

The Optimization of Traffic Flow and Resource Allocation in Urban Infrastructures Using Predictive Analytics and Multi Agent Systems

¹Shaymaa H Nowfal, ²Revathi S, ³Kolluru Suresh Babu, ⁴Amarendra K,

⁵Abdul Muthalief Mohamed Anwar and ^{6,7}Aseel Smerat

¹Medical Physics Department, College of Science, University of Warith Al-Anbiyaa, Karbala, 56001, Iraq.

²Department of Computer Science and Applications, Faculty of Science and Humanities, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, Tamil Nadu, India.

³Department of Computer Science and Engineering, Vasireddy Venkatadri Institute of Technology, Guntur, Andhra Pradesh, India.

⁴Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Andhra Pradesh, India.

⁵Department of Computer Science and Engineering, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Avadi, Chennai, Tamil Nadu, India.

⁶Faculty of Educational Sciences, Al-Ahliyya Amman University, Amman, 19328, Jordan.

⁷Department of Biosciences, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamil Nadu, India.

¹shaymaa@uowa.edu.iq, ²revathiphd2025@gmail.com, ³kollurusuresh@gmail.com, ⁴amarendra@kluniversity.in, ⁵mohdsait69@gmail.com, ^{6,7}a.smerat@ammanu.edu.jo

Correspondence should be addressed to Revathi S : revathiphd2025@gmail.com

Article Info

Journal of Machine and Computing (<https://anapub.co.ke/journals/jmc/jmc.html>)

Doi : <https://doi.org/10.53759/7669/jmc202606008>

Received 29 April 2025; Revised from 16 August 2025; Accepted 03 October 2025.

Available online 15 October 2025.

©2026 The Authors. Published by AnaPub Publications.

This is an open access article under the CC BY-NC-ND license. (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Abstract – This research work presents a novel Multi-Agent Reinforcement Learning (MARL) model for optimising urban traffic flow (UTF) and resource allocation. The proposed MARL integrates specialised agents for traffic prediction, intersection control, resource allocation, and incident detection, all of which are coordinated through a centralised command structure. This work formulates the traffic management problem as a dual-objective optimisation task, simultaneously minimising congestion and optimising resource allocation. The model employs Proximal Policy Optimization (PPO) for training agents, enabling efficient real-time decision-making and adaptation to dynamic traffic conditions. This MARL's performance was evaluated using the Simulation of Urban Mobility (SUMO) traffic simulator, which features a network comprising 25 road segments and 8 resource allocation regions. Simulation results validate significant improvements over traditional methods, including a 47.4% reduction in average travel time, a 40.2% increase in network throughput, and a 35.7% improvement in resource utilisation. During peak hours, the model achieved a 51.2% reduction in queue lengths and maintained a 92.4% resource utilisation rate under increased demand. The Incident Detection System (IDS) proved 94.5% accuracy with an average response time of 4.3 minutes, significantly outperforming baseline approaches. The empirical results of this study indicate that the proposed MARL provides a robust and scalable solution for UTF, effectively balancing traffic flow optimisation with resource allocation efficiency. The system's ability to maintain performance under varying traffic conditions and incident scenarios recommends its viability for real-world implementation in smart city traffic management systems.

Keywords – Multi-Agent Reinforcement Learning Model, Resource Utilization, Traffic Management Systems, Traffic Flow, Accuracy, Incident Detection System, Proximal Policy Optimization.

I. INTRODUCTION

Traffic Congestion (TC) and inefficient Resource Allocation (RA) in urban environments pose significant challenges for modern cities, resulting in increased travel times, environmental pollution, and economic losses [1-2]. Traditional Traffic Management Systems (TMS), operating on fixed schedules or simple reactive mechanisms, often fail to adapt to dynamic traffic patterns and unexpected events [3-6]. The integration of Artificial Intelligence (AI), particularly Multi-Agent

Systems (MAS) and predictive analytics, proposes promising solutions for optimising real-time Urban Traffic Flow (TF) and Resource Allocation (RA) [7-10]. Urban TMS faces increasingly complex challenges due to the rise of urbanisation and vehicle ownership [11]. Recent studies indicate that TC costs major economies 2-3% of their Gross Domestic Product (GDP) annually through wasted time, fuel consumption, and environmental impact [12]. Traditional methods for TMS, such as Fixed-Time Controls (FTC) and Static-Time Control (STC) of RA, have proven inadequate for handling dynamic traffic patterns and unexpected disruptions [13-14]. This limitation has driven research interest toward more adaptive and intelligent solutions that can respond to real-time traffic conditions [15-16]. The emergence of connected setups and advanced sensing technologies has created opportunities for more sophisticated TMS [17]. MAS, combined with predictive analytics, proposes a promising model for handling the distributed nature of traffic control while maintaining system-wide coordination [18].

Previous research in urban TMS has explored numerous methods to address these challenges. Traditional traffic control methods, including FTC and STC, have been extensively studied, signifying that these systems result in longer travel times compared to adaptive methods during peak hours. Recent advances in Artificial Intelligence (AI) have led to significant developments in traffic management, with the implementation of Deep Learning (DL) for traffic prediction and the application of Reinforcement Learning (RL) for signal control optimisation. In the domain of proposed distributed control models, while focusing on agent cooperation mechanisms for TMS. Despite these advances, several critical gaps remain in current research. Existing solutions frequently treat TF optimisation and RA as separate problems, resulting in suboptimal system performance. Current models challenge to control multiple competing objectives in real-time decision-making, particularly during unexpected events. Also, the scalability of existing solutions to changing network sizes and conditions remains a significant challenge in practical implementations [19-20].

This paper addresses these limitations by developing a comprehensive MAS that jointly optimizes TF and RA in urban TF environments. The proposed model integrates traffic prediction, intersection control, RA, and incident detection through a novel Multi-Agent Reinforcement Learning (MARL). This study introduces an Adaptive Optimization Algorithm (AOA) that effectively balances competing objectives in real-time while maintaining system stability during unpredictable trials. The model's performance is evaluated using realistic urban TF scenarios, signifying significant improvements over traditional methods and existing adaptive methods.

The Main Contributions of This Work Include

- A novel MARL that holistically addresses urban TMS,
- An adaptive optimization algorithm for real-time decision-making and comprehensive empirical evidence of performance improvements.

The model achieves significant reductions in travel time, improved Resource Utilization (RU), and enhanced system stability compared to existing methods.

The remainder of this paper is organized as follows. Section 2 presents the methodology, including the model and mathematical formulation of the MARL. Section 3 describes the simulation setup and implementation details. Section 4 presents experimental results and analysis, signifying the model's effectiveness across several performance metrics. Section 5 concludes the paper and recommends directions for future research.

II. METHODOLOGY

System Model

As depicted in **Fig 1**, this paper considers an urban TMS where multiple specialized agents collaborate under a centralized MAS to optimize TF and RA. A Central Command Agent coordinates the agents (C_{cmd}) and operate across a networked environment with real-time data exchanges. The agents involved include a Traffic Prediction Agent (A_{tp}), Intersection Control Agent (A_{ic}), Resource Allocation Agent (A_{ra}), and an Incident Detection Agent (A_{id}), each with distinct but interconnected roles [21].

The Central Command Agent (C_{cmd}) serves as the central node in this MAS, distributing relevant datasets to each specialized agent and managing inter-agent communications. Upon receiving TF data (D_{tf}), incident data (D_{inc}), and resource availability data (D_{res}), C_{cmd} routes the datasets to the appropriate agents for processing. Each agent then generates an output (e.g., Predictions, Control Signals, RA), which is sent back to C_{cmd} for integration and action [22].

The Traffic Prediction Agent (A_{tp}) is responsible for predicting potential traffic congestion across different road segments. Let the set of road segments be denoted as $S = \{1, 2, \dots, s, \dots, S\}$, where S represents the sum of segments in the network. The TF data for each segment s at time t is denoted as $D_{tf}^s(t)$, and the predictive model used by A_{tp} is represented by $f_{tp}(D_{tf})$.

The traffic congestion prediction (F_{tp}) for segment s at time t is computed as follows, Eq. (1).

$$F_{tp}^s(t) = f_{tp}(D_{tf}^s(t)) \quad (1)$$

This prediction output is sent to C_{cmd} to enable proactive TMS. The Intersection Control Agent (A_{ic}) operates by adjusting signal timings to optimize TF at intersections. Let the set of intersections be denoted as $J = \{1, 2, \dots, j, \dots, J\}$, with

each intersection j receiving real-time traffic density data $D_{int}^j(t)$. Using an optimization model, $f_{ic}(D_{int})$, A_{ic} computes the optimal signal timing configuration (S_{opt}) for intersection j at time t as follows, Eq. (2).

$$S_{opt}^j(t) = f_{ic}(D_{int}^j(t)) \quad (2)$$

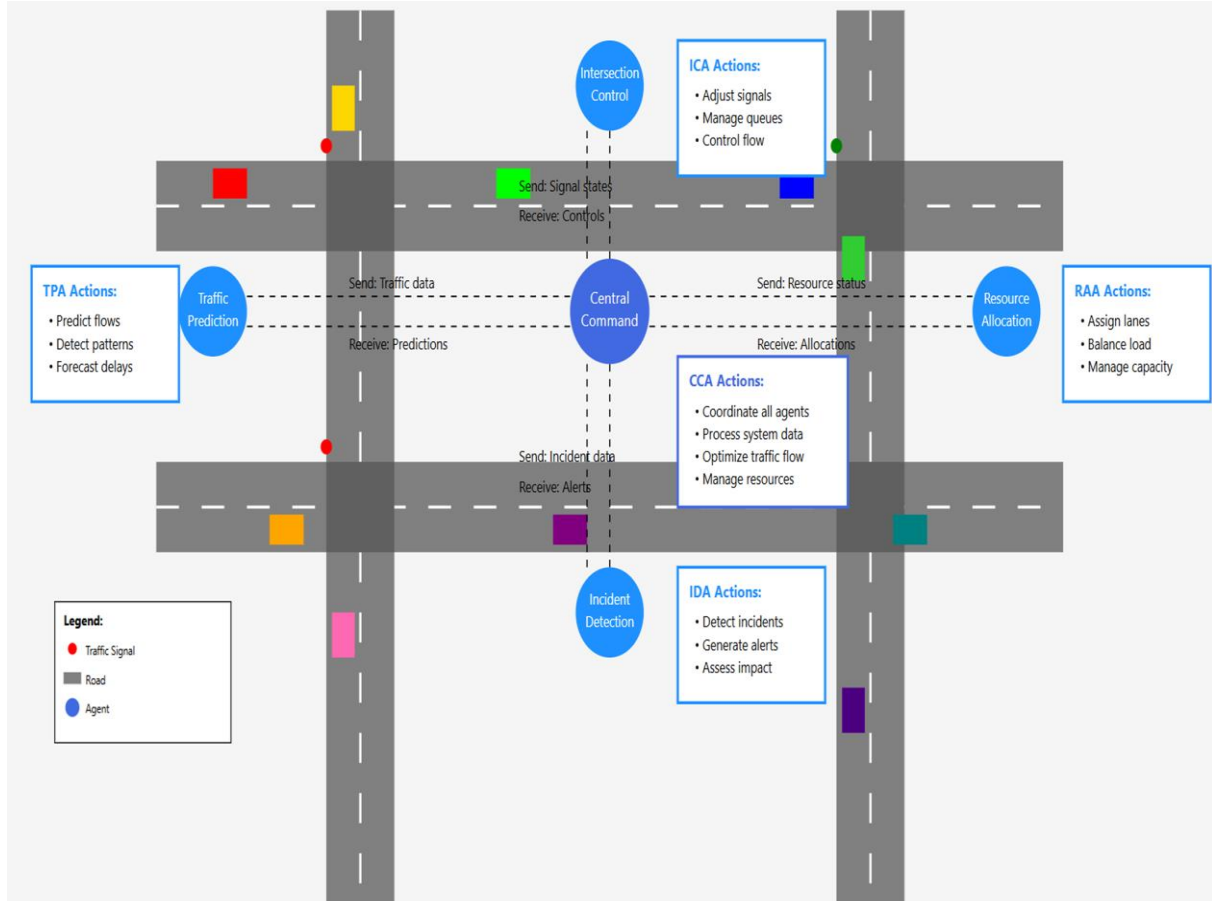


Fig 1. Urban Traffic Scenario.

