A Review of Traffic Light Control System with Reinforcement Learning

Rahul Suryawanshi [1], Om Rachalwar [2], Abhit Hedaoo [3], Ankit Thakur [4], Tanmay Rane [5]

[1.]Assistant Professor, Department of Artificial Intelligence at
G H Raisoni Institute of Engineering & Technology, Nagpur

[2,3,4,5] Students at G H Raisoni Institute of Engineering and Technology, Nagpur

Abstract: There are many different traffic congestion issues as a result of the vehicle revolution. Largescale traffic makes it impossible for people to reach their destination on time. The method used to coordinate traffic is independent of how an intersection is currently functioning. The traffic created at the junctions of multiple streets is frequently monitored and controlled by traffic light control systems with pre-set timings. However, considering the different elements involved, synchronization of several traffic signal systems at nearby junctions is a challenge. Increasing the number of roads is one option, while another is installing an intelligent adaptive traffic control system. This study describes a Traffic Light Control System that automatically adjusts signal strength based on traffic density by employing Artificial Intelligence to manage traffic. In this system, the Reinforcement Learning algorithm was used to determine optimal traffic light configuration and using Deep Neural Networks the obtained results were used to extract the features required to make a decision. This project presents a Reinforcement Learning approach to traffic lights control, coupled with an agent-based simulator (Simulation of Urban MObility - SUMO) providing a baseline simulation for the model.

Keywords: Traffic Light Control System, Artificial Intelligence, Reinforcement Learning, Deep Neural Network, SUMO, Simulation.

I. INTRODUCTION

Traffic management is a major problem with significant economic and environmental repercussions. Urbanization and motorization have caused an imbalance between demand and supply of transportation and traffic infrastructure, leading to problems such as travel delays, increase of road accidents, environmental degradation and so on. A road intersection is a shared physical space, access to this common resource must be granted intelligently to optimize the traffic throughput while ensuring safe passage of vehicles. Ever since their development at the end of the 19th century, traffic lights have been effectively used as the prime mode to grant vehicles access to the intersections, however their benefits tail off when they fail to adapt to changes in traffic flows [2]. For efficient utilization of already existing traffic-based resources, it is critically important to carry out optimization in an automated and adaptive manner, embodying characteristics such as self-configuring, self-optimizing, self-protecting and self-healing. A subfield of Machine Learning called Reinforcement Learning teaches an agent how to
interact with the environment and understand the effects of both good and negative behaviors to maximize reward.
Trial-and-error is used to gain this understanding. It uses a feedback mechanism that encourages
agents to take desirable activities and penalises them when they don't. This feedback mechanism enables agents to learn from their mistakes and stop taking unwanted acts.
The traffic micro simulator used for this paper research is Simulation of Urban MObility (SUMO).
SUMO provides a software package which includes an infrastructure editor, a simulator interface, and an application programming interface (API) [2].
These elements enable the user to design and implement custom configurations and functionalities of a road infrastructure and exchange data during the traffic simulation. SUMO - Simulation of Urban Mobility is a powerful simulator designed to handle a large load network and specified traffic demand, including a vehicle route and car following model.
It also provides a lot of useful information such as vehicle speed, model, and position. One of the major features of SUMO is the Traffic Control Interface (or TraCI for short), which is a Python API that treats the SUMO simulation as a server and allows users to gain information from a traffic simulation or modify the simulation. TraCI enables an interface to allow third party systems (or libraries) to integrate with the SUMO traffic simulation. [3] Using the SUMO simulation the project is divided in two phases:
• In Phase 1, performance-based feedback will be provided.
• In Phase 2, The trained agent will be deployed on the simulation and this time it will behave in accordance with the feedback it received during training.

