ITU Computer Science, AI and Robotics

ITU/CSAR

Control of Emergency Vehicles with Deep Q-Learning

Hasan Haydar Yıldız[1], **Furkan Güney**[2], **İlhan Tunç**[3] **and Mehmet Turan Söylemez**[4]

[1] İstanbul Teknik Üniversitesi
[2] Türk Havacılık Uzay Sanayi A.Ş.
[3] Bursa Teknik Üniversitesi
[4] İstanbul Teknik Üniversitesi

Abstract: In contemporary times, the issue of traffic congestion has become a paramount concern affecting a broad spectrum of society. However, when it comes to emergency vehicles, particularly ambulances, this matter takes on even greater significance. This study addresses a research endeavor aimed at mitigating traffic risks for emergency situations. The primary objective of the research is to employ Deep Q-Learning methodology to ensure that ambulances transport patients to hospitals in the quickest and most optimal routes. Factors such as urgency levels, traffic density, and distances between patients and ambulances are modeled using state vectors. The Deep Q-Learning algorithm utilizes these vectors to select the most effective actions, determining the most efficient routes for ambulances to transport patients. The reward function is transformed into a penalty function by prioritizing patients based on their waiting times. The study evaluates the learning outcomes of the agent created with Deep Q-Learning, demonstrating the successful completion of the learning process. This method represents a significant step in optimizing the intra-city mobility of emergency vehicles.

Keywords: Traffic Congestion, Emergency Vehicles, Deep Q-Learning, Optimization

Derin Q Temelli Ambulans Aracının Kontrolü

Özet: Günümüzde, kentlerin karşılaştığı trafik sorunu, toplumun geniş bir kesimini etkileyen öncelikli bir mesele olmuştur. Ancak acil durum araçları, özellikle ambulanslar, bu karmaşık durumu daha da karmaşık hale getirebilmektedir. Bu bağlamda, trafik riskini azaltmaya yönelik önemli bir araştırma yapılmaktadır. Bu çalışmanın odak noktası, acil durum araçları, özellikle ambulanslar için en uygun rotaları belirleyerek hastaları en hızlı şekilde hastanelere ulaştırmayı amaçlamaktadır. Araştırma, Derin Q-Öğrenimi yöntemini benimsemekte ve ambulansların rotalarını optimize etmek amacıyla bu öğrenme modelini kullanmaktadır. Çalışmanın temelini oluşturan Derin Q-Öğrenimi algoritması, acil durumları, trafik yoğunluğunu ve hasta ile ambulans arasındaki mesafeler gibi faktörleri dikkate alarak durum vektörleriyle modellenmiştir. Bu vektörler, ambulansların en etkili aksiyonları seçerek hastaları taşıması için en uygun rotaları belirlemek üzere kullanılmaktadır. Ayrıca, ödül fonksiyonu hastaların bekleme sürelerine göre önceliklendirilmiş ve bu durum ceza fonksiyonuna dönüştürülmüştür. Çalışmanın sonuçları, Derin Q-Öğrenimi ile eğitilen ajanın başarılı bir şekilde durumu öğrenip en etkili rotaları belirleme yeteneğini göstermektedir. Bu yöntem, acil durum araçlarının şehir içi hareketliliğini optimize etme açısından önemli bir adımı temsil etmektedir.

Anahtar Kelimeler: Trafik Sıkışıklığı, Acil Durum Araçları, Derin Q-Öğrenimi, Optimizasyon

RESEARCH PAPER
Corresponding Author: Hasan Haydar Yıldız, yildizh16@itu.edu.tr
Reference: H. H. Yıldız, et al., (2024), Control of Emergency Vehicles with Deep Q-Learning, ITU Computer Science, AI and Robotics, 1, (1) 32–38.
Submission Date: January, 22, 2024
Acceptance Date: May, 1, 2024
Online Publishing: July, 20, 2024

