Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 5s (2025)

Deep Reinforcement Learning for Traffic Flow Optimization in Urban Planning

Dr. Amar Bharatrao Deshmukh [1], Dr. Senthil Kumar A [2], Dr.Chethana C [3], Dr. Kavita Suresh Kumavat [4], Dr.Mallikharjuna Rao Sathuluri [5], Dr. Sarita Pundhir [6], Dr.R.Senthamil Selvan [7]

[1] Department of E & TC, ABMS'P Anantrao Pawar College of Engineering & Research, Pune, Maharashtra, India. amarbdeshmukh@gmail.com

[2] **Department of** Computer Science and engineering, School of engineering, Dayananda Sagar University, Bangalore, India. senthil.kumar-cse@dsu.edu.in

[3] Department of Information Science and Engineering, BMS Institute of Technology &Management, Bengalore, India. chethana_cse@bmsit.in

[4] Department of Computer Engineering, Vishwakarma University, Pune, India. kavita.kumavat@vupune.ac.in

[5] Department of Electronics and Communication Engineering, Andhra Loyola Institute of Engineering and Technology, Vijayawada, Andhra Pradesh, India. smr.aliet@gmial.com

[6] Department of Mathematics (ASH), Delhi Technical campus Greater Noida, India. pundhirsarita01@gmail.com

[7] Department of ECE, Annamacharya Institute of Technology and Sciences,Tirupati,India. Selvasenthamil2614@gmail.com

Article History:

Received: *12-10-2024*

Revised: *28-11-2024*

Accepted: *09-12-2024*

Abstract:
Traffic jam becomes a major problem on a worldwide scale these days. The number of vehicles has risen considerably, but the capacity of roadways and transportation infrastructure to manage the additional congestion has not yet increased proportionately. Public health and society are suffering as a result of the rise in road congestion and pollution caused by traffic. In order to minimize traffic congestion, this research develops a revolutionary reinforcement learning (RL)-based approach. Researcher have created a novel deep Q network (DQN) and seamlessly incorporated it into the system. The applied DQL model cut the length of traffic waits by 50% and made each lane more appealing by 10%. The outcomes show that the approach is successful at establishing strict standards for reducing traffic. The proposed system shows that that RL can efficiently control traffic traffic snarl-up in cities quickly without compromise in the environmentally friendly nature. Added on utilizing RL may considerably ease traffic snarl-up in cities.

Keywords: Reinforcement Learning, Deep Q Network, Vehicles, Traffic Jam, Cities, Environmental

1. Introduction:

Each year, cities are having trouble with traffic jams. This problem has a big impact on the transportation industry's travel times, fuel use, and operating costs [1,2]. Pollution from congestion is a major issue that costs the planet a lot. A lot of study has been done to develop air traffic management systems that work well in order to solve this pressing problem [3]. In recent years, most of the work has been focused on smart transportation systems (ITS). The objectives of these projects are to make traffic control systems work more effectively be safer, and be better for the environment [4–7]. Researchers are working in the field of ITS to develop novel approaches that will enhance urban sustainability and decrease traffic [8,9]. Managing critical elements that include reducing line length and delay is necessary to optimize traffic flow at a junction [10]. The structure and layout of intersecting roads have a direct impact on traffic congestion, and improving traffic management may help reduce it [11]. A better and improved traffic experience can only be achieved by putting strategic measures into place to maximize the traffic flow on certain routes [12,13]. They undoubtedly slow down traffic by minimizing traffic snarl-up on multiple routes at an intersection [15-17]. Researcher have made traffic flow smoother and balance by controlling and lowering congestion on certain roads in a planned way. This way makes the crossing work better and helps make the transit network easier and less busy [18].

The critical requirement to address traffic congestion, a major problem caused by the exponential growth in vehicle numbers without matching improvements to transportation infrastructure, is the reason that stimulated this study. Vehicle congestion and pollution have increased as a result of this imbalance, which has many negative effects on public health and society at large [19]. Researcher want to make use of cutting-edge technology, especially RL, to come up with novel approaches to alleviate tailback traffic in order to solve this major problem [20]. A possible method for successfully optimizing traffic movement and reducing congestion is presented by RL, especially DQN, which uses artificial intelligence. In order to show whether RL may revolutionize urban transportation efficacy and sustainability, the researchers are creating a complex DQL model inside a transportation structure [21-23]. The immediate requirement to implement smart, data-driven strategies to address traffic crowding and advance safer, effective & ecologically sustainable urban transportation serves as the driving force behind the research [24,25]. Modern RL techniques will be used in this project to develop new traffic reduction standards and improve urban transportation networks. The ultimate goal of this research is to help create smarter, more flexible urban transportation systems that can manage the problems imposed on through rising car counts and constrained infrastructural capacity.

