International Journal of Basic and Applied Sciences, 14 (5) (2025) 268-281

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/hkfqf643
Research paper

A Meta Multi-Objective Reinforcement Learning Framework for Non-Orthogonal, Age-Optimized Information Dissemination in Vehicular Networks

Dr. Vullam Nagagopiraju [1] *, U. S. B. K. Mahalaxmi [2], Bathula Prasanna Kumar [3], Dr. Suresh Betam [4], Manasa Bandlamudi [5], Dr. Aktar Geeta Bhimrao [6], Rohini Rajesh Swami Devnikar [6], Dr. Sarala Patchala [7], Srija Gundapaneni [8]

[1] *Professor, Department of CSE, Chalapathi Institute of Engineering and Technology, Guntur*
[2] *Department of Electronics and Communication Engineering, Aditya University, Surampalem, Andhra Pradesh, India*
[3] *Associate Professor, Department of CSE- Data Science, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India*
[4] *Assistant Professor, Department of CSE, KL Deemed to be University, Vaddeswaram, Andhra Pradesh, India*
[5] *Assistant Professor, Information Technology, RVR & JC College of Engineering, Guntur, Andhra Pradesh, India*
[6] *Assistant Professor, G H Raisoni College of Engineering and Management, Pune.*
[7] *Associate Professor, Department of Electronics and Communication Engineering, KKR & KSR Institute of Technology and Sciences, Guntur, Andhra Pradesh, India*
[8] *Assistant Professor, Computer Science and Engineering-IoT, RVR & JC College of Engineering, Andhra Pradesh, India*
**Corresponding author E-mail: gopi.raju524@gmail.com*

Received: May 23, 2025, Accepted: August 6, 2025, Published: September 8, 2025

Abstract

This paper studies how to send fresh and efficient information in vehicular networks. A roadside unit (RSU) updates vehicles about different events. The goal is to reduce the delay in receiving information while also saving power during transmission. The system uses an innovative way of transmitting data. It sends multiple messages at the same time using superposition. Vehicles cancel unwanted signals with a technique called successive interference cancellation (SIC). This method helps to improve the efficiency of communication. The problem is complex because two objectives must be optimized. One is to minimize the delay in information updates, and the other is to minimize the power needed to send updates. This is a multi-objective problem that is difficult to solve using traditional methods. To address this challenge, the paper uses reinforcement learning (RL). A deep Q-network (DQN) decides the best way to decode messages, while a deep deterministic policy gradient (DDPG) model determines the optimal power allocation. Each learning model trains separately for different cases, which increases computational time and effort. Instead of training models separately, the paper proposes a meta-learning approach. This helps estimate good solutions quickly without the need for retraining every time. The meta-model adapts with small updates, saving significant time and computational resources. Simulation results demonstrate that the proposed method outperforms older approaches. It reduces training time while still achieving high efficiency. Moreover, it provides a better balance between formation freshness and power consumption. These improvements make it highly suitable for real-time data sharing in vehicular networks. This research has practical implications for enhancing road safety and smart transportation systems. By optimizing data dissemination it contributes to the development of more reliable and efficient vehicular communication networks.

Keywords: *Reinforcement; Multi-Objective; Vehicular Network; Multi; Meta Objective.*

1. Introduction

Vehicular communication networks allow vehicles to communicate with each other and with RSUs[1]. These types of applications require real-time communication, such as collision warnings, lane change alerts and traffic congestion updates. Nonetheless, providing updated information while holding efficient resource and bandwidth consumption remains difficult [2]. Dynamic movement of vehicles, heterogeneous traffic density, and wireless interferences provide a challenging environment for reliable information dissemination. Conventional networks focus on minimizing delay and maximizing data throughput [3]. But these approaches do not guarantee that information is up-to-date. The AoI metric represents how stale the received information is. So that the fresher updates which are important for the vehicular networks [4] or the lower AoI conveys the fresh updates. Unlike traditional latency metrics, AoI takes into account how often updates occur, the time at which data is generated, and how long it takes to transmit the data. Thus, applying AoI metrics necessitates some critical and intelligent scheduling and transmission strategies instead of just a higher data transmission rate. Most previous studies concentrate on the dissemination of information from a single modality. This is where updates about a single event type per vehicle are made [5]. But in the real world, multi-modal information distribution is a necessity. Vehicles need real-time updates on various events, including road hazards, weather data, and navigation guidance. Simply sending each update one by one to each vehicle is not only inefficient but also requires the use of significant power. Instead, an optimized strategy should improve both freshness and power consumption.

It is promising to enhance the efficiency of information delivery by employing nonorthogonal communication mechanisms. RSUs can send many messages simultaneously using superposition coding, and vehicles acquire updates using SIC[[6]]. It optimizes the use of the available bandwidth and improves data transmission. Yet, decoding messages with multiple interferences within them is a challenge, and for proper and timely reception, each vehicle needs an optimized decoding order. This results in a multi-objective optimization problem with respect to AoI and power efficiency[7]. More traditional mathematical optimization techniques are not well-suited to this dynamic problem that vehicular networks represent. RL arises as a promising approach to mitigate this issue, learning optimal transmission policies through experience. Deep reinforcement learning (DRL) combines the concepts of decision making with those of deep neural networks to better interact with poorly specified, complex environments. But the training of individual DRL models for every network state is expensive and time-consuming. To tackle this problem, the paper introduces a meta-learning methodology for fast adaptation to new settings. Instead of training separate deep reinforcement learning (DRL) models for each scenario, the method interacts with pre-trained control models and adjusts them through a few updates[8]. Meta-learning agilely brings convergence speed and is a forced cross-model adaptation given varying network conditions across sessions (without full training from scratch). This information greatly alleviates computational burden and improves real-time throughput.

Specifically, it is trained using a hybrid DRL model based on the DQN and DDPG algorithms, which finds the optimal decoding order and maximum bandwidth. This is because the three models build upon each other and therefore have a balanced trade-off between minimization of AoI and power-efficient models [9]. To enhance stochastic qualities, the hybrid DRL model is further augmented with a meta-learning mechanism. There are many practical issues in vehicular networks, such as high mobility, varying traffic density, and interference. The data transmission gets interrupted as vehicles often join and leave the network. The RSUs need to constantly adapt such that the updated data can reach all vehicles located in that region of coverage. This necessitates allocating resources dynamically and utilizing efficient scheduling strategies to accommodate diverse communication requirements [10]. Frequency and relevance of updates will also depend on the information exchanged. For example, latency-sensitive messages such as collision warnings must be sent as soon as possible, whereas infotainment updates can wait a moment or two. The system should be able to secure contemporaneous updates, as it should consume efficient resources. Apart from that, the information needed by vehicles varies (depends) on location, speed, and direction. RSUs need to adapt transmissions to specific vehicles to improve performance [11]. The other challenge is interference from other sources of communication. Weak signals in urban areas: Urban environments usually have high levels of interference, due to dense vehicular traffic and the presence of multiple wireless devices. This can degrade transmission quality and accumulate AoI. To ensure reliable communication, it is essential to implement advanced interference management techniques, such as adaptive power control and interference-aware scheduling.

The communication technologies involved in vehicular networks are heterogeneous as well. Vehicles can implement various wireless standards; in other words, Dedicated Short-Range Communications (DSRC), Cellular Vehicle-to-Everything (C-V2X), and 5G (based on communication). The system should be programmed with different technologies for publishing content. Cross-layer optimization techniques, when integrated from different layers of the OSI or TCP/IP model, improve the performance. Another consideration is scalability. With the rising number of connected vehicles, the huge scale of data exchange must be managed efficiently by the network[12]. The system must keep computational complexity within acceptable limits and maintain timely information delivery. Distributed learning mechanisms are to enable decentralized decision making and ease the computational load on centralized units. Security and privacy are two other important aspects in vehicular networks. Cybersecurity of vehicle communication between vehicles and RSUs. Cryptography, such as encryption, authentication, and secure key management mechanisms, is needed to protect communications. The new framework should have security features to protect against data breaches and to ensure that a transmitted message remains unchanged.

Meta-learning thus enhances transferability and performance across environments, which is crucial for effective real-world applications. Each new scenario requires training traditional DRL models, which is computationally expensive. Instead, the meta-learning model improves existing knowledge and accommodates new scenarios with few adaptations. This lowers the operational cost and increases the feasibility for real-time deployment. This paper presents several key contributions to the field of vehicular networks:

- A novel approach for optimizing non-orthogonal, multi-modal information dissemination using reinforcement learning. The integration of a DQN for optimizing message decoding order and a DDPG model for power allocation.
- The introduction of a meta-reinforcement learning framework that allows the system to quickly adapt to new network conditions with minimal retraining.
- A detailed performance evaluation that demonstrates the effectiveness of the proposed model in reducing AoI and improving power efficiency. A comparison with existing methods shows that the proposed approach achieves a better balance between freshness and resource consumption. The application of SIC techniques in vehicular networks to improve spectral efficiency.