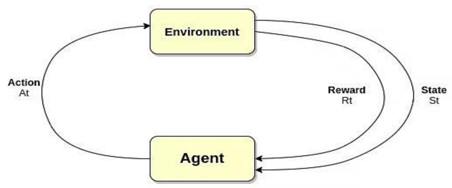

Fig 1. Reinforcement Learning [1]

II. LITREATURE REVIEW

Since this paper is a review of approaches used over the years to identify the problem and finding a optimal solution for the problem. The literature review is done using the mentioned papers:
The first paper reviewed is 'Multi-Agent Broad Reinforcement Learning for Intelligent Traffic Light Control' By-Ruijie Zhu, Lulu Li, Shuning Wu, Pei Lv, Yafai Li, Mingliang Xu.[1] Published in Cornell University, Machine Learning Journal in Mar 2022, introduces Multi-agent System for reinforcement learning along with a Multi-Agent Deep Reinforcement Learning (MADRL) exclusive to the paper and concluding with comparison between Broad Reinforcement Learning (BRL) and Multi-Agent Deep Reinforcement Learning (MADRL).

The second paper reviewed is 'Deep Reinforcement Learning based approach for Traffic Signal Control' By- Kovari Balint, Tettamanti Tamas and Becsi Tamas.[2] Published in Elsevier B.V. in September 2021 under EWGT 2021, introduces Reinforcement Learning and model design with model training for Deep Reinforcement Learning. This paper introduces its own solution in the form of Policy Gradient Algorithm with 70% accuracy using RL-agent.

The third paper reviewed is 'Traffic Light Control Using Hierarchical Reinforcement Learning and Options Framework' By-Dimitrius f. Borges , joão paulo r. R. Leite , edmilson m. Moreira, and otávio a. S. Carpinteiro.[3] Published in IEEE Access Engineering in July 2021, introduces a new approach of Hierarchical Reinforcement Learning (Training in layers). Also, the model Simulation with SUMO and concludes with H-agent being optimal for the problem.

The fourth paper reviewed is 'DRLE: Decentralized Reinforcement Learning at the Edge for Traffic Light Control in the IoV' By-Pengyuan Zhou, Xianfu Chen, Zhi Liu, Tristan Braud, Pan Hui, Jussi Kangasharju.[4] Published in IEEE transactions on intelligent transportation systems in Jan 2021, Introduces the DRLE method in IoV (Internet of Vehicles). The paper works around the central idea of decreasing the convergence time of RL-agent. In conclusion, the DRLE decreases the convergence time by 65.66% compared to PPO (proximal policy optimization) and training steps by 79.44%.

The Fifth paper reviewed is 'Fuzzy Inference Enabled Deep Reinforcement Learning-Based Traffic Light Control for Intelligent Transportation System' By-Neetesh Kumar, Syed Shameerur Rahman, and Navin Dhakad.[5] Published in IEEE transactions on intelligent transportation systems in May 2021, proposed the model DITLCS (Dynamic and Intelligent Traffic Light Control System) with three modes of operations: FM (fair mode), PM (priority mode), EM (emergency mode). The proposed DITLCS reduced 6.45% of CO_2 emission as compared to the existing FCTL (fixed cycle traffic light) model.

The sixth paper reviewed is 'PDLight: A Deep Reinforcement Learning Traffic Light Control Algorithm with Pressure and Dynamic Light Duration' By- Chenguang Zhao, Xiaorong Hu, Gang Wang.[6] Published in Cornell University, Machine Learning Journal in Sep 2020, Proposes the model PDLight with the use of PRCOL (pressure with remaining capacity of outgoing lane). as the reward function. The paper concludes with comparing the PDLight model with existing models like CoLight and PressLight. Table 1 shows comparison between there ait around time accuracies.

PDLight	CoLight	PressLight.
72.55 %	51.8 %	54.9 %

Table 1. PDLight Vs. CoLight Vs. PressLight

The seventh paper reviewed is 'STMARL: A Spatio-Temporal Multi-Agent Reinforcement Learning Approach for Cooperative Traffic Light Control' By-Yanan Wang, Tong Xu, Xin Niu, Chang Tan, Enhong Chen, Hui Xiong.[7] Published in Cornell University Multiagent Systems Journal in Nov 2020, Proposes the model STMARL (Spatio-Temporal Multi-Agent Reinforcement Learning). The STMARL is built on RNN LSTM (long-short term memory). The paper concludes with comparing the proposed model with other baseline models and STMARL outperforms the best baseline by 20.6%.