This research details the process of creating a sophisticated system to reduce traffic, one that makes use of smart technology to cut down on wait times. Researcher suggested an RL structure for the system, which is a kind of ML, that makes traffic move better and less backed up. To reduce wait times and maximize productivity, the smart traffic management system integrates complex algorithms with real-time data. This innovative method may solve the present mobility problems in cities and is an indication of great improvements in traffic management systems. The following is the outcome of this study:

- Enhancing traffic reduction with the use of an intelligent personalized layer-based technique.
- Concentrating on cutting down on tailback line duration and offering rewards at all levels.
- Reducing traffic with the use of the state-action-rewards RL approach.
- To improve a traffic-free urban environment, it is important to concentrate in particular benchmark approaches for lowering traffic snarl-up.
- Creating an improved DQN to manage the congestion minimizing system in a convergence.

The research paper is segregated in several sections. Section 2 illustrates about the methodology. The result analysis findings are presented in section 3. Finally, section 4 includes the conclusion with future scope.

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 5s (2025)

2. Methodology:

The DQN is utilized to deal traffic line-up. Researcher shown the way to use the DQL model to aid in traffic reduction in this section.

2.1. Analysis of Datasets:

Two JSON datasets with time series data were employed in the proposed method. Environmental data, like automobile number, path, exit lane, and departure speed, are included in the first dataset. In order to train the agent, these data sets are essential for constructing an intersecting environment with 206 edges that indicate distinct environment locations. The 2nd dataset, which concentrations on route data, contains characteristics like function, index, edges identification, length, shape and lane Identification. When figuring out the routes within a junction, several factors are essential. Researcher could correctly train the agent if the merging of these 2 datasets is successful. By combining route and environmental data, the model improves analysis of system characteristics.

The model created two new datasets called "plot_vehicle_queue_data" & "plot_reward_point_data" after the training. The data sets are important to the next step of testing. Researcher looked at the trained data during tests to judge the length of the wait and the reward points. The thorough testing process lets to observe whether and how often the model worked in real-life situations, which gave us important information about how lines work and the reward points the agent receives.

Using certain criteria, 31 episodes are assessed throughout the testing period. The number of automobiles built for testing stayed at 1010, but the maximum number of steps allowed was set at 248. A four-layer neural network design with a rate of learning of 0.002 was part of the testing setup. Four actions like EWA (East-West Arm), EWLA (East-West Left Arm), NSLA (north-south Left Arm), and NSA (North-South Arm) are tested. The activities were specifically chosen to shorten wait times and increase system reward points. This method gave researchers a thorough grasp of the model's potential to improve award results and queue dynamics by allowing them to assess its performance in a variety of scenarios.

2.2. RL for Mitigating Traffic:

By adjusting to actual highway conditions, reinforcement learning may enhance signal timing and route selection in traffic management. This strategy alleviates congestion and enhances transportation effectiveness by increasing driving & minimizing traffic. The system can adapt & enhance traffic jam in cities by simultaneously analysing road traffic patterns. RL may be used to design intelligent traffic control mechanisms that enhance sustainability & efficiency in urban transportation. The RL process is illustrated in Figure 1.

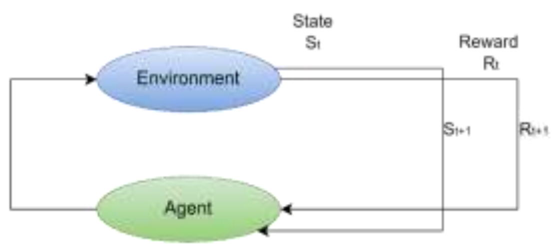

Figure 1 Schematic View of RL Process.

In this case, S_t stands for the environment's state phase, A_t for the action performed in the time level t, and R_t for the potential action's reward.

An agent learns by means of interaction & prediction in RL. Less effective measures lead to greater penalty rates, even while accurate projections show effectiveness. To improve decision-making, the agent gets rewarded to highlight favourable outcomes and downplay bad ones. Because RL is flexible, intelligent agents may navigate their surroundings in the best possible way.

2.3. Environment-based Agent Training:

For training DQL agent successfully train, there are a number of steps that must be followed. The past encounter's occurrences and learnings are get stored in a replay memory. The agent has to deal with a lot of different things in a scene that is always changing to reflect how complicated the real world is. Setting up the right settings is very important because they control how the robot acts and how fast it learns. The DQN learns a lot from its start because it helps the network learn more about what behaviors are finest as it goes along. The agent improves its decision-making abilities by using a well-designed loss function. By adjusting its weights and aiming for near-ideal Q-values in relation to this function, the agent gradually enhances its performance. As shown in Figure 2, the DQL model is used for agent training.

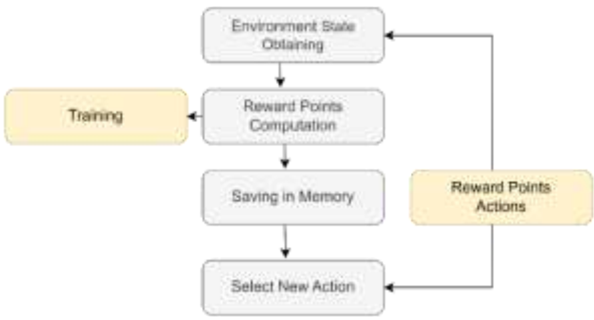

Figure 2 Training Process of Agent.