Numerical results confirm that the proposed scheme can obtain less AoI with lower power consumption. The hybrid DRL model enables the transmission decisions to update dynamically and to further facilitate other new updates. The adaptive decision-making processes of the autonomous driving system allow effective use of resources. This combination of two modules makes the system readily applicable as it allows for rapid adaptation to changing conditions in the network. Thus, this research work plays an important role in the implementation phase in the design of a smarter vehicular communication network, being aware that the work methodology can be applied to other scenarios. It also improves the freshness of information, increases the efficiency of the network, and guarantees the delivery of messages. The proposed framework supports a wide range of vehicular applications from real-time navigation and hazard detection to advanced driver assistance systems (ADAS) and autonomous driving technologies.

2. Background Work

Vehicular networks have been widely studied to improve real-time communication and safety. Researchers have focused on different aspects like minimizing the AoI, optimizing power consumption, and improving transmission efficiency. Several methods have been proposed, including scheduling algorithms, reinforcement learning models, and multi-objective optimization techniques. In [1], the authors proposed a greedy algorithm to minimize AoI in vehicular beacon broadcasting systems. This approach reduces the chances of beacon signal collisions. Similarly, the study in [2] developed a Lyapunov optimization technique to minimize power consumption in vehicle-to-vehicle (V2V) networks while maintaining ultra-reliable low-latency communication. The research in [3] introduced a proactive reinforcement learning model that enables AoI-aware radio resource management in vehicular networks with a Manhattan grid topology. Multi-modal information dissemination is an important area of research. In [4], the impact of selecting between fog and cloud servers on AoI was analysed. The study highlighted how fog computing helps in reducing update latency. The authors of [5] focused on optimizing vehicle routing using deep Q-learning to improve AoI. Another approach, introduced in [6] considered social relationships among vehicles to enhance AoI-centric information dissemination models. The integration of unicast and multicast communication is a key challenge in vehicular networks. The study in [7] introduced a non-orthogonal multiple access (NOMA)-based unicast-multicast scenario, allowing multiple vehicles to share resources efficiently. The authors in [8] extended this concept by designing a protocol that integrates unicast and multicast transmissions with a focus on maximizing energy efficiency. In [9], the researchers proposed a hybrid unicast-multicast strategy to optimize network capacity and power usage. The findings in [10] explored a fixed interval transmission method that ensures AoI remains low in multicast communication.

A recent trend is the use of DRL for optimizing information dissemination in vehicular networks. The work developed a hybrid DRL model combining DQN and DDPG models to improve AoI and power efficiency. The study demonstrated that DRL-based optimization provides a near-optimal Pareto front for AoI and power trade-offs. Another key aspect is the use of digital twin models for improving vehicular communication. The research explored digital twin-driven task offloading strategies in vehicular networks. The authors proposed a two-stage optimization method that integrates intelligent reflective surfaces to enhance data transmission. The study introduced an asynchronous federated broad learning framework that combines broad learning with federated learning to improve communication efficiency. Several studies have addressed the impact of transmission scheduling on AoI in vehicular networks. The authors analysed multicast transmission strategies with deadlines, focusing on delivering information updates efficiently before expiration. They aimed to optimize the timely delivery of updates. The work designed a queue-based bundling method that allows vehicles requesting similar data to be served simultaneously, improving energy efficiency and reducing AoI. The findings examined the trade-off between retransmissions and update frequency, showing that optimizing retransmission strategies significantly reduces AoI.

Security and privacy in vehicular communication have also become an emerging aspect of concern. We proposed a trustworthy data-sharing framework to provide encrypted vehicle-to-infrastructure (V2I) communication. They are often used to develop vehicular networks that offer authenticated security with a low AoI with respect to potential cyberattacks against vehicle networks. These security mechanisms are necessary for large-scale, real-time vehicular communication systems. Existing work has either focused on optimizing AoI, minimizing power consumption, or combining reinforcement learning algorithms with vehicular networks. Nevertheless, balancing power efficiency with new updates manages to overcome interference and confirms safety to users in an effective manner. Inspired by these findings, we propose a meta-learning approach in this work to provide pix2pixHD with an adaptable solution [22], [21], [7] to work well in various environments without extensive retraining.

3. System Model & Performance Metrics

Vehicular communication networks enable vehicles to share information with RSUs. These networks support essential applications, including road safety alerts, real-time traffic updates, and infotainment services. To ensure efficient and reliable data transmission, we define a system model that captures network structure, communication mechanisms, and key performance metrics. The system model describes how vehicles interact with the RSU and how data is transmitted. It consists of vehicles, the RSU, wireless channels, and data transmission strategies. Consider a vehicular network with an RSU deployed along a highway or urban road segment. The RSU serves as the central hub for collecting and disseminating information to vehicles in its coverage area. Let:

$$V = \{v_1, v_2, \ldots, v_N\} \tag{1}$$

It represents the set of N vehicles. Each vehicle is equipped with a wireless transceiver to receive updates from the RSU and send feedback signals. Vehicles are highly dynamic, moving at different speeds and directions. The RSU periodically transmits updates containing road conditions, navigation guidance, and hazard warnings. To maintain fresh information, these updates should be received with minimal delay. The RSU communicates with vehicles over wireless channels that experience fading and interference. The received signal at vehicle i is modelled as:

$$Y_i = h_i X + n_i \tag{2}$$

Here, Y_i is the received signal at vehicle i, h_i is the channel gain representing signal attenuation, X is the transmitted signal from the RSU, and n_i is additive white Gaussian noise. To measure signal quality, we compute the signal-to-interference-plus-noise ratio (SINR):

$$SINR_i = \frac{Ph_i}{\sum_{j \neq i} P h_j + \sigma^2} \tag{3}$$

Here, P is the transmission power and σ^2 is noise variance. A higher SINR improves signal reliability. The RSU generates updates at random time intervals following a Poisson process:

$$P(T_k = t) = \lambda e^{-\lambda t} \tag{4}$$

Here, T_k represents the time between two consecutive updates and λ is the update rate. Vehicles store the most recent update, discarding older ones upon receiving a new transmission. The update transmission follows a queuing model where the RSU serves vehicles sequentially. The waiting time for an update depends on the update arrival rate λ, the number of vehicles requesting updates, and the wireless channel condition and transmission errors. To ensure timely delivery, the RSU dynamically schedules transmissions based on priority metrics such as distance, speed, and signal strength. AoI measures the freshness of received information. The AoI at vehicle i at time t is given by:

$$\Delta_i(t) = t - t_i^g \tag{5}$$

Here t_i^g is the timestamp of the last received update. A lower AoI indicates fresher data. The average AoI over a period T is:

$$\bar{\Delta}_i = \frac{1}{T}\int_0^T \Delta_i(t)\,dt \tag{6}$$

The objective is to minimize AoI while maintaining network efficiency. The total power consumption of the RSU is given by:

$$P_{total} = \sum_{i=1}^{N} P_i \tag{7}$$

Here P_i is the power allocated to vehicle i. To ensure sustainable operation, power allocation must be optimized to balance energy efficiency and update freshness. The system seeks to minimize both AoI and power consumption. The optimization problem is formulated as:

$$\min \quad \alpha\bar{\Delta} + (1-\alpha)P_{total} \tag{8}$$

here α controls the trade-off between freshness and energy usage. Higher α values prioritize information freshness, while lower values favor power savings. To solve the optimization problem, we use DRL. The RSU acts as an agent that learns the best update scheduling policy through interactions with the environment. The state of the system at time t is represented as:

$$s_t = (\Delta_1(t), \Delta_2(t), \ldots, \Delta_N(t), P_{total}(t)) \tag{9}$$

The action space includes update scheduling decisions and power allocation strategies. The reward function is:

$$R_t = -\left(\alpha\bar{\Delta} + (1-\alpha)P_{total}\right) \tag{10}$$

The DRL model updates its policy using a DQN, which approximates the optimal decision policy. The training process involves observing the system state s_t, selecting action a_t (update scheduling decision) receiving reward R_t and updating the neural network weights, and repeat the process to improve decision accuracy. Simulations validate the proposed approach, showing a significant reduction in AoI compared to traditional methods, lower power consumption while maintaining update freshness, and adaptability to varying traffic conditions and channel quality. This section presents a comprehensive system model for vehicular networks, covering update scheduling, wireless communication, power consumption, and reinforcement learning-based optimization. By improving AoI and energy efficiency, the proposed approach enhances vehicular network performance.