The eighth paper reviewed is 'Optimised Traffic Light Management Through Reinforcement Learning: Traffic State Agnostic Agent vs. Holistic Agent with Current V2I Traffic State Knowledge' By- Johannes V. S. Busch, Vincent Latzko, Martin Reisslein and Frank H. P. Fitzek.[8]Published in IEEE open journal of intelligent transportation systems in Nov 2020 proposes Agnostic and Holistic methods to train the reinforcement learning model. Also, proposes a V2I (Vehicle to Infrastructure) communication for better traffic control. The paper concludes with a segment of Agnostic agent vs Holistic agent and the results prove that holistic agent is better.

The ninth paper reviewed is 'Multi-Agent Deep Reinforcement Learning for Urban Traffic Light Control in Vehicular Networks' By-Tong Wu, Pan Zhou, Kai Liu, Yali Yuan, Xiumin Wang, Member, Huawei Huang, Dapeng Oliver Wu. [9] Published in IEEE Transactions on Vehicular Technology in May 2020 reviews the Policy Gradient Algorithm and improves the already existing model of PPO using Vehicular Network LSTM networks.

The tenth paper reviewed is 'Comparison of Game Theoretical Strategy and Reinforcement Learning in Traffic Light Control' By-Jian Guo, István Harmati.[10] Published in Periodica Polytechnica Transportation Engineering in Jun 2020, Proposes a game theoretical strategy and SARL (single-agent reinforcement learning) model with Q-learning algorithm. In conclusion the paper compares both the models and Game theoretical strategy outperforms SARL by 11.10% accuracy.

The tenth paper reviewed is 'Traffic-signal control reinforcement learning approach for continuous-time Markov games' By-Román Aragon-Gómez, Julio B. Clempner.[11] Published in Engineering Applications of Artificial Intelligence in Mar 2020, Proposes Continuous-Time Markov Games (CTMG) model. The paper mainly focuses on slaving the problem of Triple traffic signal intersection.

The final paper reviewed is 'A Deep Reinforcement Learning Approach to Adaptive Traffic Lights Management' By-Andrea Vidali, Luca Crociani, Giuseppe Vizzari, Stefania Bandini.[12] Published in Cornell University Electrical Engineering and Systems Science in 2019, Introduces Deep Q-Learning method with the use of neural networks and introduces environment building in SUMO for a traffic signal intersection.

III. METHODODOLGY REVIEW

This section presents the review of methodology as well as our own approach to the problem

A. Reinforcement Learning

In a Reinforcement Learning (RL) problem, an autonomous agent observes the environment and perceives a state St, which is the state of the environment at time t. Then the agent chooses an action at which leads to a transition of the environment to the state St+1. After the environment transition, the agent obtains a reward Rt+1 which tells the agent how good at was with respect to a performance measure. The goal of the agent is to learn the policy $\pi *$ that maximizes the cumulative expected reward obtained as a result of actions taken while following $\pi *$. The standard cycle of reinforcement learning is shown in Figure 1 and 2.

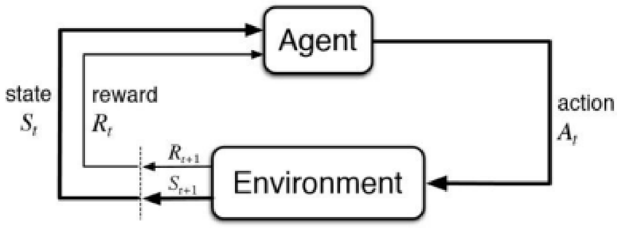

Fig 2. The Reinforcement Learning Cycle [3]

B. Learning in Traffic Signal Control

The traffic micro simulator used for this project is Simulation of Urban MObility (SUMO) SUMO provides a software package which includes an infrastructure editor, a simulator interface and an application programming interface (API). These elements enable the user to design and implement custom configurations and functionalities of a road infrastructure and exchange data during the traffic simulation.

RL techniques applied to traffic signal control address the following challenges:

• *Inappropriate traffic light sequence:* Traffic lights usually choose the phases in a static, predefined policy. This method could cause the activation of an inappropriate traffic light phase in a situation that could cause an increase in travel times.

• *Inappropriate traffic light durations*: Every traffic light phase has a predefined duration which does not depend on the current traffic conditions. This behavior could cause unnecessary waiting for the green phase.

In order to apply a RL algorithm, it is necessary to define the state representation, the available actions and the reward functions:

1) State representation:

The state is the agent's perception of the environment in an arbitrary step. In this project the environment used is shown in figure 3.

