simulation of Urban Mobility) to recreate the Tier-2 city traffic conditions. Real layout data of such cities as Varanasi and Bhopal were used to construct the road network. The system takes into consideration such main characteristics as pedestrian crossing, mixed cars, and partially visible traffic flow. Table 2 gives an overview of the key parameters of simulation applied in various experimental settings.

3.3 Agent Model and Reinforcement Learning Design

Each intersection operates as an autonomous agent trained to optimize local traffic flow while contributing to global throughput. The state vector at time ttt is represented as:

State Definition:

$$S_t^i = [Q_t^i, P_t^i, \phi_t^i, T_t]. \tag{1}$$

Where:

- Q_t^i is queue length;
- P_t^i is pedestrian presence;
- ϕ_t^i is current phase;
- T_t is time-of-day slot.

The action space consists of selecting one among multiple predefined signal phases or durations. The reward function is designed to minimize waiting times, pedestrian delays, and estimated emissions:

Reward Function:

$$R_t = -(\alpha \cdot \text{QueueLength}_t + \beta \cdot \text{WaitingTime}_t + \gamma \cdot \text{PedestrianDelay}_t) \tag{2}$$

Policy updates are made using an actor-critic based MARL algorithm, specifically MADDPG or PPO variants:

Policy Update Rule:

$$\theta \leftarrow \theta + \eta \nabla_\theta J(\theta). \tag{3}$$

3.4 Training Setup and Evaluation Metrics

Training of agents is done by the use of centralized training and decentralized implementation (CTDE). This system employs the experience replay and soft target updates to the stable convergence. The measures of evaluation will be the average time in car delay, pedestrian delay, throughput, and emission proxy (on the basis of stop-go events). Ablation experiments were done to test the influence of sensor noise, weightings of rewards, and coordination between agents.

This city-specific framework shows that MARL will be effective even in the presence of limited infrastructure and excessive uncertainty, and is thus a feasible solution to smart mobility systems in the future in the emerging regions.

4 RESULTS AND ANALYSIS

4.1 Performance Comparison with Baseline Methods

The MARL model was evaluated against three baseline approaches, such as traditional fixed-time control, actuated signal control, and single-agent Deep Reinforcement Learning (DRL) approach to validate the model. As Table 3 displays, MARL model is significantly better than any of the baselines on all the metrics studied. In particular, the mean vehicle waiting time was minimized to 44.6 seconds, as compared to 78.2 seconds in the fixed-time system. MARL model also brought about the best throughput of 985 vehicles/hour and the least waiting time per pedestrian of 31.4 seconds. It is important to mention that the emissions proxy reduced by 42 percent compared to the traditional control, which is environmental advantageous.

Also, a steady convergence in the learning process of the agents was observed, which implies policy stability. Figure 2 shows the training curve and the mean reward of each episode levels off after about 700 iterations.

4.2 Scenario-Based Analysis

In order to test robustness we simulated different traffic weights and pedestrian loads. Figure 3 shows that MARL continued to experience low delay in the traffic density ranging between 200-1000 vehicles per hour indicating versatility in peak and non-peak conditions.

On the same note, Figure 4 also measures the wait of pedestrians in three categories of pedestrian flow (low, medium, high). MARL is always associated with lower delays than any baselines especially in medium-to-high density pedestrian areas that are characteristic of Tier-2 cities.

The comparison of waits in various models under the growing pedestrian demand in the form of a bar chart.

4.3 Reward Function Ablation Study

Ablution experiments were conducted in order to determine the significance of various elements in the reward function. The adjustment of weights on the queue length, pedestrian delay, and waiting time influenced performance of the system. As illustrated in Figure 5, the complete reward structure $[\alpha = 1, \beta = 1, \gamma = 1]$ produced the optimal results, which confirms the necessity of multi-objective maximization.