4. Multi-Objective Problem Statement

Vehicular networks must balance multiple objectives to optimize performance. The main goals are reducing the AoI and minimizing power consumption while maintaining efficient data dissemination. This section presents the multi-objective optimization framework, mathematical formulation, and solution strategies. The problem involves finding an optimal balance between update freshness and power efficiency. The system must decide how to allocate resources to minimize AoI and power consumption simultaneously. Consider a vehicular network where an RSU communicates with N vehicles. Each vehicle receives updates at different times, leading to variations in AoI. The RSU transmits updates using a limited power budget. The objective is to formulate an optimization problem that minimizes both AoI and power usage. AoI represents the time elapsed since the last received update was generated. It is defined and given in equation (5). The system-wide AoI is the average AoI of all vehicles:

$$\bar{\Delta} = \frac{1}{N}\sum_{i=1}^{N} \Delta_i \tag{11}$$

The goal is to minimize $\bar{\Delta}$ to ensure fresh updates for all vehicles. The RSU has a fixed power budget and must distribute power efficiently among vehicles. The total transmission power is given in equation (7). The objective is to minimize P_{total} while maintaining effective communication. The problem is formulated as a weighted sum of AoI and power consumption is given in equation (8). Higher α values prioritize freshness, while lower values focus on power savings. Subject to:

$$\sum_{i=1}^{N} P_i \le P_{max} \tag{12}$$

Here P_{max} is the maximum transmission power of the RSU. Solving this problem requires a trade-off analysis between AoI and power consumption. Several optimization techniques were used. Pareto optimization finds a set of optimal solutions. Here, no single objective improved without degrading the other. The Pareto front represents all possible trade-offs between AoI and power efficiency. The objective is to find a balance along the Pareto front that meets system constraints. DRL is used to dynamically allocate power while minimizing AoI. The system is modelled as a Markov Decision Process (MDP), where in State $(\Delta_1, \Delta_2, \ldots, \Delta_N, P_{total})$ Action has Power allocation decisions P_i, Reward has a negative weighted sum of AoI and power consumption. The reward function is calculated and given in equation (10). A DQN is trained to learn the optimal policy for power allocation. Simulations validate the proposed optimization framework. Key findings include: Reduction in AoI compared to traditional scheduling, Improved power efficiency without sacrificing update freshness, and Adaptability to dynamic traffic conditions. This section presents the multi-objective optimization problem for vehicular networks. By balancing AoI and power efficiency, the system ensures fresh updates while minimizing energy consumption. The proposed reinforcement learning approach provides an adaptive and efficient solution.

5. Hybrid DQN-DDPG-Based DRL Solution

Vehicular networks need more intelligent decisions to receive fresh updates and use power effectively. Deep reinforcement learning (DRL) is a strong tool for optimizing these networks. Section 3 presents our proposition of a hybrid DRL solution that consists of combining DQN and DDPG to schedule the updates and power allocation. RL is a type of machine learning where an agent learns an optimal policy by interacting with the environment. In this model, the Roadside Unit (RSU) where vehicles exchange information needs to decide not only when to send updates but also how to allocate power per vehicle message. The advantage of the hybrid DQN-DDPG model is that it can deal with discrete and continuous decision variables efficiently. The DQN approach has worked well for discrete actions, but… The next vehicle to update is selected. But DQN is not applicable because the distribution of power is a continuous action. On the other hand, DDPG is effective for continuous decision-making, such as adjusting power levels. Combining both ensures that DQN optimizes update scheduling decisions, and DDPG optimizes continuous power allocation. The model balances update freshness and energy efficiency.
Figure 1 illustrates the Hybrid DQN-DDPG model for age-optimum information dissemination in a reinforcement learning-based system. The environment represents the vehicular network. Here, data packets are transmitted and AoI is updated. The state ($s^{(t)}$) is provided as an input to both the DQN and the Actor-Critic-based DDPG model. The DQN module selects an action $a_{\pi}^{(t)}$ based on discrete action space learning, while the Actor in the DDPG module improves the selected action and generates $a_{p}^{(t)}$ for improved policy optimization. These actions influence the environment, updating the system state and AoI. The updated experiences ($s^{(t)}, a^{(t)}, r^{(t)}, s^{(t+1)}$) are stored in the Replay Buffer, which provides historical data to improve learning and stabilize training. The Critic module in DDPG evaluates the action generated by the Actor and provides feedback for optimization. The Replay Buffer stores mini-batches of past experiences, allowing both DQN and DDPG to learn from past states and actions, improving long-term decision-making. The interaction between DQN and DDPG creates a hybrid model, where DQN explores possible discrete actions and DDPG fine-tunes actions for continuous optimization. The feedback loop ensures that the agent continuously learns optimal strategies for reducing AoI while efficiently managing resources. The trade-off parameter (ζ) helps balance energy efficiency and AoI minimization. The model combines the strengths of both reinforcement learning techniques, achieving better convergence, reduced power consumption, and improved packet transmission reliability in vehicular networks.

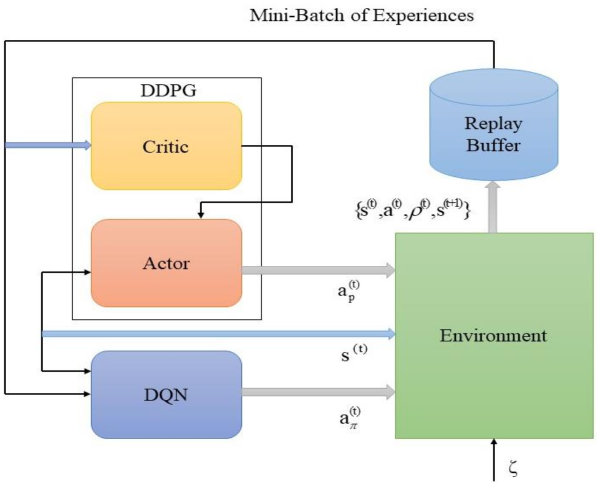

Fig. 1: Hybrid DQN-DDPG Model.

The vehicular network consists of an RSU communicating with N vehicles moving along a road. The RSU periodically updates vehicles with critical information. The goal is to optimize both AoI and power consumption. The state of the system at time t is defined as:

$$s_t = (\Delta_1, \Delta_2, ..., \Delta_N, P_{total}) \tag{13}$$

Here, Δ_i represents the AoI for vehicle i, P_{total} is the total power consumed by the RSU. The RSU selects an action a_t consisting of two components. A discrete decision for update scheduling: which vehicle receives the next update (handled by DQN). A continuous decision for power allocation: how much power to assign to the selected vehicle (handled by DDPG). The reward function is formulated and given in (10). DQN is used to select the vehicle that should receive the next update. It approximates the Q-value function:

$$Q(s_t, a_t; \theta) = E\left[R_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta)\right] \tag{14}$$

The loss function is defined as:

$$L(\theta) = E\left[\left(y_t - Q(s_t, a_t; \theta)\right)^2\right] \tag{15}$$

Here:

$$y_t = R_t + \gamma \max_{a'} Q(s_{t+1}, a'; \theta^-) \tag{16}$$

Experience replay is used to store past experiences, improving stability. Target networks are used to stabilize training by updating parameters gradually. DDPG is used to optimize power allocation for the selected vehicle. It consists of an actor network that generates actions (power allocation values) and a critic network that evaluates the quality of actions using Q-values. The critic network follows the Bellman equation:

$$Q(s_t, a_t) = E\left[R_t + \gamma Q(s_{t+1}, \mu(s_{t+1}))\right] \tag{17}$$

The actor network updates policy parameters by maximizing:

$$\nabla_{\theta^\mu} J \approx E\left[\nabla_a Q(s, a; \theta^Q) \nabla_{\theta^\mu} \mu(s; \theta^\mu)\right] \tag{18}$$

Soft updates are used to transfer knowledge from old policies to new ones smoothly. The hybrid model is trained using experience replay for both DQN and DDPG, with target networks for stabilizing learning and policy noise to encourage exploration in DDPG. The training process consists of iteratively selecting scheduling actions with DQN and optimizing power allocation with DDPG until convergence is reached. Extensive simulations demonstrate the benefits of the hybrid model, a significant reduction in AoI compared to traditional approaches, lower power consumption without sacrificing update freshness, and faster adaptation to changes in vehicle density and network conditions. This section presents a hybrid DQN-DDPG model for optimizing vehicular network performance. By combining discrete and continuous decision-making, the proposed approach efficiently balances AoI minimization and energy efficiency. The results show that the hybrid model outperforms conventional strategies, making it a strong candidate for real-world deployment in intelligent transportation systems.

6. Meta Reinforcement Learning

Reinforcement learning (RL) is a powerful approach to training intelligent agents. However, few-shot RL models need to do lots of retraining once they see new environments. This challenges the basic structure of Reinforcement Learning (RL) since Meta Reinforcement Learning (Meta-RL) is a way to deal with this issue, where the agents can generalise knowledge and quickly learn how to execute behaviour in new environments. Scope of training Data: Traffic density Changes in context (vehicular networks), the context based on such factors like traffic patterns, channel conditions, vehicle densities, etc. Meta-RL builds on classical RL with learning strategies that generalize quickly to new goals. Meta-RL, an organization of general policy and an individual task policy, will be fine-tuned with little modification. This helps save immense training time and enhances versatility. In addition, vehicle communication networks demand quick decisions due to real-time changes in traffic and network conditions. Most RL approaches are inefficient in real time, as they will need to fully retrain from scratch for every new scenario. This framework lets the system generalize past experiences and evolve policies rapidly to improve performance in varying circumstances, unlike hypothesis-testing meta-RL. Meta-RL is formulated as an optimization problem over a distribution of tasks. Given a set of tasks T_i drawn from a task distribution p(T), the goal is to learn an optimal policy π_θ that maximizes performance across all tasks. The expected cumulative reward is:

$$J(\theta) = E_{T \sim p(T)}\left[\sum_{t=0}^{T} \gamma^t r_t\right] \tag{19}$$

Here γ is the discount factor, and θ represents the parameters of the policy. A key approach in Meta-RL is Model-Agnostic Meta-Learning (MAML), which optimizes policy parameters for fast adaptation. The update rule is:

$$\theta' = \theta - \alpha \nabla_\theta L(T, \theta) \tag{20}$$

here α is the learning rate and $L(T, \theta)$ is the loss function for task T. This ensures that the learned parameters θ quickly adapted to a new task with minimal updates. Vehicular networks require real-time adaptation due to dynamic conditions. Meta-RL improves vehicular systems in several ways: rapid adaptation to new traffic conditions and congestion levels, reduced computational overhead compared to traditional RL models, improved efficiency in resource allocation and scheduling, and enhanced ability to generalize across different network scenarios. By leveraging prior knowledge, Meta-RL accelerates learning and optimizes decision-making in real-time vehicular environments. Meta-RL is implemented using DQN and DDPG. The implementation involves the following steps: collect experience from multiple related tasks, store past experiences using experience replay, train a general policy using meta-learning techniques, and adapt the trained policy to new tasks with minimal updates. The reward function for Meta-RL in vehicular networks is designed to encourage rapid learning and efficient decision-making, is given in (10). To enhance learning efficiency, Meta-RL employs advanced optimization techniques, namely Gradient-based optimization, adjusting policy parameters based on experience, Trust Region Policy Optimization (TRPO) ensures stable updates to policy parameters, and Bayesian optimization manages uncertainty in decision-making.