Applying previous information, the agent makes trained and sound choices that might result in the best outcomes as it navigates the difficulties of the natural environment. By regularly learning and applying new information, the DQL agent becomes better at handling issues that arise in the real world.

This research examines the ways for transport and city traffic administration interact to enhance traffic flow, decrease traffic jam & promote justifiable movement in urban settings. Traffic flow and operational aspects of a complicated four-direction roadway system are examined in the study. Through rigorous research, the study seeks to comprehend the road network's intricacies and foundational dynamics. The study examines the transportation system's complex linkages to find critical bottlenecks, inefficiencies, and improvement possibilities. The research proposes innovative methods for boosting traffic flow, queueing, and urban mobility. The initiative offers practical recommendations to politicians and city planners via data evaluation, modeling, and simulations. This will make cities more vibrant, sustainable, and habitable.

2.4. RL Status from Intersection:

Various node locations in the environment are represented by a state, which is represented by a large number of nodes. Here, the state includes information about node identities and item locations. The expression is as follows:

$$E = \{C_1, V_1, D_1, C_2, V_2, D_2, C_3, V_3, D_3, \dots . C_i, V_i, D_i\} \qquad (1)$$

For each vehicle i, d shows how far it is from the next vehicle in the chain, and n is the count of vehicles, v indicates the velocity of the i vehicle.

Consider lanes in which the colors red and green stand for 1 and 0, respectively. The matrix is represented as,

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} \qquad (2)$$

There is a live green light in the 1st state, which lets cars move because each of the nodes have been assigned to 1.

$$\begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \qquad (3)$$

Also, in the last state, all links are red, which means that vehicles can't move.

$$\begin{pmatrix} 0 & 0 & 0 \end{pmatrix} \qquad (4)$$

Every state has a distinct value that establishes the vehicle's permitted movements based on the given value. Figure 3 uses nodes and value representations to efficiently communicate the state. It includes 1 if a state node exists and 0 if it is null.

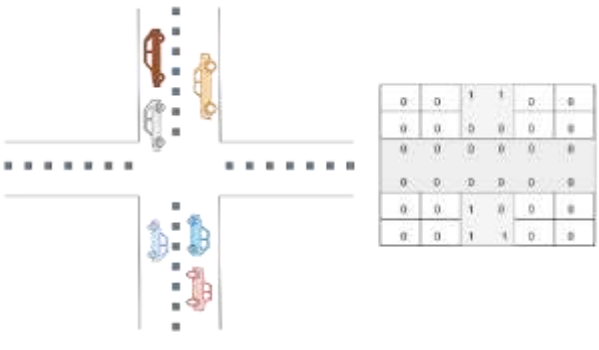

Figure 3 Indications of States of Vehicles.

2.5. RL's Actions from Intersection:

An "action" in reinforcement learning for traffic signal management refers to an option the RL agent makes about controlling the traffic lights at a particular junction. By monitoring its surroundings and striving to maximize specific targets, like minimizing queues, decreasing out delays, or improving traffic flow, the RL agent gains decision-making skills. The possible behaviors of this system include:

$$A = EWA, NSA, EWLA, NSLA$$

All RL agent actions are linked to traffic signal durations. Based on its actions, the RL agent receives reward points or penalty points.

2.6. Reward Points of RL from Intersection:

The benefit of a congestion mitigation system at a junction is the improvement of overall automobile and security. Reduced traffic, lesser travel delays, and enhanced vehicle throughput are the results of vehicles interacting smoothly and coordinating optimally. All users of the roadways benefit from a better and more sustainable urban environment due to the intersection-based incentives system, which establishes an organized and well-regulated traffic network.

The intersection's expression of reward points' function is:

$$RP[s, a, \acute{s}] \tag{5}$$

Here the existing environment is appropriately represented by s. The state provides the context for the agent's activity. The prize may be situational. The agent's current activities are indicated by a. Reward depends on agent activity. s′ the state arising from an agent's effect, represented as a. The outcome may affect incentives. Present environment, agent activity, and resultant environment s′ are used to determine rewards. All training and testing outcomes are expressed by these three variables. Valuation of reward points could be negative. The negative results of the actions are presently confronting. It describes the outcomes used to guide the RL agent's learning and behavior. By providing negative incentives, an agent learns to associate certain actions with undesirable effects, which in turn causes it to adjust its policy in order to decrease the occurrence of unwanted behaviors.

2.7. Establishment of Deep Q Networks:

The functional outline for DQL is shown in Figure 5. Through regular interactions with the situation and a persistent trial-and-error method, the agent determines the most successful course of action throughout DQN implementation. According to n steps ahead, the optimal course of action is established by taking into account both the prospective and immediate benefits in the subsequent indices. The overall reward points are indicated under a certain policy in the DQL algorithm.

$$v^{\pi}(s_I) = R_i + \gamma R_{i+1} + \gamma^2 R_{i+1} + l \tag{6}$$

The DQL algorithm determines the Q value for evaluating a certain condition & activity. Both current and discounted reward points are taken into account to arrive at this amount. Here is the formula for the Q value:

$$o_I + \gamma max q_{A_{i+1}}, \epsilon a(s_{I+1}, A_{I+1}) \longrightarrow q(S_i, A_i) \tag{7}$$

The influence that future reward points in the current action is represented by the discount coefficient, which is represented by γ ($0 < \gamma < 1$). Optimizing the total utility is the purpose of Q learning. Therefore, it uses the given values o_I & max$q_{A_{i+1}}$, ∈A (s_{I+1}, A_{I+1}) in place of the variables R_i & $V^{\pi}(s_{I+1})$ in equation (7).