1 Introduction

Traffic congestion continues to escalate incessantly worldwide, posing a significant impediment for individuals commuting to work. As a consequence of burgeoning population and urbanization, there is a persistent increase in the demand for transportation across cities globally [1]. Particularly in major metropolises, issues such as inefficient time utilization due to traffic congestion, air pollution, and the impacts of climate change are encountered [2]. The surge in transportation demand has led to a rapid proliferation of vehicles, consequently exacerbating traffic congestion. The ramifications of traffic congestion also extend to affecting the functionality of emergency vehicles [3]. In an effort to address the issue of traffic congestion, various intelligent transportation systems have been developed. Effectively managing both traffic congestion and the precision of traffic information holds significant importance in optimizing time. Enhancing the efficiency of emergency vehicles during their missions relies on accurately assessing traffic density and adjusting routes accordingly. Consequently, efforts have been made to conduct studies aimed at precise calculation and classification of traffic congestion for this purpose [4].

One of the most prevalent applications of intelligent transportation systems is traffic signal control. Given the rapid escalation in the number of vehicles in urban networks, it becomes imperative to develop effective mechanisms for traffic light control. The objective of signal control is to enhance intersection capacity, reduce delays, and concurrently ensure the safety of traffic participants. Additionally, it has the potential to reduce fuel consumption and emissions [5]. These systems typically employ pre-determined timing schedules, providing a green signal to each approach in every cycle regardless of traffic conditions. While this may be an optimal solution for areas with high traffic density, it may not be as beneficial in regions with low traffic density where vehicles do not queue due to the absence of congestion [1].

Numerous experts and academics in the field of road network route planning have conducted various studies to enhance efficiency. The focal point of these studies is the Dijkstra algorithm. Although the Dijkstra algorithm has certain limitations in finding the most optimal path between two points, it also possesses irreplaceable advantages [6]. Addressing traffic signal control problems using Deep Q-Learning is a key focus in intelligent transportation research. Researchers in this domain have proposed various solutions based on Deep Q-Learning methods, aiming to optimize traffic flow and alleviate congestion through the application of artificial intelligence in traffic signal control [7]. The optimization algorithm of Deep Q-Learning has been effectively utilized in emergency route planning, demonstrating its potential for efficient route design in emergency situations [8]. Consequently, it has been decided to undertake a study employing Deep Q-Learning techniques to address the issue of traffic congestion caused by emergency vehicles.

This study aims to optimize traffic management strategies to enhance the efficiency of emergency services and reduce emergency intervention times. This approach ensures that emergency vehicles can reach the incident site using the shortest and fastest routes, thereby increasing the effectiveness and success probability of emergency interventions. The algorithm created using Deep Q-Learning successfully achieves the goal of enabling emergency vehicles to reach their target person or location in the shortest possible time.

SUMO (Simulation of Urban Mobility) and TraCI (Traffic Control Interface) libraries are powerful tools utilized for traffic simulation and real-time data exchange. SUMO contributes to various research areas, including route selection, traffic signal algorithms, and simulating vehicle communication [9].

The second section of the paper discusses main knowledge and use of fields of deep q learning and reinforcement learning . The third section delves into the functional elements and features employed within Deep Q-Learning, while the fourth section explores the results of the implemented application and provides interpretations. The fifth section outlines the obtained results and discusses future applications.

2 Deep Q Learning and Reinforcement Learning

"Deep Learning" stands as a specialized field within machine learning, centered around artificial neural networks and crafted to execute intricate tasks inspired by the workings of the biological neural system. Recognized for its capacity to autonomously learn from extensive datasets, this method is devised to empower computer systems to tackle more complex challenges, typically employing artificial neural networks with multiple layers. Deep learning has achieved notable success in diverse domains such as visual recognition, natural language processing, and other intricate problem-solving scenarios.

Deep Q learning finds extensive application in robotics, computer simulations, the gaming industry, and is also applied in forecasting future decision-making processes. In a study addressing energy consumption in mobile application usage, researchers endeavored to anticipate the sequence of applications a user might access and predict the subsequent application the user would switch to. The study yielded highly impressive results, boasting precision and recall rates of approximately 70% and 62%, respectively [10].

