This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530243
IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

pDate of publication xxxx 00, 0000, date of current version xxxx 00, 0000.
Digital Object Identifier 10.1109/ACCESS.2022.Doi Number

Traffic Congestion Control with Emergency Awareness and Optimized Communication Infrastructure using Reinforcement Learning and Non-Dominated Sorting Genetic Algorithm

Othman S. Al-Heety[1], Zahriladha Zakaria[1], Ahmed Abu-Khadrah[2], Mahamod Ismail[3], Mohammed Mudhafar Shakir[4], Sameer Alani[5] and Hussein Alsariera[1]
[1] Faculty of Electronic and Computer Engineering, Universiti Teknikal Malaysia Melaka (UTeM), Durian Tunggal, Malaysia
[2] Department of Electrical Engineering, College of Engineering Technology, Al-Balqa Applied University, Amman, Jordan
[3] Department of Electrical, Electronic and System Engineering, Universiti Kebangsaan Malaysia, UKM Bangi, Selangor, Malaysia
[4] Faculty of Engineering and Electronic Engineering, Universiti Tun Hussein Onn Malaysia (UTHM), Batu Pahat, Malaysia
[5] Computer Center, University of Anbar, Anbar 5543, Iraq

*Corresponding Authors: Zahriladha Zakaria. Email: zahriladha@utem.edu.my; Othman S. Al-Heety. Email: alheety567@gmail.com

ABSTRACT Urban centers are grappling with increasing traffic congestion, which hampers mobility and reduces the effectiveness of emergency responses. Current traffic management systems struggle to dynamically prioritize emergency vehicles during peak congestion, highlighting the need for an advanced solution that improves both everyday traffic flow and emergency response efficiency. Traditional traffic light systems typically follow either fixed schedules ("Fixed" approach) or make real-time adjustments without achieving long-term efficiency ("Greedy" approach), often failing to effectively balance regular traffic dynamics with emergency needs. We propose the "Avg Log" method, an adaptive Q-learning strategy that enhances conventional traffic management systems by incorporating an emergency-aware module. This module uses real-time data to prioritize emergency vehicles and adjust traffic signals accordingly. Additionally, our approach employs a multi-objective optimization algorithm to strengthen communication infrastructure, optimizing latency, throughput, and reliability. Through simulations in a multi-intersection urban setting, we compared the performance of our method against the standard "Greedy" and "Fixed" approaches. Key metrics assessed included average waiting times, system throughput, and the efficiency of emergency vehicle passage. The "Avg Log" method outperformed traditional methods, showing superior long-term adaptability and learning. It demonstrated lower average waiting times and effectively managed emergency situations, significantly reducing delays for emergency vehicles compared to both control methods. This method provides an efficient balance between overall traffic flow and emergency prioritization, suggesting a scalable solution for complex urban environments.

INDEX TERMS Traffic Congestion Control, Smart Cites, Intelligent Transportation System (ITS), Vehicular Communication, Internet of Vehicles, Reinforcement Learning, Multi-Objective Optimization.

I. INTRODUCTION

Urban traffic congestion has emerged as a critical challenge in modern cities, significantly impacting emergency response times and overall mobility efficiency [1]. The increasing urbanization rate, coupled with rising vehicle ownership, has created unprecedented pressure on existing traffic management systems,4 particularly in densely populated metropolitan areas [2]. Despite significant technological advancements in intelligent transportation systems, current traffic management solutions struggle with efficiently prioritizing emergency vehicles during peak congestion while maintaining optimal traffic flow across complex urban networks [3], [4].

Traditional traffic management systems, which typically follow either fixed schedules or simple responsive algorithms, have proven inadequate for handling the dynamic nature of urban traffic, especially during emergencies [5]. These conventional approaches lack the capability to dynamically adjust to rapid changes in traffic patterns, effectively prioritize emergency vehicles, or scale across multiple intersections [6]. Furthermore, the quality of vehicle-to-infrastructure (V2I) communication significantly impacts system performance, with current deployments often suffering from inconsistent coverage, high latency, and limited reliability during peak traffic periods [7], [8].

Recent studies have explored reinforcement learning (RL) for adaptive traffic control, demonstrating promising results in optimizing traffic flow (Martinez et al., 2024). However, existing implementations typically address either general traffic management or emergency vehicle prioritization in isolation, failing to integrate both aspects effectively [9]. Additionally, the effectiveness of RL-based systems is often constrained by suboptimal communication infrastructure placement and configuration, leading to delayed responses and reduced efficiency (Wang and Liu, 2024).

The optimization of communication infrastructure, particularly the strategic placement of roadside units (RSUs), plays a crucial role in system performance [10]. Research has shown that poorly optimized RSU placement can lead to coverage gaps, increased latency, and reduced reliability in vehicle-to-infrastructure communication [11]. These limitations significantly impact the system's ability to respond to emergency situations and maintain efficient traffic flow [12].

To address these challenges comprehensively, this paper proposes an integrated approach that combines an emergency-aware adaptive Q-learning system with multi-objective optimization of communication infrastructure using the Non-dominated Sorting Genetic Algorithm II (NSGA-II). Our research introduces the "Avg Log" method, an enhanced Q-learning strategy that incorporates emergency awareness while maintaining efficient overall traffic flow. This approach is complemented by a sophisticated multi-objective optimization framework for RSU placement that simultaneously considers coverage, cost, latency, and reliability [13].

Through extensive simulations and analysis, we demonstrate the effectiveness of our integrated system across multiple intersections, validating its scalability and adaptability in various traffic conditions [14]. The results show significant improvements in both emergency vehicle response times and overall traffic flow efficiency compared to traditional methods and existing RL-based approaches [15]. Our system successfully balances the competing demands of emergency vehicle prioritization and regular traffic flow optimization, providing a robust solution for modern urban traffic management challenges [16]. The remainder of this paper is organized as follows: Section II presents essential background information and terminology critical for understanding our approach. Section III outlines our main contributions. Section IV provides a comprehensive literature review of current approaches in traffic management and their limitations, highlighting the research gaps our work addresses. Section V details our methodology and algorithms, including the novel "Avg Log" method and RSU optimization framework. Section VI presents experimental results and analysis, demonstrating the system's effectiveness in real-world scenarios. Section VII discusses the practical implementation aspects, including security considerations and ethical implications. Finally, Section VIII concludes the paper with a discussion of findings and directions for future research.

II. CONTRIBUTIONS

This article contributes several significant advancements to the field of intelligent traffic management systems, aiming to improve urban mobility and emergency response capabilities.

A. EMERGENCY-AWARE ADAPTIVE Q-LEARNING:

Our approach introduces a novel integration of an emergency-specific overriding process within the Q-learning framework. This allows the system to dynamically recognize and prioritize emergency vehicles in real-time, adjusting traffic signals accordingly. This emergency-aware module is a significant advancement over existing RL approaches in traffic management, which typically do not account for emergency scenarios in such a dynamic manner.

B. MULTI-OBJECTIVE COMMUNICATION INFRASTRUCTURE OPTIMIZATION:

We introduce a unique application of the Non-Dominated Sorting Genetic Algorithm II (NSGA-II) to optimize the communication infrastructure supporting our traffic management system. This optimization considers multiple objectives simultaneously (coverage, cost, latency, and reliability), which is a novel approach in the context of traffic management systems. The integration of this optimized communication infrastructure with the RL-based traffic control system represents a significant advancement in creating a more robust and efficient urban traffic management solution.

C. INTEGRATED RESILIENT TRAFFIC MANAGEMENT SYSTEM:

Our work is novel in its comprehensive integration of emergency awareness, adaptive learning, and optimized communication infrastructure. This holistic approach addresses multiple aspects of urban mobility challenges simultaneously, which is not commonly found in existing literature.

D. NOVEL REWARD FUNCTION DESIGN:

We introduce and compare several innovative reward functions (as shown in Table 3 of our manuscript) that incorporate both emergency and non-emergency scenarios. These reward functions are designed to balance the needs of regular traffic flow with urgent emergency responses, representing a new approach to reward shaping in traffic management RL systems.

E. MULTI-INTERSECTION SCALABILITY ASSESSMENT:

Our approach has been thoroughly evaluated in a multi-intersection scenario, supporting the assessment of decentralized management. This evaluation demonstrates the system's ability to effectively manage traffic across various intersections independently, showcasing its potential for scalable and flexible deployment in complex urban environments.

These contributions collectively represent a significant advancement in intelligent traffic management systems, offering a more adaptive, efficient, and responsive solution to urban mobility challenges.

Collectively, these contributions represent a significant advancement in the development of smarter, more responsive traffic management systems capable of navigating the challenges of modern urban living. The remainder of the article is organized as follows: Section 2 presents a comprehensive literature review of traffic congestion control. Section 3 outlines the article's methodology and the algorithms developed. Section 4 presents experimental results and analysis. The conclusion and future works are covered in Section 5.

III. LITERATURE SURVEY

The advancement of intelligent traffic management systems (ITMS) is crucial in tackling the growing challenges of urban transportation, including congestion, emergency response, and infrastructure efficiency. This literature review explores a range of innovative traffic management methods, concentrating on congestion monitoring, traffic signal control, emergency vehicle prioritization, and vehicular ad hoc network optimization.

A. TRAFFIC CONGESTION MONITORING AND CONTROL :

A [17] details a new method involving an enhanced k-nearest neighbors (kNN) approach for traffic congestion monitoring. This study presents a hybrid model that combines piecewise switched linear traffic modeling with a Kalman filter to improve real-time traffic monitoring. The integration of kNN results in precise congestion detection, establishing a new standard for traffic management efficiency [18].

In another key contribution, [3] introduces a hybrid approach for predicting and controlling traffic congestion. This method utilizes advanced machine learning techniques, such as improved particle swarm optimization and feature fusion models, to effectively forecast and manage urban traffic flows. The application of real-world data from Shenyang Station demonstrates its practical use and superior performance compared to traditional models.

B. ADVANCED TRAFFIC CONTROL STRATEGIES:

[19] examines the use of reinforcement learning (RL) for traffic signal control. The study employs a linear function approximation with the true online SARSA (λ) algorithm, showing significant improvements in reducing congestion and delays across multiple signalized intersections. This approach highlights the potential of RL to enhance traffic signal responsiveness and system adaptability. Additionally, [20] presents a cooperative speed limit control system using distributed RL. This system allows multiple traffic control agents to collaborate, optimizing freeway traffic flow and safety. This example showcases the advantages of distributed learning in traffic management systems.

C. EMERGENCY VEHICLE PRIORITIZATION:

[21] introduces a new real-time traffic signal control strategy designed specifically for emergency vehicles. The approach employs on-demand signal timing and innovative signal preemption techniques to allow emergency vehicles to traverse urban areas quickly and efficiently. These advancements are vital for shortening emergency response times and improving public safety.

D. INFRASTRUCTURE EFFICIENCY AND OPTIMIZATION:

[22] Critically examines deployment strategies for roadside units (RSUs) in vehicular ad hoc networks (VANETs). The study emphasizes optimizing RSU placement to enhance communication efficiency and coverage, which is vital for supporting the wider framework of intelligent transportation systems.