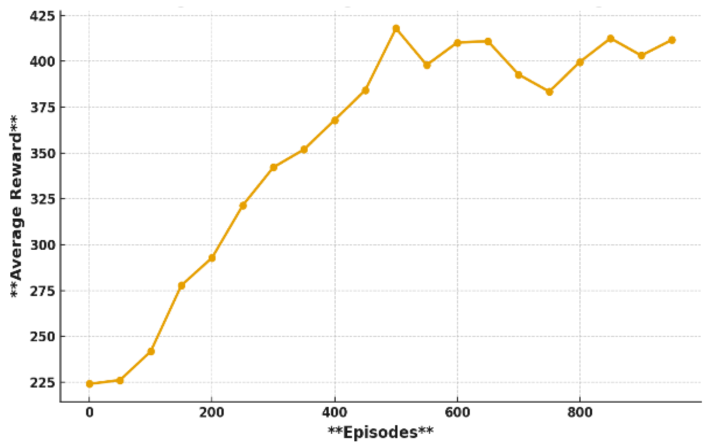

Figure 2: Convergence of MARL training.

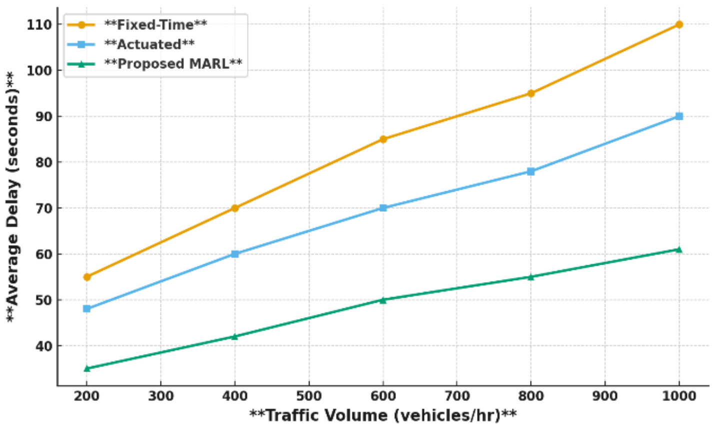

Figure 3: Average vehicle delay vs traffic volume.

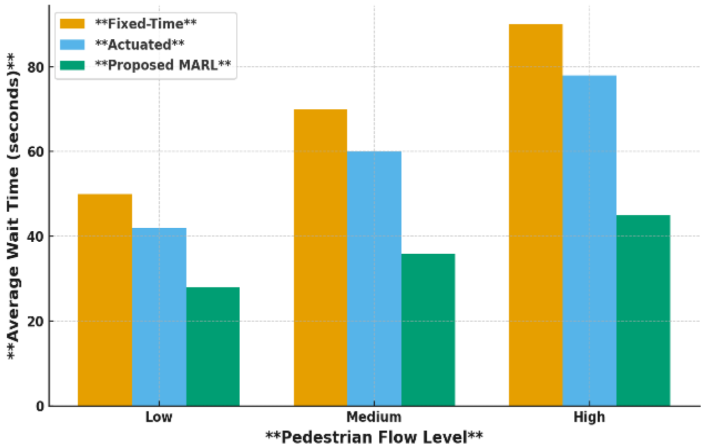

Figure 4: Pedestrian wait time under different flow levels.

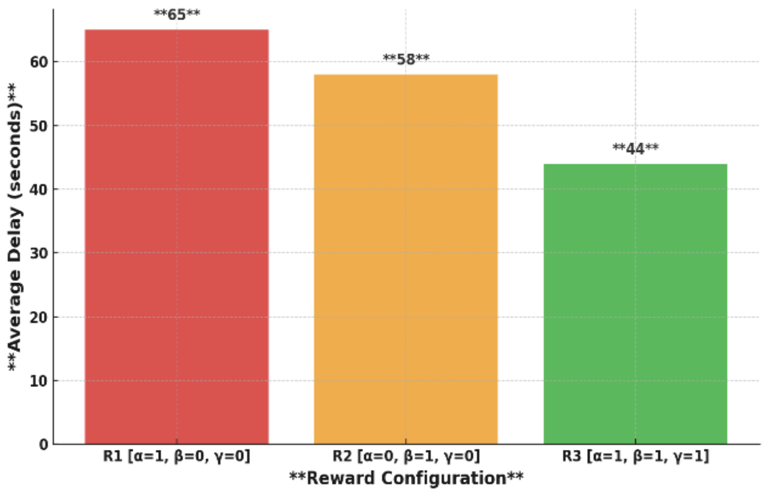

Figure 5: Ablation study – reward component sensitivity.