Reinforcement Learning is a learning paradigm that involves an agent engaging with its environment, aiming to perform a given task optimally by utilizing feedback from the environment. This learning approach concentrates on the agent acquiring an understanding of relationships among different states and utilizing this knowledge to make optimal decisions. Built upon reward and penalty systems, reinforcement learning seeks to enable the agent to learn the most effective strategy for reaching its objectives.

Q-learning, one of the most widely used approaches in reinforcement learning, is essential to the disciplines of machine learning and artificial intelligence. An agent can learn from its interactions with the environment by using this technique. In essence, the agent interacts with its surroundings as it attempts to finish a certain goal, honing its decision-making skills through the experiences it gains. Predicting a Q-value for every state-action combination in a series of actions is how Q-learning works. These Q-values help the agent create an ideal policy by showing the worth of each action in each stage. The following formula is used to update the values:

$$\begin{aligned} Q'(s_t,a_t) &= Q(st,a_t) \\ &\quad + \alpha \cdot (R_{t+1} + \gamma \cdot \max_{a_{t+1}} Q(s_{t+1},a_{t+1}) \\ &\quad - Q(s_t,a_t)) \end{aligned} \tag{1}$$

- $Q'(s_t,a_t)$*: Updated Current Q-Value*
- $Q(s_t,a_t)$*: Current Q-Value*
- α*: Learning Rate*
- R_{t+1}*: Rewards*
- γ*: Discount Factor*
- $\max_{a_{t+1}} Q(s_{t+1},a_{t+1})$*: Estimated Reward From Next Action*

The aim of the learning process is to have a Q network that accurately predicts the target values in Eq. 1, with the weights of the DQN at time step t denoted by θ_t. To achieve this, the learning algorithm seeks to minimize the loss function [11] specified in Eq. 2 :

$$\text{Loss}_t(\theta) = E\left[(y_t - Q_t(s_t,a_t;\theta_t))^2\right]$$

(2)

The primary process of updating *Q(s, a)* often depends on the maximum future rewards (anticipated rewards) as well as the existing rewards.

Deep Q Learning is the amalgamation of deep learning and reinforcement learning principles. The primary reasons for choosing this algorithm include its flexible structure with broad applicability, high performance, and autonomous decision-making capability. This methodology integrates the principles of reinforcement learning into the framework of deep learning techniques to enhance an agent's capability to perform tasks more efficiently. Deep Q Learning has exhibited considerable success, particularly in domains such as gaming, robot control, financial analysis, and autonomous driving. The agent refines its learning process by utilizing intricate structured data to comprehend and optimize its surroundings. However, despite the advantages of Deep Q-Learning, there are also some disadvantages. Factors such as large data and computational power requirements may limit the applicability of this algorithm. Additionally, challenges such as overfitting and the need for human resources may be encountered. Nevertheless, Deep Q Learning represents significant potential within the realm of machine learning and continues to be a dynamically evolving area of research

3 Deep Q-Learning for Emergency Vehicle Route Planning

In this segment of the research, the optimization of emergency vehicle control is executed using the Deep Q-Learning algorithm, taking into account the urgency and distance of patients. The emergency vehicle takes actions based on its existing state vector and undergoes training through a reward function. The primary goal is to effectively convey patients to the hospital, considering both their urgency and distances, in the most time-efficient manner possible. A subsection concludes upon the successful transportation of all patients. The simulation incorporates three distinct urgency levels for patients, categorized as 1st, 2nd, 3rd and 4th degrees, in descending order of urgency. Within the simulation, all patients emerge simultaneously, facilitating the emergency vehicle's decision-making process.

In this approach, it is stated that the ambulance selects patients as actions and simulates these choices as a step. The state vector comprises 8 elements, encompassing details such as patients' urgency levels and their respective distances to the emergency vehicle. The model makes action selections based on the state vector, with the reward function serving as a penalty mechanism when used in a negative manner. The learning rate is specified as 0.75. Each value in the state vector is recalibrated for every simulation section and continuously updated throughout the simulation. The simulation undergoes updates in sections of 50, 100, and 200, totaling a maximum of 8000 steps, where each step corresponds to a simulation progression in the program. Traffic lights maintain a fixed phase, with a green light duration of 4 seconds and a red light duration of 10 seconds. Illustrations of the road network and the emergency vehicle in simulation are depicted in Figure 1 and Figure 2.