It is a 4-way intersection where 4 lanes per arm approach the intersection from the compass directions, leading to 4 lanes per arm leaving the intersection. Each arm is 750 meters long. On every arm, each lane defines the possible directions that a vehicle can follow: the right-most lane enables vehicles to turn right or going straight, the two central lanes bound the driver to go straight while on the left-most lane the left turn is the only direction allowed.

In the center of the intersection, a traffic light system, controlled by the agent, manages the approaching traffic. In particular, on every arm the left-most lane has a dedicated traffic light, while the other three lanes share a traffic light. Every traffic light in the environment operates according to the common european regulations, with the only exception being the absence of time between the end of a yellow phase and the start of the next green phase. In this environment pedestrians, sidewalks and pedestrian crossings are not included.

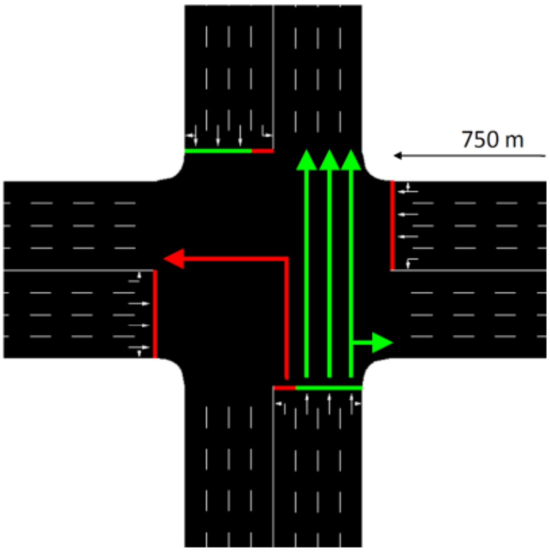

Fig 3. SUMO Environment [12]

2) Action State:
The action set identifies the possible actions that the agent can take.
The agent is the traffic light system, so doing an action translates to activate a green phase for a set of lanes for a fixed amount of time, choosing from a predefined set of green phases. In this project, the green time is set at 10 seconds and the yellow time is set at 4 seconds.
Formally, the action space is defined in the set.
The set includes every possible action that the agent can take.
A = {NSA, NSLA, EWA, EWLA}
Every action of set is described below:
• North-South Advance (NSA): the green phase is active for vehicles that are in the north and south arm and wants to proceed straight or turn right.
• North-South Left Advance (NSLA): the green phase is active for vehicles that are in the north and south arm and wants to turn left.
• East-West Advance (EWA): the green phase is active for vehicles that are in the east and west arm and wants to proceed straight or turn right.
• East-West Left Advance (EWLA): the green phase is active for vehicles that are in the east and west arm and wants to turn left.
The action sets in SUMO simulator are represented in figure 4, below:

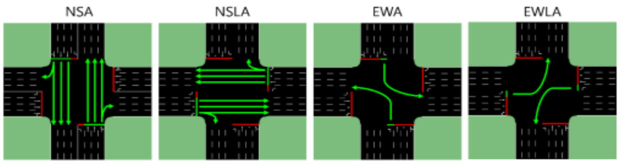

Fig 4. Action sets [12]

3) Reward Function:
In reinforcement learning, the reward represents the feedback from the environment after the agent has chosen an action. The agent uses the reward to understand the result of the taken action and improve the model for future action choices.
Therefore, the reward is a crucial aspect of the learning process. The reward usually has two possible values: positive or negative. A positive reward is generated as a consequence of good actions, a negative reward is generated from bad actions.
In this project, the objective is to maximize the traffic flow through the intersection over time. In order to achieve this goal, the reward should be derived from some performance measure of traffic efficiency, so the agent is able to understand if the taken action reduce or increase the intersection efficiency.
The reward function used in this project uses as a metric the total waiting time, defined as in equation:[6]

$$twtt = \sum_{veh=1}^{n} wt(veh,t)$$

Where *wt(veh,t)* is the amount of time in seconds a vehicle *veh* has a speed of less than 0.1 m/s at agent step *t*. *n* represents the total number of vehicles in the environment in agent step *t*.
Therefore, *twtt* is the total waiting time at agent step *t*. From this metric, the reward function can be defined as a function of *twtt* and is shown in