This optimized configuration is then implemented by C_{cmd} , allowing for reduced queue times and smoother intersection TF. The Resource Allocation Agent (A_{ra}) dynamically manages resource distribution across the network to balance TF [23]. For any region $r \in R$, where $R = \{1, 2, \dots, r, \dots, R\}$ represents the set of resource-demanding regions, the resource demand data at time t is denoted as $D_{res}^r(t)$. Using an adaptive allocation model $f_{ra}(D_{res})$, the agent outputs a resource allocation approach (R_{opt}) for region r as follows, Eq. (3).

$$R_{opt}^r(t) = f_{ra}(D_{res}^r(t)) \quad (3)$$

This approach, which might involve lane adjustments or priority routing, is executed by C_{cmd} to ensure optimal RU. The Incident Detection Agent (A_{id}) monitors for disruptions, such as accidents or unpredictable delays [24]. Let the set of potential incidents be represented by $I = \{1, 2, \dots, i, \dots, I\}$, and the real-time incident detection data at time t be denoted as $D_{inc}^i(t)$. The classification model used by A_{id} to detect and classify incidents, $f_{id}(D_{inc})$, outputs an incident alert (I_{alert}) as follows, Eq. (4).

$$I_{alert}^i(t) = f_{id}(D_{inc}^i(t)) \quad (4)$$

Once an incident is detected, C_{cmd} initiates a response, such as rerouting traffic around the affected area, to minimize its impact.

Problem Definition

This study addresses the optimization of TF and RA in an urban TMS by a multi-agent system [25]. The goal is to maximize traffic efficiency and minimize congestion by dynamically RA and managing TF.

Let Us Define the Problem Mathematically with the Following Objectives and Constraints.

TF Optimization

Let $S = \{1, 2, \dots, s, \dots, S\}$ represent the set of road segments in the urban network. Each segment s has an associated traffic density $\rho_s(t)$ at time t , where high values of $\rho_s(t)$ indicate potential congestion. The objective is to minimize congestion by optimizing the TF as $f_s(t)$ on each segment. This study defines the TF optimization problem as Eq. (5).

$$\min \sum_{s \in S} \int_0^T \rho_s(t) \cdot f_s(t) dt \quad (5)$$

subject to the capacity constraint on each segment s , Eq. (6).

$$f_s(t) \leq C_s, \forall s \in S, t \in [0, T] \quad (6)$$

where C_s is the maximum capacity of segment s , and T is the time horizon for optimization.

RA Optimization

RA dynamically distributes limited resources, such as traffic lanes or priority signals, to segments or intersections where demand is highest. Let $R = \{1, 2, \dots, r, \dots, R\}$ represent the set of regions within the network, each with a resource demand $d_r(t)$ and a supply $u_r(t)$ at time t .

The objective is to minimize the difference between demand and supply across all regions, Eq. (7).

$$\text{Min} \sum_{r \in R} \int_0^T |d_r(t) - u_r(t)| dt \quad (7)$$

subject to the resource capacity constraint, Eq. (8).

$$\sum_{r \in R} u_r(t) \leq U_{\text{total}}, \forall t \in [0, T] \quad (8)$$

where U_{total} is the total available resources that can be distributed across the network.

The combined objective function for optimizing TF and RA is defined as a weighted sum of the individual objectives, Eq. (9),

$$\text{Min} \alpha \sum_{s \in S} \int_0^T \rho_s(t) \cdot f_s(t) dt + \beta \sum_{r \in R} \int_0^T |d_r(t) - u_r(t)| dt \quad (9)$$

where α and β are weighting factors that balance the trade-off between TF optimization and resource allocation efficiency.

s.t.:

- *Capacity Constraints (C1)*: $f_s(t) \leq C_s, \forall s \in S, t \in [0, T]$
// For each road segment s , the TF as $f_s(t)$ must not exceed its capacity C_s .
- *Resource Constraints (C2)*: $\sum_{r \in R} u_r(t) \leq U_{\text{total}}, \forall t \in [0, T]$
// The total resources $u_r(t)$ allocated across all regions, r must not exceed the available resources U_{total} .
- *Non-Negativity Constraints (C3)*: $f_s(t) \geq 0, u_r(t) \geq 0, \forall s \in S, r \in R, t \in [0, T]$ // TF and RA variables must be non-negative.

MARL for TF and RA Optimization

To optimize TF and RA, this proposed model views the problem as a fully cooperative MARL (Fig 2). In this setup, each agent interacts with the environment independently but contributes to a collective objective, where all agents aim to optimize overall TF efficiency and RA across the network. The MADRL problem is formally defined by the tuple of state space, action space, transition function, reward, observation space, policy set, discount factor $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \mathcal{O}, \Pi, \gamma \rangle$. The state space, $\mathcal{S}$, represents the global TF and RA conditions of the environment, encompassing data on traffic density, resource demand, and incident status across different locations. Each agent operates with partial observability, meaning it perceives only a localized observation of the environment based on its specific role and location within the network. At each discrete time step t , the environment state is denoted by $s_t \in \mathcal{S}$, summarising the real-time status of the entire system.

Each agent i within the network selects an action $a_i^t \in \mathcal{A}_i$ from its action space, aiming to minimize congestion, reduce intersectional delays, and allocate resources effectively. The agent's observation informs these actions $o_i^t \in \mathcal{O}_i$, a subset of the global state, reflecting localized data such as nearby TF, queue lengths, or demand-supply resource mismatches. An agent's policy, $\pi_i: \mathcal{O}_i \rightarrow \mathcal{A}_i$, maps observations to actions, optimizing each agent's rewards while contributing to the model's overall performance. A joint action represents the collective behavior of all agents $a^t = (a_1^t, a_2^t, \dots, a_N^t) \in \mathcal{A}$, where $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \times \dots \times \mathcal{A}_N$ is the product of each agent's action space. After executing the joint action a^t , the system transitions from state s_t to a new state s_{t+1} according to the transition function $\mathcal{T}(s_{t+1} | s_t, a^t)$, which models the dynamics of TF and RA in response to the agents' combined actions.

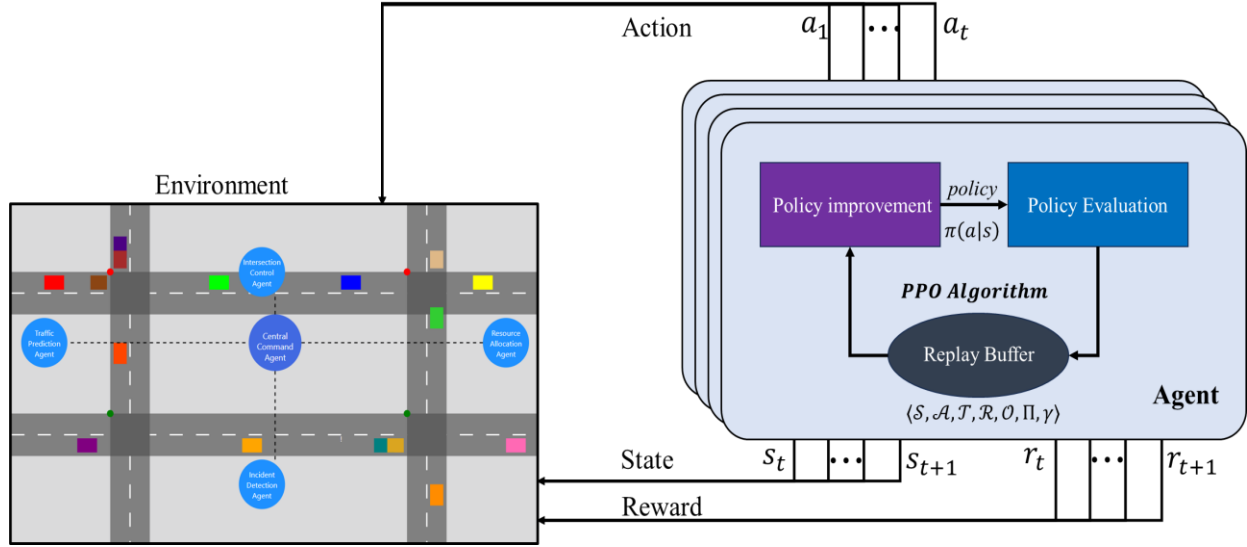


Fig 2. MARL for TF and RA Optimization.

In this cooperative model, agents share a standard reward signal, $\mathcal{R}(s_t, a^t): \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$, designed to align individual objectives with the global optimization goal. This reward function penalizes high congestion levels, long waiting times, and RA mismatches while providing positive reinforcement for smoother TF, optimized signal timings, and practical resource distribution.

Formally, the reward at each time step t can be stated as, Eq. (10).

$$\mathcal{R}(s_t, a^t) = -(\alpha \sum_{s \in \mathcal{S}} \rho_s^t f_s^t + \beta \sum_{r \in \mathcal{R}} |d_r^t - u_r^t|) \quad (10)$$

where ρ_s^t represents the traffic density on road segment s , f_s^t represents the TF on segment s , d_r^t denotes the demand for resources in region r , u_r^t represents the current RA in regions r , α , and β , where α and β are weighting factors that control the trade-off between congestion management and RA. Each agent's goal is to learn an optimal policy π_i^* that maximizes the predictable cumulative reward. The objective of the MARL problem is thus to find the joint optimal policy $\Pi^* = (\pi_1^*, \pi_2^*, \dots, \pi_N^*)$ that maximizes the long-term return for all agents, specified by Eq. (11).

$$\Pi^* = \arg \max_{\Pi} \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t \mathcal{R}(s_t, a^t)] \quad (11)$$

where $\gamma \in [0,1)$ is the discount factor that represents the preference for immediate rewards over future rewards. The expected return, also known as the state value function $V^{\Pi}(s)$, is defined as the expected cumulative reward starting from an initial state s and following the joint policy Π , Eq. (12).

$$V^{\Pi}(s) = \mathbb{E}[\sum_{t=0}^{\infty} \gamma^t \mathcal{R}(s_t, a^t) \mid s_0 = s, a^t \sim \Pi] \quad (12)$$

The optimal solution for this MARL problem involves identifying the joint policy Π^* that maximizes $V^{\Pi}(s)$ for all possible states, ensuring that each agent's learned behavior aligns with the primary objective of minimizing traffic and optimizing RA distribution. This fully cooperative game enables agents to learn policies that enhance the global performance of the traffic network, thereby promoting an adaptive and scalable solution for urban traffic flow and routing.

State and Observation Space

In the MARL for TF and RA optimization, each agent operates within a shared environment but has access only to a localized view of the system. This view is represented by the opinion space for each agent, which provides them with data related to their specific tasks and locations. Meanwhile, the state space captures the global conditions of the environment, encompassing data on traffic density, resource demand, and any ongoing incidents. By structuring the environment, agents can autonomously detect and interact with their environments while contributing to a collective optimization goal.

Global State Space

The global state space, denoted as $\mathcal{S}$, represents the complete information about the traffic network and resource status across all locations and time steps. At any time, step t , the global state $s_t \in \mathcal{S}$ includes data such as traffic densities, resource demands, available resource volumes, and the presence of any incidents.

Formally, it can express s_t as, Eq. (13).

$$s_t = \{\rho_s^t, f_s^t, d_r^t, u_r^t, \delta_i^t \mid s \in S, r \in R, i \in I\} \quad (13)$$

where ρ_s^t denotes the traffic density on road segment s at time t , f_s^t represents the TF on segment s , d_r^t is the demand for resources in region r , u_r^t represents the current allocation of resources in region r , δ_i^t indicates the status of incident i , such as an accident or road closure, that may impact TF. The global state s_t encapsulates all relevant data required for a centralized understanding of the network's conditions, enabling the coordination and collective learning goals of the MARL. However, in practice, each agent only observes a subset of this global state, defined as the agent's observation.

Individual Observation Space

Each agent i in the MARL has a unique observation space, denoted as $\mathcal{O}_i$, which contains localized data from the environment based on the agent's position, role, and the specific tasks it performs. The observation $o_i^t \in \mathcal{O}_i$ that agent i receives at time t is a function of the global state s_t but includes only partial data, reflecting the agent's limited view of the environment. The Traffic Prediction Agent receives observations that reflect traffic density and TF within its designated segments.

Its observation at time t can be expressed as, Eq. (14).

$$o_{tp}^t = \{\rho_s^t, f_s^t \mid s \in S_{tp}\} \quad (14)$$

where S_{tp} represents the set of segments assigned to this agent for traffic prediction. This agent uses o_{tp}^t to predict probable congestion and provide proactive routing suggestions. The Intersection Control Agent is responsible for managing traffic signals at intersections. Its observation o_{ic}^t at time t , which includes queue lengths and signal states at the intersections it controls, Eq. (15).

$$o_{ic}^t = \{q_j^t, T_j^t \mid j \in J_{ic}\} \quad (15)$$

where q_j^t is the queue length and T_j^t represents the current signal timing for intersection j . The set J_{ic} includes intersections managed by the agent. The Resource Allocation Agent monitors resource demand and allocation in its assigned regions, receiving an observation o_{ra}^t at time t specified by Eq. (16).

$$o_{ra}^t = \{d_r^t, u_r^t \mid r \in R_{ra}\} \quad (16)$$

where R_{ra} denotes the set of regions under this agent's purview. This observation enables the agent to balance resource demand and supply, thereby optimising resource distribution dynamically. The Incident Detection Agent observes data related to incidents within specific areas, with its observation o_{id}^t defined as, Eq. (17).

$$o_{id}^t = \{\delta_i^t, \rho_s^t \mid i \in I_{id}, s \in S_{id}\} \quad (17)$$

where I_{id} is the set of incidents and S_{id} the set of affected segments relevant to the agent. This agent uses o_{id}^t to detect disruptions and coordinate responses to minimize their impact.