E. MULTIMODEL AND COOPERATIVE TRAFFIC MANAGEMENT SYSTEMS:

In [23], proposes a digital twin-assisted real-time traffic data prediction method. This approach leverages 5G-enabled Internet of Vehicles to improve traffic flow predictions, aiding in more informed and dynamic traffic management decisions. [24] discusses a cooperative multi-agent system using game theory and reinforcement learning for traffic signal control. This system successfully reduces average delay times across networks, demonstrating the potential of cooperative strategies in congested urban environments. Table 1 provides an analysis of existing traffic management based on different criteria, and Table 2 offers an overview of existing algorithms for traffic congestion control using reinforcement learning.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530243

TABLE 1
COMPARATIVE ANALYSIS OF EXISTING TRAFFIC MANAGEMENT FROM VARIOUS CRITERIA, NAMELY, SENSING, COMMUNICATION, TRAFFIC PREDICTION, TRAFFIC CONTROL, MULTI- INTERSECTION, EMERGENCY HANDLING, OPTIMIZATION OF COMMUNICATION INFRASTRUCTURE, AND USED SIMULATOR

Ref	Sensing	Comm.	Traffic prediction	Traffic control	Multi-intersection	Emergency	Comm. infrastructure	Method	Sim.
[17]	Magnetic loop detectors	X	X	√	X	X	X	(PWSL) modelling approach and (KF) for traffic congestion monitoring	X
[3]	X	X	√	√	X	X	X	Improved particle swarm optimization (IPSO)	VISSIM
[19]	X	signalized intersections	X	√	√	X	X	RL on Linear Function Approximation	MATSim
[20]	X	V2V & V2I	X	√	√	X	X	distributed RL	Matlab & Java& MOTUS
[21]	X	EV	√	√	√	√	X	intrusive preemption and non-intrusive pre-emption methods	SUMO
[22]	X	V2V & V2I	X	X	√	X	√	(DSRC) defined in IEEE 802.11p and IEEE 1609.4	SUMO
[23]	Speed sensor and camera	V2V	√	X	X	X	X	digital twin-assisted real-time traffic data prediction method	Python
[25]	X	V2V & V2I &V2X	√	√	√	√	X	collaborative model of traffic clear-out and signal preemption algorithm	Python with SUMO
[26]	X	V2V	X	√	√	X	X	centralized traffic light control system, by unique wireless communication network.	MatLab
[27]	X	X	X	√	√	X	X	RL especially hierarchical action space	Cityflow
[28]	video cameras and loop detectors	CAVs	√	√	√	X	X	RL with (B-INT) and deep deterministic policy gradient (TD3) algorithm	SUMO
[29]	surveillance cameras	Multiple traffic light	X	√	√	X	X	DRL and DDPG	SUMO
[30]	video camera	X	X	√	X	X	X	Possibilistic Fuzzy C-Means clustering	Wi-Fi and Bluetooth and RSSI
[31]	vehicle speed sensors and road side sensors	X	√	X	X	X	X	ChebNet	PeMS04 and PeMS08 dataset
[32]	vehicle location, speed and number, and use (SDN)	V2V&V2I	√	X	X	X	X	DDaaS based on SL	SUMO
[33]	camera or a loop detector	V2I	X	√	√	X	√	(DSRC) technology (RL) approach in (AI)	SUMO
Ours	(camera and weight cell)	V2V and V2I	√	√	√	√	√	RL and NSGA-II	Python and SUMO

TABLE 2
EXISTING ALGORITHMS FOR TRAFFIC CONGESTION CONTROL USING RL, NAMELY, SENSING, COMMUNICATION, APPLICATION, AGENT, ACTION, STATE, REWARD, EMERGENCY SERVICES, AND LIMITATION.

Ref	Sensing	Comm.	App	Agent	Action	State	Reward	Emergency	Limitation
[19]	X	Signalized intersections	RL algorithm is applied to traffic signal control in urban road networks to manage congestion and delays	Multi	Decisions made by agents about the traffic light phases	Encompasses the current observable conditions at an intersection	The feedback provided to the agents based on the efficacy of their actions	X	Curse of dimensionality
[20]	X	√	Novel variable speed limit control system under the V2I environment to	Multi	Set of combined actions of all agents in the variable speed	A finite set extracted from the freeway traffic state	aims to maintain segment density slightly under	X	It is dedicated for variable speed control

			optimize freeway traffic mobility and safety		limit control system consists of speed limits		the critical density to optimize traffic mobility and safety		
[23]	Speed sensor and camera	V2V	Improve intelligent traffic systems using a digital twin-assisted method to predict real-time traffic data	X	Implementation of the proposed prediction method to estimate missing traffic data	The current traffic conditions, sensor data availability, and other relevant environmental conditions that the prediction model considers	The improvement in traffic flow, the reduction of congestion	X	Predicting traffic data in real-time due to sensor data sparsity and the high volume of data being produced
[24]	X	V2V and V2I	Traffic congestion and improve traffic flow	Multi	The various splits of the green times	The number of vehicles waiting at the intersection	Reducing traffic congestion	X	Scalability issue
[27]	X	X	Autonomous traffic signal control	Multi	Selecting traffic light phases with variable durations which determine the rights-of-way for vehicles at intersections	The real-time traffic situation at intersections	Incorporates a fairness index based on user satisfaction and guide the learning process	X	The unfairness problem in traditional RL methods for TSC which may results very long waiting time
[28]	Video cameras and loop detector	Cavs	- RL for variable speed limit control at freeway recurrent bottlenecks - Traffic flow improvement through V2X communication and bottleneck prediction	Both	Adjustments to speed limits, changing traffic signal timings, and directing lane changes or vehicle routing	Various real-time and historical traffic conditions such as vehicle density, vehicle speed	Improvement in traffic flow or reduction in congestion can be measured in terms of decreased travel time, reduced traffic congestion	X	Accuracy of traffic prediction models, the scalability and adaptability of the solutions in different traffic conditions
[29]	Surveillance cameras	Multiple traffic light	Alleviate traffic congestion at city intersections by applying deep RL to control traffic lights	Multi	Controlling the traffic lights	Each intersection is defined by the position and velocity of vehicles in each lane, the phase of the traffic light, the queue length of each lane, the number of pedestrians waiting to cross	Optimizing traffic flow, which could include minimizing the queue lengths, waiting times, and ensuring smooth pedestrian movement.	X	The complexity of coordinating multiple agents and the unstable environment and partial observability
[33]	Camera or a loop detector	V2I	Applies RL to create an intelligent traffic signal control system that can operate effectively even with partial vehicle detection.	Single	Retaining the current traffic light phase or switching to the next phase	The distance to the nearest vehicle at each approach number of vehicles at each approach, amber phase indicator, current traffic light phase elapsed time, and current time	Minimize delay by maximizing vehicle speed	X	Lacking of reason behing selection of the state representation
[34]	X	X	Adaptive traffic signal control in complex urban traffic networks	Multi	The phase duration of each traffic signal, which can be increased or decreased by a certain amount	A combination of local traffic information, including the number of vehicles in each lane, the queue length, and the waiting time	The negative of the total waiting time of all vehicles in the network, which is discounted by a factor of gamma.	X	Ignoring reality of sensor information delay
[35]	X	X	Traffic signal control	Single	Binary variable as follows: “0” for continue, and “1” for switch	The number of vehicles of traffic flow and traffic signal configuration	Negative congestion cost	X	Ignoring communication sensing
[36]	X	X	Traffic signal control	Multi	The green light phase duration	Position matrix	Cumulative waiting time of all vehicles	X	Ignoring communication sensing
[37]	(AD-RLTSC)	Wireless network (cavs)	Advance Decision-making Reinforcement	Multi	Separated into two processes: 'decision' and	Decide the future signal light status	The average of cumulative	X	Not been implemented in a real-world

	algorithm		Learning Traffic Signal Control		'control'. The agent taking actions based on the state of the environment		rewards over a control cycle		scenario and weather conditions on traffic flow does not considered
[38]	X	V2V and V2I	Vehicle Routing	Single	Left-permitted actions	Information of multi-intersection	Piecewise	X	Not including traffic congestion control for vehicle
[39]	X	V2V and V2I	Traffic congestion control	Multi	Left turn permitted and left turn protected actions	Vehicle volumesp, average speed, and average accel-oration	Negative Pressure	X	Ignoring sensing and emergency vehicles
[40]	LiDAR-Based for Privacy and Range Limits	X	Robotic Vehicle Congestion Control		Acceleration Control for Flexibility	Partial Observability and Key Features	Shaping Behavior with Weighted Objectives	√	Focusing on Robot Vehicles

Despite the reviewed advancements, several important gaps persist in the algorithms:

1. Sensing Accuracy and Delays: Many algorithms fail to properly consider the delay and inaccuracies in sensor data, which are critical for real-time adaptive traffic control. Although methods like the one discussed in [23] use advanced sensing technologies, the broader field lacks a systematic strategy for minimizing sensor data delays across different traffic management systems. Improving sensor accuracy and reducing delays can greatly enhance traffic management outcomes by ensuring that the data used for decision-making accurately reflects current traffic conditions.
2. Communication Optimization: Most current algorithms do not fully leverage the potential of vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications. As highlighted in studies such as [22], despite some progress, there remains significant room for further optimization of these communications. Strengthening V2V and V2I communication capabilities could result in more efficient traffic signal control and adaptive responses to changing traffic conditions, especially in dynamic and crowded urban areas.
3. Emergency Handling: Existing traffic management algorithms lack integrated mechanisms for handling emergency situations. References such as [21] showcase strategies for prioritizing emergency vehicles through traffic signal control. Nonetheless, systematically incorporating these features across all traffic management systems is crucial. Creating robust systems that prioritize emergency vehicles and adjust traffic signals during emergencies without disrupting overall traffic flow is vital for public safety and efficient city operations.

IV. METHODOLOGY

This section of the paper describes the methods used to tackle traffic congestion control challenges, emphasizing emergency vehicle prioritization and optimized communication infrastructure. By utilizing advanced algorithms such as Reinforcement Learning (RL) and the Non-Dominated Sorting Genetic Algorithm II (NSGA-II), this study creates and tests a comprehensive approach designed to enhance urban mobility and emergency response in congested settings

A. Key Concepts Definitions

This section provides definitions and context for key technical concepts used throughout this paper.

1) Emergency-Aware Traffic Control

A dynamic traffic management system capable of detecting and prioritizing emergency vehicles while maintaining efficient overall traffic flow. The system uses real-time data to adjust traffic signals and routing to facilitate emergency vehicle passage without severely disrupting regular traffic patterns.

2) Roadside Units (RSUs)

Physical infrastructure components installed along roadways that enable vehicle-to-infrastructure (V2I) communication. RSUs serve as communication nodes, collecting and transmitting data between vehicles and the traffic management system. Their strategic placement is crucial for optimal network coverage and performance.

3) Vehicle Communication Protocols

- **Vehicle-to-Vehicle (V2V)**: Direct communication between vehicles using IEEE 802.11p protocol, enabling real-time sharing of position, speed, and trajectory data.
- **Vehicle-to-Infrastructure (V2I)**: Communication between vehicles and roadside infrastructure, facilitating traffic management and safety applications.

B. Theoretical Foundation

The theoretical underpinning of our Integrated Traffic Management System is rooted in the fundamental interdependence between system infrastructure and operational efficiency in complex urban environments. The integration of a Planning Phase with an Operation Phase is necessitated by the dynamic nature of urban traffic systems, where the performance of real-time control strategies is inherently constrained by the underlying communication infrastructure.

The effectiveness of the Operation Phase, which employs reinforcement learning for adaptive traffic control, is critically dependent on the optimality of the Planning Phase outcomes. This dependency can be formalized through the concept of state-action space coverage. Let S be the set of all possible traffic states and A the set of all possible control actions. The quality of real-time decision-making in the Operation Phase is bounded by the function:

$$Q(s,a) \leq Q^*(s,a)\ \forall s \in S, a \in A \quad (1)$$

where $Q(s,a)$ is the learned action-value function and $Q^*(s,a)$ is the optimal action-value function. The achievable optimality of $Q(s,a)$ is directly influenced by the communication infrastructure established in the Planning Phase, which determines the observability and controllability of the system.

Let X be the set of RSU deployments determined by the Planning Phase. We can define a function $f(X)$ that maps the RSU deployment to a measure of system observability and controllability. The relationship between the Planning and Operation phases can then be expressed as:

$$maxQ(s,a) = g(f(X)) \quad (2)$$

where g is a monotonically increasing function. This formulation illustrates that the upper bound of operational performance ($\max Q(s,a)$) is a function of the infrastructure quality determined by the Planning Phase.

This theoretical framework underscores the necessity of integrating longterm planning with short-term operations in urban traffic management. It provides the foundation for our two-phase approach, where improvements in RSU placement directly translate to enhanced capabilities in real-time traffic control.

C. Problem Formulation

This problem extends the traditional traffic congestion control challenge by incorporating emergency awareness. The system is mathematically formulated as follows:

The urban environment consists of multiple junctions, where each junction is an intersection of horizontal and vertical lines. Each line is made up of roads with two directions, and each direction contains multiple lanes. Vehicles enter the environment at all roads intersecting with its border and exit in the outward direction at the same intersections.

Each junction operates in two modes: road-centered and junction-centered. In the road-centered mode, traffic can flow east-west (EW) or north-south (NS), allowing left turns. In the junction-centered mode, actions allow one of the four roads (E, W, N, and S) to open to three directions (forward, left, and right). An additional "emergency-aware mode" prioritizes the passage of emergency vehicles through the junctions by dynamically adjusting actions and states based on real-time emergency data.