3.1 State

The state vector used in the simulation consists of 8 elements. These elements carry information about the

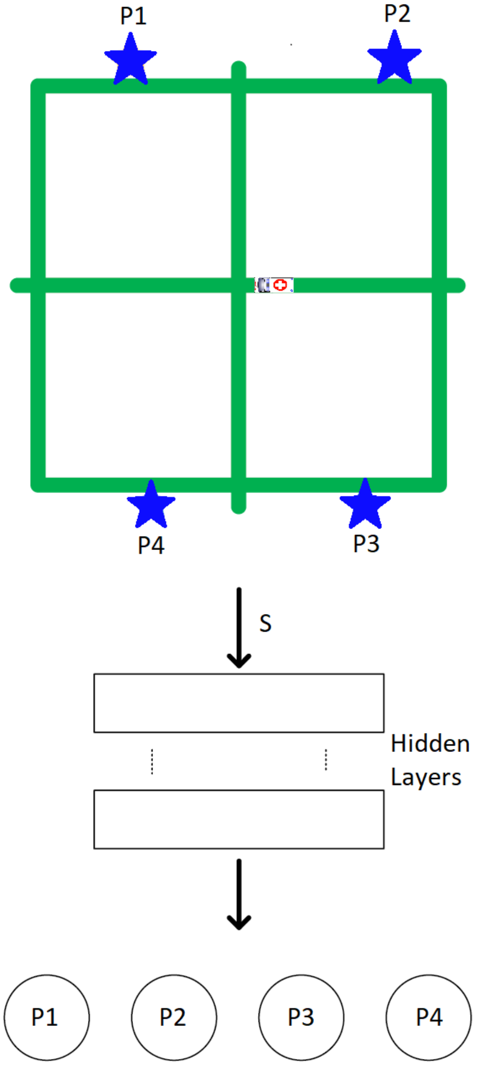

Fig. 1 Deep Leaning Algorithm

urgency levels of patients and the distance of each patient to the emergency vehicle. The elements of the state vector include the following:

- *Importance Coefficient of a Patient with Urgency Level 1*
- *Importance Coefficient of a Patient with Urgency Level 2*
- *Importance Coefficient of a Patient with Urgency Level 3*
- *Importance Coefficient of a Patient with Urgency Level 4*

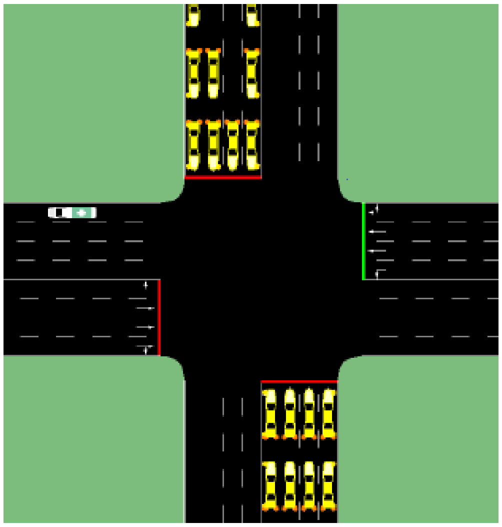

Fig. 2 Simulation Network, Emergency Vehicle

- *Distance of a Patient with Urgency Level 1 to the Hospital*
- *Distance of a Patient with Urgency Level 2 to the Hospital*
- *Distance of a Patient with Urgency Level 3 to the Hospital*
- *Distance of a Patient with Urgency Level 4 to the Hospital*

This state vector contains the essential information for the emergency vehicle to initiate the action of selecting a patient. As exemplified in Figure 3, distance information is provided. The state vector undergoes updates at each step, and an appropriate action is chosen by utilizing this information.

3.2 Action

The action involves the emergency vehicle selecting a patient at the onset of each section. At the beginning of every section, the emergency vehicle opts for one of the four patients with varying urgency levels, guided by the state vector. These actions are employed to locate the patients, and the completion of finding all patients concludes one section.