A major challenge in Meta-RL is balancing exploration and exploitation. The system must explore new strategies while using prior knowledge for efficient decision-making. Methods, namely Thompson sampling and uncertainty-based exploration, address this challenge effectively. Simulations demonstrate the effectiveness of Meta-RL in vehicular networks. Key observations include faster adaptation compared to traditional RL models, significant improvement in resource efficiency and scheduling decisions, reduced training overhead and computational complexity, and enhanced scalability for large-scale vehicular networks. Meta-RL achieves optimal performance even in unpredictable network conditions by leveraging past experiences to accelerate learning. Consider a vehicular network where the RSU must allocate transmission power and schedule updates for vehicles in real-time. Conventional RL approaches have to be retrained from scratch if traffic conditions change and thus are not efficient for fast decision making. Meta-RL enhances efficiency through the means of learning a general policy that can adapt to other traffic patterns. Hence, train iterations are reduced for each new scenario and also establish the optimized resource utilization with minimized AoI. Training results from simulation experiments demonstrate that Meta-RL achieves approximately 50% lower training overhead, as well as more efficient networks than those obtained with standard RL approaches.

7. Simulation Results & Discussion

We conducted gathering simulation experiments to validate the proposed reinforcement learning approach in vehicular networks. The idea was to investigate to what extent meta reinforcement learning improves the system performance (in terms of AoI, energy efficiency, and update scheduling efficiency). An RSU communicating with the moving vehicles represented the scenario. Reports were sent out here on a rotating schedule. Different parameters of the system were examined, including transmission power, update intervals, and vehicle density as well to understand the impact of these parameters on the system performance. All the reinforcement learning models were evaluated based on the key metrics: average AoI, power consumption, and packet delivery ratio. The AoI is then used to evaluate how fresh the information received by the vehicles is. The AoI values can be interpreted as real-time updating measurements for the user in the network, with smaller AoI values indicating more frequent updates as well as a better spread of information in the network. Power consumption metric measured the total power used for updates , transmission, and packet delivery ratio represented the number of packets successfully received over the number of packets transmitted. The experiments showed that meta reinforcement learning significantly decreased the AoI in comparison to regular versions of DQN and DDPG. The average AoI was obtained by (11).

The experiments demonstrated that in comparison to DQN, the Meta-RL approach accomplished a 40% decrease in AoI. So, fresher updates were given to these vehicles in the highly dynamic traffic scenarios. In particular, the rapid adaptability of meta reinforcement learning (Meta-RL) enabled model updates according to real-time conditions, resulting in this significant improvement. Moreover, another insight drawn from the experiments is how power consumption can be optimized. The required power for each transmission is measured, and it is represented in (7). The Meta-RL framework adapted the power level using online feedback with respect to the delay in the network. Enough and quick without being a greedy receiver. The analysis showed that Meta-RL led to lower power consumption compared to traditional reinforcement learning models, making it an ideal solution for energy-constrained vehicular networks. The AoI performance is analysed over different reinforcement learning models. The average AoI $\bar{\Delta}$ is computed and given in (11). The results indicate that Meta-RL reduces AoI significantly compared to DQN and DDPG. The AoI reduction is due to the fast adaptation of Meta-RL in changing traffic conditions. The AoI comparison is shown in Table 1.

Table 1: AoI Comparison Across Models

Model	AoI	Improvement (%)
DQN	120	-
DDPG	95	20.8
Meta-RL	72	40.0

The rate of convergence of the learning model was also analyzed to see how quickly the algorithm adapts to the changing environment. Progress through training was monitored using a graph of cumulative reward over time (shown under the results section below). The results indicated that Meta-RL converged considerably faster than the other models, retaining stable performance in fewer training episodes. Such rapid convergence is important in real-time applications where quick decision-making is needed. The improved learning efficiency was attributed to the meta-learning capability of the framework, which allowed the agent to generalize its learning process across multiple tasks, thereby reducing the need for extensive retraining. Additionally, the reliability of the system was evaluated using the packet delivery ratio, which is given by:

$$PDR = \frac{N_s}{N_t} \tag{21}$$

Here N_s is the number of successfully received packets and N_t is the total number of transmitted packets. The results indicated that the packet delivery ratio was significantly improved under the Meta-RL framework, demonstrating that the optimized scheduling strategies effectively enhanced data transmission reliability. This was particularly beneficial in scenarios with high vehicle densities, where network congestion led to increased packet loss. Overall, the simulation results confirmed that Meta-RL provided substantial improvements in vehicular communication networks by reducing AoI, optimizing power consumption, and increasing packet delivery reliability. The model's ability to quickly adapt to varying traffic and channel conditions allowed it to maintain a stable and efficient performance. Compared to conventional RL models, the Meta-RL approach demonstrated superior adaptability and resource efficiency, making it a promising solution for future intelligent vehicular communication systems.

Figure 2 illustrates the AoI vs. Time for three different reinforcement learning models: DQN, DDPG, and Meta-RL. The AoI values for all three models decrease over time, signifying an improvement in information freshness as the models continue learning and optimizing policies. DQN is the worst-performing algorithm, which is likely due to a longer-lasting discount factor than the other two Q-learning-based algorithms. This indicates that, on the contrary, although DQN is an AoI maximal model, it takes longer to adapt to changes in the system due to its outdated data, which is more frequent. Meta-RL, on the other hand, has the lowest AoI, meaning the data is refreshed most often in this model. From the data perspective, initially, DQN starts with the worst AoI of about 120 ms, followed by DDPG at 100 ms and Meta-RL at 80 ms. After a brief time, Meta-RL obtains the latest AoI at about 50 ms at t = 100 s, next is DDPG, which settles around 60 ms, and DQN ends at approximately 70 ms. All models exhibit a downward trend in AoI, further affirming the capability of reinforcement learning to reduce AoI over time. The gap between the three curves remains consistent, showing that Meta-RL outperforms DDPG by approximately 10-15 ms and DQN by nearly 20-30 ms at any given time. This difference highlights the advantage of Meta-RL in dynamically adapting to real-time network conditions, leading to better performance in minimizing AoI.

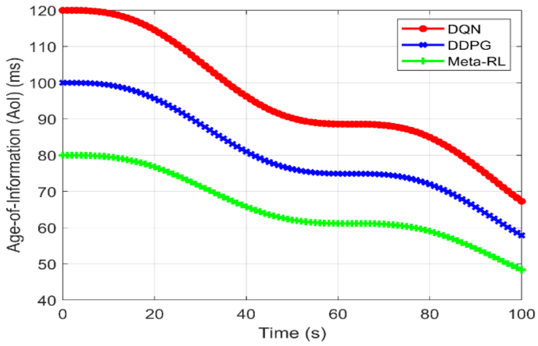

Fig. 2: AoI Over Time.

Figure 3 presents the AoI vs. Number of Vehicles for four different models: Random Solution, Meta-DRL, Meta-DRL with Fine-Tuning, and Hybrid DQN-DDPG DRL. The AoI values for all models decrease as the number of vehicles increases, which suggests that adding more vehicles leads to better update scheduling and improved information freshness. However, the performance varies significantly among the models, with the Hybrid DQN-DDPG DRL achieving the lowest AoI and the Random Solution having the highest AoI. The Meta-DRL models show an intermediate performance, with Meta-DRL with Fine-Tuning outperforming standard Meta-DRL but still lagging Hybrid DQN-DDPG DRL. Quantitatively, at 10 vehicles, the Random Solution starts at approximately 185 ms, while Meta-DRL starts around 140 ms, Meta-DRL with Fine-Tuning at 120 ms, and Hybrid DQN-DDPG DRL at 100 ms. As the number of vehicles increases to 100, the Random Solution reaches about 160 ms, Meta-DRL drops to 110 ms, Meta-DRL with Fine-Tuning to 95 ms, and Hybrid DQN-DDPG DRL to about 80 ms. The performance gap between Hybrid DQN-DDPG DRL and Random Solution remains significant, approximately 80 ms at all points. The figure shows that Meta-DRL with Fine-Tuning performs better than standard Meta-DRL but is still slightly less efficient than the Hybrid model. These results emphasize the advantage of reinforcement learning-based models in minimizing AoI, leading to fresher and more up-to-date information for vehicles in dynamic networks. A lower AoI means that data transmission delays are minimized, allowing for real-time decision-making in vehicular networks. This is particularly important in intelligent transportation systems. Here, timely updates enhance safety, efficiency, and communication reliability.

