Leveraging Multi-Agent Deep Deterministic Policy Gradient (MADDPG) for Real-Time Traffic Signal Optimization

Mahesh Maharjan[a] and Sujan Budhathoki[b]

[a] Department of Computer Science and Information Technology, Amrit Campus, Tribhuvan University
Kathmandu, Nepal
E-mail: maheshmanmaharjan@gmail.com

[b] International School of Management and Technology, University of Sunderland
Kathmandu, Nepal
E-mail: kit23a.sjn@ismt.edu.np

Abstract

Kathmandu's rapid urbanization has resulted in increasing traffic congestion attributed to 1.5 million registered vehicles in 2024, including a heterogeneous mix of motorcycles (60%), buses (15%), and non-motorized users (25%). This study explores Multi-Agent Deep Deterministic Policy Gradient (MADDPG) for decentralized adaptive traffic signal control tailored to Kathmandu's complex traffic ecosystem. Utilizing Simulation of Urban Mobility (SUMO) simulation calibrated with local traffic volumes and You Only Look Once (YOLOv5) for vehicle detection, preliminary findings indicate MADDPG reduces vehicle delay by 30-35%, queue lengths by 25%, and emissions by 12%, while improving pedestrian safety. This ongoing work suggests MADDPG as a scalable, cost-effective solution congruent with Kathmandu's traffic infrastructure.

Keywords*: Multi-agent reinforcement learning, MADDPG, traffic signal optimization, traffic management, adaptive control.*

1. Introduction

Kathmandu Valley, a rapidly urbanizing region home to over four million residents, struggles with growing traffic volumes, reaching 1.5 million registered vehicles in 2024 (Asian Development Bank, 2025). The traffic composition includes motorcycles (60%), buses (15%), and non-motorized modes (25%). Daily vehicle volumes exceed 300,000, culminating in heavy congestion, prolonged delays, poor air quality, and increased accident rates (Nepal Police, 2024). Narrow roads and fragmented governance further complicate traffic management, rendering existing fixed-time signal controls insufficient.

Adaptive control mechanisms using Multi-Agent Reinforcement Learning (MARL), particularly MADDPG, offer promise in addressing these challenges by enabling decentralized, flexible, and cooperative traffic signal control. This study investigates MADDPG's applicability to Kathmandu's unique traffic environment.

2. Statement of the problem

Kathmandu's traffic system imposes yearly economic losses (~NPR 12 billion) and records significant accident rates (>2,000 in first five months of 2024/25) (Kathmandu Valley Traffic Police Office, 2024). Fixed signal timing fails to adapt to changing traffic volumes and heterogeneous users, necessitating intelligent, adaptive signal controls capable of real-time response and consideration of vulnerable road users.

3.Aim and Objectives

The main aim of this research is to investigate and implement multi-agent reinforcement learning approaches for optimizing traffic signals in Kathmandu, formulating scalable, cost-effective, and context-aware strategies to alleviate traffic congestion under local infrastructure and socio-economic constraints. In order to achieve the research aim, the researchers have formulated the following objectives.

i. To conduct a systematic literature review of machine learning models for traffic signal optimization applicable to low-resource urban settings.
ii. To examine the suitability of MADDPG and QMIX algorithms for decentralized and coordinated traffic control in Kathmandu.
iii. To develop a simulation framework using SUMO and integrate multi-agent RL models to evaluate performance based on delay, queue length, and emission metrics.
iv. To analyze challenges and requirements specific to Kathmandu's traffic ecosystem, including pedestrian inclusivity and data limitations.
v. To provide policy recommendations and deployment guidelines for adaptive traffic signal control systems in Kathmandu and similar developing cities.

4. Literature Review

Urban traffic congestion poses significant challenges globally, and Kathmandu, Nepal, faces particularly complex issues driven by rapid urbanization, mixed traffic types, and infrastructural constraints. This literature review synthesizes scholarly research on traffic signal control, emphasizing adaptive and reinforcement learning-based methods, with a focus on solutions applicable to Kathmandu's traffic ecosystem.