Actions sets make up A. The initial step of the major obstacles in DQN at the learning phase involves managing exploring new actions with using existing knowledge. In large systems, choosing the right action may affect algorithm convergence and performance. To determine the most effective action, researcher have included an adjusted index value. The index value properly describes reward changes and swiftly adjusts exploration range to reduce selection costs.

The action index is a technique for selecting actions; it ranks activities from most to least important based on the current activity's system usefulness. The system's inclination to use its resources is shown by this strategy. On the other hand, the exploratory character of the process is on display in the uninterrupted iterative process, which tends to prefer picking the same action in the succeeding iteration if the selected amount is inadequate.

The procedure is carried out by the relay by computing the value of utility o & modifying the q value utilizing the following after the execution of action ai has been identified:

$$q_{I+1}(s_I a_I) = (1-\alpha)q_I(s_I a_I) + \alpha(o_I + \gamma max q_I(s_{I+1}), a_{I+1}) \qquad (8)$$

Here $s = s_I$ and $a = a_I$

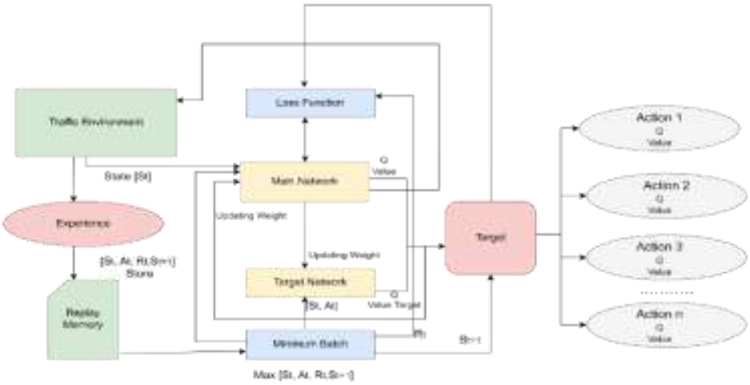

Figure 5 Schematic View of DQL Model.

2.8. Setting Hyperparameters and the DQL Model's Distinctive Characteristics:

The DQL model's hyperparameter modifying procedure is now complete. To fine-tune a machine-learning model's hyperparameters is to choose the best possible values for them. Before training starts, hyperparameters are configured as configuration settings to govern the procedure of learning and the model's structure. They have the potential to greatly impact the model's performance but are not learnt from the data like the model parameters. Finding the best settings for these parameters is called hyperparameter tuning, and it's crucial for machine learning models to run successfully. In order to optimize the accuracy of the model on both private and public data, hyperparameters are fine-tuned before training starts. Tuning is necessary to ensure that the model correctly corresponds the training data and performs well when dealing with fresh, unknown input. It might result in improved generalization, faster convergence, and higher accuracy. To avoid overfitting and underfitting, models may be adjusted. A model must have its hyperparameters adjusted to get the ideal practical result.

Numerous aspects have been examined, including batch number, epoch, stimulation, learning rates, dropout, padding, optimizer, weight delay, and kernel initializer. By specifying the hyperparameters & their potential values, first define the search space before doing hyperparameter tweaking. Select an algorithm or strategy next. Create sets for validation and training from the dataset, then use various hyperparameter combinations to train the models. Lastly, choose the hyperparameters which provide the greatest validation outcome after assessing every model's efficiency utilizing a desired measure. The field of search for the best traffic reduction outcomes is represented by a certain range of parameter values, which is specified here. Next, a value is chosen to implement the system from this range. Table 1 shows the result for hyperparameter tuning process.

Table 1 Suggested Model Hyparameter Setting.

Metrics	Chosen Values	Spaces Searched
Maximum of Steps	241	(221,231,241)
Overall Episodes	31	(16,21,31)
Layers Number	5	(5)
States Number	81	(81)
Action Number	5	(5)
Gamma	0.76	(0.81,0.77,0.76)
Size of Batch	33	(17,33,9)

Hyperparameter adjustment improves model performance by performing rigorous trials to improve prediction and robustness reliability. In AI and ML, DQL model innovates RL. The use of deep neural networks transforms how agents learn to make difficult judgments. Q-learning, a conventional RL method, integrated together with DNN for handling high-dimensional state spaces is its main novelty. Through the application of artificial neural networks to approximate the Q-function, DQL may identify optimal strategies inside extensive state spaces. This discovery has shown the applicability of reinforcement learning in autonomous driving, automated management, and complex tasks. DQL facilitates the development of sophisticated AI systems capable of immediately obtaining data derived from primary sensory data & navigating complex decision-making environments. The optimization of traffic signal timing by DQL effectively alleviates congestion in a novel manner. DQL's ability to change and adapt to new events is implemented in this method. The DQL model can correctly predict & change to road traffic trends because of real-time traffic data that includes occupancy and moving sensors on vehicles. The model changes the time of traffic lights based on the current traffic situation to make traffic move smoothly and minimize congestion. This new DQL method makes traffic flow more smoothly, cuts down on trip times, and lowers pollution, which solves problems with getting around cities.