$$rt = 0.9 \cdot twtt-1 - twtt$$

Where *rt* represents the reward at agent step *t*. *twtt* and *twtt−1* represent the total waiting time of all the cars in the intersection captured respectively at agent step *t* and *t−1*. The parameter 0.9 helps with the stability of the training process.
In a reinforcement learning application, the reward usually can be positive or negative, and this implementation is no exception. The equation is designed in such a way that when the agent chooses a bad action it returns a negative value and when it chooses a good action it returns a positive value.
A bad action can be represented as an action that, in the current agent step t, adds more vehicles in queues compared to the situation in the previous agent step $t - 1$, resulting in higher waiting times compared to the previous agent step. This behavior increases the *twt* for the current agent step *t* and consequently the equation assumes a negative value. The more vehicles were added in queues for the agent step *t*, the more negative *rt* will be and therefore the worst the action will be evaluated by the agent. The same concept is applied for good actions.

C. Model and Training

In this project, a fully connected deep neural network is used, which is composed of an input layer of 80 neurons, 5 hidden layers of 400 neurons each with rectified linear unit (ReLU) and the output layer with 4 neurons with linear activation function, each one representing the value of an action given a state. A graphical representation of the deep neural network is showed in Figure 5.

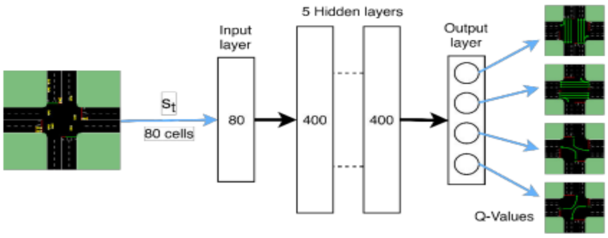

Fig 5. Neural Network Architecture [12]

A training instance consists of learning the Q-value function iteratively using the information contained in the batch of samples extracted. Every sample in the batch is used for training. From the standpoint of a single sample, which contains the elements {st, at, rt+1, st+1},
Training of the neural network. The input is the state st, while the desired output is the updated Q-values Q(st, at) that now includes the maximum expected future reward thanks to the Q-value update.

D. Flow Chart

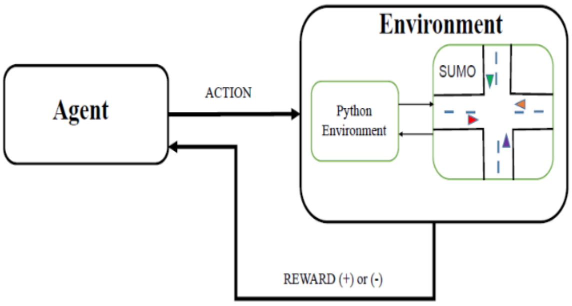

Fig 6. Flow Chart

E. RL- Agent workflow

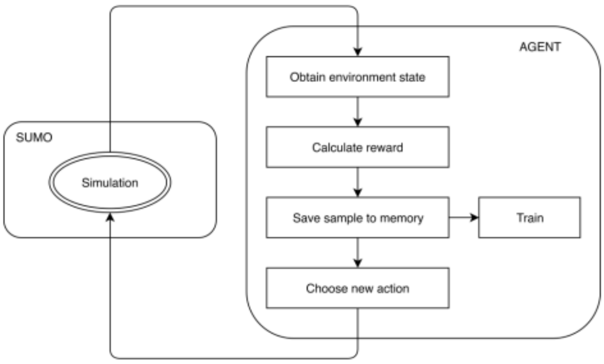

Fig 7. RL-Agent Workflow [12]

IV. RESULTS AND DISCUSSION

Results:

After the simulation is done using SUMO, the model returns some performance matrix of the agent.
The performance of the agents is assessed in two parts: initially, the reward trend during the training is analyzed. Then, a comparison between the agents and a static traffic light is discussed, with respect to common traffic metrics, such as cumulative wait time and average wait time per vehicle.