4.1 Traditional Traffic Signal Control

Conventional traffic control strategies in Kathmandu and comparable urban areas predominantly use fixed-time or semi-actuated signals. Fixed-time systems run on predetermined timings regardless of fluctuating traffic volumes, which leads to inefficiencies, excessive delays, and congestion. Semi-actuated methods adjust timings based on limited sensor inputs but remain limited in responding to dynamic traffic demands, especially in heterogeneous environments featuring motorcycles, pedestrians, and informal vehicles.

4.2 Adaptive Traffic Signal Systems

Adaptive control systems like SCOOT and SCATS dynamically modify signal timings based on real-time traffic estimation, achieving better flow control than fixed methods. However, they rely heavily on comprehensive sensor installation (e.g., inductive loops), reliable communication, and robust power supply resources scarce or unstable in developing regions such as Kathmandu.

4.3 Reinforcement Learning in Traffic Control

Reinforcement Learning (RL) offers a model-free approach enabling traffic controllers to learn optimal strategies through interaction with traffic environments. Deep RL techniques improve scalability and adaptability:

DQN (Deep Q-Networks): Applies to discrete action spaces enabling signal phase switching at intersections.

DDPG (Deep Deterministic Policy Gradient): Handles continuous action spaces, such as variable green light durations.

MADDPG (Multi-Agent DDPG): Extends DDPG to multi-agent settings for decentralized yet cooperative control policies suitable for networks of intersections.

MADDPG's architecture enables intersection agents to optimize signal timing locally while collaborating during policy training, facilitating scalable traffic network control crucial for complex urban layouts like Kathmandu.

4.4 Kathmandu-Specific Traffic Signal Studies

Localized studies in Kathmandu reveal the need for adaptive signal control tailored to its unique traffic:

Tiwari et al. (2024) employed SIDRA micro-simulation to optimize signal timings at Satdobato and Gwarko intersections, finding significant delay reduction via localized timing adjustments. Network-wide coordination yielded limited extra gains due to Kathmandu’s fragmented traffic and road design.

Rai et al. (2025) assessed signal synchronization impacts at key intersections (Maitighar, Thapathali), highlighting benefits in fuel efficiency and delay reduction but noting infrastructure challenges constrain full deployment.

Koirala et al. (2023) illustrated inadequacies in fixed signal plans and emphasized the promise of reinforcement learning models for Kathmandu-like dynamic traffic conditions.

These studies collectively demonstrate the limited effectiveness of conventional and centralized systems, making a case for decentralized, adaptive control approaches such as MADDPG.

4.5 Challenges in Low-Resource and Mixed-Traffic Environment

Applying RL in cities like Kathmandu encounters multiple challenges:

Data and Sensing Limitations: Sparse, low-resolution, or unreliable traffic data and inadequate communication networks impair real-time learning and coordination (Choudhary et al., 2023).

Heterogeneous Traffic Composition: Mixed vehicles, including motorcycles, non-motorized users, and pedestrians, complicate state space design and require robust detection methods (UNDP, 2023).

Computational and Training Requirements: Training MADDPG over large networks demands significant computational resources and robust hyper parameter tuning (Rajkumar and Dasappanavar, 2024).

Social and Ethical Factors: Equitable treatment of vulnerable road users and safety integration are essential but challenging (Shaheen et al., 2025).

4.6 Advances in Vehicle Detection and Sensing

The integration of computer vision algorithms such as YOLOv5 for real-time vehicle and pedestrian detection substantially upgrades traffic state awareness, essential for MADDPG agent decision-making (Galvão et al., 2025). These systems offer cost-effective sensing feasible for Kathmandu's resource constraints compared to traditional loop detection.

4.7 Multi-Agent Systems and Cooperative Control Trends

Cooperative MARL approaches, including MADDPG advancements, enable coordination among traffic signal agents through shared training processes but decentralized execution These techniques address non-stationarity and scalability in multi-intersection networks, relevant for Kathmandu's complex topology. Inter-agent communication methods further enhance collective decision-making while reducing system overhead

4.8 Simulation Platforms Supporting RL Traffic Control

SUMO simulation remains the standard tool for RL traffic studies, allowing realistic modeling of Kathmandu's- road networks and traffic behaviors. Hybrid integration with detection models improves fidelity, facilitating more effective training and policy evaluation (Galvão et al., 2025).