3. Results:

3.1. Development and Assessment Criteria:

Researchers first methodically map the junction's condition and calculate the reward points. Then, they assign states and activities to each lane in sequential order. Figure 6 is a picture that shows each stage of this junction along with the route each lane takes to get there. For each state, this model helps experts figure out the exact action that should be taken. This method is based on identifying the current condition, evaluating the junction's next step, and analyzing whether the reward points are becoming better.

Communications on Applied Nonlinear Analysis
ISSN: 1074-133X
Vol 32 No. 5s (2025)

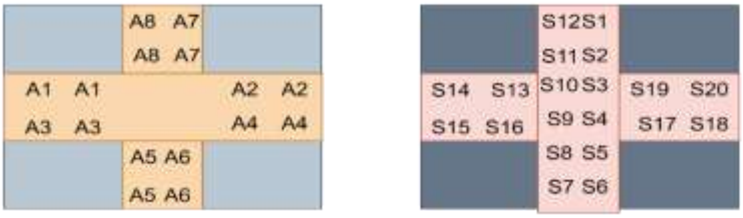

Figure 6 Various Lane Operations and Junction State Representations.

This study indicates that detecting a vehicle's present state may anticipate its future state & establish the effective phase of agent, consequently increasing performance. Table 2 displays current states and recommendations for future states. This table shows possibilities for future vehicle states, rewards, and estimated gains depending on present status. The traffic reduction system relies on this focus table for valuable information toward enhancing junction functioning.

Table 2 Estimate the Current Condition and the Next Optimal State with High Rewards.

Present State	**Succeeding State**	**Actions**	**Reward Points**
S2	(S3, S4, S5, S6, S7)	A3	2+
S2	(S3, S4, S11, S17, S16)	A6	2+
S3	(S4, S5, S6, S7, S8)	A9	2+
S3	(S4, S20, S18)	A3	2+
S8	(S5, S6, S9, S10, S11)	A6	2+
S8	(S16, S17, S19)	A2	2+
S9	(S6, S8, S10, S11, S13)	A1	2+
S9	(S7, S4, S3, S2, S18)	A5	2+
S7	(S5, S6, S4, S3, S2)	A2	2+
S7	(S16, S17, S20)	A3	2+
S4	(S5, S6, S7, S8, S9)	A6	2+
S4	(S13, S12, S11, S10, S19)	A9	2+
S13	(S1, S2, S4, S5, S6)	A3	2+
S13	(S11, S16, S17, S19, S20)	A6	2+
S12	(S5, S6, S7, S8, S9)	A2	2+
S12	(S15, S17, S19, S20, S18)	A1	2+
S15	(S8, S9, S10, S11)	A5	2+
S15	(S1, S12, S13)	A2	2+
S10	(S7, S8, S9, S11, S12)	A3	2+
S10	(S13, S14, S15)	A7	2+
S11	(S17, S18, S19)	A4	2+
S11	(S20, S3, S4, S5, S7)	A6	2+
S18	(S17, S16, S17, S14, S13)	A2	2+
S18	(S1, S2, S3, S4, S5)	A1	2+

Both throughout training and testing, the results of the model have shown significant amounts of potential. The queue time decreased to 853 during the training set, but the corresponding incentives, which were negative at -94493, show improved comprehension & adaptation to the RL environment. During the testing period, the queue length dropped to 419, and an 8521 positive testing reward points was earned. When compared to the original condition, these data together show a considerable reduction in wait time and an enhancement in rewards. This shows positive improvements in both training and testing scenarios, demonstrating the model's effectiveness in learning and enhancing its performance.

$$Queue_d = \frac{Queue\ Length\ Testing}{Queue\ Length\ Training} * 100 \tag{9}$$

$$Reward_m = \frac{Reward\ Testing}{Reward\ Training} * 100 \tag{10}$$

Reduced queue length is represented by $queue_d$, and maximum outcomes are represented by $reward_m$, in this scenario.

Queue duration dropped 50% and reward points increased 10%. This methodology effectively reduced wait length by over 50%, resulting in considerable improvements. Effective execution of the formula for limiting traffic congestion is supported by this outcome. The model's successful decrease in line time and rise in reward points confirms its usefulness in reducing traffic congestion.

3.2. Results of RL Training:

During the training iteration for the RL agent, researcher was able to accomplish 5401 steps. Notably, at the conclusion of this training session, the line length had increased to 853. Note that the training data came from a variety of sources, including the environment, the edge, and the vehicle. The 5401 steps revealed a wealth of information on the environment's periphery. The variations in the line length across the phases are clearly shown in Figure 7, which shows the training development.