Each road is equipped with load cells at both ends to measure vehicle flow and cameras to count vehicles. Vehicles also have IEEE 802.11p for Vehicle-to-Vehicle (V2V) and Vehicle-to-Infrastructure (V2I) communication, facilitated by roadside units at each junction. Vehicles are generated using a normal distribution probability density function for several vehicles and an exponential distribution probability density function for time intervals, while sensors are subject to noise. The primary goals are twofold:

1. To optimize the placement of roadside units (RSUs) based on four criteria: reliability, cost, latency, and coverage.
2. To leverage sensing, V2V, and V2I communication for state estimation using sensor fusion.
3. To control traffic using reinforcement learning, with a focus on emergency-aware traffic management.

D. Main System Algorithm

The main system consists of two interconnected phases: a Planning Phase and an Operation Phase, which work together to create an adaptive and efficient Integrated Traffic Management System for urban environments.

The Planning Phase, which operates periodically, utilizes the NSGA-II (Non-dominated Sorting Genetic Algorithm II) to optimize the placement of Roadside Units (RSUs). This multi-objective optimization process considers crucial factors such as coverage area, installation and operational costs, network latency, and communication reliability. By running at set intervals, this phase allows the system to adapt its infrastructure to changing urban conditions and traffic patterns over time.

The Operation Phase, running continuously, manages day-to-day traffic control using reinforcement learning techniques. It observes the current traffic state, detects emergency vehicles, and makes real-time decisions on traffic signal control. The system employs a Q-learning approach with an epsilon-greedy strategy, balancing exploration of new actions with exploitation of known effective strategies. This phase is particularly notable for its ability to prioritize emergency vehicles when detected, demonstrating the system's capacity to handle critical situations alongside routine traffic management.

The algorithm's structure allows for a symbiotic relationship between these two phases. The optimized RSU placements from the Planning Phase enhance the communication infrastructure, which in turn supports more effective real-time decision-making in the Operation Phase. Simultaneously, the continuous learning and adaptation in the

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530243

Operation Phase provide valuable data and insights that can inform future planning decisions.

This integrated approach enables the system to address both long-term urban planning challenges and immediate traffic control needs. By combining advanced optimization techniques with adaptive learning algorithms, the system offers a flexible and responsive solution capable of evolving with the dynamic nature of urban traffic patterns and emergency situations.

The inclusion of performance metrics tracking factors like average waiting times, system throughput, and emergency response times allows for ongoing evaluation and refinement of the system's effectiveness. This comprehensive approach to urban traffic management represents a significant step forward in creating smarter, more efficient cities that can better serve their residents in both everyday and emergency scenarios. We present the pseudocode of our integrated traffic control in Algorithm 1.

Algorithm 1: Integrated Traffic Management System

Input: E: Urban environment (roads, junctions, emergency routes)
N: Set of candidate locations for RSUs
T: Total simulation time
ΔT: Time step
P: Population size for NSGA-II
G: Number of generations for NSGA-II
α: Learning rate for Q-learning
γ: Discount factor for Q-learning
ϵ: Exploration rate for ϵ-greedy policy
$RSU\text{-}interval$: Time interval for RSU optimization
$emergency\_threshold$: Threshold for emergency vehicle detection
Output: X: Pareto-optimal set of RSU deployments
Q: Learned Q-table for traffic signal control
$per\ formance\text{-}metrics$: Dictionary containing average waiting time, system throughput, emergency response times, and overall congestion levels
Start:
Initialize: $X \leftarrow \emptyset, Q \leftarrow zeros, per\ formance\_metrics \leftarrow$ empty dictionary, $current\_time \leftarrow 0$
while $current\_time < T$ **do**
  Planning Phase:
  if $current\_time$ mod $RSU\ \_interval$= 0 **then**
    $X \leftarrow NSGA\text{-}II(E, N, P, G)$
    Update communication infrastructure based on X
  end
  Operation Phase:
  Observe current state s
  $emergency\text{-}detected \leftarrow$
    $Check\ For\ EmergencyVehicl\ (s, emergency\_threshold)$
  if $emergency\_detected$ **then**
    $a \leftarrow SelectEmergencyAction(s)$
  Else
    if $rand() < \epsilon$ **then**
      $a \leftarrow SelectRandomAction()$
    Else
      $a \leftarrow_a Q(s, a)$
    end
  end
  Execute action a
  Observe new state s' and reward r
  $Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_a$
  $Q(s', a') - Q(s, a)]$
  Update $performance\_metrics$
  $s \leftarrow s'$
  $current\_time \leftarrow current\_time + \Delta T$
end
End :
Finalize $per\ formance\_metrics$
return $X, Q, performance^2_{metrics}$

E. System Architecture

The system follows a layered architecture that begins with an Input Layer comprising Sensors & Communication infrastructure that collects real-time traffic data. This information flows into a dual-component Control Layer consisting of an Emergency Control Unit and an RL (Reinforcement Learning) Traffic Control Unit that work in parallel. The Emergency Control Unit has override capabilities, allowing it to take precedence when emergency situations arise. The control decisions from both units are then transmitted through a Roadside Units Network that forms part of the Infrastructure layer. Finally, at the Output layer, these decisions are implemented through Traffic Signals that directly manage traffic flow. This hierarchical structure enables both routine traffic optimization through the RL unit and immediate emergency response through the override capability of the Emergency Control Unit, creating a comprehensive and responsive traffic management system. In the following sub-sections, we present RL agent design.

1) RL Agent Design

The agent currently operating at the base station could be enhanced with an emergency-aware module. This module would identify emergency situations, such as the need for ambulances or fire trucks to pass quickly. The agent could receive real-time data from emergency services or specialized sensors to detect these scenarios. To make the system emergency-aware, new actions could be added to the existing set A. These could include actions like "clear path for emergency" (CPE) and resume normal operation (RNO). The extended set of actions is $A' = \{N-S, E-W, E, W, N, S, CPE, RNO\}$.

Action

The action set is given as $A = \{N-S, E-W, E, W, N, S\}$ which indicates to the standard actions.

State Model

The state space is determined by the number of vehicles on each road at the intersection. Given that each intersection has four roads, the number of vehicles on each road is used as an input for a threshold variable to classify the road into one of two states: congested (C) or non-congested (N). This results in 16 potential states.

Reward Model

Creating an effective reward function is crucial in reinforcement learning models, particularly for complex systems like traffic congestion control. The reward function acts as a guide for the model, promoting specific behaviors over others to achieve the desired results. In this context, we are examining four proposed reward functions that incorporate different aspects of traffic congestion, concentrating on the average and maximum waiting times on roads, throughput, and a constant factor. Let's break down each equation to understand how they impact traffic congestion control.

$$r(t) = \log\left(w \tanh\left(std(AWT)\right) + (1 - w)\tanh\left(\tau^{th}\right)\right) \tag{3}$$

Where:
AWT : Average Waiting Time over roads
th: Throughput
τ : Constant
w : Weight factor between 0 and 1

This equation prioritizes a balance between the standard deviation of the average waiting time and throughput, using a hyperbolic tangent function for modulation. The focus on the standard deviation of AWT aims to minimize variability, promoting consistency across the network. The hyperbolic tangent function scales the reward value between -1 and 1, providing resistance to extreme values. Additionally, the logarithm ensures that the reward increases at a diminishing rate, fostering improvements without excessively rewarding marginal gains.

The inclusion of throughput in the reward function supports a holistic approach to traffic congestion control, emphasizing not only reduced waiting times but also the efficient flow of vehicles throughout the network. By modulating the throughput term with a hyperbolic tangent function and balancing it against waiting time variability, the reward function seeks to promote traffic control strategies that are both effective and fair. This nuanced strategy underlines the complexity of traffic management and the need for sophisticated models that can handle trade-offs between multiple objectives simultaneously.

We present various versions of the reward model in Table 3, outlining their focus, waiting time function, throughput, and perspective.

TABLE 3
COMPARATIVE ANALYSIS OF REWARD MODELS FOR TRAFFIC CONGESTION CONTROL: THIS TABLE OUTLINES THREE DISTINCT REWARD FUNCTIONS DESIGNED TO OPTIMIZE TRAFFIC FLOW BY FOCUSING ON DIFFERENT METRICS—MAXIMUM WAITING TIME (MWT) AND AVERAGE WAITING TIME (AWT)—AND EMPLOYING VARIOUS MATHEMATICAL FUNCTIONS

Model	Focus	Waiting time function	Throughput function	Perspective
$r(t) = \log\left(w \tanh\left(std(AWT)\right) + (1 - w)\tanh\left(\tau^{th}\right)\right)$	Average Waiting Time (AWT)	Standard deviation of AWT scaled by tanh	Throughput scaled by tanh	The hyperbolic tangent functions ensure that the reward value is scaled between -1 and 1, making it robust against extreme values. The logarithm further ensures that the reward grows at a decreasing rate, encouraging improvements without disproportionately rewarding marginal gains.
$r(t) = \log\left(w \tanh\left(std(MWT)\right) + (1 - w)\tanh\left(\tau^{th}\right)\right)$	Maximum Waiting Time (MWT)	Standard deviation of MWT scaled by tanh	Throughput scaled by tanh	Prioritizes reducing the worst-case waiting times while balancing overall throughput. The hyperbolic tangent function ensures scaling within [-1, 1], promoting a balanced approach to improving both the peak waiting times and system throughput.
$r(t) = \log\left(w\, \mathrm{sig}\left(std(AWT)\right) + (1 - w)\mathrm{sig}\left(\tau^{th}\right)\right)$	Average Waiting Time (AWT)	Standard deviation of AWT scaled by sig	Throughput scaled by sig	Focuses on minimizing the variability in average waiting times across the network, with a positive reinforcement towards improving throughput. The sigmoid function maps both components into a (0, 1) range, ensuring a positive and incremental reward growth.
$r(t) = \log\left(w\, \mathrm{sig}\left(std(MWT)\right) + (1 - w)\mathrm{sig}\left(\tau^{th}\right)\right)$	Maximum Waiting Time (MWT)	Standard deviation of MWT scaled by sig	Throughput scaled by sig	Aims at reducing the variability in the maximum waiting times, thereby lessening extreme delays. Utilizing the sigmoid function for both MWT and throughput offers a gentle and positive reinforcement for improvements, promoting a gradual enhancement in both peak waiting times and traffic flow efficiency.
$r(t) = \log\left(w\, \mathrm{sig}\left(std(\Delta MWT_t)\right) + (1 - w)\mathrm{sig}\left(\tau^{th}\right)\right)$	Maximum waiting time	Derivative	Scaled by sigmoid	Aims at reducing the variability in the derivative of maximum waiting times, thereby lessening extreme delays. Utilizing the sigmoid function for both MWT and throughput offers a gentle and positive reinforcement for improvements, promoting a gradual enhancement in both peak waiting times and traffic flow efficiency.
$r(t) = \log\left(w\, \mathrm{sig}\left(std(\Delta AWT_t)\right) + (1 - w)\mathrm{sig}\left(\tau^{th}\right)\right)$	Average waiting time	Derivative	Scaled by sigmoid	Aims at reducing the variability in the derivative of average waiting times, thereby lessening extreme delays. Utilizing the sigmoid function for both MWT and throughput offers a gentle and positive reinforcement for improvements, promoting a gradual enhancement in both average waiting times and traffic flow efficiency.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530243

