IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 4 February 2023, accepted 1 March 2023, date of publication 6 March 2023, date of current version 22 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3253503

RESEARCH ARTICLE

Emergency Vehicle Aware Lane Change Decision Model for Autonomous Vehicles Using Deep Reinforcement Learning

AHMED ALZUBAIDI[1], AMEENA SAAD AL SUMAITI[2], (Senior Member, IEEE), YOUNG-JI BYON[3], AND KHALIFA AL HOSANI[2], (Senior Member, IEEE)

[1]Electrical Engineering and Computer Science Department, Khalifa University, Abu Dhabi, United Arab Emirates
[2]Advanced Power and Energy Center, Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi, United Arab Emirates
[3]Department of Civil Infrastructure and Environmental Engineering, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates

Corresponding authors: Ahmed Alzubaidi (ahmedamz@outlook.com) and Ameena Saad Al Sumaiti (ameena.alsumaiti@ku.ac.ae)

This work was supported by Khalifa University under Award kkjrc-2019-trans2.

ABSTRACT Autonomous Vehicles (AVs) have advanced rapidly in recent years as they promise to be safe and minimize the burden coming from the driving task. AVs share the road with various categories of vehicles as Emergency Vehicles (EMVs) (e.g police and ambulance vehicles). When being approached by an active EMV, it is natural to expect all vehicles to cooperate with EMV, such that the EMV travel time is minimized. The decision-making block of an AV includes the responsibility of instructing the AV to change lanes, which is typically handled by the Lane Change Decision (LCD) model. A typical LCD model tends to overlook the presence of EMVs around, as they neglect the impact of the lane change on the EMV utility. To address this challenge, this paper proposes an Emergency Vehicle Aware LCD via utilizing Deep Reinforcement Learning. To our best knowledge, this is one of the pioneering works that propose a DRL solution for the problem, addressing important limitations that have been identified. The proposed solution was evaluated against a rule-based LCD known as MOBIL in terms of safety and level of cooperativeness with the EMV. Some key results found from the comparison between the proposed solution and MOBIL are (1) identical safety levels,(2) proposed solution is takes far less time to give up the lane when being approached by an EMV, and (3) proposed solution never blocks the path of the EMV, whereas MOBIL occasionally block the path.

INDEX TERMS Deep reinforcement learning, autonomous vehicles, lane changes.

I. INTRODUCTION

There are various types of vehicles on roads such as Human-driven Vehicles (HVs), AVs, and heavy trucks. There has been huge interest from technical experts and researchers on AVs due to the potential benefits they are capable of delivering, in terms of safety and efficiency [1], [2]. EMVs belong to HVs and include police and ambulance vehicles. Naturally, EMV deserves the highest priority when being in operation. Thus, all other road users should prioritize the efficiency of EMVs, with cooperative efforts by all vehicles to minimize the travel times of EMVs to reach their desired destination.

The associate editor coordinating the review of this manuscript and approving it for publication was Abderrahmane Lakas.

Evidence of vehicles prioritizing the EMVs is manifested by vehicles giving up the lane and avoiding making lane changes when being approached by an EMV, especially when making lane changes that directly hinder the travel path of the EMV. Fig. 1 and Fig. 2 elaborate on two specific road situations, where the existence of EMV changes the desired optimal behavior. In the incident an EMV gets blocked by any vehicle on the road, the consequence can be damaging. Several countries have targeted reducing the travel times for EMVs, as it helps significantly in rescuing lives and protecting people's possessions [3]. In the UK, health authorities have set a target of 8 minutes for the rescue time [4]. Dubai which is a major state in the UAE, is aiming for a target of 4 minutes [5]. The importance of prioritizing EMVs over all other vehicles

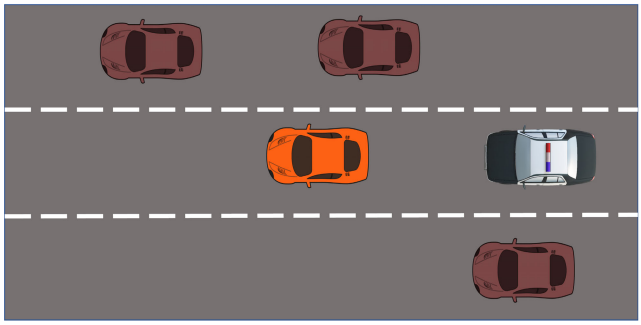

FIGURE 1. Red vehicles are HVs and the orange vehicle is the AV (or ego-vehicle). In the depicted road situation above, the AV is expected to give up the lane to prioritize the EMV. The optimal behavior for the AV in this situation is to execute the LC left while avoiding any collision with any vehicles.

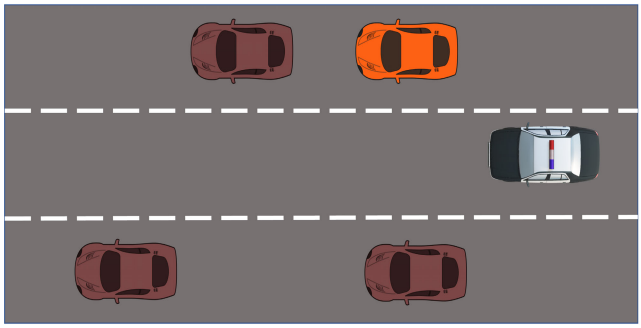

FIGURE 2. Red vehicles are HVs and the orange vehicle is the AV (or ego-vehicle). In the depicted road situation above, assume the AV is preceded by an HV which is driving at a slow speed, which is affecting its efficiency. Despite the fact that AV remaining in its current lane is an inefficient decision, however, it is optimal as it would avoid blocking the way on the EMV which is going to affect the efficiency of the EMV.

has been emphasized and established. Therefore, it should be noted that AV's decision-making must be aware of an EMV that exists in the vicinity. The presence of an EMV around can be classified as an edge case to the AV decision-making, which it must be able to handle. This is a specific edge case that cannot be compromised by the AV, as it only takes a single vehicle being uncooperative with the EMV, to hinder the prioritized operation of the EMV.