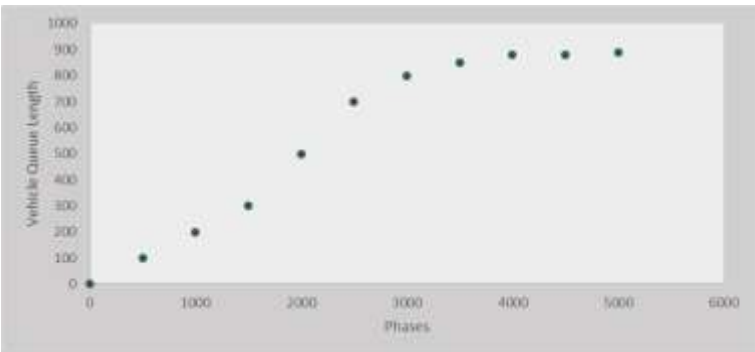

Figure 7 Agents' Queue Length in Training.

The model used queue length performance to quantify training success and award rewards accordingly. Negative reward value (-944993) suggests a penalizing aspect in RL settings. Negative cultivate incentives work as a feedback mechanism to penalize undesired actions. Negative cultivate incentives, unlike conventional ones, aim to improve behavior by punishing acts that result in less optimum results. This method improves traffic flow by punishing acts that create delays or congestion. In the traffic management system, negative incentives discourage sudden lane variations & extreme speeding, promoting smooth and effective traffic management. Putting fines allows the system to improve traffic flow & reduce congestion. Consider Figure 8 for the relationship between rewards & epochs in the DQL model's training rewards. It reveals changes in training efficacy over time.

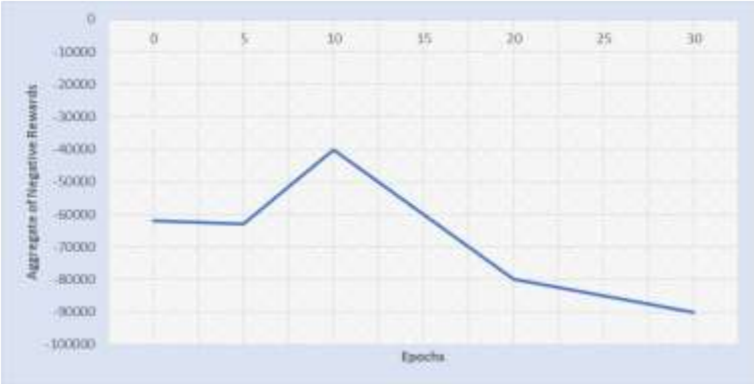

Figure 8 Agents' Produces Reward Points in Training.

3.3. Results of RL Testing:

The length of the testing line has reached 419 after 31 epochs of evaluating on the DQL model. This figure shows that the length of the line has been shortened compared to the training line. The shorter wait time in line shows that traffic jams are getting better, which proves that this approach works. 419 items are in the queue at the end of the testing phase. It's easy to see the improved queue management has been done by looking at Figure 9, which shows the duration of the testing line over several epochs. From training to testing, this method cut the length of the lines by almost half. In the training phase, the last number in the index line was 853. In the testing stage, it went down to 419. The fact that the length of the lines dropped so much during tests shows that this method works to ease traffic. In this way, the testing results show that the method can shorten wait times and ease traffic.

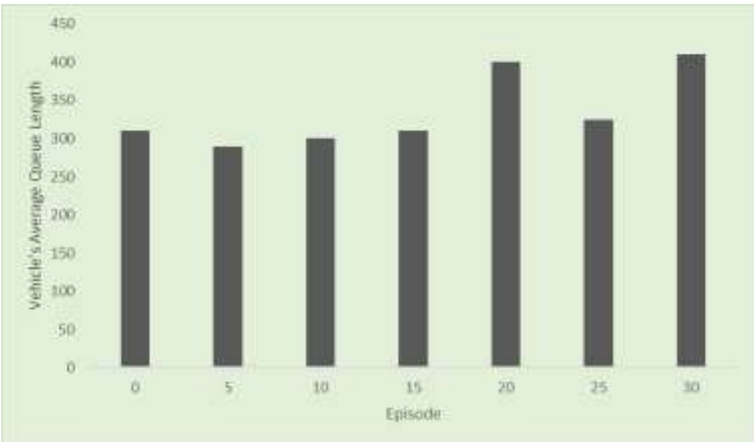

Figure 9 Evaluation of Queue Length Representation across Episodes.

Researcher examined 536 steps during rewards calculating testing. The training reward depends on queue length, which is measured here. The training penalty is high for negative reward points, but lower negative values increase reward points. Incentives totalling − 8521.0 are included in the final phases. Visually, Figure 10 shows the relationship between testing data incentives and action steps. The reward index increased dramatically − 944993 during training to − 8521 during testing. This large disparity shows performance improvement. Testing rewards are maximized, indicating the agent can make better judgments. Evaluation in the testing environment shows considerable improvement in the agent's decision-making method.