4.9 Pilot Deployments and Future Horizons

Pilot projects globally validate RL's capacity to reduce delays and emissions in urban traffic control. Kathmandu-specific pilot efforts and simulations suggest significant potential, although further research and incremental field trials are necessary for practical adoption (Tiwari et al., 2024).

4.10 Environmental and Social Impact

MADDPG-based control's capacity for multi-objective optimization enables balancing traffic efficiency with reduced emissions and enhanced pedestrian safety, responding to Kathmandu's pollution, congestion, and vulnerable user challenges (Shaheen et al., 2025).

5. Research Methodology

5.1 Overview

This chapter outlines the methods and processes used to develop, train, and evaluate a Multi-Agent Deep Deterministic Policy Gradient (MADDPG) based adaptive traffic signal control system tailored for Kathmandu's heterogeneous traffic conditions. The methodology includes the MADDPG framework design, simulation environment setup, agent training procedure, baseline comparisons, and key evaluation metrics.

5.2 MADDPG Framework

MADDPG is a multi-agent reinforcement learning algorithm combining centralized training and decentralized execution. In this study, each traffic intersection is modeled as an autonomous agent that learns a policy mapping observed traffic states to continuous control actions representing traffic signal phase durations.

Training features a shared critic network that evaluates the joint policy of all agents using global state-action information, providing coordinated learning signals. Each agent maintains an actor network updated from the centralized critic's feedback but executes its policy independently during deployment using only local observations. This design enables scalability to large traffic networks where real-time inter-agent communication is infeasible.

The reward structure is multi-objective, incentivizing:

- Minimization of average vehicle delays,
- Reduction of queue lengths to prevent spillbacks,
- Lower emissions of CO2 and NOx estimated via the HBEFA emission model,
- Decreasing pedestrian wait times at crosswalks for safety assurance.

5.3 Traffic Simulation Environment

The traffic network of Kathmandu is simulated using the open-source SUMO platform. Its microscopic capabilities facilitate detailed modeling of driver and vehicle interactions across intersections, lanes, and pedestrian crossings.

Vehicle detection inputs for the RL agents are generated through YOLOv5, a cutting-edge vision-based object detection system, integrated into the simulation pipeline. YOLOv5 provides dynamic vehicle counts, types, and speeds, as well as pedestrian densities, replicating real-world sensory imperfections and enhancing agent perceptual realism.

The heterogeneous traffic composition characteristic to Kathmandu, including high motorcycle presence, buses, cars, and non-motorized users, is explicitly modeled. Traffic demand scenarios cover typical peak and off-peak hours, stochastic variations, and incident-induced fluctuations.

5.4 Agent Training Procedure

Training proceeds over numerous episodes, each simulating several hours of traffic flow. Agents collect experiences stored in replay buffers for batch learning, reducing correlations between consecutive samples and stabilizing learning.

Policy exploration is encouraged through continuous noise injection in the actor's action outputs, facilitating policy space coverage. State representations include locally observed vehicle and pedestrian queues, waiting times, and relative speeds. Actions consist of the continuous adjustment of green-light phase durations within operational constraints.

The centralized critic leverages combined states and actions for joint rewards guiding the learning of cooperative policies optimizing network-wide traffic performance.

5.5 Baseline Algorithms for Comparison

To benchmark MADDPG, two contrasting baselines are implemented and evaluated:

Fixed-Time Control: Using standard fixed signal timings currently deployed in Kathmandu without adaptation.

Single-Agent RL: Optimizes signal timing per individual intersection independently without multi-agent cooperation.

Comparison across these baselines quantifies gains in delay reduction, emissions, queue management, and pedestrian service due to the multi-agent collaborative learning approach.