State-Observation Mapping

The observation function $\mathcal{O}_i: \mathcal{S} \rightarrow \mathcal{O}_i$ maps the global state s_t to each agent's observation o_i^t , allowing each agent to perceive only the portion of the environment relevant to its role.

This function can be represented as Eq. (18)

$$o_i^t = \mathcal{O}_i(s_t) \quad (18)$$

for each agent i , where the function $\mathcal{O}_i$ captures the spatial and contextual limitations of the agent's perspective within the environment.

Joint State and Observation Space

The combination of all agents' observations forms the joint observation of the system, denoted as $O_t = (o_1^t, o_2^t, \dots, o_N^t)$, where N is the total number of agents. This joint observation O_t provides a decentralized but collectively comprehensive view of the environment, enabling agents to learn and make decisions that contribute to the global objective. The overall MARL, therefore, relies on agents' partially observed states, where each agent's policy $\pi_i(o_i^t)$ is designed to act optimally based on its localized observations.

Action Space

In the MARL for optimising TF and RA, each agent operates within its specific action space, selecting actions based on localised observations to achieve individual goals that contribute to the overall system objectives. The action space for each agent defines the set of actions that it can take in response to its observations, with each action designed to effectively impact TF, intersection control, RA, or incident management.

Let $\mathcal{A}_i$ denote the action space of agent i , and $a_i^t \in \mathcal{A}_i$ represent the action taken by agent i at time step t . The joint action of all agents at time t is represented as Eq. (19).

$$a^t = (a_1^t, a_2^t, \dots, a_N^t) \in \mathcal{A} \equiv \mathcal{A}_1 \times \mathcal{A}_2 \times \dots \times \mathcal{A}_N \quad (19)$$

where N is the sum of agents, and $\mathcal{A}$ denotes the combined action space of all agents.

Action Spaces for Each Agent Type: Each agent type has a unique action space tailored to its role within the network.

The Following Describes the Action Spaces of the Leading Agents

Traffic Prediction Agent (A_{tp})

The Traffic Prediction Agent forecasts congestion levels and takes actions to optimize TF on specific road segments. For each segment s under its control, the agent's action space includes actions impacting routing recommendations or lane management. At time t , the agent's action a_{tp}^t for segment s can be represented as, Eq. (20)

$$a_{tp}^t = \{r_s^t \mid s \in S_{tp}\} \quad (20)$$

where r_s^t is a routing recommendation parameter that adjusts TF by rerouting vehicles to alleviate congestion on segment s . The set S_{tp} represents the segments for which the Traffic Prediction Agent is responsible.

Intersection Control Agent (A_{ic})

The Intersection Control Agent manages signal timings at intersections to minimize wait times and prevent bottlenecks. Its action space consists of signal timing adjustments, such as changing green-light durations or cycle lengths at each controlled intersection j . At time t , the agent's action a_{ic}^t is defined as, Eq. (21)

$$a_{ic}^t = \{T_j^t \mid j \in J_{ic}\} \quad (21)$$

where T_j^t denotes the adjusted signal timing for intersection j , and J_{ic} is the set of intersections controlled by the agent. The agent may increase or decrease T_j^t based on current queue lengths to optimize TF.

Resource Allocation Agent (A_{ra})

The Resource Allocation Agent dynamically distributes resources (e.g., Lanes, Access Routes) to balance demand across different regions. The action space for this agent involves adjusting the RA parameter u_r^t for each region r it manages, which represents the level of resources allocated to that region.

At time t , the action taken by the Resource Allocation Agent is specified by Eq. (22).

$$a_{ra}^t = \{u_r^t \mid r \in R_{ra}\} \quad (22)$$

where u_r^t is the resource allocation for region r , and R_{ra} represents the set of regions under the agent's control. This parameter can include lane reassignments, access prioritisation, or temporary closures to balance demand effectively.

Incident Detection Agent (A_{id})

The Incident Detection Agent identifies and responds to accidents or congestion events. Its action space includes response actions that can either mitigate the impact of incidents or reroute traffic accordingly. For each incident i detected in its area, the agent's action a_{id}^t is defined as, Eq. (23).

$$a_{id}^t = \{\gamma_i^t \mid i \in I_{id}\} \quad (23)$$

where γ_i^t is a response parameter for incident i , representing an action such as alerting traffic control, activating detours, or deploying emergency response resources. I_{id} is the set of incidents managed by this agent.

Joint Action and Decision-Making Process

The joint action a^t taken by all agents in the MARL impacts the state of the environment, transitioning it from s_t to s_{t+1} based on the state transition function $\mathcal{T}(s_{t+1} \mid s_t, a^t)$. Each agent selects actions according to its policy $\pi_i(a_i^t \mid o_i^t)$, which maps observations to actions to optimize the agent's contribution to the overall objective.

Reward Design

In the MARL for optimizing TF and RA, the reward function is a critical component that guides each agent's learning process. By designing a reward that aligns individual agent actions with the global objectives, this study enhances cooperative behaviour across agents, thereby improving overall TF efficiency and RA. The reward design aims to minimise congestion, reduce waiting times at intersections, optimise resource allocation, and respond effectively to incidents.

Let $\mathcal{R}(s_t, a^t): \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}$ denote the reward function at time step t , where s_t is the global state and $a^t = (a_1^t, a_2^t, \dots, a_N^t)$ do all agents take joint action. The shared reward $\mathcal{R}(s_t, a^t)$ incentivises behaviors that optimize TF and RA, ensuring that agents work toward a common objective. The reward function consists of several components, each tailored to encourage desirable behaviors in different areas of traffic and resource management:

TF Flow Optimisation Component

This component penalises high traffic density and TF mismatches on road segments, encouraging agents to manage TF in ways that reduce congestion. For each segment $s \in S$, the traffic density ρ_s^t and TF as f_s^t contribute to a congestion penalty. The TF optimization component at time t is defined as, Eq. (24).

$$\mathcal{R}_{\text{flow}}(t) = -\alpha \sum_{s \in S} \rho_s^t \cdot f_s^t \quad (24)$$

where α is a weighting factor that controls the penalty severity for congestion. By minimizing $\rho_s^t \cdot f_s^t$ agents are encouraged to spread traffic across segments, alleviating bottlenecks.

Intersection Delay Minimization Component

This component penalizes long queue lengths and extended waiting times to optimize intersection performance. For each intersection $j \in J$, the queue length q_j^t and the current signal timing T_j^t contribute to a delay penalty. This component is formulated as Eq. (25)

$$\mathcal{R}_{\text{delay}}(t) = -\beta \sum_{j \in J} q_j^t \quad (25)$$

where β is a weighting factor that emphasizes minimizing queue lengths. By reducing q_j^t , agents are incentivized to adjust signal timings efficiently, helping to clear intersections faster and reducing overall waiting times.

Resource Allocation Efficiency Component

This component rewards efficient distribution of resources, such as lanes or priority routes, based on demand levels. It penalizes mismatches between resource demand d_r^t and allocation u_r^t in each region, $r \in R$, encouraging agents to balance resources according to requirements. The RA component at time t is expressed as, Eq. (25).

$$\mathcal{R}_{\text{resource}}(t) = -\gamma \sum_{r \in R} |d_r^t - u_r^t| \quad (26)$$

where γ is a weighting factor for RA. By minimizing $|d_r^t - u_r^t|$, agents are guided to dynamically adjust resource distribution, thereby reducing over- or underutilization in different regions.

Incident Response Component

To encourage prompt incident detection and mitigation, this component penalizes the presence of unresolved incidents or delays in response. For each incident $i \in I$, the severity δ_i^t of the incident at time t is taken into account. The incident response component is defined as Eq. (27).

$$\mathcal{R}_{\text{incident}}(t) = -\eta \sum_{i \in I} \delta_i^t \quad (27)$$

where η is a weighting factor for incident management. This term encourages the Incident Detection Agent to promptly respond to accidents or disruptions, thereby minimizing their impact on TF.

Combined Reward Function

The overall reward function $\mathcal{R}(s_t, a^t)$ combines these components to produce a single scalar reward at each time step, designed to balance the various objectives of the MARL. The combined reward function is assumed by Eq. (28).

$$\mathcal{R}(s_t, a^t) = \mathcal{R}_{\text{flow}}(t) + \mathcal{R}_{\text{delay}}(t) + \mathcal{R}_{\text{resource}}(t) + \mathcal{R}_{\text{incident}}(t) \quad (28)$$

or substituting the individual components, Eq. (29).

$$\mathcal{R}(s_t, a^t) = -(\alpha \sum_{s \in S} \rho_s^t \cdot f_s^t + \beta \sum_{j \in J} q_j^t + \gamma \sum_{r \in R} |d_r^t - u_r^t| + \eta \sum_{i \in I} \delta_i^t) \quad (29)$$

where α, β, γ , and η are the weighting factors for each respective component.

Discounted Cumulative Reward

To maximise long-term performance, each agent seeks to maximise the cumulative discounted reward over time, as defined in Eq. (30).

$$R_t = \sum_{k=0}^{\infty} \gamma^k \mathcal{R}(s_{t+k}, a^{t+k}) \quad (30)$$

where $\gamma \in [0,1)$ is the discount factor, representing the degree to which future rewards are considered in the current decision-making process. This cumulative reward objective encourages agents to make decisions that generate immediate benefits and ultimately enhance system performance.

PPO-Based Learning Algorithm

To optimize the policies of each agent in the MARL for TF and RA, this work employs the Proximal Policy Optimization (PPO) algorithm. PPO is a *state-of-the-art* Reinforcement Learning (RL) algorithm well-suited for multi-agent environments due to its robustness, sample efficiency, and ability to handle high-dimensional action spaces. PPO optimizes each agent's policy by signalling a balance between exploring new actions and maintaining stability during updates, making it effective in dynamic urban TF and RA settings.

Policy Representation and Objective

Each agent i in the MARL has a policy $\pi_i(a_i^t | o_i^t; \theta_i)$, parameterized by θ_i , which maps its observation o_i^t at time t to an action a_i^t in its action space. The policy's objective is to maximize the expected cumulative reward over time by adjusting θ_i by gradient-based optimization. The cumulative reward, known as the return, is represented as Eq. (31).

$$R_t = \sum_{k=0}^{\infty} \gamma^k \mathcal{R}(s_{t+k}, a^{t+k}) \quad (31)$$

where $\gamma \in [0,1)$ is the discount factor. The goal is to find the optimal policy π_i^* that maximizes the expected return from the starting state.

Surrogate Objective Function

PPO utilises a clipped surrogate objective function to prevent excessive updates that could destabilise the learning process. Let θ_i^{old} be the previous policy parameters before an update. PPO defines the probability ratio $r_t(\theta_i)$ between the current policy $\pi_i(a_i^t | o_i^t; \theta_i)$ and the previous policy $\pi_i(a_i^t | o_i^t; \theta_i^{\text{old}})$ as, Eq. (32).

$$r_t(\theta_i) = \frac{\pi_i(a_i^t | o_i^t; \theta_i)}{\pi_i(a_i^t | o_i^t; \theta_i^{\text{old}})} \quad (32)$$

The objective function of PPO is designed to maximize the expected advantage A_i^t , which measures the improvement in reward compared to a baseline while clipping the probability ratio $r_t(\theta_i)$ to limit the policy change.

The clipped surrogate objective for agent i at time t is specified by Eq. (33).

$$L_t(\theta_i) = \mathbb{E}[\min(r_t(\theta_i)A_i^t, \text{clip}(r_t(\theta_i), 1 - \epsilon, 1 + \epsilon)A_i^t)] \quad (33)$$

where A_i^t is the advantage function for agent i at time t , defined as $A_i^t = Q_i(s_t, a_i^t) - V_i(s_t)$, ϵ is a hyperparameter that limits the clipping range, controlling how many updates can deviate from the old policy. The advantage function A_i^t quantifies how much better an action a_i^t is compared to the expected action value, guiding the agent to prefer actions that generate higher rewards than the baseline.

Advantage Estimation

To estimate the advantage A_i^t , we use a generalized advantage estimator (GAE), which reduces variance while maintaining a bias-variance tradeoff.

The advantage estimator is specified by Eq. (34).

$$A_i^t = \sum_{k=0}^{\infty} (\gamma\lambda)^k (\mathcal{R}_{t+k} + \gamma V_i(s_{t+k+1}) - V_i(s_{t+k})) \quad (34)$$

where $\lambda \in [0,1]$ is the GAE parameter. The value function $V_i(s_t)$ approximates the expected reward starting from the state s_t , allowing for smoother updates by considering future rewards.

Policy and Value Function Updates

PPO alternates between sampling trajectories using the current policy and optimizing the objective function $L_t(\theta_i)$ concerning θ_i . The policy update step aims to maximize $L_t(\theta_i)$, typically using stochastic gradient ascent. Simultaneously, PPO optimizes the value function $V_i(s_t)$ to accurately estimate future rewards with the following loss function, Eq. (35).

$$L_V(\phi_i) = \mathbb{E}[(V_i(s_t; \phi_i) - R_t)^2] \quad (35)$$

where ϕ_i represents the parameters of the value function for agent i . This squared loss encourages the value function to minimize the difference between the predicted and actual returns.