The model is then tested on the simulation by letting itself run for some time without any interference.
Figure 9 and 10 shows graph of the tested model:

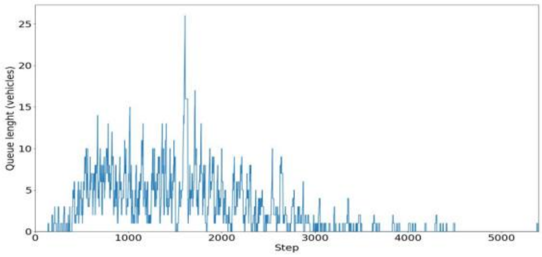

Fig 8. Queue Plot

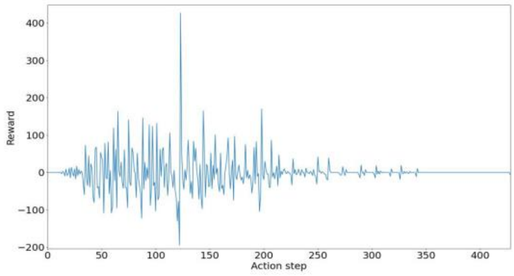

Fig 9. Reward Plot

SUMO SIMULATION:
This is how the traffic simulation looks like in SUMO:

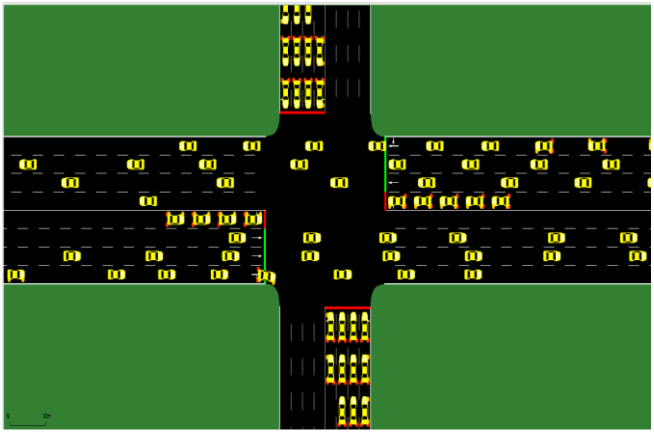

Fig 10. SUMO Simulation

Discussion:
This project has presented a exploration of a RL approach to the problem of traffic lights adaptation and management. The work has employed a realistic and validated traffic simulator to provide an environment in which training and evaluating a RL agent.
Two metrics for the reward of agent' actions have been investigated, clarifying that a proper description of the application context is just as important as the competence in the proper application of machine learning approaches for achieving proper results. Future works are aimed at further improving achieved results, but also, within a longer term, at investigating what would be the implications of introducing multiple RL agents within a road network and what would be the possibility to coordinate their efforts for achieving global improvements over local ones, and also the implications on the vehicle population, that could perceive the change in the infrastructure and adapt in turn to exploit additional opportunities and potentially negating the achieved improvements due to an additional traffic demand on the improved intersections. It is important to perform analyses along this line of work to understand the plausibility, potential advantages or even unintended negative implications of the introduction in the real world of this form of self-adaptive system.

V. CONCLUSION AND FUTURE SCOPE

Conclusion:
In this project, we address the traffic light control problem using reinforcement learning approach. This project has conducted simulations using SUMO and demonstrate the performance of our model using plots and results. In addition, we have conducted an extensive literature review for the project to show in-depth case studies and observations to understand how the agent adjust to the changing traffic, as a complement to quantitative measure on rewards. These in-depth case studies can help generate traffic rules for real world application.
Reinforcement learning is in use for a decade now, this approach has some potential to it for solving the Traffic problem.
In this project we have not discussed the impact on environment from this approach which something to be studied in more depth.
Future Scope:

We also acknowledge the limitations of our current approach and would like to point out several important future directions to make the method more applicable to real world.

1) The project can be extended from two-phase traffic light to multi-phase traffic light, which will involve more complicated but more realistic state transition.

2) The project addresses a simplified one intersection case, whereas the real-world road network is much more complicated than this. Although some studies have tried to solve the multi-intersection problem by using multiple reinforcement learning agents, they do not explicitly consider the interactions between different intersections.

3) This approach is still tested on a simulation and thus the feedback is simulated. A field study should be conducted to learn the real-world feedback and to validate the proposed reinforcement learning approach.