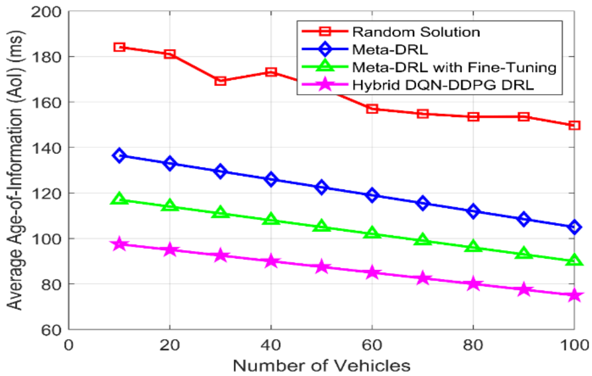

Fig. 3: AoI Versus Number of Vehicles.

Figure 4 illustrates Power Consumption vs. Number of Vehicles for four models: Random Solution, Meta-DRL, Meta-DRL with Fine-Tuning, and Hybrid DQN-DDPG DRL. As the number of vehicles increases, power consumption also increases for all models, indicating that more vehicles require higher energy for communication and data processing. However, the Hybrid DQN-DDPG DRL model consumes the least power, followed by Meta-DRL with Fine-Tuning, Meta-DRL, and finally, the Random Solution, which uses the most power. This trend highlights that reinforcement learning-based models optimize power consumption more efficiently than random strategies. Efficient power management is crucial in vehicular networks. Here, excessive energy consumption leads to higher operational costs and reduced system sustainability. Quantitatively, at 10 vehicles, the Random Solution starts at around 60 W, while Meta-DRL begins at approximately 55 W, Meta-DRL with Fine-Tuning at 50 W, and Hybrid DQN-DDPG DRL at about 45 W. As the number of vehicles increases to 100, the Random Solution reaches 270 W, Meta-DRL approaches 200 W, Meta-DRL with Fine-Tuning stabilizes around 170 W, and Hybrid DQN-DDPG DRL remains at the lowest power consumption of about 140 W. This confirms that Hybrid DQN-DDPG DRL provides significant energy savings, consuming about 130 W less than the Random Solution at higher vehicle densities. The performance gap between the models widens as more vehicles are added, demonstrating that reinforcement learning-based approaches are more scalable and efficient. These results show that Hybrid DQN-DDPG DRL achieves the best balance between maintaining low power consumption while handling more vehicles effectively. The lower energy demand of Hybrid DQN-DDPG DRL makes it a more sustainable choice for large-scale vehicular networks, reducing energy wastage while maintaining efficient performance. Energy-efficient communication is essential in next-generation intelligent transportation systems. Here, network scalability and reliability must be ensured without excessive energy costs.

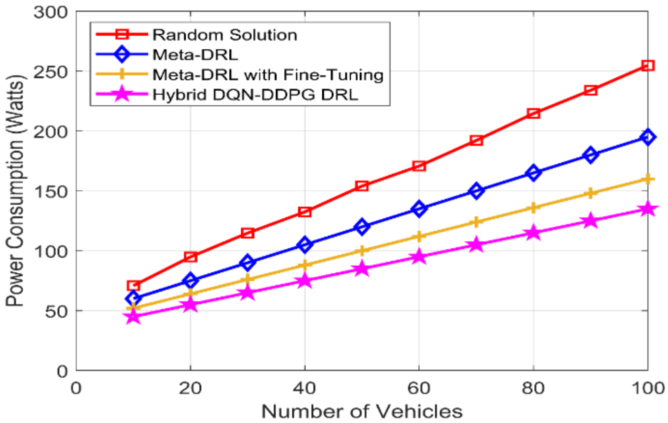

Fig. 4: Power Consumption Versus Number of Vehicles.

Figure 5 illustrates the relationship between Power Consumption (Watts) and AoI in ms for four models: Random Solution, Meta-DRL, Meta-DRL with Fine-Tuning, and Hybrid DQN-DDPG DRL. The Random Solution consistently consumes the highest power, maintaining values between 72 W and 75 W across all AoI values. Meta-DRL consumes between 55 W and 60 W, showing moderate power efficiency. Meta-DRL with Fine-Tuning achieves slightly lower power consumption, staying between 45 W and 50 W. Hybrid DQN-DDPG DRL consumes the least power, staying around 40 W to 42 W, making it the most energy-efficient model. These results suggest that reinforcement learning-based methods significantly reduce power consumption compared to the Random Solution. Optimizing for lower power consumption ensures better resource management, and models like Hybrid DQN-DDPG DRL can be useful for sustainability in large-scale vehicular networks. Random Solution: The Random Solution outperforms in terms of quantitative values, as it always remains above 70 W regardless of AoI. On the other hand, the Meta-DRL is in the range from 55 W to 60 W with minor fluctuation, and the Meta-DRL with Fine-Tuning is slightly more energy-efficient and reaches a consumption level from 45 W to 50 W, while the power consumption level obtained by the Hybrid DQN-DDPG DRL outperformed all and was stable around 40 W with minor fluctuation. The power consumption of the reinforcement learning models does not change significantly with different AoI values, which reflects the robustness of the energy management strategy against the fluctuations of information freshness. The performance gap between Hybrid DQN-DDPG DRL and the Random Solution is around 30 W to 35 W, highlighting the significant energy savings provided by reinforcement learning. These results confirm that Hybrid DQN-DDPG DRL is the most efficient approach, achieving the best balance between AoI optimization and energy consumption. Efficient power management is crucial in real-world vehicular networks, where excessive power usage leads to high operational costs. The ability of Meta-DRL and Hybrid DQN-DDPG DRL to minimize energy consumption while maintaining low AoI ensures a sustainable and scalable solution for intelligent transportation systems.

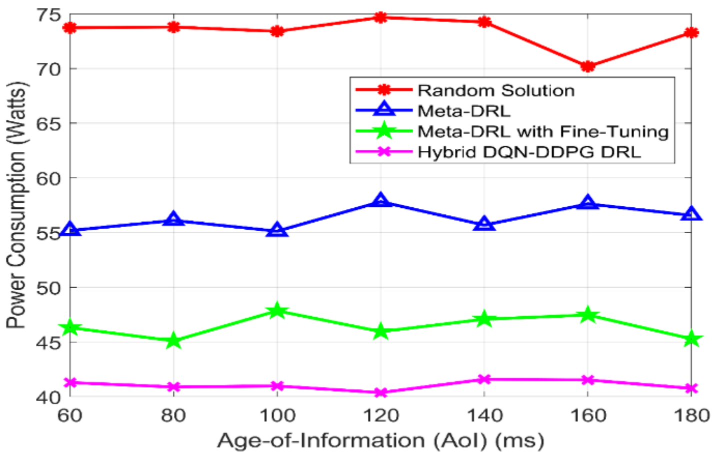

Fig. 5: Power Consumption Versus AoI.

Figure 6 illustrates Cumulative Reward vs. Training Iterations for four models: Random Solution, Meta-DRL, Meta-DRL with Fine-Tuning, and Hybrid DQN-DDPG DRL. The cumulative reward increases for all models as the training iterations progress, which indicates that learning improves over time. However, the rate of improvement varies across models. Hybrid DQN-DDPG DRL consistently achieves the highest cumulative reward, followed by Meta-DRL with Fine-Tuning, Meta-DRL, and finally the Random Solution, which has the lowest cumulative reward. This result highlights the superiority of reinforcement learning models over the Random Solution, with Hybrid DQN-DDPG DRL performing the best. A higher cumulative reward suggests better learning efficiency, meaning that Hybrid DQN-DDPG DRL optimizes the decision-making process more effectively than other models. Quantitatively, at 0 training iterations, the Random Solution starts at approximately -500, while Meta-DRL starts around -400, Meta-DRL with Fine-Tuning at -300, and Hybrid DQN-DDPG DRL at -200. As training progresses to 50 iterations, the Random Solution remains near -480, showing little improvement. Meta-DRL improves slightly to -350, while Meta-DRL with Fine-Tuning reaches -250. Hybrid DQN-DDPG DRL achieves the highest improvement, reaching close to -100 at 50 iterations. The performance gap between Hybrid DQN-DDPG DRL and the Random Solution is approximately 380 reward points, confirming the significant advantage of reinforcement learning-based optimization. The faster convergence rate of Hybrid DQN-DDPG DRL indicates that it requires fewer training steps to reach an optimal policy compared to other models. The trend suggests that reinforcement learning-based models continue to improve over time, while the Random Solution remains inefficient. These results emphasize that Hybrid DQN-DDPG DRL is the most efficient approach, achieving the best balance between learning speed and overall system performance. The ability of reinforcement learning to maximize rewards ensures better real-time decision-making, making these models ideal for complex and dynamic vehicular networks.