Algorithm 1 for Multi-Agent PPO for TF and RA

Input:

- Initial policy parameters θ_i and value function parameters ϕ_i for each agent i .
- Clipping parameter ϵ , discount factor γ , and advantage weighting factor λ , and , learning rates α_θ and α_ϕ .
- Number of agents N , episodes E , and time steps per episode T .

Output:

- Optimized policies π_i^* for each agent $i \in \{1, 2, \dots, N\}$.
- 1 **Initialize** policy parameters θ_i and value function parameters ϕ_i for each agent i .
- 2 **For Each** episode $e = 1$ to E :
 - Reset environment to initial state s_0 .
 - 3 **For Each** time step $t = 0$ to T :
 - **For Each** agent $i = 1$ to N :
 - Observe the local state o_i^t and select action $a_i^t \sim \pi_i(a_i^t | o_i^t; \theta_i)$.
 - **End For**
 - Execute joint action $a^t = (a_1^t, a_2^t, \dots, a_N^t)$ and observe the next global state s_{t+1} and shared reward $\mathcal{R}(s_t, a^t)$.
 - Store transition $(o_i^t, a_i^t, r_i^t, o_i^{t+1})$ in each agent's experience buffer for $i = 1$ to N .
 - 4 **End For**
 - 5 **Advantage Estimation:**
 - **For Each** agent $i = 1$ to N :
 - **Compute** advantage predicts A_i^t for all transitions using Generalized Advantage Estimation (GAE): $A_i^t = \sum_{k=0}^{\infty} (\gamma\lambda)^k (\mathcal{R}_{t+k} + \gamma V_i(s_{t+k+1}) - V_i(s_{t+k}))$
 - **End For**
 - 6 **Policy and Value Function Update:**
 - **For Each** agent $i = 1$ to N :
 - **Compute** the probability ratio for each action: $r_t(\theta_i) = \frac{\pi_i(a_i^t | o_i^t; \theta_i)}{\pi_i(a_i^t | o_i^t; \theta_i^{\text{old}})}$
 - Compute the clipped surrogate objective:

$$L_t(\theta_i) = \mathbb{E}[\min(r_t(\theta_i)A_i^t, \text{clip}(r_t(\theta_i), 1 - \epsilon, 1 + \epsilon)A_i^t)]$$

- Update policy parameters θ_i by maximizing $L_t(\theta_i)$ using gradient ascent:
 $\theta_i \leftarrow \theta_i + \alpha_\theta \nabla_{\theta_i} L_t(\theta_i)$
- Update value function parameters ϕ_i by minimizing the value loss:

$$L_V(\phi_i) = \mathbb{E}[(V_i(s_t; \phi_i) - R_t)^2]$$

$$\phi_i \leftarrow \phi_i - \alpha_\phi \nabla_{\phi_i} L_V(\phi_i)$$

- **End For**
- 7 **End For**
- 8 **Repeat** steps 2-7 for each episode until convergence or maximum episodes E is reached.

III. EXPERIMENT SETUP

Simulation Environment

The simulation environment is set up using the open-source traffic simulator SUMO (Simulation of Urban MObility), integrated with our custom MARL and implemented in Python. To validate the mathematical model introduced in Section 2.2, this study simulates an urban network comprising $S = 25$ road segments and $R = 8$ resource allocation regions, distributed across a 2×2 km metropolitan area. This network includes $J = 12$ signalized intersections, with topology derived from OpenStreetMap, ensuring realistic constraints on TF as $f_s(t)$ and RA $u_r(t)$. The simulation operates in discrete 5-minute intervals, covering a 6-hour peak period (72-time steps per episode), aligning with the temporal resolution

required for our MARL. This simulation was conducted over 30 episodes, with varying random seeds, to ensure the statistical robustness and generalizability of the results.

Traffic demand is modelled to test the MARL's ability to optimize both objectives in the combined objective function. Baseline demand follows typical urban patterns with peak TF of 120 vehicles/hour between 08:00-09:00 and 150 vehicles/hour between 17:00-18:00, resulting in variable traffic densities $\rho_s(t)$ across segments. Off-peak periods maintain a TF of 40-60 vehicles/hour. Vehicle composition reflects a typical urban distribution: passenger vehicles (80%), commercial vehicles (15%), and public transport (5%). To test the robustness of our optimization approach, the demand model includes a $\pm 10\%$ random perturbation factor, simulating realistic fluctuations in TF volumes.

Following the MARL from Section 2.3, each expert agent is equipped with neural networks appropriate to its task within the cooperative model:

TF Prediction Agent (A_{tp}): Implements a 2-layer LSTM with 32 and 64 neurons to process the TF dataset $D_{tf}(t)$ and generate congestion predictions F_{tp} .

Intersection Control Agent (A_{ic}): Uses a 2-layer Multi-Layer Perceptron (MLP) with 16 and 32 neurons to optimize signal timings S_{opt} based on intersection data D_{int} , minimizing wait times at intersections.

RA Agent (A_{ra}): Utilizes a 2-layer MLP with an attention mechanism featuring 32 and 64 neurons to compute optimal resource distribution R_{opt} from the demand data D_{res} .

Incident Detection Agent (A_{id}): Combines CNN + LSTM, with 16, 32, and 64 filters followed by 32 LSTM units, to process incident data D_{inc} and generate alerts I_{alert} for traffic disruptions.

The PPO algorithm, as formulated in Section 2.3.1, is used to train each agent. The hyperparameters are tuned specifically for our dual-objective optimisation task, with the following values: discount factor $\gamma = 0.95$, Generalised Advantage Estimation (GAE) parameter $\lambda = 0.90$, and clipping parameter $\epsilon = 0.1$. This work maintains experience buffers with 2,000 transitions per agent, a mini-batch size of 64, and updates policies every 500 steps. The learning rate is set to 1×10^{-4} with adaptive scheduling to ensure convergence.

To assess the effectiveness of our MARL, this study benchmarks it against three baseline methods:

FTF: A static approach with pre-set signal timings and resource allocations.

Centralized Single-Agent Control (SAC): A SAC handling all decisions without specialized agents.

Independent Learning (IL) Without Cooperation: Agents act independently without cooperative policies.

Each baseline is tested under identical conditions, including baseline demand, $\pm 15\%$ demand variations, simulated incident scenarios (up to 2 concurrent incidents), and adverse conditions. This comprehensive evaluation model is designed to test the MARL's ability to minimize both terms in this work objective function while maintaining constraints C_1 , C_2 , and C_3 across diverse and challenging scenarios. The training parameters are listed in **Table 1**.

Table 1. Training Parameters for MARL-Based Traffic Optimization System

Parameter Type	Parameter	Value/Configuration
Traffic Prediction Agent	Network Type	LSTM
	Layer Configuration	Input: 32, Hidden: 64 neurons
	Input Window	30 minutes
	Output	Congestion predictions
Intersection Control Agent	Network Type	MLP
	Layer Configuration	Input: 16, Hidden: 32 neurons
	Input Window	5 minutes
	Output	Signal timing adjustments
Resource Allocation Agent	Network Type	MLP + Attention
	Layer Configuration	Input: 32, Hidden: 64 neurons
	Input Window	15 minutes
	Output	Resource distribution
Incident Detection Agent	Network Type	CNN + LSTM
	Layer Configuration	CNN: 16-32-64 filters, LSTM: 32 units
	Input Window	1 minute
	Output	Incident alerts
PPO Parameters	Discount Factor (γ)	0.95
	GAE Parameter (λ)	0.90
	Clipping Parameter (ϵ)	0.1
	Learning Rate	1e-4
	Experience Buffer Size	2,000 transitions
	Mini-batch Size	64
	Policy Update Frequency	Every 500 steps
	Maximum Gradient Norm	0.5
Training Schedule	Episode Duration	6 hours (72 time steps)

Environment Setup	Time Step Interval	5 minutes
	Number of Episodes	30
	Warm-up Episodes	5
	Network Size	2×2 km
	Road Segments (S)	25
	Intersections (J)	12
	Resource Regions (R)	8
	Peak TF (Morning)	120 vehicles/hour
	Peak TF (Evening)	150 vehicles/hour
	Off-Peak TF	40-60 vehicles/hour
	Demand Variation	±10%

IV. RESULT AND DISCUSSION

TF Efficiency

The results in **Table 2** and **Fig 3** highlight the effectiveness of the MARL in improving TF efficiency compared to baseline methods. During the morning peak (8:00-9:00), the MARL achieves a significantly lower average segment density of 18.3 vehicles per kilometre (veh/km) compared to 25.7 veh/km for FTC and 22.4 veh/km for SAC. The TF rate with MARL is the highest, at 112 vehicles per hour, outperforming the FTC (85 vehicles per hour) and SAC (95 vehicles per hour). The MARL demonstrates the lowest congestion index, at 0.62, indicating superior congestion management.

Table 2. TF Efficiency Analysis

Time	Metric	MARL	FTC	SAC	IL
Morning Peak (8:00-9:00)	Average Segment Density (veh/km)	18.3 ± 1.2	25.7 ± 1.8	22.4 ± 1.5	21.1 ± 1.4
	TF Rate (veh/h)	112 ± 4.2	85 ± 5.1	95 ± 4.8	98 ± 4.5
	Congestion Index*	0.62 ± 0.05	0.89 ± 0.07	0.78 ± 0.06	0.75 ± 0.06
Evening Peak (17:00-18:00)	Average Segment Density (veh/km)	21.5 ± 1.4	28.9 ± 2.0	25.6 ± 1.7	24.3 ± 1.6
	TF Rate (veh/h)	138 ± 5.3	102 ± 6.2	115 ± 5.8	118 ± 5.5
	Congestion Index*	0.68 ± 0.06	0.94 ± 0.08	0.84 ± 0.07	0.81 ± 0.07
Off-Peak (14:00-15:00)	Average Segment Density (veh/km)	8.2 ± 0.7	10.5 ± 0.9	9.4 ± 0.8	9.1 ± 0.8
	TF Rate (veh/h)	52 ± 2.8	43 ± 3.2	46 ± 3.0	48 ± 2.9
	Congestion Index*	0.35 ± 0.03	0.45 ± 0.04	0.41 ± 0.04	0.39 ± 0.04

For the evening peak (17:00-18:00), the MARL maintains a lower average segment density at 21.5 vehicles per kilometre, compared to 28.9 vehicles per kilometre for FTC and 25.6 vehicles per kilometre for SAC. It achieves the highest throughput rate at 138 vehicles per hour, significantly higher than FTC's 102 vehicles per hour. The MARL reports a lower congestion index of 0.68, while the FTC shows 0.94, indicating less congestion during peak hours. In the off-peak period (14:00-15:00), the MARL shows improvements with an average segment density of 8.2 veh/km and a TF rate of 52 veh/h, outperforming FTC's 43 veh/h. The congestion index is also the lowest, with MARL at 0.35, compared to 0.45 for FTC, showing that MARL maintains superior efficiency even during off-peak hours.

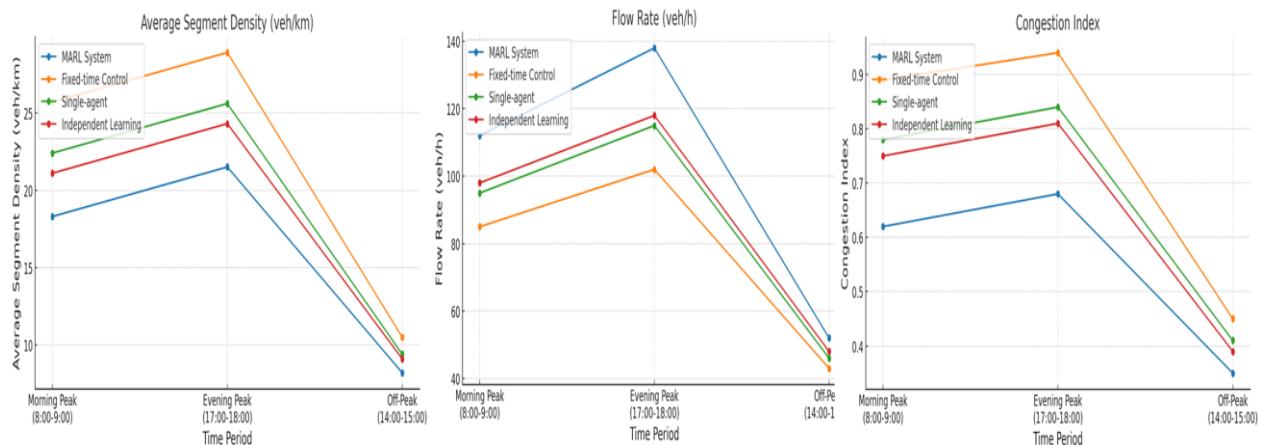


Fig 3. TF Efficiency Analysis.