3.3 Reward

The reward serves as feedback to the learning algorithm, ensuring improved results in the vehicle's subsequent steps. In this study, the reward function is utilized as a penalty function with a negative sign. Equation 1 demonstrates the calculation process of the reward algorithm.

- *Instantaneous Step:* The current step count

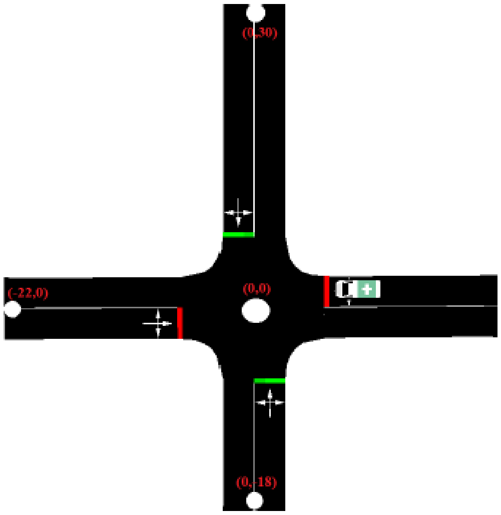

Fig. 3 Location Coordinates

- *Maximum Step Count:* The maximum number of steps in the simulation, set to 8000 in this project.
- *Selected Patient Coefficients:* Assigned as 100, 50, 25, 1 in descending order of urgency, from most urgent to least urgent.

In the reward function, the urgency level of patients in an emergency is a fundamental factor. However, to incorporate this influence proportionally, importance coefficients are introduced. These coefficients are defined in a descending order from the most urgent to the least urgent as 100, 50, 25, 1. The aim of these coefficients is to steer the emergency vehicle's patient selection towards those with higher urgency levels. The most urgent patient incurs a higher penalty if chosen late, motivating the vehicle to avoid repeating such instances. Moreover, instead of having a solely independent relationship with scalar numbers, the reward function is scaled by the ratios of the instantaneous step and maximum step counts to be proportional to both the initial and final steps of the simulation. According to this adjustment, errors made in later steps have a more significant impact on optimization compared to errors in the initial steps. The (-) sign is used to convert the reward function into a penalty function. The reward function is computed at each action stage and simulation step, accumulated at the end of each section, and stored.

$$Reward = -1 * \frac{InstantaneousStep}{MaximumStep} * (PatientCoefficients) \quad (3)$$

4 Simulation Results

The computations were carried out on an AMD Ryzen 5 2600 64-bit machine with a clock speed of 3.40 GHz and 16 GB RAM. The entire training process, comprising a total of 200 sections, took 2 hours and 14 minutes.