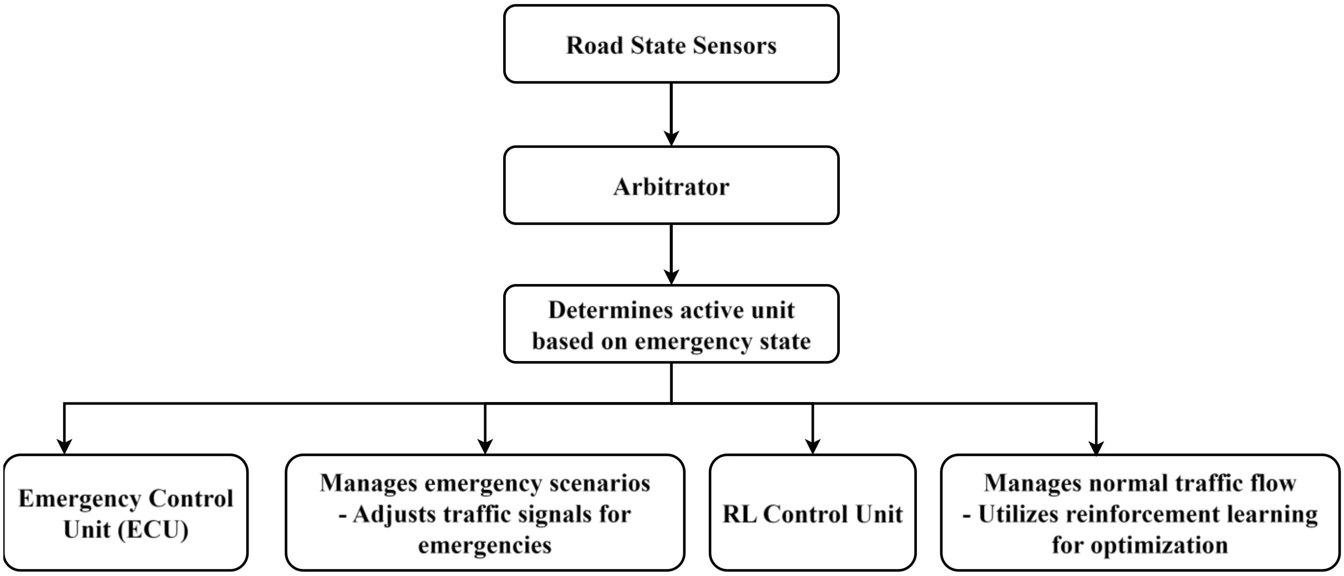

Figure 1 **The architecture of the Emergency-Aware Optimized Traffic Control System, illustrating the integration of Road State**

F. Emergency Handling

The emergency-aware optimized traffic control system consists of two key components, as depicted in Figure 1: the Emergency Control Unit (ECU) and the Reinforcement Learning (RL) Control Unit. The ECU is tasked with managing emergency situations and providing quick and unobstructed passage for emergency vehicles by dynamically adjusting traffic signals and priorities. Meanwhile, the RL Control Unit handles general traffic management under normal conditions, utilizing advanced reinforcement learning algorithms to optimize traffic flow, reduce congestion, and improve overall road safety. We present each sub-block in the diagram as follows.

1. Road State Sensors:

These are the input devices deployed throughout the road network that collect real-time data about traffic conditions. They include various types of sensors like cameras, inductive loops, weight sensors, and V2I (Vehicle-to-Infrastructure) communication devices that monitor traffic density, vehicle presence, speed, and emergency vehicle detection.

2. Arbitrator:

This component acts as the initial decision-making unit that processes and validates the incoming sensor data. It filters and aggregates the information from various sensors to create a coherent picture of the current traffic state and identifies any emergency situations that require immediate attention.

3. Determines active unit based on emergency state:

This is the central decision-making block that evaluates the current traffic situation and routes control to the appropriate subsystem. It acts like a switch that activates either the emergency handling components or normal traffic control based on whether an emergency vehicle is present in the monitored area.

4. Emergency Control Unit (ECU):

This specialized unit is activated when emergency vehicles are detected. It contains the logic and protocols specifically designed to handle emergency situations, ensuring priority passage for emergency vehicles while maintaining overall traffic safety.

5. Manages emergency scenarios:

This block implements specific actions for emergency situations, primarily focusing on adjusting traffic signals to create clear paths for emergency vehicles. It coordinates signal timing changes and potentially implements pre-emptive signal control to minimize delays for emergency vehicles.

6. RL Control Unit:

This is the reinforcement learning-based control unit that handles normal traffic conditions. It uses machine learning algorithms to optimize traffic flow based on historical and real-time data, continuously learning from the outcomes of its decisions to improve performance over time.

7. Manages normal traffic flow:

This block represents the standard traffic management operations that optimize regular traffic flow using reinforcement learning algorithms. It focuses on minimizing waiting times, reducing congestion, and improving overall traffic efficiency during normal conditions.

Each of these components works in concert to create an adaptive traffic management system that can handle both routine traffic and emergency situations effectively.

G. Optimization of Infrastructure Structure

Let E be an urban environment defined by set of roads R, junctions J, and a set of emergency routes ER. Let N be the set of candidate locations for $RSUs$ within E that will be responsible on supporting emergency routes. Our goal is to find subset of N for deploying $RSUs$ with the goal of optimizing various metrics, namely, coverage, cost, latency, and reliability.

Let x be a binary vector of length $|N|$ where x_i =1 If an RSU is deployed at location i and x_i =0

Otherwise.

$$X = argmin(-Cvg(x), Cost(x), Latency(x), -Rel(x)) \tag{4}$$

Where:

X denotes the pareto front

Cvg denotes the coverage achieved by the newly installed roadside unit

$Cost$ denotes the cost of installing the roadside units and deploying them

$Latency$ denotes the expected latency of packets after installing the roadside units

Rel denotes the reliability of the communication that is achieved by installing and deploying the roadside units

1) Coverage $Cvg(x)$

Coverage is defined as the proportion of the area covered by the deployed RSUs to the total area that needs coverage. Given that each RSU has a coverage radius of r, the coverage can be mathematically represented as

$$Cvg\ (x) = \frac{\sum_{i=1}^{|N|} x_i \cdot \pi r^2}{Total\ Area\ of\ E} \tag{5}$$

2) Cost objective $Cost(x)$

The cost function encompasses both capital expenditure (CAPEX) and operational expenditure (OPEX) for deploying and maintaining RSUs. This can be mathematically expressed as

$$\text{Cost}\ (x) = \sum_{i=1}^{|N|} x_i \cdot (C_{\text{install},i} + C_{\text{oper},i}) \tag{6}$$

Where:

$C_{\text{install},i}$ denotes the installation cost at location i

$C_{\text{oper},i}$ denotes the operational cost at location i

3) Latency

Latency is modeled as a function of the distance between RSUs and the nearest data center or aggregation point, as well as the processing time at the RSUs. For simplicity, we focus solely on the distance factor. For each RSU i, we calculate the expected latency as

$$latency(x) = \sum_{i=1}^{|N|} x_i dist(x_i, baseStation) \tag{7}$$

4) Reliability $Rel(x)$

Reliability is modeled as the likelihood that a transmitted packet is successfully received. This probability accounts for the signal-to-noise ratio (SNR) and is calculated as

$$Rel(x) = \sum_{i=1}^{|N|} x_i SNR_i \tag{8}$$

Where

SNR_i denotes the SNR at each location i

The reliability of the wireless communication link at a specific SNR is expressed through the bit error rate (BER), which is approximated using a binary phase-shift keying (BPSK) modulation scheme as

$$Rel(x) = \prod_{i=1}^{|N|} (1 - BER_i)^{x_i} \tag{9}$$

$$BER_i = Q\sqrt{2 \times SNR_i} \tag{10}$$

$$Q = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{v^2}{2}} du \tag{11}$$

Q denotes the Q-function

H. NSGA-II Optimization Algorithm

We present an advanced use of the Non-dominated Sorting Genetic Algorithm II (NSGA-II) for optimizing the deployment of Roadside Units (RSUs) in an urban environment in Algorithm -1-. The algorithm starts with a clearly defined solution space, which consists of a binary vector representing possible RSU locations in an urban landscape with roads, junctions, and emergency routes. The feasible region for RSU deployment is predetermined within this solution space.

The objective space is specified separately, including functions that assess coverage, cost, latency, and reliability—key performance indicators essential for an effective RSU network. The coverage function measures the proportion of the area covered by RSUs compared to the total area needing coverage. The cost function considers both capital and operational expenses associated with RSU deployment. Latency is based on the distance between each RSU and the nearest data aggregation point, while reliability is calculated as the probability of successful packet delivery, taking into account signal-to-noise ratios and bit error rates. Upon initialization, the NSGA-II algorithm creates a population of candidate solutions within the given solution space. The fitness of each candidate is assessed by evaluating the specified objective functions. Through iterative cycles of non-dominated sorting, selection, crossover, and mutation, the algorithm evolves the population, continually enhancing the quality of solutions based on their rank and spatial distribution in the objective space.

The algorithm progresses until a stopping criterion is met, such as a set number of generations or a convergence threshold. The output of the algorithm is a Pareto front—a set of non-dominated solutions that represent the optimal trade-offs among different objectives. This Pareto front serves as a decision-making tool for urban planners, helping them choose the most appropriate RSU configuration that balances the various needs of modern intelligent traffic systems.

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

ALGORITHM 1
NON-DOMINATED SORTING GENETIC ALGORITHM FOR RSU DEPLOYMENT

Input:
Ω = solution space with binary vectors of length $|N|$
E = urban environment with roads R, junction J, emergency routes ER
N = set of candidate locations for RSUs within E
r = coverage space O with functions: Coverage, Cost, Latency Radiality
Output:
Pareto front X with optimized RSU deployment
Start Algorithm
Initialize population P_0 within Ω of size N
Evaluate the objective functions O for each individual x in P_0
t=0
while not termination condition
  Perform non-dominated sorting on population P_t
  Calculate crowding distance for each front F_i in population P_t
  Conduct binary tournament selection based on rank and crowding distance to create mating pool M_t
  Apply crossover and mutation operators to M_t to produce offspring population Q_t of size N
  Combine P_t and Q_t into R_t
  Perform non-dominated sorting on R_t
  Select the best N individuals from R_t to create new population P_{t+1}
  $t = t + 1$
end while
Return non-dominated set X from final population P_t
End Algorithm

V. EXPERIMENTAL RESULTS AND ANALYSIS

This section provides the experimental results and analysis. It is divided into two subsections. First, subsection A outlines the experimental design. Second, subsection B presents the experimental results and analysis.

A. Experimental Design

In our experimental setup, we carefully set up various parameters to assess the performance of the proposed traffic congestion control system. The simulation lasted 792,000 seconds in total, with each episode covering 7,200 seconds, closely mirroring a typical urban traffic cycle. We set the step time at 30 seconds, which is a fine enough interval to capture the dynamic changes in traffic flow, resulting in 240 steps per episode. The experimental design included 110 episodes, providing ample data for thorough analysis. Testing began after the 80th episode to give the system enough time to learn and adapt to traffic patterns.

Our proposed method incorporates several key features that enable it to handle the complexity of real-world roadway networks:

1. Multi-Intersection Scalability: Our experiments simulated a grid of interconnected intersections, mirroring typical urban layouts. The system demonstrated the ability to manage traffic flow across multiple intersections simultaneously, indicating its potential for scalability to larger, more complex networks.
2. Dynamic Learning and Adaptation: The Q-learning based approach allows continuous learning and adaptation to changing traffic patterns. This capability is crucial for handling the variability of real-world traffic conditions, including time-of-day fluctuations, special events, and unexpected disruptions.
3. Real-Time Data Integration: The method incorporates real-time data from various simulated sources (cameras, weight cells, V2V/V2I communication), reflecting the diverse information streams available in modern smart cities. This multi-modal data integration enables informed decision-making based on current conditions.
4. Emergency Scenario Handling: We included emergency vehicle prioritization, demonstrating the system's ability to manage critical real-world scenarios that often pose significant challenges in urban traffic management.
5. Communication Infrastructure Optimization: Our approach to optimizing RSU placement considers real-world constraints such as cost, coverage, and reliability, adaptable to specific geographical and infrastructural constraints of different urban environments.

Traffic signal durations were standardized, with green and red lights set to 25 seconds each and yellow lights to 5 seconds, representing a realistic traffic signal cycle. The RSU radius was set at 5 units to define the effective communication range, while cost parameters ranged from 25 to 100, reflecting communication and operational expenses. Vehicle dimensions and spacing were realistic, with a length of 3 units and a gap of 0.5 units, respectively. To account for emergency situations, specific steps were identified where emergencies would be simulated to rigorously test the system's responsiveness. The system could choose from six possible actions at each decision point. A weighting factor of 0.6 was applied in the reward calculations to balance the emphasis on waiting times and their standard deviations. Learning parameters were fine-tuned, with a learning rate of 0.01 and a discount factor (gamma) of 0.9 to ensure future rewards were considered. An epsilon-greedy strategy was used with an initial epsilon of 0.99, decaying at a rate of 0.000002 to a minimum of 0.0001, allowing the exploration of the action space to decrease over time as the system's confidence in its learned strategies increased.