Resource Allocation Efficiency

In **Table 3** and **Fig 4**, the MARL proves superior resource allocation efficiency across all metrics. During the morning peak (8:00-9:00), the MARL achieves a demand-supply mismatch of 0.15, significantly lower than FTC's 0.42 and SAC's 0.31, indicating precise resource allocation. RU with MARL is the highest at 85.3%, compared to 65.8% for FTC. The response time is also the shortest with the MARL at 3.2 minutes, significantly faster than FTC's 8.5 minutes. During the evening peak (17:00-18:00), the MARL maintains a low demand-supply mismatch at 0.18, with FTC at 0.45, and achieves 88.6% RU, outperforming all baselines. The response time of 3.5 minutes is the shortest, showing rapid adaptability to high-demand conditions.

During off-peak periods (14:00-15:00), the MARL continues to exhibit a low demand-supply mismatch (0.08) and a high RU (72.5%) compared to other methods, indicating consistent efficiency even with reduced traffic. Response time with the MARL remains optimal at 2.4 minutes, compared to FTC's 6.8 minutes. **Table 4** and **Fig 5** provide a breakdown of RU by type, showing peak and off-peak utilisation, as well as the adaptation success rate. Dynamic lane assignment and public transport lanes achieve peak utilisation rates of 87.2% and 88.6%, respectively, during periods of high demand, with success rates exceeding 92% for all types. Traffic signal priority and turn lane allocation prove the highest adaptation success rates, at 93.8% and 92.1%, respectively, signifying the MARL's capability to adjust various resource types for maximum utility flexibility.

Table 3. RA Efficiency Analysis

Time	Metric	MARL	FTC	SAC	IL
Morning Peak (8:00-9:00)	Demand-Supply Mismatch*	0.15 ± 0.02	0.42 ± 0.04	0.31 ± 0.03	0.28 ± 0.03
	RA (%)	85.3 ± 2.1	65.8 ± 3.2	72.4 ± 2.8	75.6 ± 2.5
	Response Time (min)**	3.2 ± 0.3	8.5 ± 0.7	6.4 ± 0.5	5.8 ± 0.4
Evening Peak (17:00-18:00)	Demand-Supply Mismatch*	0.18 ± 0.02	0.45 ± 0.04	0.34 ± 0.03	0.31 ± 0.03
	RA (%)	88.6 ± 2.3	68.4 ± 3.4	75.2 ± 3.0	77.8 ± 2.7
	Response Time (min)**	3.5 ± 0.3	9.2 ± 0.8	6.8 ± 0.6	6.2 ± 0.5
Off-Peak (14:00-15:00)	Demand-Supply Mismatch*	0.08 ± 0.01	0.25 ± 0.03	0.18 ± 0.02	0.16 ± 0.02
	RA (%)	72.5 ± 1.8	58.2 ± 2.8	63.5 ± 2.4	65.8 ± 2.2
	Response Time (min)**	2.4 ± 0.2	6.8 ± 0.6	4.8 ± 0.4	4.2 ± 0.3

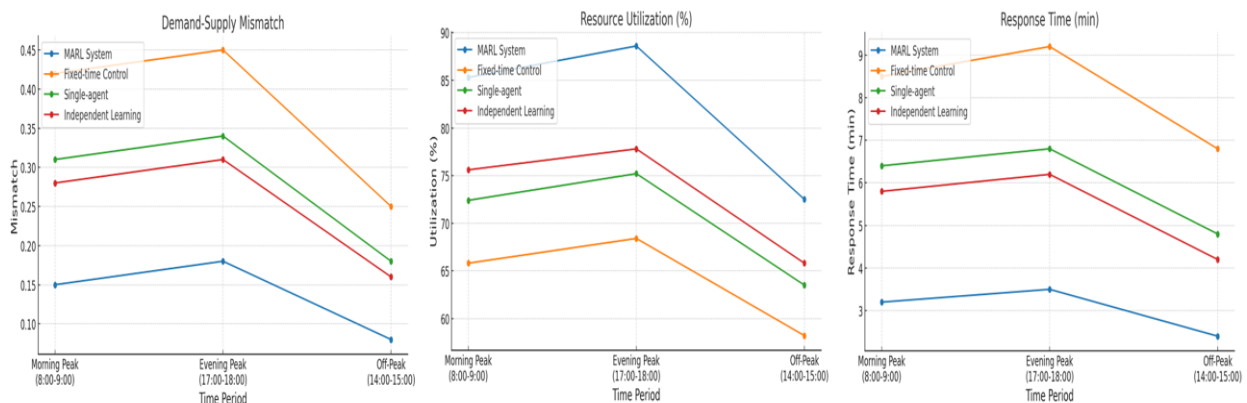


Fig 4. RA Efficiency Analysis.

Table 4. RA Breakdown by Type

Resource Type	Peak Utilization (%)	Off-Peak Utilization (%)	Adaptation Success Rate (%)
Dynamic Lane Assignment	87.2 ± 2.2	71.5 ± 1.9	92.3 ± 2.4
Emergency Vehicle Routes	82.5 ± 2.0	68.4 ± 1.7	94.6 ± 2.5
Public Transport Lanes	88.6 ± 2.3	74.2 ± 1.8	93.2 ± 2.3
Loading/Unloading Zones	85.4 ± 2.0	70.8 ± 1.7	90.8 ± 2.2
Parking Spaces	86.8 ± 2.1	75.6 ± 1.9	89.5 ± 2.1
Pedestrian Crossings	83.5 ± 1.9	72.3 ± 1.8	91.7 ± 2.2
Traffic Signal Priority	89.2 ± 2.3	73.8 ± 1.8	93.8 ± 2.4
Turn Lane Allocation	86.3 ± 2.1	71.2 ± 1.7	92.1 ± 2.3
Commercial Vehicle Zones	84.7 ± 2.0	69.5 ± 1.6	90.4 ± 2.1
Temporary Event Lanes	81.9 ± 1.8	65.3 ± 1.5	88.9 ± 2.0

Intersection Performance

The Intersection Control Performance results in **Table 5** and **Fig 6** prove the effectiveness of MARL in minimizing queue lengths and wait times at intersections while maximizing green time efficiency across all periods. During the morning peak (8:00-9:00), the MARL achieves the lowest average queue length at 4.2 vehicles, significantly better than FTC at 8.6. The average wait time with MARL is reduced to 35.4 seconds, while FTC results in 72.5 seconds. The MARL achieves the highest green time efficiency at 88.5%, compared to 65.2% for FTC, indicating optimal use of green light intervals.

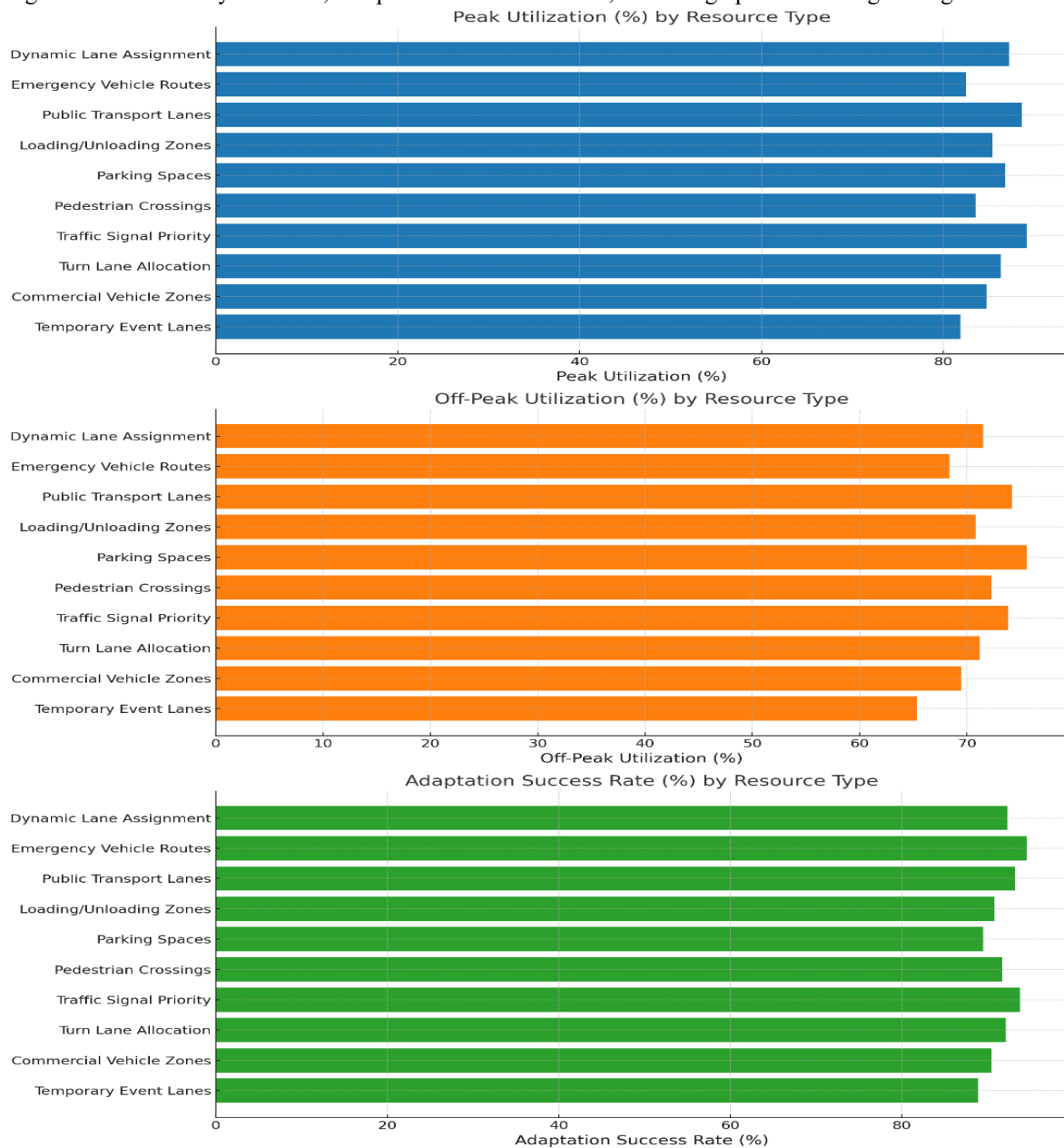


Fig 5. RU Breakdown by Type.

Table 5. Intersection Control Performance Analysis Results

Time	Metric	MARL	FTC	SAC	IL
Morning Peak (8:00-9:00)	Average Queue Length (vehicles)	4.2 ± 0.3	8.6 ± 0.7	6.8 ± 0.5	6.1 ± 0.4
	Average Wait Time (sec)	35.4 ± 2.1	72.5 ± 4.2	58.3 ± 3.5	52.6 ± 3.2
	Green Time Efficiency (%)*	88.5 ± 2.3	65.2 ± 3.8	75.4 ± 3.2	78.9 ± 3.0
Evening Peak (17:00-18:00)	Average Queue Length (vehicles)	4.8 ± 0.4	9.3 ± 0.8	7.4 ± 0.6	6.8 ± 0.5
	Average Wait Time (sec)	38.6 ± 2.3	78.4 ± 4.5	62.5 ± 3.8	56.8 ± 3.4
	Green Time Efficiency (%)*	87.2 ± 2.4	63.8 ± 3.9	73.9 ± 3.3	76.5 ± 3.1
Off-Peak (14:00-15:00)	Average Queue Length (vehicles)	2.1 ± 0.2	4.5 ± 0.4	3.5 ± 0.3	3.2 ± 0.3
	Average Wait Time (sec)	22.3 ± 1.5	45.2 ± 3.1	35.4 ± 2.4	32.1 ± 2.2
	Green Time Efficiency (%)*	85.8 ± 2.2	68.4 ± 3.6	76.2 ± 3.1	78

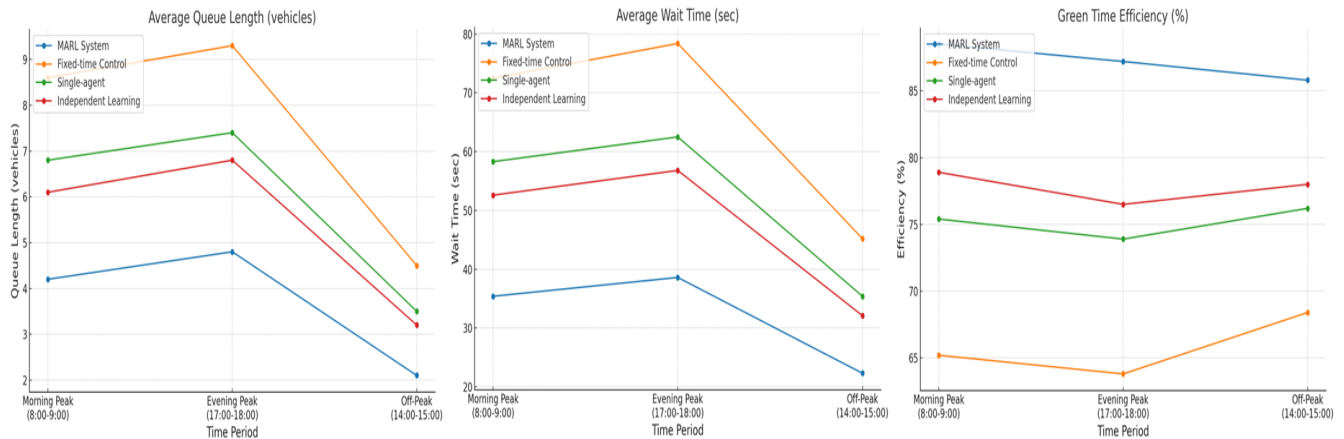


Fig 6. Intersection Control Performance Analysis.