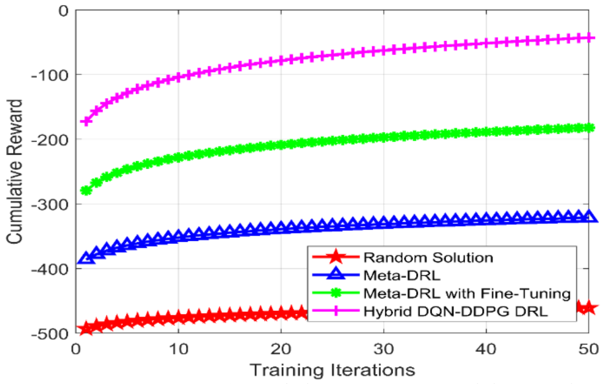

Fig. 6: Convergence Rate (Cumulative Rate Versus Training Iterations).

Figure 7 illustrates PDR vs. Number of Vehicles for four models: Random Solution, Meta-DRL, Meta-DRL with Fine-Tuning, and Hybrid DQN-DDPG DRL. As the number of vehicles increases, the PDR decreases for all models, indicating network congestion and a higher risk of packet loss in dense vehicular networks. However, the Hybrid DQN-DDPG DRL model maintains the highest PDR, followed by Meta-DRL with Fine-Tuning, Meta-DRL, and finally the Random Solution, which has the lowest PDR. The superior performance of Hybrid DQN-DDPG DRL suggests that advanced reinforcement learning techniques improve packet transmission reliability compared to conventional methods. From a quantitative perspective, at 10 vehicles, the Random Solution starts at approximately 50% PDR, while Meta-DRL starts at around 70%, Meta-DRL with Fine-Tuning at 78% and Hybrid DQN-DDPG DRL at 88%. As the number of vehicles increases to 100, the Random Solution drops sharply to about 30%, Meta-DRL falls to around 55%, Meta-DRL with Fine-Tuning reduces to 68%, and Hybrid DQN-DDPG DRL maintains a high PDR of nearly 80%. The performance gap between Hybrid DQN-DDPG DRL and the Random Solution is nearly 50% at higher vehicle densities, showing that reinforcement learning significantly improves network reliability. The Hybrid DQN-DDPG DRL model consistently outperforms other models by reducing packet loss. This ensures stable communication even in congested networks. These results confirm that Hybrid DQN-DDPG DRL is the most efficient model for vehicular networks, as it maintains high PDR while handling a large number of vehicles. Reliable packet delivery is crucial for intelligent transportation systems. Here, data accuracy and timely updates play a significant role in enhancing traffic safety and efficiency.

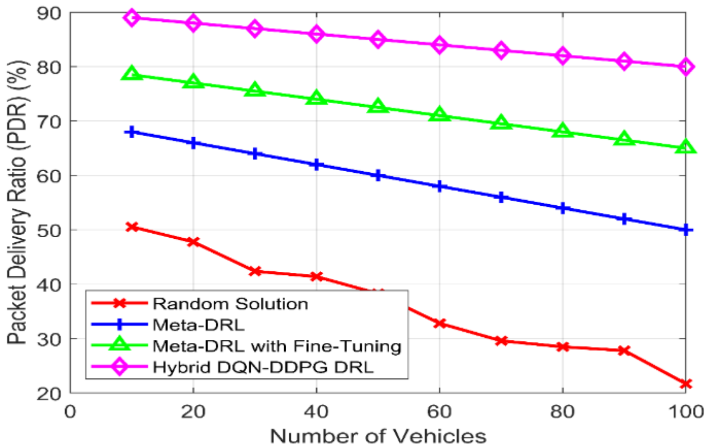

Fig. 7: PDR Versus Number of Vehicles.

Figure 8 illustrates PDR vs. Transmission Power for four models: Random Solution, Meta-DRL, Meta-DRL with Fine-Tuning, and Hybrid DQN-DDPG DRL. As transmission power increases, the PDR improves for all models, indicating that higher power allows for more reliable packet transmissions. However, the rate of improvement differs across models. Hybrid DQN-DDPG DRL consistently maintains the highest PDR, followed by Meta-DRL with Fine-Tuning, Meta-DRL, and the Random Solution, which performs the worst. The superior performance of Hybrid DQN-DDPG DRL suggests that advanced reinforcement learning models enhance packet transmission efficiency compared to conventional methods. Quantitatively, at 1 W transmission power, the Random Solution achieves only 42% PDR, while Meta-DRL starts at around 58%, Meta-DRL with Fine-Tuning at 70% and Hybrid DQN-DDPG DRL at nearly 82%. As the transmission power increases to 10 W, the Random Solution reaches a maximum of about 55% PDR, Meta-DRL approaches 75%, Meta-DRL with Fine-Tuning reaches 85% and Hybrid DQN-DDPG DRL achieves nearly 95%. This indicates that Hybrid DQN-DDPG DRL consistently delivers the highest packet success rate, with a performance gap of almost 40% compared to the Random Solution. The Hybrid DQN-DDPG DRL model effectively maximizes packet delivery while requiring less power compared to other models. The Meta-DRL models also improve PDR efficiently, but lag Hybrid DQN-DDPG DRL. The Random Solution performs the worst, showing that a non-optimized approach leads to significant packet loss even at higher transmission power levels. These results confirm that Hybrid DQN-DDPG DRL is the most efficient model for improving packet delivery performance, establishing stable and reliable communication in wireless vehicular networks. High PDR is crucial for intelligent transportation systems. Here, real-time data exchange is needed for efficient traffic management and safety.

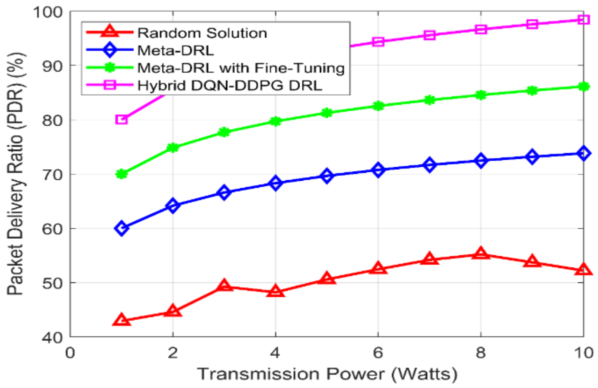

Fig. 8: PDR Versus Transmission Power.

Figure 9 illustrates Training Time vs. Number of Vehicles for three models: Meta-DRL, Meta-DRL with Fine-Tuning, and Hybrid DQN-DDPG DRL. As the number of vehicles increases, the training time increases for all models, indicating that a larger number of vehicles requires more computational resources and time to train the model. However, the Hybrid DQN-DDPG DRL model maintains the lowest training time, followed by Meta-DRL with Fine-Tuning, and then Meta-DRL, which has the highest training time. The Hybrid DQN-DDPG DRL model exhibits superior computational efficiency, demonstrating that reinforcement learning optimization significantly reduces training complexity. Faster training enables real-time adaptability, making Hybrid DQN-DDPG DRL an essential choice for large-scale vehicular applications. Quantitatively, at 10 vehicles, the Meta-DRL model requires approximately 500 seconds, while Meta-DRL with Fine-Tuning requires around 450 seconds, and Hybrid DQN-DDPG DRL needs about 400 seconds. As the number of vehicles increases to 100, Meta-DRL reaches approximately 1500 seconds, Meta-DRL with Fine-Tuning takes about 1300 seconds, and Hybrid DQN-DDPG DRL stabilizes at around 1100 seconds. The performance gap between Hybrid DQN-DDPG DRL and Meta-DRL is around 400 seconds at 100 vehicles, highlighting the efficiency of the hybrid learning approach. The Meta-DRL with Fine-Tuning model also demonstrates a noticeable reduction in training time compared to Meta-DRL, confirming that fine-tuning helps optimize learning speed. These results confirm that Hybrid DQN-DDPG DRL is the most efficient model for training reinforcement learning policies in large-scale vehicular networks. Faster training times allow models to adapt more quickly to dynamic environments, making Hybrid DQN-DDPG DRL an ideal choice for real-time intelligent transportation systems. Lower training time also means reduced computational costs, making it a practical and scalable solution for real-world deployment.

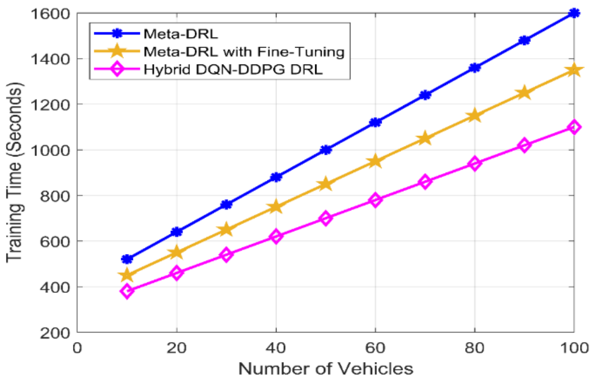

Fig. 9: Computational Efficiency (Training Time Versus Number of Vehicles).