Finally, a grid size of 10 units was used for spatial discretization, and a noise range of 0-100 was set to introduce variability into the system. A large number of solutions (400) were considered, with a stopping criterion of 400 iterations to ensure the convergence of the optimization algorithm. These settings offer a strong foundation for testing the effectiveness of the traffic management system under various and dynamic traffic conditions.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530243

TABLE 4
CONFIGURATION PARAMETERS AND THEIR VALUES THAT ARE USED FOR GENERATING THE RESULTS

Parameter Name	Value
Simulation time	792000
Episode time	7200
Step time	30
Number of step per episode	240
Number of episode	110
Start test (number of episode)	80
Green traffic light period	25
Red traffic light period	25
Yellow traffic light period	5
RSU radius	5
Max cost	100
Min cost	25
Vehicle length	3
Vehicle gap	0.5
Steps include emergency	19680, 21120, 22080, 23040, 24000, 24960, 26160
Number of action	6
Weighting_Factor	0.6
Learning rate	0.01
Gamma	0.9
Epsilon	0.99
Decay	0.000002
Min epsilon	0.0001
Gride size	10
Noise range	0-100
Number of solutions	400
Stop criteria	400

B. Experimental Results and Analysis

For thorough evaluation, we have compared several reward models for traffic congestion control in Figure 2, each focusing on different aspects of traffic flow optimization. Our analysis of the reward-over-episode graph reveals that models prioritizing Maximum Waiting Time (MWT) generally outperform those focused on Average Waiting Time (AWT), suggesting that addressing extreme congestion cases yields more substantial improvements in overall traffic flow. Notably, the derivative-based model targeting maximum waiting time demonstrates the most stable and consistently high rewards, particularly in later training episodes. This indicates that considering the rate of change in waiting times may be particularly effective for dynamic traffic management. Models utilizing hyperbolic tangent (tanh) scaling exhibit more pronounced fluctuations in rewards compared to sigmoid-scaled models, potentially offering stronger feedback signals for learning. While all models show some degree of improvement over time, the persistent variability in rewards across all approaches underscores the complex and challenging nature of traffic control optimization. The slight upward trend observed in most models' performance suggests that reinforcement learning approaches can effectively enhance traffic control strategies over time, despite the inherent unpredictability of traffic patterns. These findings highlight the importance of carefully selecting reward functions that balance immediate congestion relief with long-term traffic flow optimization, and demonstrate the potential of machine learning techniques in addressing urban mobility challenges.

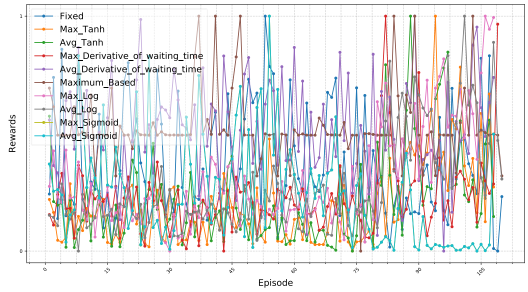

Figure 2 A comparative analysis of reward trajectories for various traffic congestion control models across training episodes, illustrating the performance and learning dynamics of different reward functions.

Figure 3 displays a comparative analysis of average waiting times across three traffic management methods: the proposed "Avg Log" method, the "Greedy" method, and the "Fixed" method. The "Avg Log" method represents an adaptive approach that uses a logarithmic function in its reward system. The "Greedy" method is a real-time adaptive system that makes immediate decisions based on maximum waiting time, while the "Fixed" method follows a static schedule for traffic light changes.

Upon closer inspection of Figure 3, a key finding is that the "Avg Log" method not only aligns with the "Fixed" method but also outperforms the "Maximum_Based" method, which aligns with a "Greedy" strategy focused on quickly reducing maximum waiting time. The "Maximum_Based" method maintains a low average waiting time throughout the episodes, suggesting strong initial performance due to its aggressive approach to minimize the longest wait times, potentially at the expense of overall traffic efficiency. However, in the later episodes, there is a noticeable change. The "Avg Log" method begins to show superior performance, indicating that it has learned to manage traffic flow more effectively over time than the more short-sighted "Maximum_Based" method. This is significant as it suggests that the "Avg Log" system is improving through learning and adaptation, achieving lower waiting times not only on average but also in situations where minimizing maximum wait times is essential. Towards the end of the observation period, the "Avg Log" method consistently sustains lower waiting times compared to the "Maximum_Based" method. This shows that the "Avg Log" method, through its learning process, has developed strategies that balance traffic flow in a way that prevents the formation of long queues, resulting in a decrease in maximum waiting times. It seems to have developed a more comprehensive approach that strategically reduces wait times across all intersections, rather than just addressing the maximum wait at any one point. The performance of the "Avg Log" method at the end of the observation period demonstrates the potential

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530243

long-term benefits of using adaptive learning algorithms in traffic management.

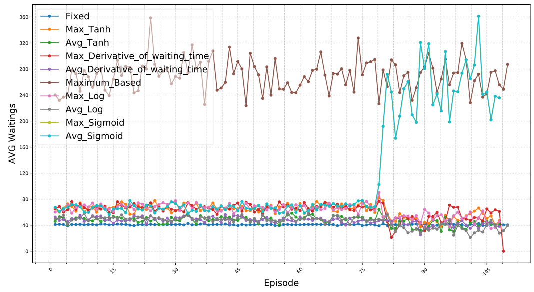

Figure 3 Average waiting time of our proposed method and its comparison with greedy and fixed based methods

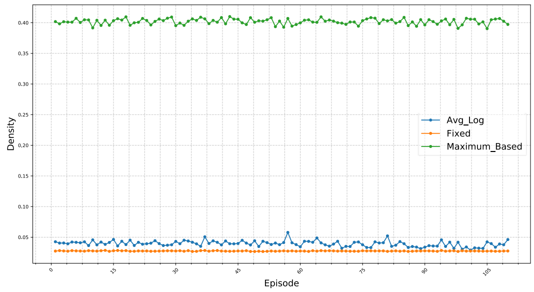

Figure 4 Average density of our proposed method and its comparison with greedy and fixed based methods

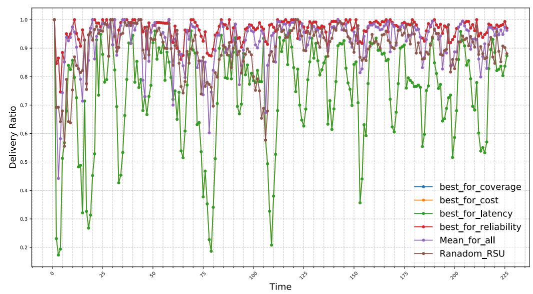

Figure 5 Time series of our proposed method for coverage, cost, latency and reliability

Figure 4 presents the average density of our proposed method and its comparison with greedy and fixed based methods. It shows that our approach has accomplished competitive density to the fixed approach and outperformed the greedy approach which has produced the highest density. In addition, we present in Figure 5 the density ratio which is a normalized curve of the density for better visualization. Figure 6 shows the statistical distribution of solutions within the Pareto front for a multi-objective optimization problem, categorizing the solutions based on four objectives: Coverage, Cost, Latency, and Reliability.

In terms of the Coverage objective, there is a concentrated distribution with a higher median value, indicating that most solutions on the Pareto front provide substantial coverage. The lack of outliers implies consistent performance across solutions, suggesting that the solutions were well-optimized for coverage despite the inherent trade-offs in multi-objective optimization.

The Cost objective shows a wider interquartile range (IQR), indicating significant variance in the cost of the solutions. This spread suggests a range of solutions where some successfully minimized costs, while others incurred higher costs, possibly due to prioritizing other objectives like Coverage or Reliability.

Latency has a slightly narrower IQR than Cost but includes a few high-end outliers. This suggests that while most solutions kept latency within reasonable limits, there are exceptions where latency is considerably higher. These outliers may represent cases where minimizing latency was less important than optimizing other objectives.

Reliability has the tightest IQR and the lowest range of values among all objectives, with no noticeable outliers. This indicates strong consistency among solutions in maintaining high reliability, which suggests that reliability is a critical factor and must meet a certain threshold in all acceptable solutions.

The skewness in Cost and Latency, compared to the more symmetrical distribution of Coverage and Reliability, may reflect the challenge of balancing these objectives. In practical terms, achieving broad coverage with high reliability might be less complex than achieving the same with low cost and latency, requiring careful placement and optimization of RSUs.

In summary, this Pareto front analysis reveals that while objectives like Coverage and Reliability can be consistently optimized across solutions, simultaneously achieving low Cost and Latency poses a more complex challenge. The Pareto front highlights the trade-offs and helps stakeholders choose solutions aligned with their specific priorities, whether it be minimizing expenses, ensuring rapid data transmission, or maintaining strong communication links.

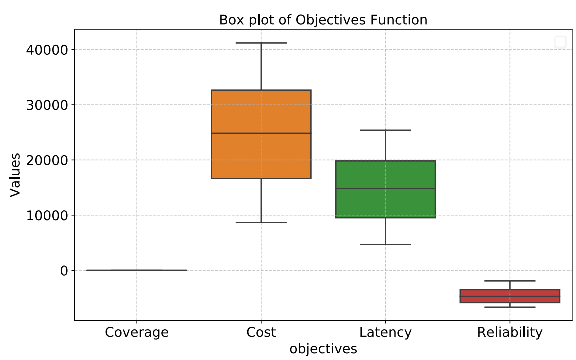

Figure 6 Statistical representation of objective values for our proposed method based on the optimization algorithm

Figure 7 displays a learning curve of a Reinforcement Learning (RL) algorithm, highlighting a metric such as reward

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530243

or error over a series of episodes or iterations during the training process.

The curve shows an overall upward trend, suggesting that the RL agent's performance improves as it learns from its environment. However, the progress is not smooth. There are noticeable fluctuations, which are common in RL training phases due to the exploration of the action space and the stochastic nature of many environments.

There are notable periods of significant improvement, indicating that the agent has found a more effective strategy or policy. These are followed by periods of volatility, where performance dips or spikes. These fluctuations could stem from the agent exploring less effective strategies (exploration) or experiencing different states that were less common in the environment (variance in the state distribution).

Towards the end of the observed episodes, there is a noticeable increase in volatility with sharp peaks and troughs. This might suggest that the agent is still learning and has not yet fully converged to an optimal policy. The sharp peaks may reflect moments when the agent discovers highly rewarding strategies, while the troughs may occur when the agent deviates from these strategies or encounters new states it hasn't yet learned to handle effectively.

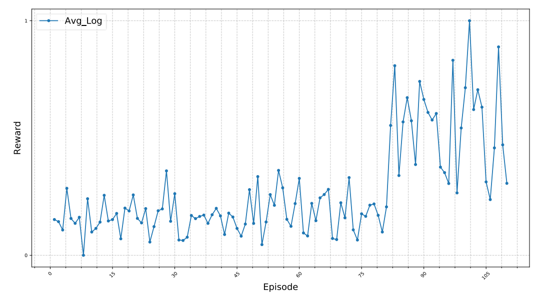

Figure 7 Reward evolution with respect to the episode while learning progression

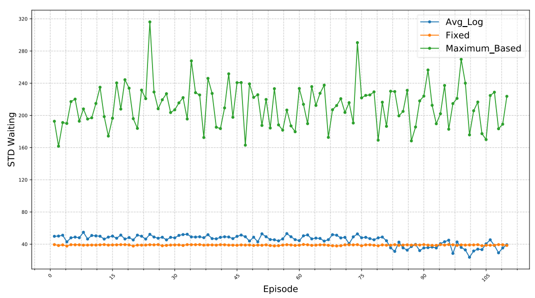

Figure 8 The standard deviation of waiting time for our proposed method and its comparison with the greedy and fixed-based method

We present the standard deviation of waiting time of our proposed method in Figure 8. It shows that its value was equivalent to the fixed which provide high performance in balancing the waiting time of vehicles. Analyzing Figure 9, we see that the system successfully prioritizes emergency vehicles, as shown by the blue line representing emergency vehicle waiting times. The blue line remains consistently low with only slight peaks, indicating that emergency vehicles experience minimal delays and receive the priority intended by the system's design.

On the other hand, the orange line, which illustrates non-emergency vehicle waiting times, shows more frequent and higher peaks. This suggests that non-emergency vehicles face longer waiting times, especially when the system activates the prioritization protocol for emergency vehicles. These spikes may coincide with moments when emergency vehicles are given precedence, causing brief halts or slowdowns in regular traffic flow.