4) Lastly, as said earlier this project can be more optimized from the point of view of controlling CO2 release in the environment and making it more eco-friendly.

ABBREVATIONS

SUMO: Simulator for Urban Mobility.
RL: Reinforcement Learning.
RL - agent: Reinforcement Learning agent.
TraCI: Traffic Command Interface.
LSTM: Long Short-Term Memory.
RNN: Recurrent Neural Network

DECLARATIONS

This paper is a review paper and all the information mentioned in this paper is well cited and researched.
The authors of this paper give full consent to publish this paper across any journal.
All the data and material are gathered from various research papers and publications which are cited in the references section.
This is paper is a review on the topic of reinforcement learning in traffic light control, therefore the authors have no interest in competing with this paper.
All the authors have deliberately provided their time and efforts to research the topic and produce this research paper.

Conflict of Interest

The authors declare that they have no conflict of interest.

AKNOWLEDGEMENT

The authors would like to acknowledge the authors of the various research papers used in this review paper and in this research. The authors would also like to acknowledge the authors of all the programming libraries as well as the developers of SUMO.
The authors would also like to acknowledge the authors behind free to use resources such as Anaconda Navigator and TensorFlow.

REFERENCES

1. Zhu, Ruijie, Lulu Li, Shuning Wu, Pei Lv, Yafai Li, and Mingliang Xu. "Multi-Agent Broad Reinforcement Learning for Intelligent Traffic Light Control." arXiv preprint arXiv:2203.04310 (2022).
2. Bálint, Kővári, Tettamanti Tamás, and Bécsi Tamás. "Deep Reinforcement Learning based approach for Traffic Signal Control." Transportation Research Procedia 62 (2022): 278-285.
3. Borges, Dimitrius F., João Paulo RR Leite, Edmilson M. Moreira, and Otávio AS Carpinteiro. "Traffic light control using hierarchical reinforcement learning and options framework." IEEE Access 9 (2021): 99155-99165.
4. Zhou, Pengyuan, Xianfu Chen, Zhi Liu, Tristan Braud, Pan Hui, and Jussi Kangasharju. "DRLE: Decentralized reinforcement learning at the edge for traffic light control in the IoV." IEEE Transactions on Intelligent Transportation Systems 22, no. 4 (2020): 2262-2273.
5. Kumar, Neetesh, Syed Shameerur Rahman, and Navin Dhakad. "Fuzzy inference enabled deep reinforcement learning-based traffic light control for intelligent transportation system." IEEE Transactions on Intelligent Transportation Systems 22, no. 8 (2020): 4919-4928.
6. Zhu, Ruijie, Lulu Li, Shuning Wu, Pei Lv, Yafai Li, and Mingliang Xu. "Multi-Agent Broad Reinforcement Learning for Intelligent Traffic Light Control." *arXiv preprint arXiv:2203.04310* (2022).
7. Bálint, Kővári, Tettamanti Tamás, and Bécsi Tamás. "Deep Reinforcement Learning based approach for Traffic Signal Control." *Transportation Research Procedia* 62 (2022): 278-285.
8. Borges, Dimitrius F., João Paulo RR Leite, Edmilson M. Moreira, and Otávio AS Carpinteiro. "Traffic light control using hierarchical reinforcement learning and options framework." *IEEE Access* 9 (2021): 99155-99165.
9. Abdulhai, Baher, Rob Pringle, and Grigoris J. Karakoulas. "Reinforcement learning for true adaptive traffic signal control." *Journal of Transportation Engineering* 129, no. 3 (2003).
10. Wiering, Marco A. "Multi-agent reinforcement learning for traffic light control." In *Machine Learning: Proceedings of the Seventeenth International Conference (ICML'2000)*, pp. 1151-1158. 2000.
11. Li, Chenghao, Xiaoteng Ma, Li Xia, Qianchuan Zhao, and Jun Yang. "Fairness control of traffic light via deep reinforcement learning." In *2020 IEEE 16th International Conference on Automation Science and Engineering (CASE)*, pp. 652-658. IEEE, 2020.
12. Vidali, Andrea, et al. "A Deep Reinforcement Learning Approach to Adaptive Traffic Lights Management." WOA. 201