One component of the AV decision-making is the LCD model, which has the main objective of instructing the AV to make Lane Changes (LCs) as required [6]. Usually, the output of the LCD model is Lane Keep (LK), Lane Change Right (LCR), and Lane Change Left (LCL), where the lateral direction control is managed [7]. The LCD model can instruct mandatory or discretionary LCs, with the former being LC triggered as a result of traffic regulations, destination, or any other factor forcing the AV to execute LC [8], [9]. An AV deploying an LCD model that does not incorporate the whereabouts of an EMV is bound to have the shortcoming of not being able to handle the aforementioned edge case. Typically, an LCD model will be developed with the objective of having to maximize its own individual reward including efficiency, comfort, and safety. In the presence of an EMV around, we expect the AV to prioritize the EMV's interest. The typical objectives a regular LCD model aspires to achieve can hinder the efficiency of the EMVs. Thus, when being approached by an EMV, the commonly known LCD models are not reliable as EMVs are not considered.

One of the first works on developing LCD models can be seen in Gipps models which are classed as rule-based models [6]. A common theme in all Gipps' and other similar models is they are known to be deterministic which provides an interpretable LCD model. The set of rules to instruct an LC, based on necessity and safety. A variant of the Gipps model, known as ''Minimizing Overall Braking Induced by Lane Changes'' (MOBIL), has been suggested by Kesting et al. [10]. MOBIL is an acceleration-based that instructs LCs upon considering safety and desirability rules. By modifying MOBIL parameters, various driving behaviors can be simulated from conservative to aggressive types of driving. Due to the rigid nature of MOBIL, they are prone to struggle and fail to perform as expected in unanticipated road situations. Another line of techniques that exist in the literature is the use of Deep Reinforcement Learning (DRL) [11] to obtain LCD models that can be used by the AV. The DRL offers a set of approximation algorithms that is capable of achieving near-optimal performance on various problems [12], such as DQN and DDQN. DRL has shown its potential in various domains such as game mastering [13], [14], robotics control [15], algorithmic trading [16], and autonomous driving [11]. Ye et al. proposed an LCD model using Proximal Policy Optimization-based DRL mainly for mandatory LCs. The study considered the safety, efficiency, and comfort of the ego-vehicle were considered. The evaluation demonstrated a high success rate in performing the required mandatory LC in heavy traffic while maintaining high safety standards. Wang et al. [17] proposed a harmonious LCD model where the welfare of neighboring vehicles is considered. This work opted to exploit DQN to obtain an LCD model that has demonstrated a high cooperative level among AVs. Further existing research works have employed DRL and have demonstrated huge potentials [18], [19], [20], [21], [22].

As regular LCD models are not designed or trained to handle the presence of an EMV, we suggest aiding the AV decision-making with an Emergency Vehicle Aware LCD (EMV-LCD). The EMV-LCD is a special type of LCD model that overrides the decisions of the main LCD model deployed by the AV, once an EMV is detected around. In other words, the EMV-LCD will take over the LCD model responsibilities as an EMV comes under the AV radar. It should be clear to the reader, that EMV-LCD is not intended to replace the main LCD model; however, it is indented to address the edge case raised since it is considered as a shortcoming of existing LCD models. In this paper, we refer to the AV using the EMV-LCD or LCD as the ego-vehicle. Shoaraee et al. [23] have discussed this edge case and proposed a DRL solution that addresses it. They used the Dueling Deep Q-networks (DDQN) to train a model which is similar to the concept of

the EMV-LCD. The objective was to produce similar human driving behavior, where the AV will give the lane to an EMV behind it. They attempted to minimize an objective function that comprised of a number of accidents, time steps taken until the EMV-LCD gives way to the EMV, frequency of leaving the road boundaries, and speed violations. The results obtained by Shoaraee et al. [23] demonstrated their proposed solution outperformed MOBIL in terms of safety and level of cooperativeness with the EMV, which indicates the DRL has the potential of handling the discussed edge case. Despite the success, the existing work is still short to be complete and contains several shortcomings, such as (1) their training and evaluation only considered when the ego-vehicle resides in the center lane, whereas far-end lanes were neglected under evaluation and (2) considering various types of EMVs such as an ambulance and police vehicles where they vary in dimensions. The contributions for this paper are listed below:

- Proposes an EMV-LCD that can be deployed by an AV where EMV interest is considered using a DRL approach. The decision-making of the proposed EMV-LCD provided a high level of safety and demonstrated evidence of being highly cooperative as it gets approached by an EMV.
- Addresses shortcomings found in [23] where our solution can maintain consistent performance over all lanes and is capable of handling different types of EMVs with varying dimensions.

This paper is organized as follows[1]: The problem statement is presented in section II. Preliminaries required for this work are covered briefly in section III. Then, we present our proposed DRL solution in section IV. Next, the training details including the simulation setup are discussed in section IV-C. In section VI, evaluation results are presented coupled with an extensive discussion. Finally, section VII concludes the paper.