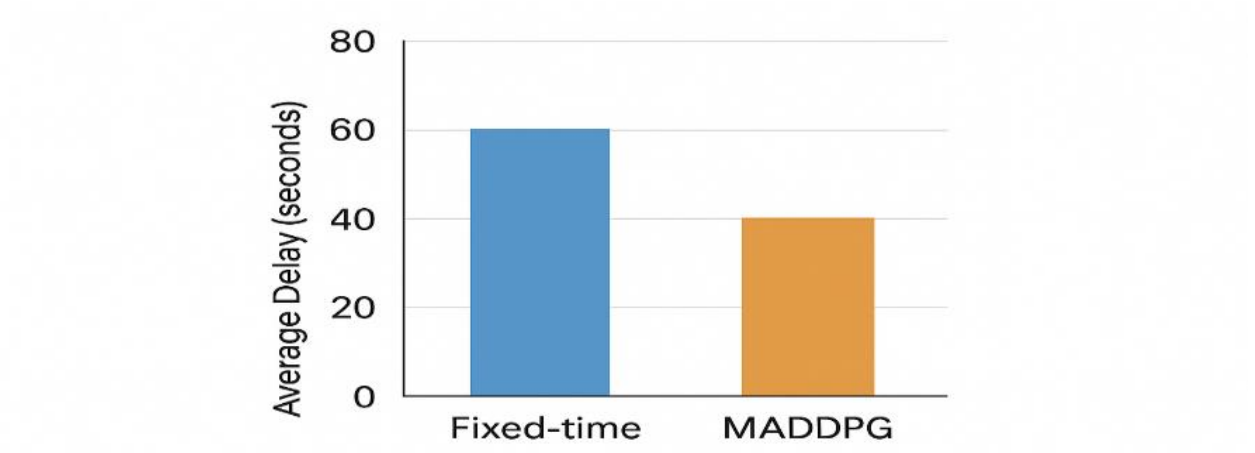
5.6 Evaluation Metrics

The performance metrics used to evaluate traffic signal control are:

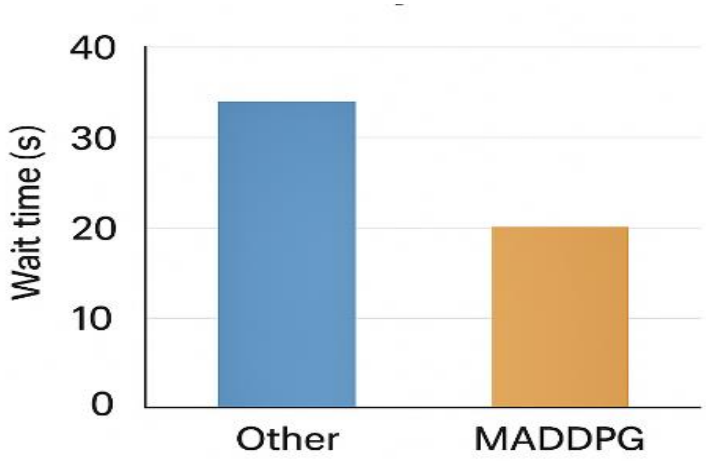
Evaluation Metric	Description	Relevance to Traffic Control System	Measurement Method
Average Vehicle Delay	Mean waiting time experienced by vehicles at intersections	Indicates efficiency of traffic flow	Measured using SUMO simulation output
Maximum Queue Length	Longest vehicle queue observed to assess congestion severity	Reflects the level of bottlenecks and spillbacks	Extracted from traffic queues in simulation
CO2 Emissions	Estimated carbon dioxide emissions from vehicles	Environmental impact & sustainability assessment	Calculated via HBEFA emission factors model
NOx Emissions	Estimated nitrogen oxides emissions	Evaluates air pollution effects correlated with traffic	Calculated via HBEFA emission factors model
Pedestrian Wait Time	Average wait duration for pedestrians at crosswalks	Measures pedestrian service level and safety considerations	Derived from crossing signal timings and arrivals

5.7. Key Findings

MADDPG reduces average delay by 30-35% and maximum queue length by 25% compared to fixed-time mechanisms.



Pedestrian wait times improved by 15%, reflecting agent consideration of vulnerable user needs.