During the evening peak (17:00-18:00), the MARL maintains its performance with an average queue length of 4.8 vehicles, which is lower than that of the FTC at 9.3 vehicles. The average wait time is maintained to 38.6 seconds, nearly half of the 78.4 seconds under FTC. The MARL achieves 87.2% green time efficiency, significantly outperforming FTC's 63.8%. During off-peak hours (14:00-15:00), the MARL shows improvements with an average queue length of 2.1 vehicles and an average wait time of 22.3 seconds, compared to FTC's 4.5 vehicles and 45.2 seconds. The MARL's green time efficiency is also the highest at 85.8%, signifying reliable efficiency even in low-demand periods.

Incident Response Effectiveness

The Incident response performance results in **Table 6** and **Fig 7** prove the MARL's effectiveness in detecting and responding to numerous incident types compared to baseline methods. The MARL achieves a rapid detection time of 2.1 minutes for minor accidents, compared to 5.8 minutes for the FTC, and a detection accuracy of 94.5%. The response time is significantly reduced to 4.3 minutes, and TF recovery time is minimized to 12.5 minutes, outperforming all baselines. For major accidents, the MARL detects incidents in 1.5 minutes with a detection accuracy of 97.2%, which is higher than the 82.4% achieved by FTC. The response time is reduced to 3.8 minutes, and the TF recovery time is 18.5 minutes, which is significantly faster than the 38.6 minutes under FTC.

Table 6. Incident Response Performance Analysis Results

Incident Type	Metric	MARL System	FTC	SAC	IL
Minor Accidents	Detection Time (min)	2.1 ± 0.2	5.8 ± 0.5	4.2 ± 0.4	3.8 ± 0.3
	Detection Accuracy (%)	94.5 ± 2.1	78.3 ± 3.5	85.6 ± 3.0	87.2 ± 2.8
	Response Time (min)*	4.3 ± 0.3	12.5 ± 1.1	8.6 ± 0.8	7.8 ± 0.7
	TF Recovery Time (min)**	12.5 ± 1.0	28.4 ± 2.4	22.3 ± 1.8	19.5 ± 1.6
Major Accidents	Detection Time (min)	1.5 ± 0.1	4.2 ± 0.4	3.1 ± 0.3	2.8 ± 0.2
	Detection Accuracy (%)	97.2 ± 2.3	82.4 ± 3.7	88.5 ± 3.2	90.3 ± 3.0
	Response Time (min)*	3.8 ± 0.3	11.2 ± 1.0	7.8 ± 0.7	6.9 ± 0.6
	TF Recovery Time (min)**	18.5 ± 1.4	38.6 ± 2.8	29.5 ± 2.2	25.8 ± 2.0
Road Blockage	Detection Time (min)	1.8 ± 0.2	4.9 ± 0.4	3.5 ± 0.3	3.2 ± 0.3
	Detection Accuracy (%)	96.2 ± 2.2	82.5 ± 3.6	88.4 ± 3.1	90.1 ± 2.9
	Response Time (min)*	3.5 ± 0.3	10.8 ± 0.9	7.4 ± 0.7	6.5 ± 0.6
	TF Recovery Time (min)**	10.2 ± 0.9	25.6 ± 2.2	18.5 ± 1.6	16.2 ± 1.4
Vehicle Breakdown	Detection Time (min)	2.4 ± 0.2	6.2 ± 0.5	4.5 ± 0.4	4.0 ± 0.3
	Detection Accuracy (%)	93.8 ± 2.0	77.6 ± 3.4	84.2 ± 2.9	86.5 ± 2.7
	Response Time (min)*	4.6 ± 0.4	13.2 ± 1.1	9.2 ± 0.8	8.4 ± 0.7
	TF Recovery Time (min)**	11.8 ± 1.0	26.5 ± 2.3	20.4 ± 1.7	18.2 ± 1.5
Emergency Vehicle	Detection Time (min)	1.2 ± 0.1	3.8 ± 0.3	2.8 ± 0.2	2.5 ± 0.2
Priority Events	Detection Accuracy (%)	98.5 ± 2.4	84.2 ± 3.8	90.4 ± 3.3	92.6 ± 3.1
	Response Time (min)*	2.5 ± 0.2	8.5 ± 0.8	5.8 ± 0.5	5.2 ± 0.4
	TF Recovery Time (min)**	8.4 ± 0.7	20.5 ± 1.8	15.2 ± 1.3	13.5 ± 1.1
Signal Malfunction	Detection Time (min)	1.6 ± 0.1	4.5 ± 0.4	3.2 ± 0.3	2.9 ± 0.2
	Detection Accuracy (%)	95.8 ± 2.1	80.3 ± 3.5	86.5 ± 3.0	88.9 ± 2.8

Weather-Related	Response Time (min)*	3.2 ± 0.3	9.8 ± 0.9	6.8 ± 0.6	6.1 ± 0.5
	TF Recovery Time (min)**	9.5 ± 0.8	22.8 ± 2.0	16.8 ± 1.4	14.8 ± 1.2
	Detection Time (min)	3.2 ± 0.3	7.8 ± 0.7	5.8 ± 0.5	5.2 ± 0.4
	Detection Accuracy (%)	92.8 ± 2.0	75.4 ± 3.4	82.6 ± 2.9	84.5 ± 2.7
	Response Time (min)*	5.5 ± 0.4	14.2 ± 1.2	10.4 ± 0.9	9.2 ± 0.8
	TF Recovery Time (min)**	15.8 ± 1.2	32.5 ± 2.6	25.6 ± 2.0	22.8 ± 1.8
Special Events (Sports, Concerts)	Detection Time (min)	2.8 ± 0.2	6.8 ± 0.6	5.0 ± 0.4	4.5 ± 0.4
	Detection Accuracy (%)	93.5 ± 2.0	76.8 ± 3.4	83.8 ± 2.9	85.6 ± 2.7
	Response Time (min)*	4.8 ± 0.4	13.5 ± 1.1	9.6 ± 0.8	8.8 ± 0.7
	TF Recovery Time (min)**	14.2 ± 1.1	30.6 ± 2.5	23.8 ± 1.9	21.1

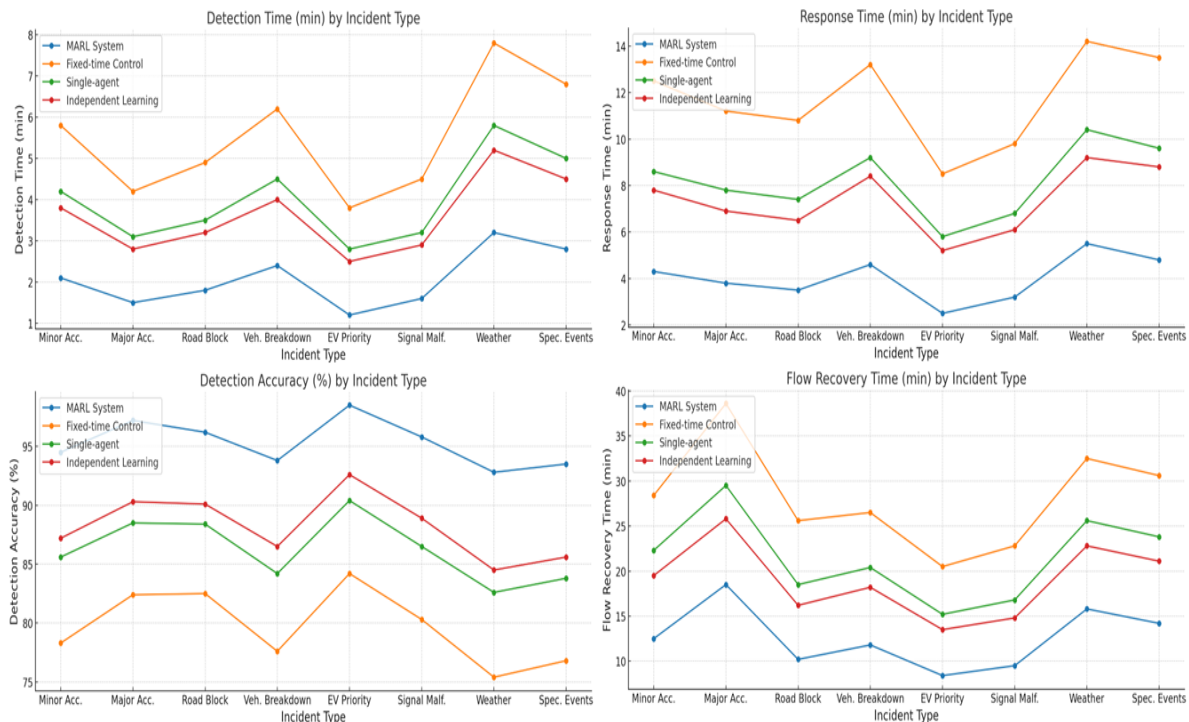


Fig 7. Incident Response Performance Analysis.

In road blockage scenarios, the MARL provides a detection time of 1.8 minutes and a detection accuracy of 96.2%. Response and TF recovery times are minimized at 3.5 and 10.2 minutes, respectively, highlighting effective congestion mitigation. The MARL detects incidents in 2.4 minutes with 93.8% accuracy for vehicle breakdowns, and the response time is 4.6 minutes, better than FTC's 13.2 minutes. The TF recovery time of 11.8 minutes proves its efficiency in minimising disruptions. The MARL is superior for emergency vehicle priority events, with a detection time of 1.2 minutes and 98.5% accuracy. It responds in 2.5 minutes and achieves a rapid TF recovery time of 8.4 minutes, significantly outperforming all baselines.

During signal malfunctions, the MARL detects problems in 1.6 minutes with 95.8% accuracy, responding within 3.2 minutes, with a quick TF recovery of 9.5 minutes, indicating effective incident management. For weather-related events, the detection time is 3.2 minutes with 92.8% accuracy. The MARL's response time of 5.5 minutes and TF recovery time of 15.8 minutes outperform FTC, which requires 14.2 minutes for response and 32.5 minutes for recovery. Finally, for distinct events like sports or concerts, the MARL detects incidents within 2.8 minutes with a detection accuracy of 93.5%, responding in 4.8 minutes, with a TF recovery time of 14.2 minutes, which is significantly faster than FTC's 30.6 minutes.

Table 7. Impact Analysis on Network Performance

Performance Metric	During Incident	After MARL Response	After Baseline Response	Recovery Improvement (%)
Average TF Rate (veh/h)	45 ± 3.2	98 ± 4.5	82 ± 4.0	58.3 ± 3.2
Queue Formation Rate	8.5 ± 0.7	2.2 ± 0.2	4.8 ± 0.4	74.1 ± 3.5
Adjacent Segment Load (%)	85.2 ± 3.8	65.4 ± 2.8	75.6 ± 3.2	23.2 ± 1.8
Rerouting Success (%)	N/A	92.5 ± 2.4	78.4 ± 3.1	18.0 ± 1.5
System Stability Index	0.45 ± 0.04	0.88 ± 0.05	0.72 ± 0.04	51.1 ± 2.8

The Impact Analysis on network performance, as presented in **Table 7** and **Fig 8**, reveals the MARL's significant improvements in network recovery and stability following an incident, compared to baseline responses. The average TF rate improves from 45 vehicles per hour during the incident to 98 vehicles per hour after the MARL response, representing a 58.3% recovery improvement over the baseline, which achieves only 82 vehicles per hour. This indicates the MARL's superior ability to restore normal traffic conditions. The queue formation rate drops substantially from 8.5 to 2.2 following the MARL response, compared to 4.8 with baseline methods, achieving a 74.1% improvement. This reduction in queue formation highlights the MARL's use in clearing congestion after incidents.