II. PROBLEM STATEMENT

Typical DRL-based LCD models deployed by AVs are designed and trained to achieve efficiency, safety, and comfort for the AVs deploying it [17]. Naturally, these objectives drive the ego-vehicle to become individual-oriented, which makes the LCD model not suitable to handle the edge case when being approached by an EMV. Thus, there is a need to complement the AV decision-making with an EMV-LCD to be able to handle this edge case, where the AV cooperates with the EMV, behaving in the best interest of the EMV while maintaining safety. To our best knowledge, only a single work [23] has been found where DRL was leveraged to propose an EMV-LCD, where they have demonstrated huge potential. However, this work did contain a number of limitations and shortcomings as earlier discussed in section I. The primary aim of this work is to further explore the use of DRL in developing an EMV-LCD and address some of the existing limitations of the literature.

[1]This work is based on [24].

III. PRELIMINARIES

A. IDM

A well-known car following the model, proposed by Treiber et al. [25], is the Intelligent Driver Model (IDM) which is considered to be a rule-based model. It handles the vehicle acceleration based on the current leading vehicle. The IDM is known to be accident-free and easily interpretable which makes us avoid accidents occurring as a result of longitudinal control. At each time step, Eq. (1) is used to compute the modified acceleration.

$$a_{t+1} = a_t \left[1 - \left(\frac{v_t}{v_{exp}} \right)^{\delta} - \left(\frac{d^*}{d} \right)^2 \right] \tag{1}$$

$$d^* = d_0 + T v_t + \frac{v_t \Delta v}{2\sqrt{ab}} \tag{2}$$

The variables define the behavior of the model, and these variables are target velocity (v_{exp}), desired time headway which is the desired time gap to the leading vehicle (T), jam distance which is the required minimum distance to the leading vehicle (d_0), acceleration limits (a and b), and velocity component (δ). In addition to the model parameters, a set of other variables are required that represent the current situation. The variables are (d) distance to the front vehicle and (Δv) difference in velocity with the leading vehicle. Note that (d^*) is defined in Eq. (2) to compute the desired gap distance based on the model parameters.

B. MOBIL

For the LCD model, we make use of the rule-based model, Minimizing Overall Braking Induced by Lane Change (MOBIL) [10]. Note, MOBIL will only be used by Human-driven vehicles on the road to make LC decisions. Recall that the model checks if making LC, from the current to a specific target lane is possible and desired.

$$\tilde{a_{t_s}} \geq -b_{safe} \tag{3}$$

$$\tilde{a} - a + p\left[\tilde{a_{t_f}} - a_{t_f} + \tilde{a_{c_f}} - a_{c_f}\right] > \Delta_{th} \tag{4}$$

Several useful notations are introduced as they are used in the remaining of this subsection. (a) and ($\tilde{a}$) refers to ego-vehicle current acceleration before and after making the LC to the target lane, respectively. Regarding the following vehicle, (a_{t_f}) and ($\tilde{a_{t_f}}$) are associated with follower vehicle acceleration in the target lane before and after the LC gets executed, respectively. Similarly, the leading vehicle in the target lane is represented using a_{c_f} and $\tilde{a_{c_f}}$. MOBIL states two criteria, which both must hold true, in order for the model to instruct an LC to a particular target lane which is explained below:

- **Safety** check if as a result of this ego-vehicle making an LC, the new following vehicle will need to make an abrupt stopping with a high deceleration value. Eq. (3) defines the inequality that must hold for a specific deceleration constant b_{safe}.
- **Incentive** determine if executing an LC has the potential to improve the local traffic situation around the vehicle.

The inequality in Eq. (4) must hold true for the current vehicle to have the incentive to execute an LC. In the LHS, the first term indicates the utility of the ego-vehicle, and the second term denotes the utility of both followers on the current and target lane. p is called the politeness factor, which is a constant that states the weight significance of neighbors' utility if an LC is made. In RHS, Δ_{th} is the switching threshold, which sets the minimum local traffic utility for an LC to be desired.

C. DQN

A subfield of machine is known as Reinforcement Learning (RL) where an agent is trained to make certain actions in a specific environment to maximize its reward over time [26]. The reward signal received following each action is defined in the reward formula that is defined in terms of the objectives the agent must achieve. Therefore, the agent will utilize the reward formula to learn how to arrive at the goal state or maximize its total cumulative reward over time. Most RL algorithms target obtaining a policy that always selects the action that maximizes its expected accumulative reward given the current agent's state. Such a policy is known as optimal policy as it always selects the optimal action in every state, providing the optimal performance for the agent. Formally, a policy is mapping between states and action $\pi(a|s)$ where a is the action and s is the current state for the agent, with π^* being the optimal policy. Typically, the environment is modeled as a Markov Decision Process (MDP), which is composed of state space, action space, and reward formula. In order to obtain the optimal policy, we need to have the true values of the Q-function as seen Eq. 5. A well-known RL algorithm called Q-learning is capable of obtaining the optimal policy [27]. Q-learning works by gradually converging to true values of the Q-function that is associated with the optimal policy. However, if the tackled problem is large or infinite in terms of the state or action space, the huge required computation time and memory makes Q-learning fail to be a reasonable algorithm choice. In such problems, we opt to use approximation methods, such as the ones under the umbrella of Deep Reinforcement Learning (DRL), where it is possible to obtain near-optimal policies [12]. Deep Q-networks (DQN) is an example of a DRL algorithm, that is capable of obtaining near-optimal policies by approximating the values of the Q-function [28].

$$\pi(s)^* = argmax_a Q(s, a) \tag{5}$$

IV. PROPOSED SOLUTION

The core of the EMV-LCD purpose is the decision-making problem, which must be formulated to be able to proceed with the solving task where an agent is trained. In this work, we chose to formulate the decision-making problem as a Markov Decision Process (MDP), which is composed of state space, action space, and reward formula. Formally put, an MDP is defined as tuple (S, A, R, T, γ). S defines the state space, A defines the action space, T is a transition probability function, and R is a reward formula. For every time step t, s_t is observed by the agent according to the state definition, where the agent will have a set of possible actions at its disposal, $a_t \in A$, executed based on T. The state defines the agent's observation, based on the current state of the environment, which must be tailored to the current problem at hand.