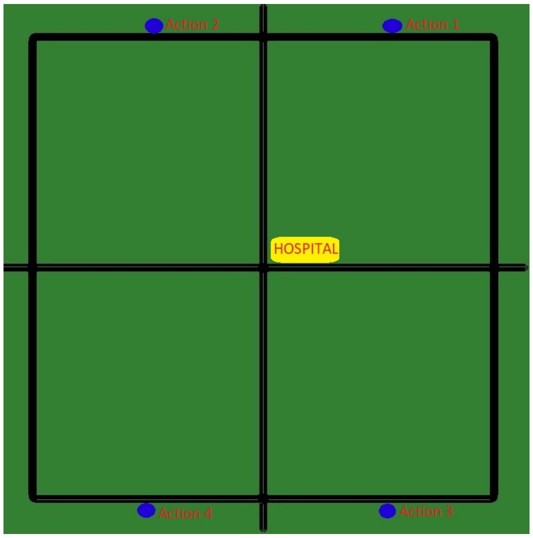

Fig. 4 Representation of Actions

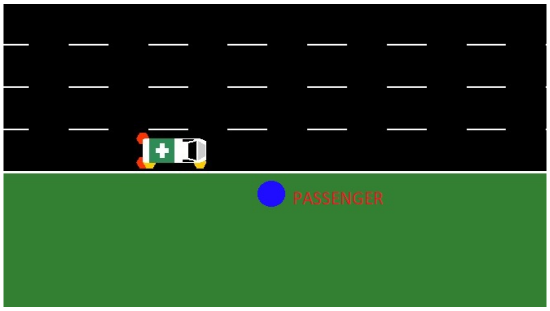

Fig. 5 Fetching the Selected Patient

The features of the Deep Q-Learning model are as follows: The model has a four-layer deep neural network structure. Each layer has a width of 400, indicating that the model has a wide and complex structure, enhancing its ability to solve complex problems. The model has a batch size of 100 data samples per training iteration, specifying the amount of data used in each iteration. The learning rate is set to 0.001, determining how much the weights are adjusted in each update step. The model is trained for a total of 8000 epochs, indicating how long it is trained to learn patterns in the dataset. These features are critical factors that determine the structure, training process, and performance of the model.

In this study, the Deep Q-Learning algorithm was utilized for the route planning of emergency vehicles. The agent is guided to deliver patients to the hospital in the shortest time possible, considering the urgency and distances of patients. The state vector incorporates information about patients' urgency levels and distances to the hospital. By utilizing the information in the state vector, the agent selects appropriate actions and undergoes self-training with the reward function. Simulation results indicate that the agent successfully reaches the desired goal and accomplishes effective route planning. As the number of sections increases, the agent achieves the goal more efficiently with fewer penalties.

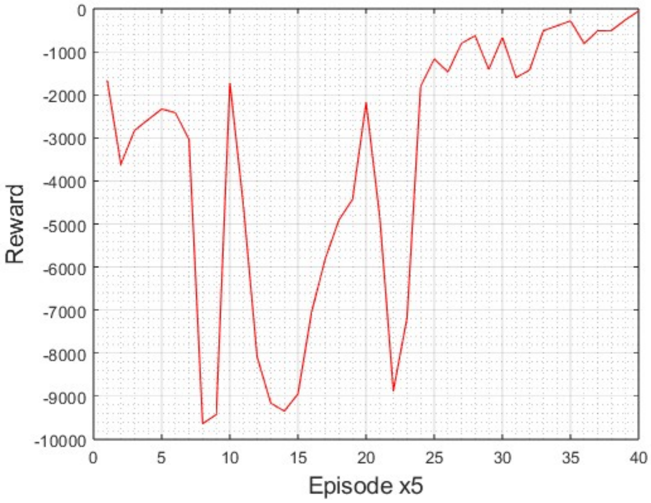

Fig. 6 Reward Graph for 200 Episode

The outcomes of the simulation presented in Figure 6 are provided. While labeled as "Reward" in the figure, the primary objective in Deep Q-Learning is actually to minimize penalties. Hence, the aim is to approach a value close to 0. In essence, the agent strives to complete the designated number of sections with minimal penalties. As per the results in Figure 6, it is evident that the simulations utilizing the Deep Q-Learning algorithm have reached the intended outcome, and the agent has successfully concluded the learning process.

In the reward function graph observed at the conclusion of the simulation, fluctuations are evident. These fluctuations primarily stem from the number of patients. The emergency vehicle takes action to locate the most urgent patient, updating its strategy based on the state matrix and reward function for each patient. With a total of four patients, it spent more time locating the first three patients, who were in the most urgent condition, than the last patient. Consequently, the penalty function exerted greater influence. However, after attending to the first three patients, the emergency vehicle found the last patient more swiftly. It rapidly completed the learning process within the departments, ultimately converging the reward function to zero.

Drawing insights from the visualization, it can be asserted that the agent has attained the desired result and accomplished effective learning. Maintaining an optimal number of sections for the agent throughout the learning process can yield superior results, indicating an enhancement in the agent's learning speed and efficiency. Visual analysis plays a pivotal role in evaluating the agent's performance. Upon scrutiny of the results, it is clear that the agent has effectively met its goal and made accurate decisions. This affirms the effective operation of the Deep Q-Learning algorithm, with the agent having gathered adequate data to comprehend its task. In summary, the developed Deep Q-Learning algorithm has successfully met its intended objective, and the learning process has been executed triumphantly. With the advancement of the number of sections, the agent has approached a value close to 0 with fewer penalties, showcasing commendable results.