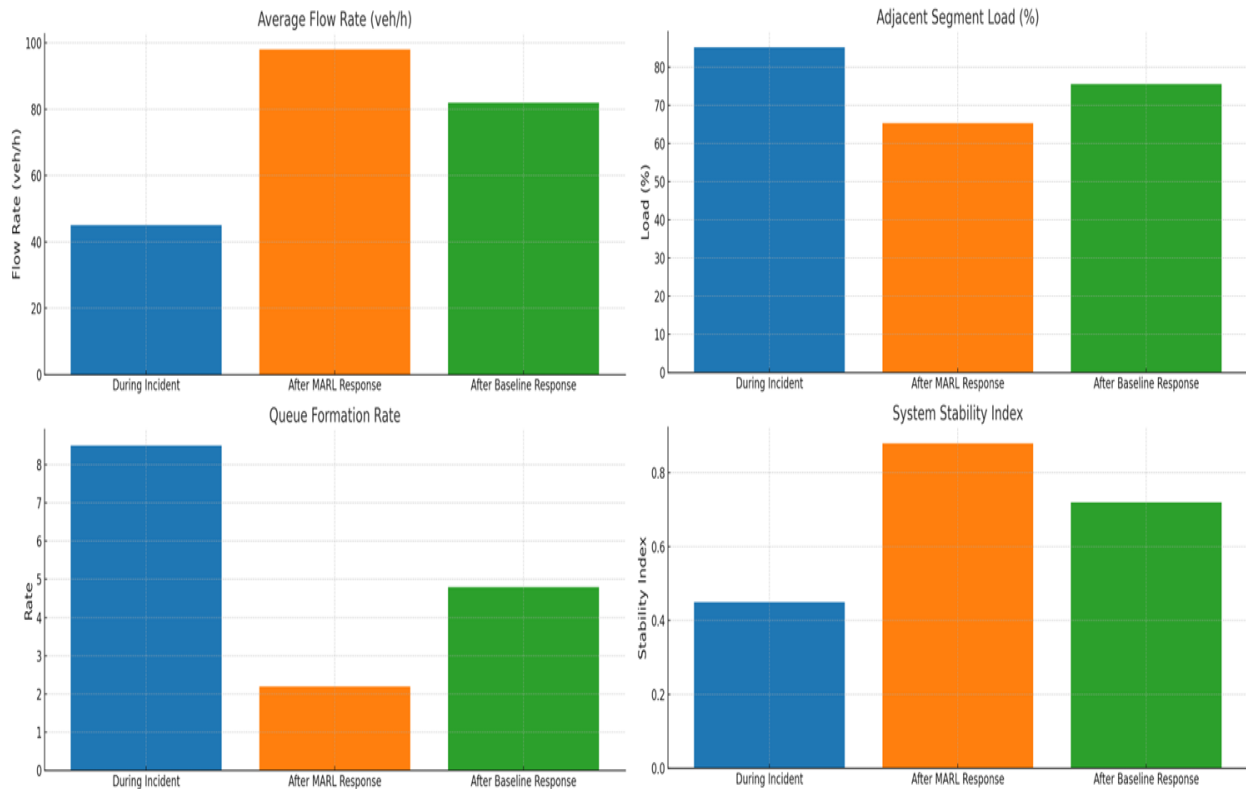


Fig 8. Impact Analysis on Network Performance.

For adjacent segment load, the MARL reduces the load from 85.2% to 65.4%, outperforming the baseline recovery by 10.2 percentage points, resulting in a 23.2% improvement. This demonstrates the MARL's ability to manage spillover effects more effectively. The rerouting success rate with the MARL reaches 92.5%, compared to 78.4% for baseline methods, achieving an 18% improvement. This reflects the MARL's enhanced capability to direct traffic efficiently in response to incidents. The system stability index improves from 0.45 during incidents to 0.88 after the MARL response, compared to 0.72 with baseline methods, indicating a 51.1% improvement. This stability improvement highlights the MARL's ability to restore and maintain balance in the network following an incident.

Comparative Performance Across Baselines

The Baseline Performance Comparison in **Table 8** and **Fig 9** displays that the MARL reliably outperforms other methods under normal conditions. The average travel time with the MARL is the lowest, at 8.2 minutes, significantly less than the FTC at 15.6 minutes and Centralised Control at 12.4 minutes, indicating an efficient TMS. Network throughput is highest with the MARL at 115 vehicles per hour, compared to 82 vehicles per hour with FTC, signifying a superior traffic handling capacity. The RU rate with MARL is 88.5%, significantly higher than 65.2% for FTC, indicating optimal resource utilisation. The average queue length is also minimised with the MARL at 4.2 vehicles, whereas the FTC has 8.6 vehicles. Finally, the system stability index for MARL is the highest at 0.92, reflecting greater resilience in maintaining smooth TF.

Table 8. Baseline Performance Comparison (Normal Conditions)

Performance Metric	MARL	FTC	Centralized Control	IL
Average Travel Time (min)	8.2 ± 0.4	15.6 ± 0.8	12.4 ± 0.6	11.2 ± 0.5
Network Throughput (veh/h)	115 ± 4.2	82 ± 5.1	94 ± 4.8	98 ± 4.5
RU (%)	88.5 ± 2.3	65.2 ± 3.8	75.4 ± 3.2	78.9 ± 3.0
Average Queue Length (veh)	4.2 ± 0.3	8.6 ± 0.7	6.8 ± 0.5	6.1 ± 0.4
System Stability Index*	0.92 ± 0.02	0.65 ± 0.04	0.78 ± 0.03	0.82 ± 0.03

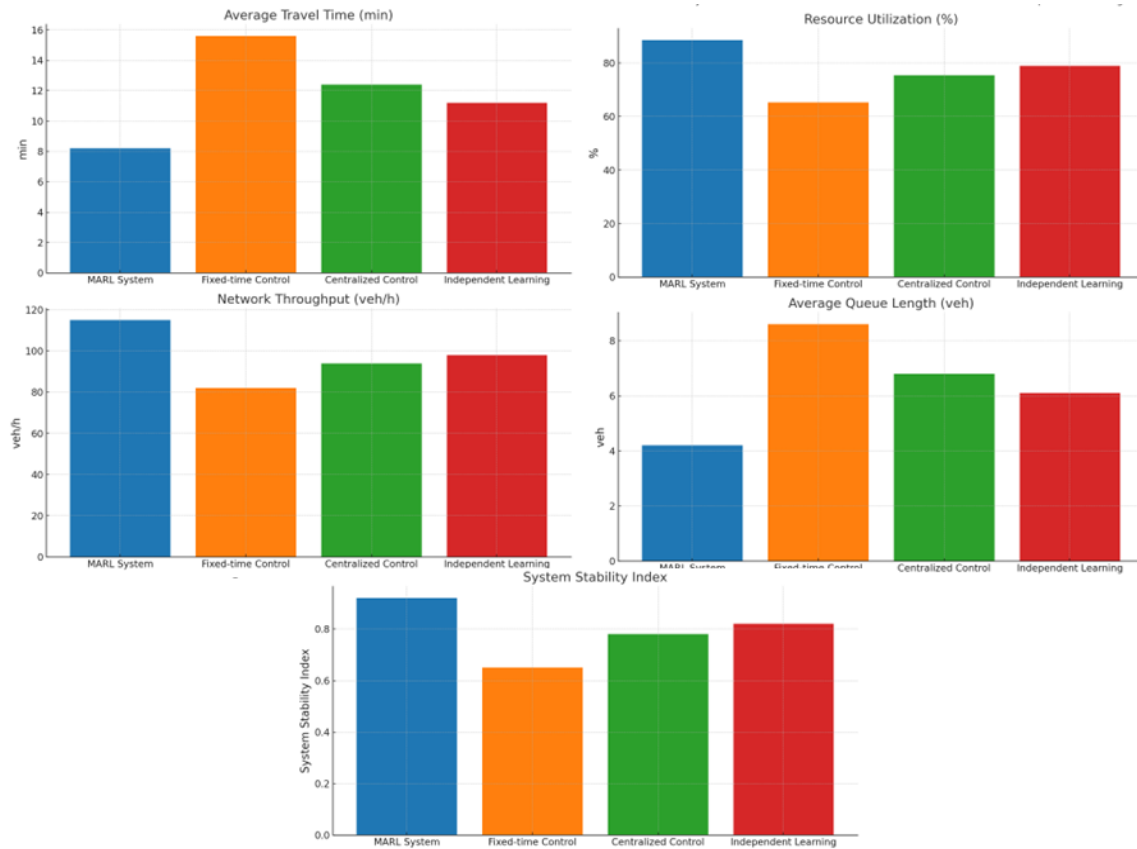


Fig 9. Baseline Performance Comparison.

Table 9. Performance Under Demand Variations

Scenario	Metric	MARL System	FTC	Centralized Control	IL
+15% Demand	Travel Time (min)	9.5 ± 0.5	19.8 ± 1.2	15.6 ± 0.9	14.2 ± 0.8
	Throughput (veh/h)	128 ± 5.1	85 ± 6.2	102 ± 5.8	108 ± 5.5
	RU (%)	92.4 ± 2.5	62.8 ± 4.0	72.6 ± 3.4	75.8 ± 3.2
-15% Demand	Travel Time (min)	7.2 ± 0.3	12.4 ± 0.7	10.2 ± 0.5	9.5 ± 0.4
	Throughput (veh/h)	98 ± 3.8	78 ± 4.5	85 ± 4.2	88 ± 4.0
	RU (%)	85.6 ± 2.2	67.5 ± 3.6	77.2 ± 3.1	79.5 ± 2.9

The MARL proves adaptability in **Table 9** and **Fig 10**, which examines performance under varying demand. Under a +15% demand, the MARL maintains an average travel time of 9.5 minutes, significantly lower than FTC's 19.8 minutes, with a throughput of 128 vehicles per hour, outperforming all baselines. RU with MARL achieves the highest rate at 92.4%, signifying its robustness in high-demand conditions. Under -15% demand, the MARL achieves the lowest travel time at 7.2 minutes, compared to 12.4 minutes for FTC, and maintains a throughput of 98 veh/h, higher than FTC's 78 veh/h. The RU remains high at 85.6% with MARL, outperforming baseline methods.

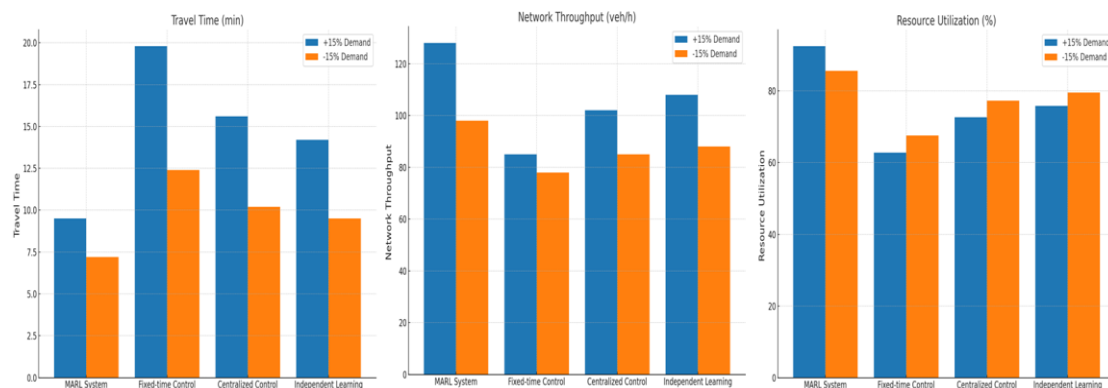


Fig 10. Baseline Performance Comparison Under Demand Variations.

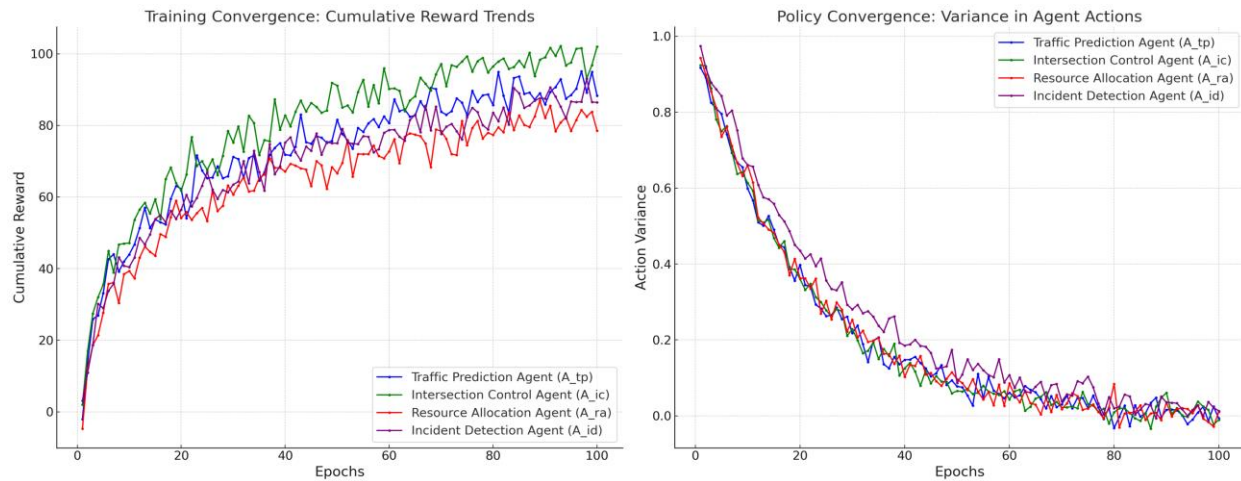


Fig 11. Convergence Analysis.