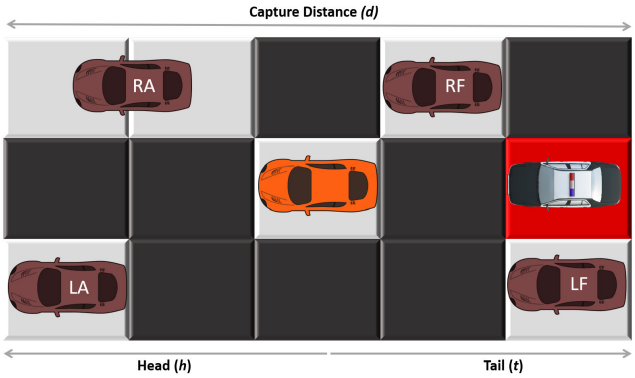

FIGURE 3. Snapshot constructed based on the road situation. The black, white, and red cells correspond to 0, 1, and 2, respectively. Additionally, the name on top of each of HVs (Red vehicles) defines the current relative direction to ego-vehicle (Orange vehicle).

A. STATE

The components of the state definition are tailored according to the problem EMV-LCD tries to solve. The components are snapshots, relative speed, and EMV-info. Each of the components is explained below in detail:

- **Snaphot** is a binary 2D array or an occupancy grid depicting the current road situation around the vehicle. The 2D array aims to represent the grid around the vehicle, having the vehicle centered along the y-direction. Fig. 3 demonstrates the construction of the grid from the current road situation. Cells are set to one (white), indicating the existence of the vehicle, otherwise set to 0 (black). A cell occupied by an EMV is given a value of 2 (red). Note that the top and bottom rows in the grid, represent neighboring lanes, relative to the ego-vehicle current lane. The grid construction requires defining a set of parameters that defines the level of detail needed. The parameters are cell size (dim), capture distance (d), ego-vehicle head (h), and tail (t) distances (see Fig. 3). A snapshot at t, is denoted as M_t.
- **Relative speed** is a four elements vector composed of relative speeds with respect to all neighbor vehicles. The considered vehicles are ahead and behind in both neighboring lanes which are Left Follower (LF), Left Ahead (LA), Right Follower (RF), and Right Ahead (RA). For further clarity, refer to Fig. 3. The relative speed is defined by taking the difference in speed between the considered vehicle and the ego-vehicle, as seen in Eqs. (6)-(9). $RS_t(A)$ denotes the relative speed with respect to vehicle A, and $V_t(A)$ returns vehicle A's current speed at time step t. The entire

TABLE 1. Function descriptions.

Function	Description
LC(e)	Return *True* if the vehicle e is currently executing a lane change. Otherwise, return *False*.
DET(e)	Return *True* if the EMV is around within 70 m range around the vehicle e. Otherwise, return *False*.
LAN(e)	Return the *index* of the lane resided by vehicle e. As we have three lanes, lanes indices are 1,2, and 3.
TAR(e)	As the vehicle makes an LC, it sets a specific target lane that it aims to reach. This function returns this *index* of the current target lane for vehicle e.

relative speed vector is denoted as RS_t and formed by $[RS_t(LF),\ RS_t(LA),\ RS_t(RF),\ RS_t(RA)]$. Note that the vehicle that is not within 30 m around the ego-vehicle is disregarded.

$$RS_t(LF) = V_t(LF) - V_t(EGO) \tag{6}$$
$$RS_t(LA) = V_t(LA) - V_t(EGO) \tag{7}$$
$$RS_t(RF) = V_t(RF) - V_t(EGO) \tag{8}$$
$$RS_t(RA) = V_t(RA) - V_t(EGO) \tag{9}$$

- **EMV-Info** denoted as E_t, which attempts to capture the state of the ego-vehicle specifically in relation to the EMV. The EMV-Info is composed of two elements, namely, E_{det_t} and E_{same_t} which are binary values defined in Eqs. (10) and (11), respectively. Table 1 describes the function definitions used in the equations. E_{det_t} checks if the EMV is detected within the detection range, whereas E_{same_t} checks if the detected EMV is located in the same lane as the ego-vehicle.

This finishes defining the entire information needed to define the full state. Note that RS_t and E_t are formed of a number of elements according to the state definition discussed in this section.

$$E_{det_t} = \begin{cases} 1, & \textbf{DET}(EMV) \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

$$E_{same_t} = \begin{cases} 1, & \textbf{DET}(EMV) \ \wedge\ \textbf{LAN}(EGO) = \textbf{LAN}(EMV) \\ 0, & \text{otherwise} \end{cases} \tag{11}$$

B. ACTION

At any t, the agent has a set of three actions to instruct the ego-vehicle to execute. Namely, $A = [LK, LCL, LCR]$ representing lane keep, lane change left, and lane change right. As we select $a_t \in A$, the high-level decision is passed down to the lower-level controller for the execution phase. The a_t chosen by the agent will be executed over a specific horizon time called the execution time (*step*). In this work, $step = 1s$ is used.