5 Conclusion

In emergency scenarios where time is a critical factor, the swift and efficient arrival of emergency vehicles at the incident scene holds vital importance. This study introduces an innovative approach to optimize the routes of emergency vehicles and guide their flow in traffic, resulting in a reduction in the arrival time at the incident scene and expediting emergency interventions.

The problem in this study was addressed using the Deep Q-Learning method. Factors such as urgency levels, distances of patients, and ambulances were utilized as the state vector in experiments. The reward function was proportionally scaled based on the waiting times of patients, ensuring priority treatment for the most urgent cases. Patients themselves were considered as action options. Experiments with different learning episodes were conducted, and the most impressive results were achieved with 200 learning episodes. The magnitude of the penalty function decreased after each episode, and the graphs converged towards zero. These outcomes indicate the successful applicability of the Deep Q-Learning method in addressing traffic-related issues.

The achieved results are promising and can serve as guidance for future research endeavors. In subsequent studies, it is envisioned that the system can be further enhanced to model complex scenarios involving multiple ambulances, allowing for more effective emergency management. Additionally, the dynamic adjustment of traffic lights based on the routes taken by ambulances could be implemented to ensure unimpeded progress. In real-world applications, the use of methods granting priority passage rights to ambulances and raising awareness among drivers to yield to ambulances remains crucial.

References

[1] K. Shingate, K. Jagdale, and Y. Dias, "Adaptive traffic control system using reinforcement learning," *International Journal of Engineering Research and Technology*, vol. 9, 2020.

[2] İ. Tunç, Ö. Elmas, A. Edem, A. Köroğlu, S. Akmeşe, and M. Söylemez, “Derin q öğrenme tekniği ile trafik işık sinyalizasyonu,” 2023.

[3] M. Vardhana, N. Arunkumar, E. Abdulhay, *et al.*, “Iot based real time traffic control using cloud computing,” *Cluster Computing*, vol. 22, no. Suppl 1, pp. 2495–2504, 2019.

[4] X. Yin, G. Wu, J. Wei, Y. Shen, H. Qi, and B. Yin, “Deep learning on traffic prediction: Methods, analysis, and future directions,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 4927–4943, 2022. DOI: 10.1109/TITS.2021.3054840.

[5] M. Abdoos, N. Mozayani, and A. Bazzan, “Traffic light control in non-stationary environments based on multi agent q-learning,” in *IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC*, Oct. 2011.

[6] D. Fan and P. Shi, “Improvement of dijkstra’s algorithm and its application in route planning,” in *2010 seventh international conference on fuzzy systems and knowledge discovery*, vol. 4, 2010, pp. 1901–1904.

[7] G. Han, Q. Zheng, L. Liao, P. Tang, Z. Li, and Y. Zhu, “Deep reinforcement learning for intersection signal control considering pedestrian behavior,” *Electronics (Basel)*, vol. 11, no. 21, p. 3519, 2022.

[8] S. A. El-Tantawy, H. Abdelgawad, and R. A. Ramadan, “Cooperative deep q-learning for traffic signal control,” *Transportation Research Part C: Emerging Technologies*, vol. 71, pp. 1–16, 2016.

[9] M. Behrisch, L. Bieker, J. Erdmann, D. Krajzewicz, and C. Rössel, “Sumo – simulation of urban mobility: An overview,” *International Journal on Advances in Systems and Measurements*, vol. 4, no. 34, pp. 308–316, 2011.

[10] Z. Shen, K. Yang, W. Du, X. Zhao, and J. Zou, “Deepapp: A deep reinforcement learning framework for mobile application usage prediction,” Nov. 2019, pp. 153–165, ISBN: 978-1-4503-6950-3. DOI: 10.1145/3356250.3360038.

[11] T. Pan, “Traffic light control with reinforcement learning,” pp. 4–5, Aug. 2023.