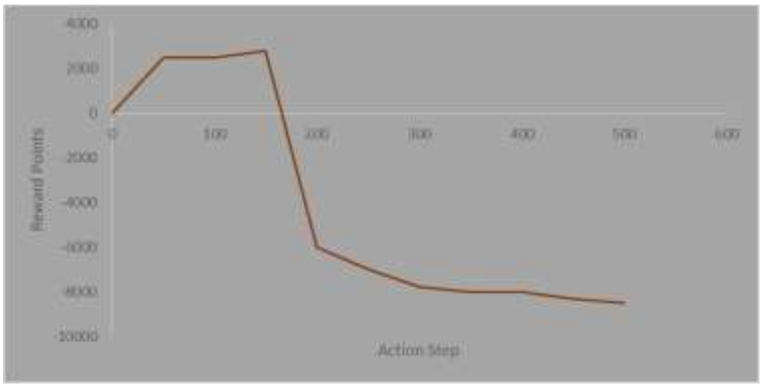

Figure 10 Assessing testing incentives using model action steps.

3.4. Research's Performance and Results:

This paper proposes a novel urban traffic reduction strategy. Researchers created a reward-based system using RL approaches, comprising an enhanced DQL algorithm. The idea is simple: As wait length lowers, agents reward traffic optimization. The findings demonstrate this approach's effectiveness. By reducing queue length by 50% and waiting times by almost 50%, so significant progress is made in traffic management. This significant decrease in wait duration throughout testing demonstrates improved system efficiency. The suggested technology reduces traffic congestion by lowering wait duration, demonstrating its capacity to perform jobs efficiently. Significant ramifications result from this advancement. Reducing queue length indicates quicker and more efficient work management and processing. Implementing strategies to alleviate traffic congestion might have positive effects on productivity, operational costs, and user satisfaction in real-life situations. As this result shows, the plan executed and the system was able to reduce traffic congestion. Due to the ongoing optimization and improvement brought about by the proposed method, the system's efficiency and performance will continue to improve. A 10% incentive boost across the board is also indicative of consistent development and improvement in the system. The objective is to raise collective incentives, that might be used to enhance traffic control and decision-making. This is achieved by teaching an RL agent to operate within a traffic simulation situation, in which it may be tasked with controlling traffic flow, managing road use, and altering traffic signals. The agent acquires feedback in the form of awards depending on how successfully its actions reduce traffic, eliminate delays & improve overall traffic efficiency. Through examination and manipulation, the agent ultimately learns to take choices that stabilizes the immediate rewards with long-term benefits, improving traffic flow. Though trial & error, the RL agent could recognize patterns and plans that help traffic move more smoothly & reduce traffic jams. This method makes sure that decisions are always the most effective they can be in different situations by helping to handle traffic in real time and adapting to fluctuating conditions & traffic trends. The traffic reduction method, shown in Figure 11, uses a set of rules to find the next best state.

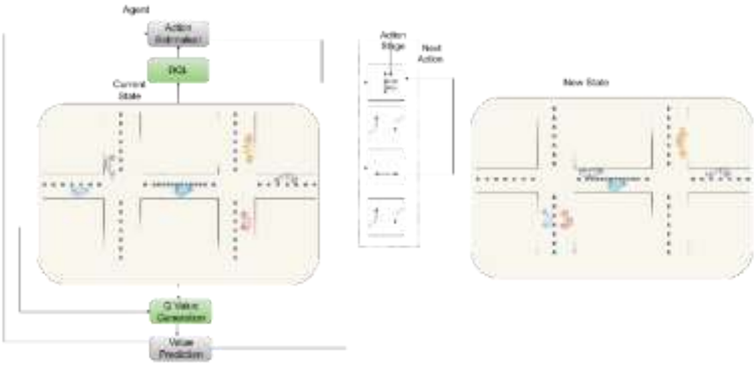

Figure 11 Traffic Congestion System.

Measures to shorten wait times may impact traffic reduction by increasing overall flow and lowering congestion. Employing strategies like improved traffic management, ITS implementation, and traffic signal timing may result in shorter wait times. These strategies reduce commuter travel times and increase productivity by reducing wait times at intersections and other areas of congestion. In urban areas, such programs have a considerable impact on traffic flow, pollution from parked automobiles, and road safety. On the other hand, these strategies often only work if the plan is done ahead, communicate with the proper local authorities, and stay alert since traffic conditions may change at any moment.

4. Conclusion:

The importance for real-time adaptive management for enhancing the efficiency of transportation systems makes reinforcement learning (RL) a valuable tool. In contrast to traffic management methods that rely on predetermined frameworks of these processes, the ability to evolve through dynamic interaction with the environment is considered as a major benefit. This article introduces a revolutionary RL method that effectively reduces traffic congestion. It makes use of the DQL algorithm. The design of the system is depend on an intersection-centered traffic model which illustrates how it can improve reward systems and cut down on wait times. The study results give us an effective strategy to keep traffic as low as possible, which is a big step forward in managing traffic. The current method controls road crossings well and makes the right decisions so that traffic jams are kept to a minimum. This new technology possesses a lot of promise and is an important part of managing traffic.