C. REWARD

The design of the reward formula is a crucial step in this work and in general in DRL solutions. Thus, it must be defined carefully, to both contribute to generating reliable EMV-LCD behavior and speed up the training time. Upon formulating the reward formula, a number of desired objectives were targeted which are discussed below:

- It is desired that an ego-vehicle that happens to be followed by an EMV should minimize the duration of time it takes until it prioritizes the EMV by giving up the lane to the EMV. This objective aims to contribute to improving the efficiency of the EMV, by being able to drive at its maximum speed.
- While we want the ego-vehicle to cooperate with the EMV, there is a safety concern for the ego-vehicle and the neighboring vehicles cannot be compromised. Therefore, the LCs instructed by the EMV-LCD should consider the safety of the lane changes.
- Recall that the EMV-LCD is only active while the EMV is around the ego-vehicle. Thus, we also expect the EMV-LCD to avoid executing LCs that would block the path of the EMV as explained in Fig. 2.
- We also expect the EMV-LCD to consider the boundaries of the road such that it does not leave the ego-vehicle to become outside the road boundaries. For instance, the EMV-LCD should not instruct the ego-vehicle to make the right LC in the situation depicted in Fig. 2, as the ego-vehicle is currently residing in the far right lane.

$$r_t = \begin{cases} -300 & \text{collision} \\ r_{lc_t} + r_{emv_t} & \text{otherwise} \end{cases} \tag{12}$$

$$r_{emv_t} = \begin{cases} -\alpha_1, & \textbf{DET}(EMV) \wedge \textbf{LAN}(EGO) = \textbf{LAN}(EMV) \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

$$r_{lc_t} = \begin{cases} -\alpha_2 & \textbf{LC}(EGO) \wedge \neg(\textbf{DET}(EMV)) \\ -\alpha_3 & \textbf{LC}(EGO) \wedge \textbf{DET}(EMV) \wedge \textbf{TAR}(EGO) \\ & = \textbf{LAN}(EMV) \\ -\alpha_4 & \textbf{LC}(EGO) \wedge \textbf{DET}(EMV) \wedge \textbf{TAR}(EGO) \\ & \neq \textbf{LAN}(EMV) \\ -\alpha_5 & \textbf{LC}(EGO) \wedge 0 < \textbf{LAN}(EGO) < 4 \\ 0, & \text{otherwise} \end{cases} \tag{14}$$

Eqs. (12)-(14) define the reward formula used in this work. Regarding the functions used in the reward formula, **LC** checks if the vehicle is making an LC, **DET** checks if the EMV is detected by the vehicle, **LAN** return the lane the vehicle is currently residing, and **TAR** return the current target lane for the vehicle. Refer to Table 1 for the full definitions of the functions used in the reward formula. In Eq. (12), if the ego-vehicle gets involved in a collision with other vehicles on the road, a penalty of 300 is returned. Otherwise, the reward is composed of two components, which are the

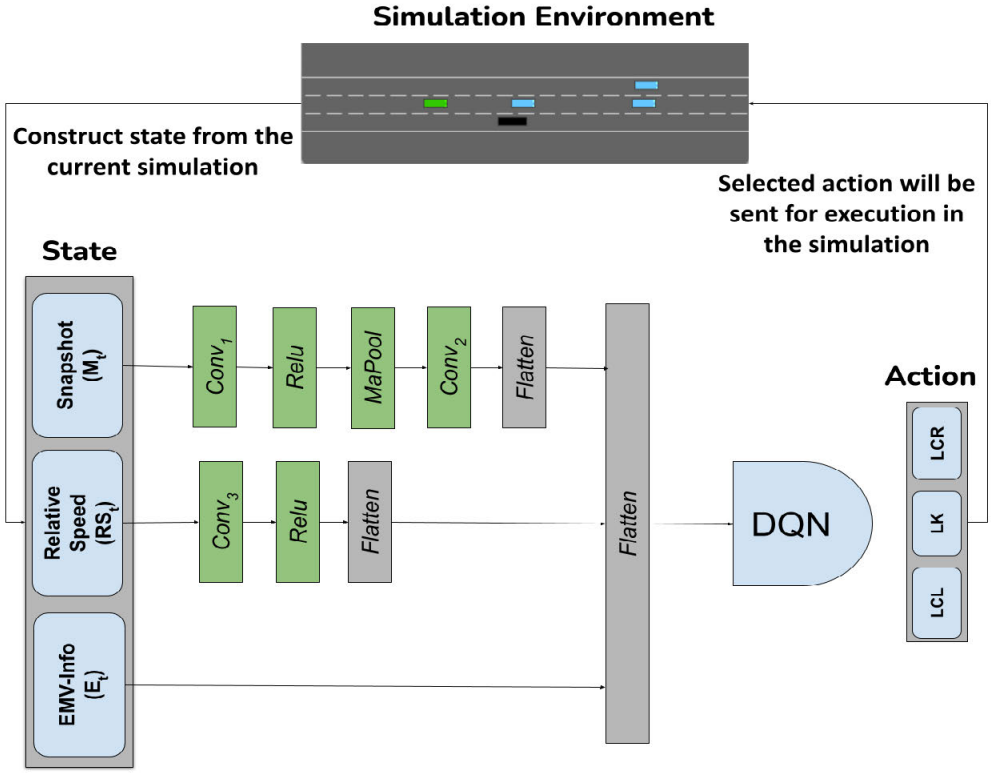

FIGURE 4. Training model architecture.