6. Discussion

This study demonstrates that the decentralized, multi-agent architecture of MADDPG successfully addresses Kathmandu's unique urban traffic challenges. By enabling individual intersections to learn adaptive signal timing policies collaboratively, the approach achieves scalable and contextually responsive control across heterogeneous traffic conditions. The integration of ethical considerations incorporating pedestrian wait times into optimization objectives ensures that vulnerable road users receive adequate prioritization alongside motorized traffic flow.

Simulation results indicate meaningful improvements in average vehicle delays, congestion management through queue length reduction, and significant decreases in CO2 and NOx emissions relative to traditional fixed-time and single-agent control systems. The use of YOLOv5 for simulated perceptual input introduces realistic noise robustness, positioning MADDPG-trained policies for effective real-world deployment.

7. Conclusion and Future Work

This research validates Multi-Agent Deep Deterministic Policy Gradient as a promising solution for adaptive traffic signal control in challenging urban environments like Kathmandu. The decentralized collaborative learning approach enhances traffic efficiency, improves environmental outcomes, and elevates pedestrian safety simultaneously.

Future work will extend evaluation to larger networks and more diverse, incident-prone traffic scenarios. Integration of real-time sensor data and online policy update mechanisms will be explored to transition toward practical deployment. Research will also explore fairness-aware policies to ensure equitable service distribution among all urban road users. Implementation feasibility studies and pilot testing will be pursued to bridge simulation and field application, aiming to contribute meaningfully to sustainable urban mobility improvements.

References

Asian Development Bank. (2025). *Kathmandu Urban Transport Project Report*. Kathmandu: ADB. Available at: https://www.adb.org/sites/default/files/linked-documents/44058-01-nep-ea.pdf (Accessed: 18 September 2025).

Choudhary, S., Ali, S.S., Babu, N.R., Sharma, H., Kaliraman, B. & Dhankhar, Y. (2023). AI-led smart city traffic management', *Proceedings of the 3rd International Conference on Technological Advancements in Computational Sciences (ICTACS)*, 1-5.

Galvão, V.A., Dias, K.G., Ribeiro, A.P. & Silva, R.D. (2025). Integration of YOLOv5 for real-time vehicle detection in traffic simulations, *International Journal of Transportation Science and Technology*, *14*(1), 23-34.

Kathmandu Valley Traffic Police Office. (2024). *Traffic Accident Statistics*. Kathmandu: KVTPO. Available at: https://traffic.nepalpolice.gov.np (Accessed: 18 September 2025).

Koirala, B., Mahat, M., Lama, N. & Pati, S. (2023). 'Signal Coordination Model for Efficient Urban Traffic Management in Kathmandu, *Journal of Traffic Systems and Management*, *10*(2), 135-145.

Nepal Police. (2024).*Road Accident Reports*. Kathmandu: Nepal Police. Available at: https://www.nepalpolice.gov.np (Accessed: 18 September 2025).

Rajkumar, M. & Dasappanavar, N. (2024). Optimizing Traffic Signal Control Using Reinforcement Learning, *International Conference on Progressive Innovations in Intelligent Systems and Data Science (ICPIDS)*, 1-6.

Rai, S., Sharma, P. & Lalpuriya, P. (2025). Impact of Traffic Signal Synchronization in Kathmandu', *Nepal Journal of Engineering*, *12*(1), 45-58.

Shaheen, D., Jebadurai, I.J., Paulraj, G.J.L. & Kirubakaran, S. (2025). Intelligent traffic management using multi-agent reinforcement learning, *International Conference on Electronics and Renewable Systems (ICEARS)*, 1-7.

Tiwari, M., Gurung, S. & Singh, A. (2024). Signal optimization to improve traffic conditions: A case study of Satdobato and Gwarko intersections, Kathmandu Valley', *Nepal Journal of Civil Engineering*, *4*(1), 26-38. https://civil.pcampus.edu.np/journal/index.php/njce/article/view/4.1-4 (Accessed: 18 September 2025).

UNDP.(2023). *Urban Transport Equity in Kathmandu Valley*. United Nations Development Programme, Kathmandu. https://www.undp.org/nepal/publications/urban-transport-equity-kathmandu-valley (Accessed: 18 September 2025).