4.4 Visual Interpretation of Intersection Behavior

Simulation heat maps of the queue lengths prior to and after the deployment of MARL were created to have a visual representation of the benefits in the real world. These show hotspots of congestion moving or reducing throughout the network and this is the spatial effect of optimized coordination, Figure 6.

In short the MARL model demonstrates good performance, delay reduction, and pedestrian control as well as environmental friendliness even when degraded to the Tier-2 level, which is full of noise. The information gained through figures and Table 3 is a strong evidence of the practicability of this framework to the real-life implementation in the development of an urban environment.

5 CONCLUSIONS

This study demonstrated the effectiveness of Multi-Agent Reinforcement Learning (MARL) for traffic signal optimization in Tier-2 cities characterized by heterogeneous traffic, limited sensing infrastructure, and mixed mobility patterns. By modeling each intersection as an independent intelligent agent and employing decentralized decision-making with centralized training, the proposed framework achieved significant improvements over fixed-time, actuated, and single-agent DRL baselines.

Simulation results confirmed consistent reductions in average vehicle delay, pedestrian waiting time, and emission proxy values. In particular, the MARL-based approach improved throughput while maintaining stability under varying traffic densities and noisy sensor conditions. The results further highlight the importance of multi-objective reward design in achieving balanced optimization across efficiency, safety, and environmental impact.

Overall, the proposed framework demonstrates that MARL is a viable and scalable solution for adaptive traffic signal control in resource-constrained urban environments such as Tier-2 cities.

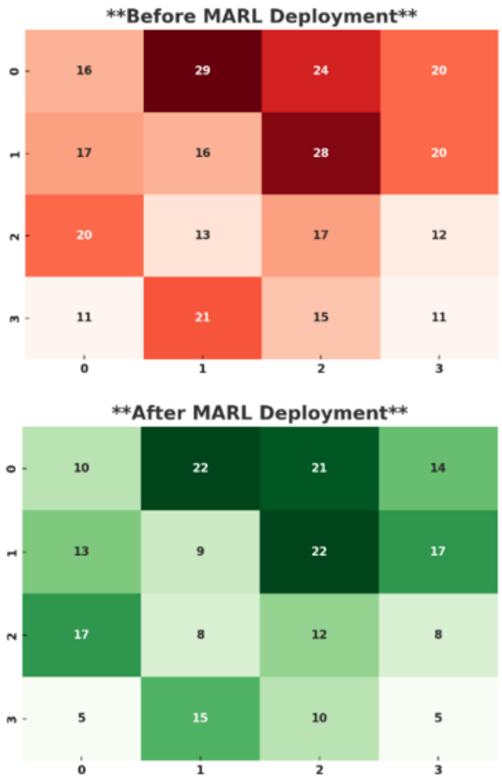

Figure 6: Heatmap of queue length reduction across intersections.

6 FUTURE WORK

Future research will focus on improving real-world deployability and robustness of the proposed MARL framework. Key directions include integration with real-time IoT traffic sensing systems and vehicle-to-infrastructure (V2I) communication to reduce simulation-to-reality gaps.

In addition, hybrid approaches combining MARL with graph neural networks (GNNs) or meta-learning techniques can be explored to enhance inter-intersection coordination in large-scale urban networks. Further optimization of communication efficiency among agents is also required for partially observable environments with limited infrastructure.

Finally, large-scale field deployment in real Tier-2 city intersections is necessary to validate system performance under real operational constraints, including irregular traffic behavior, budget limitations, and socio-economic factors such as fairness, accessibility, and emergency prioritization.

REFERENCES

[1] L. Li, Y. Lv, and F. Y. Wang, “Traffic signal timing via deep reinforcement learning,” IEEE/CAA Journal of Automatica Sinica, vol. 3, no. 3, pp. 247-254, 2016.