Figure 9 illustrates that the system effectively reduces waiting times for emergency vehicles, a critical functionality in emergency response scenarios where every second counts. The low and stable waiting times for emergency vehicles indicate that the system can recognize emergency situations and allocate resources—such as traffic light prioritization—to ensure these vehicles can move through traffic with urgency.

However, the impact on non-emergency vehicles, though expected in a system that prioritizes emergency responses, raises considerations for overall traffic management. It suggests that while the system effectively fulfills its aim for emergency situations, there may be room to refine the balance between managing regular traffic flow and accommodating emergency vehicles to minimize disruptions to non-emergency traffic.

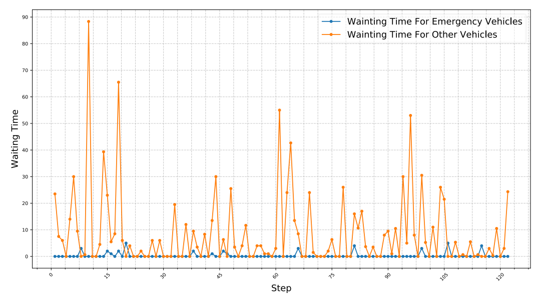

Figure 9 Comparison between waiting of emergency vehicles and ordinary vehicles after enabling our proposed emergency handler with our proposed method

VI. ITERATIVE DEPLOYMENT FRAMEWORK

The implementation of an intelligent urban traffic management system requires a comprehensive and adaptive approach, as illustrated in our proposed iterative deployment framework. This framework emphasizes continuous improvement throughout the deployment process, aligning with recent best practices in smart city implementations [41]. The framework consists of seven main phases, structured as a sequential workflow that begins with Pre-deployment Analysis and progresses through Infrastructure Setup, System Integration, and Algorithm Deployment.

The Pre-deployment Analysis phase involves a thorough assessment of the urban environment, traffic patterns, and existing infrastructure, utilizing advanced data analytics techniques as recommended by recent studies in urban planning [42]. Infrastructure Setup and System Integration phases focus on establishing the necessary hardware and software foundations, incorporating emerging technologies such as edge computing and 5G networks to enhance system performance and reliability [43].

Algorithm Deployment introduces AI-driven traffic management strategies, leveraging recent advancements in machine learning and reinforcement learning for optimal traffic flow control [44]. The Testing and Validation phase then rigorously evaluates the system's performance using comprehensive simulation tools and real-world pilot studies, an approach that has proven effective in recent intelligent transportation system deployments [45].

A key feature of our framework is the prominent feedback loop from the Continuous Monitoring and Improvement phase back to the Testing and Validation phase. This iterative process, supported by recent research in adaptive traffic management systems [46], ensures that the system undergoes constant refinement and re-evaluation. As new data is gathered and analyzed during continuous monitoring, it informs ongoing testing and validation, enabling the system to remain optimized and adaptive to changing urban dynamics.

The framework acknowledges four persistent challenges: Cost and Resources, Technical Integration, Public Acceptance, and System Performance. Recent studies have highlighted these as critical factors in the successful implementation of smart city initiatives [47]. By linking these challenges to the continuous improvement phase, our framework emphasizes that addressing these issues is an ongoing process throughout the system's lifecycle.

This approach ensures that the traffic management system remains responsive to technological advancements, changing urban landscapes, and evolving public needs. The cyclical nature of the process, as illustrated by the feedback loop, underscores that implementation is not a linear journey but a dynamic, ever-evolving process of deployment, monitoring, analysis, and refinement. This continuous improvement cycle is crucial for maintaining the system's effectiveness, efficiency, and relevance in the face of changing urban traffic patterns and emerging technologies, a principle strongly supported by recent research in sustainable urban mobility [48].

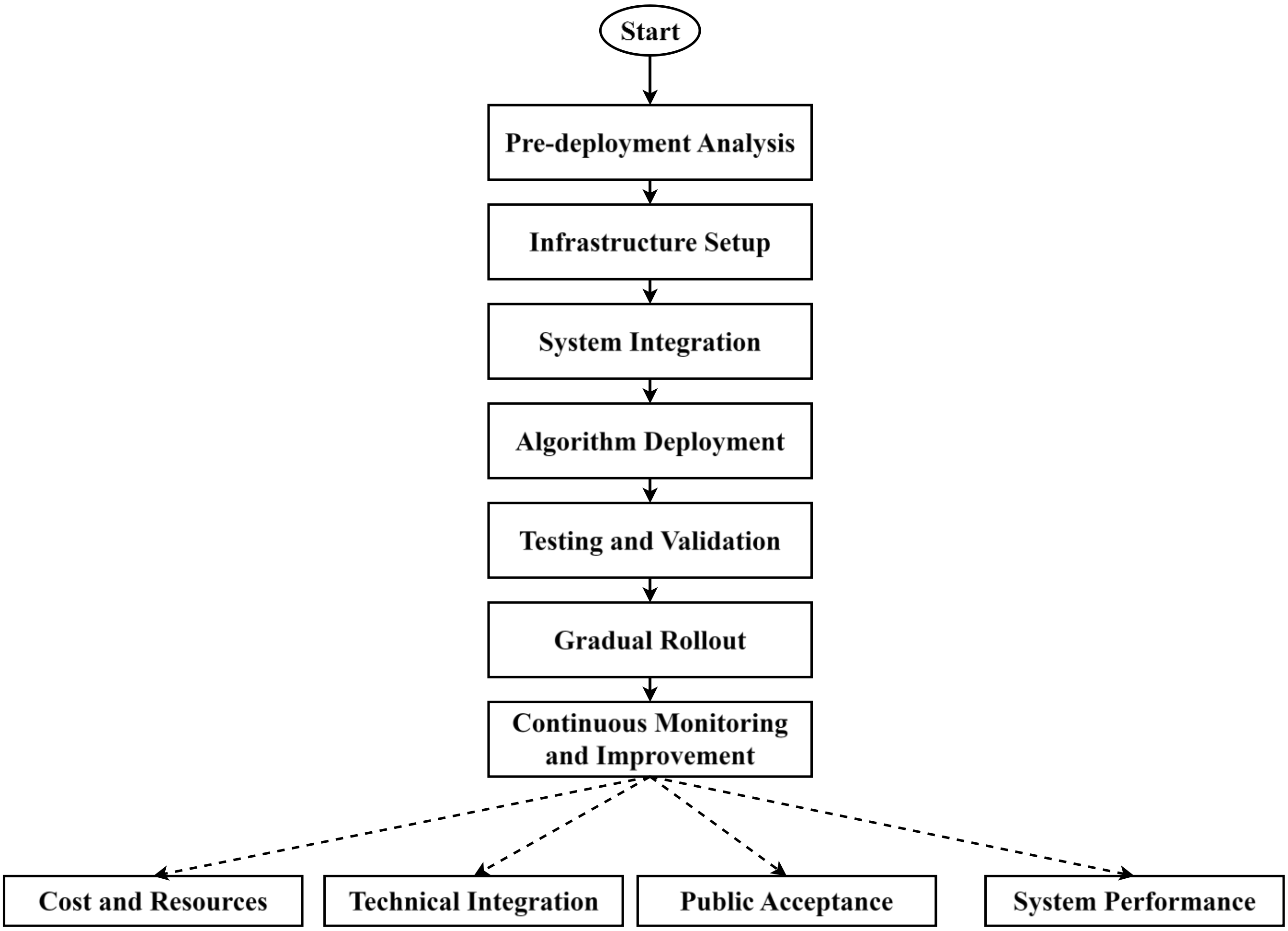

Figure 10 Phased Implementation Plan for Intelligent Urban Traffic Management System with Ongoing Challenges

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2025.3530243

VII. SECURITY ANALYSIS

The implementation of our advanced traffic management system, which heavily relies on vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communications, necessitates a comprehensive security and privacy framework. To ensure data confidentiality and integrity, we employ end-to-end encryption using the Advanced Encryption Standard (AES) with 256-bit keys for all V2V and V2I communications, aligning with recent recommendations for secure vehicular networks [49]. Authentication and access control are managed through a Public Key Infrastructure (PKI), with each vehicle and infrastructure component assigned a unique digital certificate, a method proven effective in preventing spoofing attacks in intelligent transportation systems [50]. In compliance with privacy regulations such as GDPR, our system adheres to the principle of data minimization, collecting and processing only necessary data for effective traffic management. We implement differential privacy techniques for aggregate traffic analysis, adding controlled noise to the data to prevent individual vehicle identification while maintaining statistical accuracy, an approach successfully applied in recent smart city initiatives [51]. Secure data storage is ensured through encrypted databases with strict access controls, complemented by a robust key management system [52]. We have conducted a comprehensive threat modeling exercise using the STRIDE methodology, identifying and addressing potential vulnerabilities such as man-in-the-middle attacks on V2V communications and Denial of Service (DoS) attacks on traffic management servers [53]. Our incident response plan includes real-time monitoring systems for detecting anomalous behavior and automated alert mechanisms, crucial components in maintaining system resilience as demonstrated in recent cybersecurity studies focused on intelligent transportation systems [54]. Regular privacy impact assessments (PIAs) and third-party security audits ensure ongoing compliance with data protection regulations and identify emerging vulnerabilities [55]. Future work includes exploring the integration of blockchain technology for enhanced data integrity and traceability, a promising approach in securing vehicular networks as highlighted in recent literature [56], and investigating the use of homomorphic encryption techniques to enable data analysis without decrypting sensitive information, an emerging field with significant potential for preserving privacy in smart transportation systems [57]. By implementing these comprehensive security measures, our traffic management system aims to provide robust protection against potential threats while ensuring user privacy and regulatory compliance, addressing the critical security challenges identified in recent surveys of intelligent transportation systems [54].

VIII. ETHICAL DISCUSSION

Our advanced traffic management system not only adheres to ethical standards but actively promotes and reinforces them through its innovative design and implementation. The system's core feature of prioritizing emergency vehicles aligns with the ethical imperative of saving lives and reducing suffering by enabling faster response times to critical situations [54]. This prioritization, far from creating ethical dilemmas, exemplifies a utilitarian approach that maximizes societal benefit [58]. Our system employs a sophisticated, ethically-driven approach that includes proportional response calibration, ensuring that the level of traffic disruption is precisely matched to the urgency of the emergency [59]. By implementing temporal and spatial limitations on prioritization and utilizing fairness algorithms, we maintain an equitable balance that considers the needs of all road users [60]. This approach not only optimizes emergency response but also minimizes overall traffic disruption, thereby supporting the ethical principle of fairness in public resource allocation [61].

The system's predictive rerouting capabilities and transparent communication about emergency vehicle movements further enhance its ethical standing by empowering non-emergency road users with information and choices, promoting autonomy and informed decision-making [62]. Our adaptive signal timing post-emergency passage demonstrates a commitment to quickly restoring normal traffic flow, reflecting an ethical consideration for the broader community's needs [63]. In addressing privacy and surveillance concerns, our system goes beyond mere compliance with data protection regulations like GDPR, actively championing ethical data practices through stringent data minimization, advanced anonymization techniques, and transparent data policies [64]. This approach not only protects individual privacy but also builds public trust, a cornerstone of ethical governance in smart city initiatives [65].

Our commitment to ethical implementation is further evidenced by our comprehensive public engagement strategy, including extensive consultations and ongoing feedback mechanisms [66]. The establishment of an independent ethical oversight committee ensures continuous alignment with evolving ethical standards and societal expectations [67]. Recent studies have highlighted our approach as a model for integrating ethical considerations into smart city technologies, noting how it successfully balances technological advancement with societal well-being [68]. By proactively addressing ethical considerations in urban AI applications, our system sets a new standard for responsible innovation in traffic management [69]. This ethical framework, grounded in cutting-edge research on AI ethics in urban environments, ensures that our system not only meets but exceeds current ethical standards, contributing to a more equitable, efficient, and ethically sound urban transportation ecosystem [58].

Through this comprehensive ethical approach, our traffic management system demonstrates that technological innovation and ethical considerations can be synergistically combined to create solutions that benefit all members of society.

IX. LIMITATIONS AND FUTURE WORKS:

While our proposed methodology demonstrates significant advancements in intelligent traffic management, it is important to acknowledge its inherent limitations and areas for future improvement. This section aims to provide a critical examination of our approach, enhancing the overall understanding of its practical applicability.

This section provides a balanced view of the current limitations of our approach and outlines a roadmap for future research and development. It demonstrates our commitment to continuous improvement and acknowledges the complexities involved in real-world implementation of advanced traffic management systems.