EMV reward r_{emv_t} and the lane change reward r_{lc_t}. Eq. (13) defines r_{emv_t}, which is the reward associated with the position of the ego-vehicle in regard to an approaching EMV. In the case the EMV is noticed to be within the detection range and the ego-vehicle happens to be driving in the same lane as the EMV, the agent will be penalized by α_1. However, if the EMV is not within the detection area of the ego-vehicle, r_{emv_t} is set to zero. The second component of the reward is r_{lc_t} which is defined in Eq. (14). It evaluates the impact of any LC executed by the ego-vehicle, with the intention to push the ego-vehicle to make LCs only when it is required and safe, while being cooperative with an approaching EMV. As seen in Eq. (14), there are multiple conditions where the agent is penalized with some α_i. The conditions are: (1) EMV is not detected, and the ego-vehicle is executing an LC, (2) EMV is detected, and the ego-vehicle is changing lanes where the EMV currently resides, (3) EMV is detected, and the ego-vehicle is making an LC to a lane that EMV is not currently at, and (4) if the agent instructs the ego-vehicle to make an LC that will lead the vehicle to leave the road boundaries regardless if the EMV is detected or not. The consequence of this reward formula is that the ego-vehicle being in different lane relative to the detected EMV is encouraged, as it will not return any penalties.

V. TRAINING

The details of how the training step was conducted are explained in this section. This section presents the training model architecture used, the simulation platform and setup adopted, and discusses the parameters used in training.

A. MODEL ARCHITECTURE

The model used during the training can be seen in detail in Fig. 4. The model was selected following experimenting with various model architectures and we selected the one that produced the maximum performance. In every training step, the state is constructed as explained in section IV. Whenever possible, features in the constructed state are normalized using Max-Min normalization before starting the feature extraction step. Then, a feature extraction step is done on both RS_t and M_t, using a set of convolution layers. As seen in Fig. 4, M_t goes through two convolution layers, whereas RS_t has a single convolution layer. The details of each component used in the feature extractor can be found in Table 2. Following the extraction step, the extracted features will be combined using a flatten layer. In turn, they would be fed into the DQN model which is expected to output a specific action. Over the training steps, it is expected that the DQN model decision-making will improve as the DQN weights will be tuned to optimize the performance.

B. SIMULATION SETUP

In this subsection, we present the environment configuration and setup set under the training stage. For the simulation task, this work has utilized an open-source implementation,

TABLE 2. Feature extraction layers.

Component	kernel size	# of filters	stride	padding
$Conv_1$	(2,2)	8	1	0
$Conv_2$	(2,1)	4	1	0
$MaxPool$	(2,1)	n/a	1	0
$Conv_3$	(1,2)	8	1	0

known as Highway-env [29]. The simulation tool facilitates easy setup for custom observation, action definition, and easy modification to vehicle macroscopic and microscopic behaviors for all vehicles on the highway. Each training episode includes a single EMV, a single ego-vehicle, and set of HVs. The considered road environment is a 3-lane highway, with each lane having a length of 4 m. A training episode terminates in two conditions which are either reaching the maximum episode length or in the case the EMV passes the ego-vehicle with a distance of 50 m. The maximum length of an episode is set to 30 time steps or seconds. These limits are set with the intention of exposing the agent to as many different road scenarios as possible. During training, the agent will encounter two different episodes with each having a particular desired behavior. *Eps-1* is intended to teach the agent to give the way when it is being followed by an EMV. Whereas *Eps-2* primary purpose is to avoid blocking the way on the EMV and discourage making LCs when the EMV is detected. The episode type is selected from uniform distribution, with weights of 0.85 and 0.15 set for *Eps-1* and *Eps-2*, respectively. *Eps-2* is set a lower probability to be selected due to the less complexity and difficulty of learning the desired behavior. Fig. 1 represents how would *Eps-1* start, whereas the *Eps-2* starting situation is depicted in Fig. 2.

EMV can be either a police vehicle or an ambulance vehicle where they vary in dimensions as defined in Table 3. The EMV starts behind all other vehicles along the longitudinal direction with the desired speed of 150 km/hr which is the highest speed aspired by any vehicle in the road portraying an emergency situation where the EMV is in operation. In both episode types, the EMV will start at lane L_i, evenly chosen among the set of lanes. As the EMV expects other vehicles to give up the lane, the EMV will not be making any LCs and remain located in L_i during the entire length of the episode. Regarding the ego-vehicle, its desired speed is chosen from the uniform distribution defined by 125 km/hr and 140 km/hr. In *Eps-1*, the ego-vehicle will start at lane L_i to be located in the same lane as EMV with a specific gap distance ahead of it. The gap distance is selected from a uniform distribution, ranging from 10 m and 75 m. Whereas in *Eps-2*, the ego-vehicle starts at lane L_j given that $i \neq j$, which implies that the ego-vehicle starts at a different lane compared to the EMV. Additionally, the episode will include a set of HVs on the road with their quantity ranging from 4 to 8, selected at beginning of every training episode. Each HV in the simulation will be set a specific driving behavior as human driving attitudes varies among drivers. All HVs can be located in any of the lanes apart from L_i. To expose the trained agent to various types of road situations under training, several parameters are selected from distributions such as gap distance, desired speed, number of HVs, and human driving styles.