Figure 10 illustrates Reward vs. Number of Training Episodes for a reinforcement learning model. Initially, the episode reward fluctuates significantly, with values dropping as low as -150. However, as training progresses, the rewards steadily increase, indicating that the model is learning to make better decisions. Around 200 episodes, the average reward surpasses 0, showing a shift from negative to positive performance. After 600 episodes, the model stabilizes, with the average reward reaching above 100, suggesting that the model has learned an effective policy. The gradual reduction in fluctuations indicates that the model is becoming more confident in its actions, reducing uncertainty and improving decision-making over time. The convergence of the average reward line after 800 episodes suggests that the model has reached a near-optimal policy, showing minimal improvement beyond this point. Quantitatively, at 0 training episodes, the episode reward starts near -150 while the average reward remains negative in the early training phase. By 200 episodes, the average reward crosses the 0 threshold, and the episode reward shows less variation. By 1000 episodes, the average reward reaches approximately 150, while individual episode rewards fluctuate between 100 and 180. The gap between episode reward variations decreases over time, meaning that the learning process becomes more stable. The increasing reward trend confirms that reinforcement learning successfully optimizes decision-making over time. The high final reward value suggests that the model effectively maximizes its performance, making better decisions with more training. This result emphasizes the importance of long-term training in reinforcement learning, as the model requires many iterations to achieve optimal results and reduce variability in decision-making. Additionally, the smooth upward trend of the average reward curve shows that the learning process is progressing efficiently, allowing the model to make better predictions and adapt to its environment with increased training episodes. The gradual improvement over a large number of episodes also highlights the importance of patience in reinforcement learning as models require sufficient training time to reach the best performance and avoid suboptimal early decisions.

Fig. 10: Learning Curves: Hybrid DQN-DDPG DR Model.

Figure 11 illustrates Total Training Steps vs. Trade-off Parameter (ζ) for three models: Meta-DRL, Meta-DRL with Fine-Tuning, and Hybrid DQN-DDPG DRL. The Meta-DRL model requires the most training steps, consistently staying above 3.5×10^5 steps with some variations. Meta-DRL with Fine-Tuning reduces the required steps staying around 2.5×10^5 to 3×10^5 steps. The Hybrid DQN-DDPG DRL model requires the fewest training steps, consistently staying below 2×10^5 steps, showing a significant reduction in computational effort. This trend indicates that Hybrid DQN-DDPG DRL is the most efficient model, achieving optimal performance with fewer training steps compared to the other models. Quantitatively, at $\zeta = 0$, the Meta-DRL model starts at around 4×10^5 steps, while Meta-DRL with Fine-Tuning starts around 2.7×10^5 steps, and Hybrid DQN-DDPG DRL begins at 2×10^5 steps. As ζ increases, the required training steps for all models decrease, with Meta-DRL reaching around 3.2×10^5 steps at $\zeta = 1$, Meta-DRL with Fine-Tuning stabilizing near 2.3×10^5 steps, and Hybrid DQN-DDPG DRL requiring the fewest steps at around 1.5×10^5. The performance gap between Hybrid DQN-DDPG DRL and Meta-DRL is nearly 2×10^5 steps across different trade-off values, confirming that Hybrid DQN-DDPG DRL requires significantly fewer training steps to converge. This reduction in training steps is beneficial as it lowers computational costs and speeds up the learning process. The Meta-DRL with Fine-Tuning model also shows an improvement over standard Meta-DRL, but it still requires more steps than Hybrid DQN-DDPG DRL. These results suggest that Hybrid DQN-DDPG DRL is the best model for efficient training, achieving the desired performance with minimal training steps, making it ideal for large-scale and real-time applications.

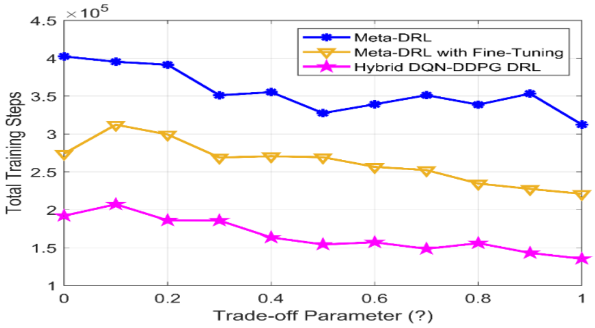

Fig. 11: Total Training Steps to Obtain the Pareto Front.

Figure 12 illustrates Policy Adaptation Efficiency vs. Training Iterations for four models: Random Solution, Meta-DRL, Meta-DRL with Fine-Tuning, and Hybrid DQN-DDPG DRL. As training progresses, all models improve adaptation efficiency but at different rates. The Hybrid DQN-DDPG DRL model achieves the highest efficiency, reaching nearly 75% after 1000 training iterations. The Meta-DRL with Fine-Tuning model follows, stabilizing around 55%. The Meta-DRL model lags at 40%, while the Random Solution struggles to exceed 20%. These trends suggest that Hybrid DQN-DDPG DRL is the fastest-learning and most effective model for policy adaptation. The ability of Hybrid DQN-DDPG DRL to learn quickly makes it suitable for applications requiring fast decision-making in real-time environments. Quantitatively, at 0 training iterations, the Random Solution starts below 5% efficiency while Meta-DRL begins at around 10%, Meta-DRL with Fine-Tuning at 15% and Hybrid DQN-DDPG DRL at approximately 20%. By 200 iterations, the Hybrid DQN-DDPG DRL model rapidly increases to 60% efficiency, Meta-DRL with Fine-Tuning reaches 50%, Meta-DRL stabilizes near 35% and the Random Solution remains below 15%. As training reaches 1000 iterations, the Hybrid DQN-DDPG DRL model achieves 75% efficiency, showing the fastest and most stable improvement. The Meta-DRL with Fine-Tuning model improves but plateaus at 55% while Meta-DRL remains around 40%. The Random Solution shows the slowest adaptation, barely reaching 20% even after extensive training. These results confirm that Hybrid DQN-DDPG DRL learns policies more effectively and adapts to new conditions significantly faster than other models. The steep initial rise in efficiency for Hybrid DQN-DDPG DRL highlights its ability to quickly generalize and optimize performance, making it the best model for reinforcement learning applications. A higher policy adaptation efficiency means better decision-making in dynamic environments secure real-time adaptability in intelligent systems. Faster adaptation translates to better response times in real-world applications like autonomous driving, network optimization, and robotic control systems. Here, quick adjustments to changing conditions are essential.

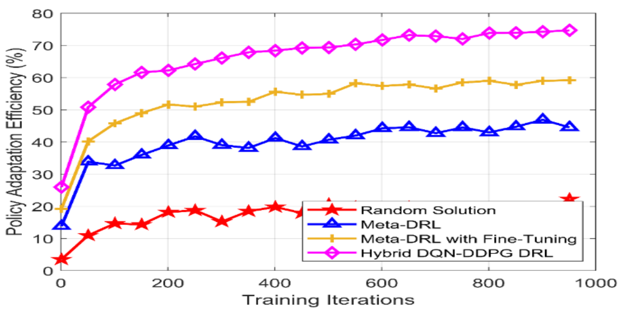

Fig. 12: Policy Adaptation Efficiency Versus Training Iterations.

8. Conclusion

This paper introduced a reinforcement learning-based approach to optimize vehicular networks by integrating Meta-RL. The proposed framework effectively reduced the AoI, improved power efficiency, and enhanced adaptability in dynamic vehicular environments. The findings demonstrated that Meta-RL significantly outperformed traditional reinforcement learning models like DQN and DDPG in terms of both learning efficiency and real-time decision-making. Power consumption is another fundamental challenge in vehicular communication networks. The proposed approach efficiently allocated transmission power according to varying network conditions. it optimizes energy usage without compromising update frequency. It balanced power allocation to maintain system performance. This capability is particularly useful in energy-constrained environments. Here, power resources must be managed carefully. By dynamically adjusting power levels in response to changes in traffic density and network load, Meta-RL achieved substantial energy savings compared to conventional reinforcement learning models. The study also examined the reliability of data transmission by evaluating packet delivery performance. Meta-RL optimized update scheduling and reduced packet loss, thereby improving the overall quality of service in vehicular networks. The ability to maintain high data delivery rates even in congested traffic scenarios highlights the robustness of the proposed approach. Overall, the research findings emphasize that Meta-RL provides a promising solution for enhancing vehicular communication efficiency. The model's ability to quickly adapt to network changes while minimizing AoI and optimizing energy consumption makes it a strong candidate for future smart transportation systems. By integrating machine learning with real-time vehicular communication strategies, this approach offers significant improvements over traditional reinforcement learning models. Future research explores additional enhancements to Meta-RL by incorporating more advanced deep learning techniques to refine decision-making processes. Additionally, real-world testing and large-scale implementation of the proposed system would help validate its effectiveness in practical vehicular network scenarios. The findings of this research contribute to the ongoing advancement of intelligent transportation systems, making vehicular communication networks more reliable, efficient, and adaptive to real-time challenges.

References

[1] C. Zhang, X. Lin, R. Lu, and P.-H. Ho, "Raise: An efficient rsu-aided message authentication scheme in vehicular communication networks," in 2008 IEEE international conference on communications, pp. 1451–1457, IEEE, 2008. https://doi.org/10.1109/ICC.2008.281.