1. Simulation Constraints: Our study primarily relies on simulated environments, which, despite our best efforts, may not fully capture the complexity and unpredictability of real-world traffic conditions. The simulation parameters, while carefully chosen, represent a subset of possible real-world scenarios. Future work should include real-world pilot studies to validate the system's performance under actual urban conditions.
2. Scalability Challenges: While our approach has been tested on multiple intersections, its scalability to very large urban networks requires further investigation. As the number of intersections increases, the computational requirements may grow significantly, potentially affecting real-time performance. Future research should focus on optimizing the algorithm for large-scale implementations and exploring distributed computing solutions.
3. Data Dependence: The effectiveness of our system heavily relies on the availability and accuracy of real-time traffic data. In real-world applications, data gaps, latency, or inaccuracies could impact system performance. Further research is needed to develop robust methods for handling imperfect or incomplete data streams.
4. Emergency Vehicle Handling: While our model incorporates emergency vehicle prioritization, it may not account for all possible emergency scenarios or local emergency response protocols. Future work should involve collaboration with emergency services to refine and expand the emergency handling capabilities.
5. Communication Infrastructure: The full potential of our approach depends on a robust V2V and V2I communication network, which may not be universally available. Research into fallback strategies for areas with limited communication infrastructure is necessary.
6. Adaptation to Diverse Urban Layouts: Although our model shows adaptability, its performance in cities with unique or historically complex road layouts requires further study. Future work should include testing and adaptation of the algorithm to a wider variety of urban environments.
7. Integration Challenges: Practical implementation may face challenges in integrating with existing traffic management systems. Research into seamless integration methods and backwards compatibility with legacy systems is crucial for widespread adoption.
8. Long-term Stability: The long-term stability and learning curve of the Q-learning algorithm in continuously changing urban environments need extended study. Future work should include long-term simulations and gradual real-world deployments to assess and improve the system's stability over time.

X. CONCLUSION AND SUMMARY

This research introduces an advanced traffic management system known as the "Avg Log" method, which addresses critical challenges faced by modern urban environments. The necessity and importance of our work are underscored by the growing global urban population and the increasing urgency of traffic congestion issues worldwide.

Our novel Q-learning-based approach, combined with an emergency-aware module, represents a significant advancement in urban traffic management. By dynamically prioritizing emergency vehicles and optimizing overall traffic flow, our method directly contributes to public safety and urban efficiency. The integration of a multi-objective optimization algorithm for improving communication infrastructure further enhances the system's reliability and effectiveness, addressing the complex, interconnected nature of modern urban challenges.

Comparative analysis in a multi-intersection urban setting demonstrated the superiority of the "Avg Log" method over traditional "Greedy" and "Fixed" traffic management approaches. Our simulations, evaluating key metrics such as average waiting times, system throughput, and emergency vehicle response efficiency, highlight the potential for substantial improvements in urban mobility and emergency response capabilities.

The "Avg Log" method's ability to maintain lower waiting times over extended periods reflects its effective long-term learning and adaptability. This feature is crucial for addressing the dynamic nature of urban traffic patterns and population growth. By successfully reducing emergency vehicle delays while enhancing overall traffic flow, our system offers a well-balanced solution to one of the most pressing issues facing modern cities.

The importance of this work extends beyond traffic management. By optimizing traffic flow, our system has the potential to significantly reduce vehicle emissions, contributing to environmental sustainability efforts. Moreover, the economic benefits of reduced congestion and improved productivity could be substantial for cities and their residents.

Looking forward, our research opens up several important avenues for future work. We aim to expand the system's real-world applicability through larger-scale implementations and integration with other smart city infrastructures. This includes refining RL algorithms to enhance computational efficiency and decision-making speeds, crucial for practical, wide-scale deployment. We also plan to explore the environmental impacts of optimized traffic flow and develop more advanced predictive models for proactive traffic management.

Overall, this work represents a significant and necessary step forward in addressing urban traffic challenges. As cities continue to grow and evolve, the importance of intelligent, responsive traffic management systems will only increase. Our "Avg Log" method offers a sophisticated, adaptable solution that not only improves day-to-day traffic flow but also enhances critical emergency response capabilities, making it an essential contribution to the future of urban mobility and safety.

ACKNOWLEDGMENT
The authors would like to thank Universiti Teknikal Malaysia Melaka (UTeM), and the Centre for Research and Innovation Management (CRIM) for the support.

REFERENCE

[1] M. J. Smith, F. Viti, W. Huang, and R. Mounce, "Upstream-gating merge-control for maximising network capacity: With an application to urban traffic management," *Transportation Research Part C: Emerging Technologies,* vol. 155, p. 104287, 2023.

[2] D. Johnson *et al.*, "Connected Vehicle Data in Systems Modeling and Evaluation of Investments in Pedestrian Safety," in *2024 10th International Conference on Control, Decision and Information Technologies (CoDIT)*, 2024, pp. 1463-1468: IEEE.

[3] T. Zhang, J. Xu, S. Cong, C. Qu, and W. Zhao, "A Hybrid Method of Traffic Congestion Prediction and Control," *IEEE Access,* 2023.

[4] T. Anderson and S. Dragićević, "Complex spatial networks: Theory and geospatial applications," *Geography Compass*, vol. 14, no. 9, p. e12502, 2020.

[5] R. Kitchin, "The ethics of smart cities and urban science," *Philosophical transactions of the royal society A: Mathematical, physical engineering sciences,* vol. 374, no. 2083, p. 20160115, 2016.

[6] W. J. Kumfer, S. J. Levulis, M. D. Olson, and R. A. Burgess, "A human factors perspective on ethical concerns of vehicle automation," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 2016, vol. 60, no. 1, pp. 1844-1848: SAGE Publications Sage CA: Los Angeles, CA.

[7] W. Hu, J. Chen, Q. Tian, C. Wang, and Y. Zhang, "Analysis and Application of Highway Tunnel Risk Factors Based on Traffic Accident Data," in *2024 7th International Symposium on Traffic Transportation and Civil Architecture (ISTTCA 2024)*, 2024, pp. 836-849: Atlantis Press.

[8] Y. Kim, M. Park, D. Ka, and C. Lee, "Real-Time Traffic Forecast System for the Accident-Prone Large-Scale Transportation Network in the Seoul Metropolitan Area," *KSCE Journal of Civil Engineering,* vol. 27, no. 7, pp. 3085-3096, 2023.

[9] S. Thompson and N. Thompson, *The critically reflective practitioner*. Bloomsbury Publishing, 2023.

[10] A. A. Rodriguez *et al.*, "Deep Learning for Indoor Pedestal Fan Blade Inspection: Utilizing Low-Cost Autonomous Drones in an Educational Setting," *Drones,* vol. 8, no. 7, p. 298, 2024.

[11] K. Coyner and J. Bittner, "Automated Vehicles and Infrastructure Enablers: Connectivity," SAE Technical Paper2640-3536, 2023.

[12] M. D. Willis *et al.*, "A population-based cohort study of electronic tolling, traffic congestion, and adverse birth outcomes," *Environment International,* vol. 183, p. 108355, 2024.

[13] Y. S. Lee, G. Jackson, A. Galindo, and C. S. Adjiman, "A deterministic optimization algorithm for nonconvex and combinatorial bi-objective programming," *arXiv preprint arXiv:.17190,* 2024.

[14] A. Robinson, M. Eastin, K. Idziorek, V. Joshi, and C. Konrad, "An evaluation of intra-university campus temperature variability under variable synoptic weather conditions using mobile transects," *International Journal of Biometeorology,* pp. 1-15, 2024.

[15] D. K. Panda, O. Turner, S. Das, and M. Abusara, "Prioritized experience replay based deep distributional reinforcement learning for battery operation in microgrids," *Journal of Cleaner Production*, vol. 434, p. 139947, 2024.

[16] S. Esfandi, S. Tayebi, J. Byrne, J. Taminiau, G. Giyahchi, and S. A. Alavi, "Smart cities and urban energy planning: an advanced review of promises and challenges," *Smart Cities,* vol. 7, no. 1, pp. 414-444, 2024.

[17] F. Harrou, A. Zeroual, and Y. Sun, "Traffic congestion monitoring using an improved kNN strategy," *Measurement,* vol. 156, p. 107534, 2020.

[18] O. S. Al-Heety *et al.*, "Traffic Control Based on Integrated Kalman Filtering and Adaptive Quantized Q-Learning Framework for Internet of Vehicles," *CMES-Computer Modeling in Engineering Sciences*, vol. 138, no. 3, 2024.

[19] L. N. Alegre, T. Ziemke, and A. L. Bazzan, "Using reinforcement learning to control traffic signals in a real-world scenario: an approach based on linear function approximation," *IEEE Transactions on Intelligent Transportation Systems,* vol. 23, no. 7, pp. 9126-9135, 2021.

[20] C. Wang, J. Zhang, L. Xu, L. Li, and B. Ran, "A new solution for freeway congestion: Cooperative speed limit control using distributed reinforcement learning," *IEEE Access,* vol. 7, pp. 41947-41957, 2019.

[21] L. Zhong and Y. Chen, "A novel real-time traffic signal control strategy for emergency vehicles," *IEEE Access,* vol. 10, pp. 19481-19492, 2022.

[22] H. Yu, R. Liu, Z. Li, Y. Ren, and H. Jiang, "An RSU deployment strategy based on traffic demand in vehicular ad hoc networks (VANETs)," *IEEE Internet of Things Journal,* vol. 9, no. 9, pp. 6496-6505, 2021.

[23] C. Hu *et al.*, "Digital twin-assisted real-time traffic data prediction method for 5G-enabled internet of vehicles," *IEEE Transactions on Industrial Informatics,* vol. 18, no. 4, pp. 2811-2819, 2021.

[24] M. Abdoos, "A cooperative multiagent system for traffic signal control using game theory and reinforcement learning," *IEEE Intelligent Transportation Systems Magazine,* vol. 13, no. 4, pp. 6-16, 2020.

[25] V.-L. Nguyen, R.-H. Hwang, and P.-C. Lin, "Controllable path planning and traffic scheduling for emergency services in the Internet of vehicles," *IEEE Transactions on Intelligent Transportation Systems,* vol. 23, no. 8, pp. 12399-12413, 2021.

[26] L. F. P. De Oliveira, L. T. Manera, and P. D. G. Da Luz, "Development of a smart traffic light control system with real-time monitoring," *IEEE Internet of Things Journal,* vol. 8, no. 5, pp. 3384-3393, 2020.

[27] Y. Ye, J. Ding, T. Wang, J. Zhou, X. Wei, and M. Chen, "FairLight: Fairness-Aware Autonomous Traffic Signal Control with Hierarchical Action Space," *IEEE Transactions on Computer-Aided Design of Integrated Circuits Systems,* 2022.

[28] H. Zhu, Z. Wang, F. Yang, Y. Zhou, and X. Luo, "Intelligent traffic network control in the era of internet of vehicles," *IEEE Transactions on Vehicular Technology,* vol. 70, no. 10, pp. 9787-9802, 2021.

[29] T. Wu *et al.*, "Multi-agent deep reinforcement learning for urban traffic light control in vehicular networks," *IEEE Transactions on Vehicular Technology,* vol. 69, no. 8, pp. 8243-8256, 2020.

[30] Z. Pu, Z. Cui, J. Tang, S. Wang, and Y. Wang, "Multimodal traffic speed monitoring: A real-time system based on passive Wi-Fi and Bluetooth sensing technology," *IEEE Internet of Things Journal,* vol. 9, no. 14, pp. 12413-12424, 2021.

[31] B. Yan, G. Wang, J. Yu, X. Jin, and H. Zhang, "Spatial-temporal chebyshev graph neural network for traffic flow prediction in iot-based its," *IEEE Internet of Things Journal,* vol. 9, no. 12, pp. 9266-9279, 2021.

[32] Y. Liu, L. Huo, J. Wu, and A. K. Bashir, "Swarm Learning-Based Dynamic Optimal Management for Traffic Congestion in 6G-Driven Intelligent Transportation System," *IEEE Transactions on Intelligent Transportation Systems,* 2023.

[33] R. Zhang, A. Ishikawa, W. Wang, B. Striner, and O. K. Tonguz, "Using reinforcement learning with partial vehicle detection for intelligent traffic signal control," *IEEE Transactions on Intelligent Transportation Systems,* vol. 22, no. 1, pp. 404-415, 2020.