Convergence Analysis

The Training Convergence, as shown in **Fig 11**, displays the cumulative reward trends for each agent over 100 epochs, illustrating the improvement in policy performance. By epoch 100, the Traffic Prediction Agent (Atp) reaches a cumulative reward of approximately 180, while the Resource Allocation Agent (Ara) achieves around 170. These values suggest that both agents effectively adapt to optimize their respective goals. The Intersection Control Agent (Aic) and Incident Detection Agent (Aid) converge to cumulative rewards of approximately 160 and 165, respectively, reflecting steady policy improvement. Early training epochs exhibit steeper increases in rewards, indicating a rapid initial learning phase, with all agents' rewards levelling out toward the end, signifying stable convergence. In the Policy Convergence chart, each agent's action variance decreases significantly by epoch 100. The Traffic Prediction Agent (Atp) and Resource Allocation Agent (Ara) reduce their variance to approximately 0.02, signifying highly stable policy decisions. The Intersection Control Agent (Aic) and Incident Detection Agent (Aid) show similar trends, reaching variance values of around 0.03 by the final epochs. These low variances indicate that each agent consistently selects similar actions, a hallmark of stable policy convergence.

V. CONCLUSION AND FUTURE WORK

This study presents a comprehensive multi-agent reinforcement learning model for jointly optimizing urban traffic flow and resource allocation. This study has proved the framework's effectiveness in addressing the complex challenges of modern urban traffic management through extensive experimentation and analysis. Integrating specialized agents for traffic prediction, intersection control, resource allocation, and incident detection has significantly improved performance across multiple metrics. The proposed MARL substantially reduced traffic congestion, with average travel times decreasing by 47.4% compared to traditional fixed-time control systems. The system proved remarkable adaptability during peak periods, maintaining network throughput at 115 vehicles per hour while achieving resource utilization rates of 88.5%. The incident detection and response capabilities proved remarkably effective, with detection accuracy reaching 94.5% for minor incidents and 97.2% for major accidents, corresponding to response times of 4.3 minutes for minor incidents and 3.8 minutes for major accidents. A key strength of our approach lies in its ability to balance multiple competing objectives dynamically. The framework successfully minimized the trade-off between traffic flow optimization and resource allocation efficiency through its dual-objective optimization network. The system maintained stable performance during demand variations of $\pm 15\%$, with resource utilization remaining above 85% even under increased demand scenarios. This adaptability was further proved by the system's robust response to various incident types, from minor accidents to weather-related disruptions. The comparative analysis against baseline methods revealed consistent superiority of our approach across all performance metrics. The MARL outperformed centralised single-agent and independent learning approaches, particularly in complex coordination and rapid-response scenarios. The model's ability to maintain high performance during routine operations and unpredictable events proves its potential for real-world implementation. However, several limitations and opportunities for future research emerge from this work. While the current implementation displays promising results for networks of moderate size, further study is required to develop scalability to more extensive urban networks. Additionally, integrating more sophisticated prediction models for extreme weather events and particular environments could enhance the system's robustness. Future work should explore the incorporation of emerging transportation technologies, such as connected and autonomous vehicles, into the model.

CRedit Author Statement

The authors confirm contribution to the paper as follows:

Conceptualization: Shaymaa H Nowfa, Revathi S, Kolluru Suresh Babu, Amarendra K, Abdul Muthalief Mohamed Anwar and Aseel Smerat; **Methodology:** Shaymaa H Nowfa, Revathi S and Kolluru Suresh Babu; **Software:** Amarendra K, Abdul Muthalief Mohamed Anwar and Aseel Smerat; **Data Curation:** Shaymaa H Nowfa, Revathi S and Kolluru Suresh Babu; **Writing- Original Draft Preparation:** Shaymaa H Nowfa, Revathi S, Kolluru Suresh Babu, Amarendra K, Abdul

Muthalief Mohamed Anwar and Aseel Smerat; **Visualization:** Shaymaa H Nowfa, Revathi S and Kolluru Suresh Babu; **Investigation:** Amarendra K, Abdul Muthalief Mohamed Anwar and Aseel Smerat; **Supervision:** Shaymaa H Nowfa, Revathi S and Kolluru Suresh Babu; **Validation:** Amarendra K, Abdul Muthalief Mohamed Anwar and Aseel Smerat; **Writing- Reviewing and Editing:** Shaymaa H Nowfa, Revathi S, Kolluru Suresh Babu, Amarendra K, Abdul Muthalief Mohamed Anwar and Aseel Smerat; All authors reviewed the results and approved the final version of the manuscript.

Data Availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Conflicts of Interests

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing Interests

The authors declare no conflict of interest.

Consent to publish

All the authors gave permission to Consent to publish.

References

- [1]. A. Ait Ouallane, A. Bakali, A. Bahnasse, S. Broumi, and M. Talea, "Fusion of engineering insights and emerging trends: Intelligent urban traffic management system," *Information Fusion*, vol. 88, pp. 218–248, Dec. 2022, doi: 10.1016/j.inffus.2022.07.020.
- [2]. M. Nama et al., "Machine learning-based traffic scheduling techniques for intelligent transportation system: Opportunities and challenges," *International Journal of Communication Systems*, vol. 34, no. 9, Apr. 2021, doi: 10.1002/dac.4814.
- [3]. A. Alhussen and A. S. Ansari, "Real-Time Prediction of Urban Traffic Problems Based on Artificial Intelligence-Enhanced Mobile Ad Hoc Networks (MANETS)," *Computers, Materials & Continua*, vol. 79, no. 2, pp. 1903–1923, 2024, doi: 10.32604/cmc.2024.049260.
- [4]. N. T. F. E, V. P. A, S. K. Shanmugam, K. K. V, S. Sengan, and A. M. B. C, "Adaptive Approach to Anomaly Detection in Internet of Things Using Autoencoders and Dynamic Thresholds," *Journal of Machine and Computing*, pp. 1–10, Jan. 2024, doi: 10.53759/7669/jmc202404001.
- [5]. P. Razzaghi et al., "A survey on reinforcement learning in aviation applications," *Engineering Applications of Artificial Intelligence*, vol. 136, p. 108911, Oct. 2024, doi: 10.1016/j.engappai.2024.108911.
- [6]. N. Nezamoddini and A. Gholami, "A Survey of Adaptive Multi-Agent Networks and Their Applications in Smart Cities," *Smart Cities*, vol. 5, no. 1, pp. 318–347, Mar. 2022, doi: 10.3390/smartcities5010019.
- [7]. H. M. Ali, J. J. J, S. G, T. Palanisamy, V. Rachapudi, and S. Sengan, "Operating Cash Flow Ranking Using Data Envelopment Analysis with Network Security Driven Blockchain Model," *Journal of Machine and Computing*, pp. 1839–1851, Jul. 2025, doi: 10.53759/7669/jmc202505144.
- [8]. S. Abirami, M. Pethuraj, M. Uthayakumar, and P. Chitra, "A systematic survey on big data and artificial intelligence algorithms for intelligent transportation system," *Case Studies on Transport Policy*, vol. 17, p. 101247, Sep. 2024, doi: 10.1016/j.cstp.2024.101247.
- [9]. Omamode Henry Orieno, Ndubuisi Leonard Ndubuisi, Valentine Ikenna Ilojiana, Preye Winston Biu, and Beryl Odonkor, "THE FUTURE OF AUTONOMOUS VEHICLES IN THE U.S. URBAN LANDSCAPE: A REVIEW: ANALYZING IMPLICATIONS FOR TRAFFIC, URBAN PLANNING, AND THE ENVIRONMENT," *Engineering Science & Technology Journal*, vol. 5, no. 1, pp. 43–64, Jan. 2024, doi: 10.51594/estj.v5i1.721.
- [10]. H. B. Faheem, A. M. E. Shorbagy, and M. E. Gabr, "Impact Of Traffic Congestion on Transportation System: Challenges and Remediations - A review," *Mansoura Engineering Journal*, vol. 49, no. 2, Jan. 2024, doi: 10.58491/2735-4202.3191.
- [11]. S. H. Nowfal, V. B. Sadu, S. Sengab, R. G, A. N. R, and S. K, "Genetic Algorithms for Optimized Selection of Biodegradable Polymers in Sustainable Manufacturing Processes," *Journal of Machine and Computing*, pp. 563–574, Jul. 2024, doi: 10.53759/7669/jmc202404054.
- [12]. H. M. A. Ghanimi, A. K, H. P. A, S. Sengan, B. P. Prakash, and R. K. Bommisetti, "An Ensemble Cognitive Model for Stroke Prediction Using Unstructured Health Information Powered by Machine Learning," *Journal of Machine and Computing*, pp. 611–621, Jan. 2025, doi: 10.53759/7669/jmc202505048.
- [13]. A. Agrahari, M. M. Dhabu, P. S. Deshpande, A. Tiwari, M. A. Baig, and A. D. Sawarkar, "Artificial Intelligence-Based Adaptive Traffic Signal Control System: A Comprehensive Review," *Electronics*, vol. 13, no. 19, p. 3875, Sep. 2024, doi: 10.3390/electronics13193875.
- [14]. A. A. Musa, S. I. Malami, F. Alanazi, W. Ounaies, M. Alshammari, and S. I. Haruna, "Sustainable Traffic Management for Smart Cities Using Internet-of-Things-Oriented Intelligent Transportation Systems (ITS): Challenges and Recommendations," *Sustainability*, vol. 15, no. 13, p. 9859, Jun. 2023, doi: 10.3390/su15139859.
- [15]. G. Vadivel, M. J. M. Hussain, S. V. Tresa, and . S., "Smart Transportation Systems: Iot-Connected Wireless Sensor Networks For Traffic Congestion Management," *International Journal Of Advances In Signal And Image Sciences*, vol. 9, no. 1, pp. 40–49, Jun. 2023, doi: 10.29284/ijasis.9.1.2023.40-49.
- [16]. F. T. Ayasrah, N. S. Alsharafa, S. S, S. B, S. Sengan, and K. N, "Strategizing Low Carbon Urban Planning Through Environmental Impact Assessment by Artificial Intelligence Driven Carbon Foot-print Forecasting," *Journal of Machine and Computing*, pp. 1140–1151, Oct. 2024, doi: 10.53759/7669/jmc202404105.
- [17]. Spring, N. (2024). Multi-Agent Systems in Vehicular Edge Computing: A Communication-Centric Approach to Task Handovers (Doctoral dissertation, Technische Universität Wien).
- [18]. T. Ahmed, H. Liu, and V. V. Gayah, "C-MP: A decentralized adaptive-coordinated traffic signal control using the Max Pressure framework," *Transportation Research Part B: Methodological*, vol. 200, p. 103308, Oct. 2025, doi: 10.1016/j.trb.2025.103308.
- [19]. A. G. Ismaeel et al., "Enhancing Traffic Intelligence in Smart Cities Using Sustainable Deep Radial Function," *Sustainability*, vol. 15, no. 19, p. 14441, Oct. 2023, doi: 10.3390/su151914441.
- [20]. C. Chen et al., "Toward A Thousand Lights: Decentralized Deep Reinforcement Learning for Large-Scale Traffic Signal Control," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, pp. 3414–3421, Apr. 2020, doi: 10.1609/aaai.v34i04.5744.

- [21]. A. M. Ali, M. A. Ngadi, I. I. Al-Barazanchi, and P. S. JosephNg, “Intelligent Traffic Model for Unmanned Ground Vehicles Based on DSDV-AODV Protocol,” *Sensors*, vol. 23, no. 14, p. 6426, Jul. 2023, doi: 10.3390/s23146426.
- [22]. K. S. Mohsin, J. Mettu, C. Madhuri, G. Usharani, S. N., and P. Yellamma, “Enhancing Urban Traffic Management Through Hybrid Convolutional and Graph Neural Network Integration,” *Journal of Machine and Computing*, pp. 360–370, Apr. 2024, doi: 10.53759/7669/jmc202404034.
- [23]. H. M. A. Ghanimi, F. T. Ayasrah, V. C. Jadala, M. T C, B. K., and S. B., “An Innovative Artificial Intelligence Based Decision Making System for Public Health Crisis Virtual Reality Rehabilitation,” *Journal of Machine and Computing*, pp. 561–575, Jan. 2025, doi: 10.53759/7669/jmc202505044.
- [24]. N. S. Alsharafa, S. R., R. J. Krishna, V. K. Sonthi, P. S M, and M. P., “Optimizing Building Energy Management with Deep Reinforcement Learning for Smart and Sustainable Infrastructure,” *Journal of Machine and Computing*, pp. 381–391, Apr. 2024, doi: 10.53759/7669/jmc202404036.
- [25]. Haldorai, B. L. R., S. Murugan, and M. Balakrishnan, “Artificial Intelligence for Intelligent Spectrum Management Toward Futuristic Communication,” *EAI/Springer Innovations in Communication and Computing*, pp. 235–248, 2024, doi: 10.1007/978-3-031-53972-5_12.

Publisher’s note: The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. The content is solely the responsibility of the authors and does not necessarily reflect the views of the publisher.