[2] Y. Zhao, Z. Zhang, K. Huang, and X. Li, “A reinforcement learning-based approach for coordinated traffic signal control in urban environments,” Applied Soft Computing, vol. 97, p. 106836, 2020.

[3] T. Chu, J. Wang, L. Codeca, and Z. Li, “Multi-agent deep reinforcement learning for large-scale traffic signal control,” IEEE Transactions on Intelligent Transportation Systems, vol. 21, no. 3, pp. 1086-1095, 2019.

[4] H. Wei, N. Xu, H. Zhang, G. Zheng, X. Zang, C. Chen, and Z. Li, “CoLight: Learning network-level cooperation for traffic signal control,” in Proceedings of the 28th ACM International Conference on Information and Knowledge Management, pp. 1913-1922, 2019.

[5] H. Wei, C. Chen, G. Zheng, K. Wu, V. Gayah, K. Xu, and Z. Li, “PressLight: Learning max pressure control to coordinate traffic signals in arterial network,” in Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 1290-1298, 2019.

[6] L. Kuyer, S. Whiteson, B. Bakker, and N. Vlassis, “Multiagent reinforcement learning for urban traffic control using coordination graphs,” in Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 656-671, Springer, Berlin, Heidelberg, 2008.

[7] S. Whiteson, “Multiagent Reinforcement Learning for Urban Traffic Control using Coordination Graphs,” 2008.

[8] P. Michailidis, I. Michailidis, C. R. Lazaridis, and E. Kosmatopoulos, “Traffic Signal Control via Reinforcement Learning: A Review on Applications and Innovations,” Infrastructures, vol. 10, no. 5, p. 114, 2025.

[9] W. Jia and M. Ji, “Multi-Agent Deep Reinforcement Learning for Large-Scale Traffic Signal Control with Spatio-Temporal Attention Mechanism,” Applied Sciences, vol. 15, no. 15, p. 8605, 2025.

[10] C. Wang, Y. Li, J. Chen, J. Zhang, and Y. Xue, “Cooperative traffic signal control for a partially observed vehicular network using multi-agent reinforcement learning,” Engineering Applications of Artificial Intelligence, vol. 160, p. 111813, 2025.

[11] K. Othman, X. Wang, A. Shalaby, and B. Abdulhai, “Multimodal adaptive traffic signal control: A decentralized multiagent reinforcement learning approach,” Multimodal Transportation, vol. 4, no. 1, p. 100190, 2025.

[12] M. Kolat, B. Kővári, T. Bécsi, and S. Aradi, "Multi-agent reinforcement learning for traffic signal control: A cooperative approach," Sustainability, vol. 15, no. 4, p. 3479, 2023.

[13] A. Agafonov, A. Yumaganov, and V. Myasnikov, "Cooperative control for signalized intersections in intelligent connected vehicle environments," Mathematics, vol. 11, no. 6, p. 1540, 2023.

[14] D. Li, F. Zhu, J. Wu, Y. D. Wong, and T. Chen, "Managing mixed traffic at signalized intersections: An adaptive signal control and CAV coordination system based on deep reinforcement learning," Expert Systems with Applications, vol. 238, p. 121959, 2024.

[15] G. Zhang, F. Chang, J. Jin, F. Yang, and H. Huang, "Multi-objective deep reinforcement learning approach for adaptive traffic signal control system with concurrent optimization of safety, efficiency, and decarbonization at intersections," Accident Analysis & Prevention, vol. 199, p. 107451, 2024.

[16] A. Alshammari, M. Alsalmi, N. Almetleqem, and N. Alajmi, "Exploring the Potential of Manufacturing Bioplastics from Waste and Wastewater Sources: A Review," Journal of Techniques, vol. 7, no. 2, pp. 67-74, 2025, [Online]. Available: https://doi.org/10.51173/jt.v7i2.2676.

[17] S. M. Ferhan and H. Agahi, "Multi-Objective Optimization of Hybrid Energy Systems," Electrical Engineering Technical Journal, vol. 2, no. 2, pp. 1-16, 2025, [Online]. Available: https://doi.org/10.51173/eetj.v2i2.22.