References:

[1] Mushtaq, Anum, et al. "Traffic flow management of autonomous vehicles using deep reinforcement learning and smart rerouting." *IEEE Access* 9 (2021): 51005-51019.

[2] Ye, Qiming, et al. "Spatial-temporal flows-adaptive street layout control using reinforcement learning." *Sustainability* 14.1 (2021): 107.

[3] Coşkun, Mustafa, Abdelkader Baggag, and Sanjay Chawla. "Deep reinforcement learning for traffic light optimization." *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2018.

[4] Merbah, Amal, and Jalel Ben-Othman. "Optimizing Traffic Flow With Reinforcement Learning: A Study on Traffic Light Management." *IEEE Transactions on Intelligent Transportation Systems* (2024).

[5] Dzmanashvili, Demetre. Optimizing urban traffic flow: reinforcement learning-based traffic light control. MS thesis. Universitat Politècnica de Catalunya, 2023.

[6] Zheng, Yu, et al. "Spatial planning of urban communities via deep reinforcement learning." *Nature Computational Science* 3.9 (2023): 748-762.
[7] BramahHazela, J. Hymavathi, T. Rajasanthosh Kumar, S. Kavitha, D. Deepa, Sachin Lalar, and Prabakaran Karunakaran. "Machine Learning: Supervised Algorithms to Determine the Defect in High-Precision Foundry Operation." *Journal of Nanomaterials* 2022, no. 1 (2022): 1732441.
[8] Schultz, Laura, and Vadim Sokolov. "Deep reinforcement learning for dynamic urban transportation problems." *arXiv preprint arXiv:1806.05310* (2018).
[9] SANJANA, KB SAI, K. SAI MOUNIKA, SANTHOSH KUMAR T RAJA, and DV SRIKANTH. "THERMAL ANALYSIS OF ADVANCED IC ENGINE CYLINDER."
[10] Swapno, S. M., et al. "A reinforcement learning approach for reducing traffic congestion using deep Q learning." *Scientific Reports* 14.1 (2024): 1-20.
[11] Casas, Noé. "Deep reinforcement learning for urban traffic light control." (2017).
[12] Wu, Tong, et al. "Multi-agent deep reinforcement learning for urban traffic light control in vehicular networks." *IEEE Transactions on Vehicular Technology* 69.8 (2020): 8243-8256.
[13] Garg, Deepeka, Maria Chli, and George Vogiatzis. "A deep reinforcement learning agent for traffic intersection control optimization." *2019 ieee intelligent transportation systems conference (itsc)*. IEEE, 2019.
[14] Xie, Peng, et al. "Urban flow prediction from spatiotemporal data using machine learning: A survey." *Information Fusion* 59 (2020): 1-12.
[15] Yang, Jiachen, Jipeng Zhang, and Huihui Wang. "Urban traffic control in software defined internet of things via a multi-agent deep reinforcement learning approach." *IEEE Transactions on Intelligent Transportation Systems* 22.6 (2020): 3742-3754.
[16] Zeynivand, Ahmad, et al. "Traffic flow control using multi-agent reinforcement learning." *Journal of Network and Computer Applications* 207 (2022): 103497.
[17] Khan, AB Feroz, and P. Ivan. "Integrating Machine Learning and Deep Learning in Smart Cities for Enhanced Traffic Congestion Management: An Empirical Review." *J. Urban Dev. Manag* 2.4 (2023): 211-221.
[18] Zhao, Xiangxue, et al. "Deep reinforcement learning for the co-optimization of vehicular flow direction design and signal control policy for a road network." *IEEE Access* 11 (2023): 7247-7261.
[19] Patil, Priyadarshan. "Machine Learning for Traffic Management in Large-Scale Urban Networks: A Review." *Sage Science Review of Applied Machine Learning* 2.2 (2019): 24-36.
[20] Yan, Longhao, et al. "Refined path planning for emergency rescue vehicles on congested urban arterial roads via reinforcement learning approach." *Journal of Advanced Transportation* 2021.1 (2021): 8772688.
[21] Lin, Tuxiang, and Rongliang Lin. "Smart City Traffic Flow and Signal Optimization Using STGCN-LSTM and PPO Algorithms." *IEEE Access* (2024).
[22] Mohebifard, Rasool, and Ali Hajbabaie. "Deep Reinforcement Learning Technique for Traffic Metering in Connected Urban Street Networks." *Transportation Research Record* (2024): 03611981241253581.
[23] Karimzadeh, Mostafa, et al. "RL-CNN: Reinforcement learning-designed convolutional neural network for urban traffic flow estimation." *2021 International Wireless Communications and Mobile Computing (IWCMC)*. IEEE, 2021.
[24] Noaeen, Mohammad, et al. "Reinforcement learning in urban network traffic signal control: A systematic literature review." *Expert Systems with Applications* 199 (2022): 116830.
[25] Li, Yuanyuan, et al. "A hierarchical deep reinforcement learning method for solving urban route planning problems under large-scale customers and real-time traffic conditions." *International Journal of Geographical Information Science* (2024): 1-24.