TABLE 3. Dimensions set associated with each vehicle type.

Type	Width	Length
Ambulance	2.5 m	8 m
Police	2 m	6 m
HV/ego-vehicle	2 m	5 m

TABLE 4. DQN Parameters for EMV-LCD.

Parameter	Value
Learning Rate	0.005
Discount factor	0.99
Mini batch size	32
Maximum Memory size	15000
Update Target Network	5

TABLE 5. Agent state parameters for EMV-LCD.

Parameter	Value
d	40 m
dim	(20,3)
h	20 m
t	20 m

TABLE 6. Agent reward parameters for EMV-LCD.

Parameter	Value
α_1	3
α_2	3
α_3	60
α_4	3
α_5	10

C. TRAINING SETUP

This work opted to use the DRL model known as DQN, since the tackled problem has an infinite state space. Following optimizing the hyperparameters of the model, the values selected can be found in Table 4. Additionally, the agent includes optimizing a number of parameters found in the state and reward. The agent parameters of this work can be found in tables 5 and 6. The model was trained for 100k training time steps. For the training task, the DQN implementation provided by the DRL library SB-3 [30] was leveraged.

VI. EVALUATION

In this section, we present the performed evaluation experiments in this work to assess the applicability and safety of the proposed solution. The main purpose of the evaluation is to see how the ego-vehicle deploying the proposed EMV-LCD performs when being approached by an EMV. A similar simulation setup is used under training as it is sufficient to produce a broad range of road scenarios. The only major difference is that we increase the maximum length of an episode to 60 seconds. Under evaluation, the episode is

defined by a specific expected speed for the ego-vehicle. This allows evaluating the EMV-LCD when being approached by an EMV against different desired ego-vehicle behavior. The considered desired speeds are 125, 133, and 140, in km/hr.

This work has opted to perform the evaluation into two stages. In the first stage, experiments were performed with the initial lane for the ego-vehicle is chosen randomly. Whereas, in the second stage, an additional parameter is added into the episode definition which specifies the starting lane of the ego-vehicle. The second stage is intended to evaluate how the EMV-LCD performance varies, among all three lanes. This can give us further insight into the performance of the considered EMV-LCD. We refer to the first stage as *Random Lane*, whereas, the second stage is called *Specific Lane*. The following subsection proceeds with the used metrics and benchmarks for comparison purposes, followed by presenting the obtained results.

A. METRICS

Several metrics were chosen to assess this work and compare it against benchmarks. Each metric selection is associated with a particular property that we desire the EMV-LCD to possess. The metrics used in this work are as follows:

- **Collision-free episodes:** The percentage of evaluation episodes that are free of any collision involving any of vehicles on the road, including HVs, EMV, and the ego-vehicle. This metric allows evaluating the safety of the EMV-LCD which is an important feature the EMV-LCD must provide. The metric will be evaluated using *Eps-1* configuration.
- **Steps-Sharing:** This counts the number of seconds it takes the ego-vehicle to give the way to the EMV. To be further specific, the counts start from the beginning of the episode until the EMV passes the ego-vehicle, with a distance of 50 m, which can only occur if the ego-vehicle gives the way to the EMV. In the case the ego-vehicle was not able to give up the lane to the EMV, the corresponding value should be the maximum length of the episode which is 60 seconds. The metric will evaluate how cooperative the EMV-LCD and the level of emergency awareness when being followed by an active EMV. As with the previous metric, *Eps-1* is used.
- **Blocks-free episodes:** The percentage of evaluation episodes that are free of any blocks done by the ego-vehicle. A block occurs when EMV-LCD instructs an LC that ends ups blocking the way ahead of the EMV which contributes to delaying the arrival time for the EMV. This also allows for capturing how cooperative the EMV-LCD is when being approached by an EMV in a different lane. Naturally, these metrics were evaluated using the *Eps-2* setup.

B. BENCHMARKS

To demonstrate the effectiveness of the proposed solution, we compare the results of the proposed solution in this work with two different benchmarks. Recall that we are only evaluating the lateral control of the ego-vehicle, when being approached by an EMV. Therefore, to ensure a fair comparison, all considered benchmarks and the solution proposed will be using a rule-based car-following model with identical parameter values.

- **Detect-LC** This LCD is naive and simple, where it always makes an LC the moment it detects an approaching EMV, while being in the same lane. Despite it being extremely simple, it is useful to see how much improvement the proposed solution offers over it.
- **MOBIL** As done by [23], the proposed solution is compared against a well-known rule-based LCD model in the literature, with parameters seen in Table 7. The defined LCD model parameters simulate the maximum level of politeness with neighboring vehicles. Consequently, when being approached by an EMV, the parameters defined will push the ego-vehicle to cooperate with EMV. Additionally, the maximum braking that can be imposed on other vehicles is high to produce as aggressive LC as possible when being approached by an EMV.

TABLE 7. MOBIL parameters.

Parameter	Value
Politeness	1
Maximum braking imposed	4 m/s^2
Gain threshold	0.1 m/s^2

TABLE 8. *Random Lane* collision-free percentage results.

Desired Speed(km/hr)	Our solution	Detect-LC	MOBIL
125	100%	61%	97.5%
133	99.5%	61%	97.5%
140	98.5%	57%	97.5%
Average	**99.3%**	**59%**	**97.5%**

TABLE 9. *Random Lane* Steps-Sharing results.

Desired Speed(km/hr)	Our solution	MOBIL
125	14.02s	39.23s
133	13.57s	44.54s
140	13.91s	54.91s
Average	**13.83**s	**46.22**s

TABLE 10. *Random Lane* block-free percentage results.