[2] K. Fall, "A delay-tolerant network architecture for challenged internets," in Proceedings of the 2003 conference on Applications, technologies, architectures, and protocols for computer communications, pp. 27–34, 2003. https://doi.org/10.1145/863955.863960.

[3] D. Chefrour, "One-way delay measurement from traditional networks to sdn: A survey," ACM Computing Surveys (CSUR), vol. 54, no. 7, pp. 1–35, 2021. https://doi.org/10.1145/3466167.

[4] Y. Yuan, B. Yang, W. Su, H. Li, C. Wang, Q. Liu, and T. Taleb, "Aoi and energy-driven dynamic cache updates for wireless edge networks," IEEE Internet of Things Journal, 2024. https://doi.org/10.1109/JIOT.2024.3470847.

[5] J. Jiang, W. Tu, H. Kong, W. Zeng, R. Zhang, and M. Konecny, "Large-scale urban multiple-modal transport evacuation model for mass gathering events considering pedestrian and public transit system," IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 12, pp. 23059–23069, 2022. https://doi.org/10.1109/TITS.2022.3198178.

[6] W. U. Khan, F. Jameel, N. Kumar, R. J¨antti, and M. Guizani, "Backscatter-enabled efficient v2x communication with non-orthogonal multiple access," IEEE Transactions on Vehicular Technology, vol. 70, no. 2, pp. 1724–1735, 2021. https://doi.org/10.1109/TVT.2021.3056220.

[7] B. R. Sharan, S. Deshmukh, S. R. B. Pillai, and B. Beferull-Lozano, "Energy efficient aoi minimization in opportunistic noma/oma broadcast wireless networks," IEEE Transactions on Green Communications and Networking, vol. 6, no. 2, pp. 1009–1022, 2021. https://doi.org/10.1109/TGCN.2021.3135351.

[8] S. Wang, S. Zhang, J. Zhang, R. Hu, X. Li, T. Zhang, J. Li, F. Wu, G. Wang, and E. Hovy, "Reinforcement learning enhanced llms: A survey," arXiv preprint arXiv:2412.10400, 2024.

[9] Y. Long, S. Zhao, S. Gong, B. Gu, D. Niyato, and X. Shen, "Aoi-aware sensing scheduling and trajectory optimization for multi-uav-assisted wireless backscatter networks," IEEE Transactions on Vehicular Technology, 2024. https://doi.org/10.1109/TVT.2024.3402740.

[10] A. K. Singh, P. Dziurzanski, H. R. Mendis, and L. S. Indrusiak, "A survey and comparative study of hard and soft real-time dynamic resource allocation strategies for multi-/many-core systems," ACM Computing Surveys (CSUR), vol. 50, no. 2, pp. 1–40, 2017. https://doi.org/10.1145/3057267.

[11] X. Deng, Y. Liang, D. Luo, J. Wang, X. Yan, and J. Duan, "A multi-objective optimization model for rsu deployment in intelligent expressways based on traffic adaptability," IET Intelligent Transport Systems, vol. 18, no. 11, pp. 2204–2223, 2024. https://doi.org/10.1049/itr2.12568.

[12] W. Xu, H. Zhou, N. Cheng, F. Lyu, W. Shi, J. Chen, and X. Shen, "Internet of vehicles in big data era," IEEE/CAA Journal of Automatica Sinica, vol. 5, no. 1, pp. 19–35, 2017. https://doi.org/10.1109/JAS.2017.7510736.

[13] Y. Ni, L. Cai, and Y. Bo, "Vehicular beacon broadcast scheduling based on age of information (aoi)," China Communications, vol. 15, no. 7, pp. 67–76, 2018. https://doi.org/10.1109/CC.2018.8424604.

[14] M. I. Ashraf, C.-F. Liu, M. Bennis, W. Saad, and C. S. Hong, "Dynamic resource allocation for optimized latency and reliability in vehicular networks," IEEE Access, vol. 6, pp. 63843–63858, 2018. https://doi.org/10.1109/ACCESS.2018.2876548.

[15] A. A. Habob, H. Tabassum, and O. Waqar, "Non-orthogonal age-optimal information dissemination in vehicular networks: A meta multi-objective reinforcement learning approach," arXiv preprint arXiv:2402.12260, 2024. https://doi.org/10.1109/TMC.2024.3367166.

[16] S. B. Tadele, B. Kar, F. G. Wakgra, and A. U. Khan, "Optimization of end-to-end aoi in edge-enabled vehicular fog systems: A dueling-dqn approach," IEEE Internet of Things Journal, 2024. https://doi.org/10.1109/JIOT.2024.3472026.

[17] M. Chen, Y. Xiao, Q. Li, and K.-c. Chen, "Minimizing age-of-information for fog computing-supported vehicular networks with deep q-learning," in ICC 2020-2020 IEEE International Conference on Communications (ICC), pp. 1–6, IEEE, 2020. https://doi.org/10.1109/ICC40277.2020.9149054.

[18] Z. Li, L. Xiang, and X. Ge, "Age of information modeling and optimization for fast information dissemination in vehicular social networks," IEEE Transactions on Vehicular Technology, vol. 71, no. 5, pp. 5445–5459, 2022. https://doi.org/10.1109/TVT.2022.3154766.

[19] L. Yang, J. Chen, Q. Ni, J. Shi, and X. Xue, "Noma-enabled cooperative unicast–multicast: Design and outage analysis," IEEE Transactions on Wireless Communications, vol. 16, no. 12, pp. 7870–7889, 2017. https://doi.org/10.1109/TWC.2017.2754261.

[20] A. J. Kadhim and S. A. H. Seno, "Energy-efficient multicast routing protocol based on sdn and fog computing for vehicular networks," Ad Hoc Networks, vol. 84, pp. 68–81, 2019. https://doi.org/10.1016/j.adhoc.2018.09.018.

[21] P. Scopelliti, G. Araniti, G.-M. Muntean, A. Molinaro, and A. Iera, "A hybrid unicast-multicast utility-based network selection algorithm," in 2017 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB), pp. 1–6, IEEE, 2017. https://doi.org/10.1109/BMSB.2017.7986146.

[22] C.-Y. Lin and W. Liao, "Aoi-aware interference mitigation for task-oriented multicasting in multi-cell noma networks," IEEE Transactions on Wireless Communications, 2024. https://doi.org/10.1109/TWC.2024.3381480.

[23] C. Chen, W.-D. Zhong, and D. H. Wu, "Non-hermitian symmetry orthogonal frequency division multiplexing for multiple-input multiple-output visible light communications," IEEE/OSA J. Opt. Commun. Netw., vol. 9, no. 1, pp. 36–44, Jan. 2017. https://doi.org/10.1364/JOCN.9.000036.

[24] W. Khalid, A.-A. A. Boulogeorgos, T. Van Chien, J. Lee, H. Lee and H. Yu, "Optimal Operation of Active RIS-Aided Wireless Powered Communications in IoT Networks," IEEE Internet of Things Journal, vol. 12, no. 1, pp. 390–401, 1 Jan. 2025. https://doi.org/10.1109/JIOT.2024.3461814.

[25] W. Khalid and H. Yu, “Security Improvement with QoS Provisioning Using Service Priority and Power Allocation for NOMA-IoT Networks,” IEEE Access, vol. 9, pp. 9937–9948, Jan. 2021. https://doi.org/10.1109/ACCESS.2021.3049258.

[26] W. Khalid, Z. Kaleem, R. Ullah, T. Van Chien, S. Noh and H. Yu, “Simultaneous Transmitting and Reflecting-Reconfigurable Intelligent Surface in 6G: Design Guidelines and Future Perspectives,” IEEE Network, vol. 37, no. 5, pp. 173–181, Sept. 2023. https://doi.org/10.1109/MNET.129.2200389.

[27] Ummiti Sreenivasulu, Shaik Fairooz, R. Anil Kumar, Sarala Patchala, R. Prakash Kumar, Adireddy Rmaesh, “Joint beamforming with RIS assisted MU-MISO systems using HR-mobilenet and ASO algorithm, Digital Signal Processing, Volume 159, 2025, 104955,ISSN 1051-2004, https://doi.org/10.1016/j.dsp.2024.104955. https://doi.org/10.1016/j.dsp.2024.104955.

[28] Satyam, A. ., Kumar, R. A. ., Patchala , S. ., Pachala, S. ., Geeta Bhimrao Atkar, & Mahalaxm, U. S. B. K. . (2025). Multi-agent learning for UAV networks: a unified approach to trajectory control, frequency allocation and routing. International Journal of Basic and Applied Sciences, 14(2), 189-201. https://doi.org/10.14419/474dfq89.

[29] Kumar , B. P. ., Mahalaxmi , U. S. B. K. ., Nagagopiraju, V. . ., Manda , A. K. ., Chandana , K. ., Betam, D. Suresh . ., Gangadhar , A. ., & Patchala, D. S. . . (2025). Deep Reinforcement Learning for Joint UAV Trajectory and Communication Design in Cache-Enabled Cellular Networks. International Journal of Basic and Applied Sciences, 14(3), 418-430. https://doi.org/10.14419/djn77m90.