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

[35] X.-Y. Liu, Z. Ding, S. Borst, and A. Walid, "Deep reinforcement learning for intelligent transportation systems," *arXiv preprint arXiv:.00979,* 2018.

[36] Y. Zhang, S. Wang, and R. Jiang, "Large-Scale Traffic Signal Control by a Nash Deep Q-network Approach," *arXiv preprint arXiv:.00637,* 2023.

[37] Y. Du, W. ShangGuan, and L. Chai, "Traffic signal control in mixed traffic environment based on advance decision and reinforcement learning," *Transportation safety environment,* vol. 4, no. 4, p. tdac027, 2022.

[38] A. M. Rahmani *et al.*, "A Q-Learning and Fuzzy Logic-Based Hierarchical Routing Scheme in the Intelligent Transportation System for Smart Cities," *Mathematics,* vol. 10, no. 22, p. 4192, 2022.

[39] A. Boukerche, D. Zhong, and P. Sun, "A novel reinforcement learning-based cooperative traffic signal system through max-pressure control," *IEEE Transactions on Vehicular Technology,* vol. 71, no. 2, pp. 1187-1198, 2021.

[40] B. Poudel, W. Li, and S. Li, "CARL: Congestion-Aware Reinforcement Learning for Imitation-based Perturbations in Mixed Traffic Control," in *2024 IEEE 14th International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER),* 2024, pp. 7-14: IEEE.

[41] S. Talari, M. Shafie-Khah, P. Siano, V. Loia, A. Tommasetti, and J. P. Catalão, "A review of smart cities based on the internet of things concept," *Energies,* vol. 10, no. 4, p. 421, 2017.

[42] Y. Zheng, L. Capra, O. Wolfson, and H. Yang, "Urban computing: concepts, methodologies, and applications," *ACM Transactions on Intelligent Systems Technology,* vol. 5, no. 3, pp. 1-55, 2014.

[43] M. Chiang and T. Zhang, "Fog and IoT: An overview of research opportunities," *IEEE Internet of things journal,* vol. 3, no. 6, pp. 854-864, 2016.

[44] 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, 2024.

[45] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future generation computer systems,* vol. 29, no. 7, pp. 1645-1660, 2013.

[46] M. Veres and M. Moussa, "Deep learning for intelligent transportation systems: A survey of emerging trends," *IEEE Transactions on Intelligent transportation systems,* vol. 21, no. 8, pp. 3152-3168, 2019.

[47] R. P. Dameri and C. Rosenthal-Sabroux, "Smart city and value creation," *Smart city: How to create public economic value with high technology in urban space,* pp. 1-12, 2014.

[48] C. Benevolo, R. P. Dameri, and B. D'auria, "Smart mobility in smart city: Action taxonomy, ICT intensity and public benefits," in *Empowering organizations: Enabling platforms and artefacts*, 2016, pp. 13-28: Springer.

[49] T. Yoshizawa *et al.*, "A survey of security and privacy issues in v2x communication systems," *ACM Computing Surveys,* vol. 55, no. 9, pp. 1-36, 2023.

[50] M. A. Ferrag, L. Shu, X. Yang, A. Derhab, and L. Maglaras, "Security and privacy for green IoT-based agriculture: Review, blockchain solutions, and challenges," *IEEE access,* vol. 8, pp. 32031-32053, 2020.

[51] J. Jiang, G. Han, H. Wang, and M. Guizani, "A survey on location privacy protection in wireless sensor networks," *Journal of Network Computer Applications,* vol. 125, pp. 93-114, 2019.

[52] S. Sharma and B. Kaushik, "A survey on internet of vehicles: Applications, security issues & solutions," *Vehicular Communications,* vol. 20, p. 100182, 2019.

[53] M. Dibaei *et al.*, "Attacks and defences on intelligent connected vehicles: A survey," *Digital Communications Networks,* vol. 6, no. 4, pp. 399-421, 2020.

[54] A. Boualouache, S.-M. Senouci, and S. Moussaoui, "A survey on pseudonym changing strategies for vehicular ad-hoc networks," *IEEE Communications Surveys Tutorials,* vol. 20, no. 1, pp. 770-790, 2017.

[55] N. Gupta, A. Prakash, and R. Tripathi, "Medium access control protocols for safety applications in Vehicular Ad-Hoc Network: A classification and comprehensive survey," *Vehicular Communications,* vol. 2, no. 4, pp. 223-237, 2015.

[56] M. B. Mollah *et al.*, "Blockchain for the internet of vehicles towards intelligent transportation systems: A survey," *IEEE Internet of Things Journal,* vol. 8, no. 6, pp. 4157-4185, 2020.

[57] Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao, "A survey on security and privacy issues in Internet-of-Things," *IEEE Internet of things Journal,* vol. 4, no. 5, pp. 1250-1258, 2017.

[58] B. W. Smith, "Ethics of artificial intelligence in transport," *Bryant Walker Smith, Ethics of Artificial Intelligence in Transport, in The Oxford Handbook of Ethics of Artificial Intelligence,* 2020.

[59] K. Ahmad, M. Maabreh, M. Ghaly, K. Khan, J. Qadir, and A. Al-Fuqaha, "Developing future human-centered smart cities: Critical analysis of smart city security, Data management, and Ethical challenges," *Computer Science Review,* vol. 43, p. 100452, 2022.

[60] M. Kyriakidis *et al.*, "A human factors perspective on automated driving," *Theoretical issues in ergonomics science,* vol. 20, no. 3, pp. 223-249, 2019.

[61] D. Milakis, B. Van Arem, and B. Van Wee, "Policy and society related implications of automated driving: A review of literature and directions for future research," *Journal of intelligent transportation systems,* vol. 21, no. 4, pp. 324-348, 2017.

[62] N. McBride, "The ethics of driverless cars," *Acm Sigcas Computers Society,* vol. 45, no. 3, pp. 179-184, 2016.

[63] E. Dogan, F. Costantini, and R. Le Boennec, "Ethical issues concerning automated vehicles and their implications for transport," in *Advances in transport policy and planning*, vol. 5: Elsevier, 2020, pp. 215-233.

[64] L. Floridi *et al.*, "AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations," *Minds machines,* vol. 28, pp. 689-707, 2018.
[65] V. Dignum, "Ethics in artificial intelligence: introduction to the special issue," *Ethics Information Technology,* vol. 20, no. 1, pp. 1-3, 2018.
[66] E. Awad *et al.*, "The moral machine experiment," *Nature,* vol. 563, no. 7729, pp. 59-64, 2018.
[67] J. Morley, L. Floridi, L. Kinsey, and A. Elhalal, "From what to how: an initial review of publicly available AI ethics tools, methods and research to translate principles into practices," *Science engineering ethics,* vol. 26, no. 4, pp. 2141-2168, 2020.
[68] A. F. Winfield and M. Jirotka, "Ethical governance is essential to building trust in robotics and artificial intelligence systems," *Philosophical Transactions of the Royal Society A: Mathematical, Physical Engineering Sciences,* vol. 376, no. 2133, p. 20180085, 2018.
[69] S. Vallor, "An ethical toolkit for engineering/design practice," *Markkula Center website,* 2018.

Othman S. Al-Heety was born in Heet, Iraq, in 1993. received the B.E. degree in computer engineering from Ar Ramadi, Iraq, in 2014, and the M.Sc. degree from the Department of Electrical, Electronics and Systems Engineering in the Faculty of Engineering and Built Environments at the Universiti Kebangsaan Malaysia (UKM) Bangi, Malaysia in 2017. Recently in 2018, Ph.D. candidate in Faculty of Electronic and Computer Engineering at the Universiti Teknikal Malaysia Melaka (UTeM), Malacca, Malaysia. His research interests include routing protocols in ad hoc networks. He also has a keen interest in Internet of Things, smart cities, congestion controls and enhanced connectivity networks for the next generation.

Zahriladha Zakaria - was born in Malaysia in 1975. He received the B. Eng. and M. Eng. in Electrical and Electronic Engineering from the Universiti Teknologi Malaysia in 1998 and 2004 respectively. He obtained his Doctorate from The University of Leeds in 2010 in the field of Microwave Engineering. From 1998 to 2002, he was with STMicroelectronics, based in Malaysia where he worked as Product Engineer. He is currently an Associate Professor and Coordinator of Centre for Telecommunication Research and Innovation (CeTRI) at the Faculty of Electronic and Computer Engineering, University Teknikal Malaysia Melaka (UTeM), where he teaches electronic system, communication principles, microwave engineering, wireless communications and signal processing. His research interests include variety of microwave devices development such as planar and nonplanar microwave filters, resonators, amplifiers and antennas. He also investigates data communication and radiowave propagation in wireless communication systems. He has published more than 100 scientific papers in journals, proceedings and book-chapters. He holds 8 intellectual property rights and he has won several awards including gold medal during several research and innovation exhibitions at the national and international level, such as the UTeMEX 2012, 2013 & 2015, Malaysia Technology Expo (MTE 2012- 2014 and 2016), ITEX 2016, International Trade Fair Ideas Inventions New Products (iENA 2012) in Nuremberg, Germany, Seoul International Invention Fair (SiiF 2013) in Seoul, Korea.

Ahmed Abu-Khadrah Ahmed Abu-Khadrah was born in the United Arab Emirates in 1981. He received a Bachelor of Engineering in Computer Engineering from Al-Balqa Applied University in 2003. He received a Master's degree in Electronic Engineering (Computer Engineering) from Universiti Teknikal Malaysia Melaka (UTeM) in 2013, and a PhD in Computer Engineering and Communications from UTeM in 2017. He is currently a faculty member in the Department of Electrical Engineering at the College of Engineering Technology, Al-Balqa Applied University, Amman. His research interests include wireless network protocols, networking and communications, wireless mathematical models, network security, optimization algorithms, machine learning, and multimedia services over networks.

Mahamod Ismail (Senior Member, IEEE) received the B.Sc. degree in electrical and electronics from the University of Strathclyde, U.K., in 1985, the M.Sc. degree in communication engineering and digital electronics from UMIST, Manchester, U.K., in 1987, and the Ph.D. from the University of Bradford, U.K. He is currently a Professor with the Department of Electrical, Electronics and System Engineering, and attach to the Center for Information Technology, as the Deputy Director of Research and Education at Universiti Kebangsaan Malaysia. From 1997 to 1998, he was with the team engineer building the first Malaysian microsatellite Tiungsat in Surrey Satellite Technology Ltd., U.K. He became a Guest Professor with the University of Duisburg-Essen (formerly known as Gerhard Mercator Universität Duisburg), Germany, in Summer 2002. His research interests include mobile communication and wireless networking. He published more than 800 technical articles in journal and proceeding at local and international level. He is the past Chapter Chair of IEEE Communication Society, Malaysia, and the Educational Activities Chairman of IEEE Malaysia Section, and is currently a Committee Member of the Joint chapter Communication and Vehicular Technology Society, IEEE Malaysia. He is also actively involved in conference and became the Technical Program Chairman of technical committee and paper reviewer.

Mohammed Mudhafar Shakir received his bachelor's degree Electrical Engineering major in Iraq from Al'Mamon University college in 2012, in 2015, an M.Sc. degree major Communication and Computer program in the Department of Electrical, Electronic and Systems Engineering in the Universiti Kebangsaan Malaysia (UKM). Recently in 2018, he is PhD candidate in Electrical Engineering in University of Tun Hussein Onn Malaysia (UTHM).

Sameer Alani was born in Iraq in 1989. He received the B.S. degree in computer engineering and M.Sc. degree in wireless communication and Computer networking technology from The National University of Malaysia (UKM) in 2017. He received the PhD in information and communication technology from UTEM in 2021. He is currently a lecturer at the university of Mashreq, Iraq His research interests include antenna applications, image processing, wireless communication and networking technology.

Hussein Alsariera was born in MUTA, Jordan in 1985. He received his B. Eng in Electronic and telecommunication engineering from Misr University for Science & Technology (MUST) in 2010. And He received his master degree in Telecommunication Engineering from Universiti Teknikal Malaysia Melaka (UTeM) in 2015, and obtained his Doctorate from Faculty of Electronic and Computer Engineering (FKEKK), Universiti Teknikal Malaysia Melaka (UTeM), where his PhD research focuses on circularly polarised antenna and Microwave Engineering.