Desired Speed(km/hr)	Our solution	MOBIL
125	100%	63%
133	100%	72%
140	100%	87.5%
Average	**100%**	**74.1%**

C. RESULTS

This subsection presents the results obtained for both evaluation stages. Note that 200 episodes were run to generate the evaluation results.

TABLE 11. ***Specific Lane*** **steps-sharing.**

Desired Speed(km/hr)	Left Lane		Center Lane		Right Lane	
	Our Solution	*MOBIL*	*Our Solution*	*MOBIL*	*Our Solution*	*MOBIL*
125	13.94s	41.33s	12.86s	32.85s	15.39s	41.07s
133	13.3s	47.89s	12.61s	41.28s	15.48s	47.61s
140	13.29s	56.05s	13.03s	47.98s	14.79s	55.42s
Average	**13.51**s	**48.89**s	**12.83**s	**40.7**s	**15.22**s	**48.03**s

1) RANDOM LANE

Tables 8, 9, and 10 present the results obtained during the first stage of evaluation. Table 8 compares our solution against benchmarks in terms of the collision-free evaluation episodes of *Eps-1* type. The proposed solution scored an average percentage of 99.3%. Furthermore, the proposed solution scored 100% when the desired speed was 125 km/hr, whereas the lowest percentage was obtained on 140 km/hr with 98.5%. Moving to MOBIL, the achieved percentages were all 97.5% for all desired speeds. Detect-LC results were below MOBIL and the proposed solution standards, as Detect-LC ranged from 57% to 61%. Table 9 compares the proposed solution against MOBIL in terms of Steps-Sharing, by averaging the results of 200 episodes of *Eps-1* type. The difference in average score is 32.39s to the advantage of the proposed solution. The proposed solution results range from 13.57s to 14.02s, whereas MOBIL scores range from 39.23s to 54.91s with an increasing trend noticed as the speed increases. Finally, Table 10 compares the proposed solution against MOBIL in terms of the percentage of block-free *Eps-2* type episodes. The proposed solution obtained 100% for all desired speeds, whereas MOBIL scores were all below 88% for all scenarios. Regarding MOBIL, the obtained results ranged from 63% to 87.5%, with an increasing trend noticed as the speed increases.

2) SPECIFIC LANE

Figs. 5-6 and Table 11, present the results obtained following *Specific Lane* experiments. Regarding the figures, each speed on the x-axis is associated with three pairs of bars. The pair located on the left is associated with results obtained on the left-lane, the center pair corresponds to the center-lane, and the right pair relates to the right-lane. In all of the considered figures and table, the solution proposed and MOBIL are compared considering varying ego-vehicle desired speeds and starting lanes. Fig. 5 depict the result with collision-free being the considered metric. The proposed solution scored the highest percentage when used in the left-lane with an average percentage of 99.8%. In the other far end of the road, the proposed solution obtains a 98.16% average, with the percentage ranging from 97% to 99.5%. The lowest average percentage occurred in the center-lane, with an average of 97.83%. Comparing the proposed solution with MOBIL, the maximum average difference occurred in the left-lane with a magnitude of 1.47%. Whereas 0.17% percentage difference is seen in the center-lane, which is the least average obtained among all lanes. Table 11 present the results associated with the Steps-sharing metric. Both solutions obtained the best Steps-sharing in the center-lane, with 12.83s and 40.7s associated with the proposed solution and MOBIL, respectively. MOBIL scores in both far ends of the road were in the range of 48s. However, the proposed solution varied in result between the two far-ends with 13.51s and 15.22s scored for the left-lane and right-lane, respectively. Comparing the solutions, the difference in average is significant ranging from 27.87s to 35.5s to the advantage of the proposed solution. Fig. 6 show the results obtained regarding the Blocks-free metric. In all scenarios considered, the proposed solution obtains a 100% percentage. Regarding MOBIL, the average percentage ranges from 70.5% to 75.16%.

D. DISCUSSIONS

The safety of the proposed EMV-LCD is evaluated in terms of the results obtained in Table 8, where the proposed solution obtained similar results when compared to MOBIL percentages. This was further bolstered by the result seen in Fig. 5, where high safety is maintained among the lanes. From the same table, one is able to observe that our proposed solution is the safest when used in the left-lane. Furthermore, the proposed solution performed the worst when used in the center-lane. However, as evident in the results, the difference is insignificant. The authors attribute this difference to the set of traffic situations observed and experienced during the training stage. Overall, it can be concluded that MOBIL and the proposed solution maintain similar levels of safety. Additionally, the obtained results by the proposed solution are very close to 100% which indicates a high level of safety is promised to be achieved. The percentages scored by Detect-LC demonstrated that executing an LC immediately as detecting the EMV produced unacceptable levels of safety. This illustrates that neglecting the safety of the LC when trying to cooperate with EMVs can be damaging.

The results seen in Tables 9 and 11 show that the proposed solution is superior to MOBIL in terms of the duration spent sharing the same lane with EMV. In both tables, the difference is noticeable and significant which indicates that the proposed solution will react sooner than MOBIL while ensuring the safety of LC being executed, as seen in Table 8. From the results seen in Table 11, one can notice that for both solutions, less amount of time is taken to give up the lane in the center-lane. This is a natural outcome of having two possible lanes when the ego-vehicle is in the center-lane, as opposed to one when residing at either end of the road. Looking at Table 10, it can be concluded that the proposed solution cooperates

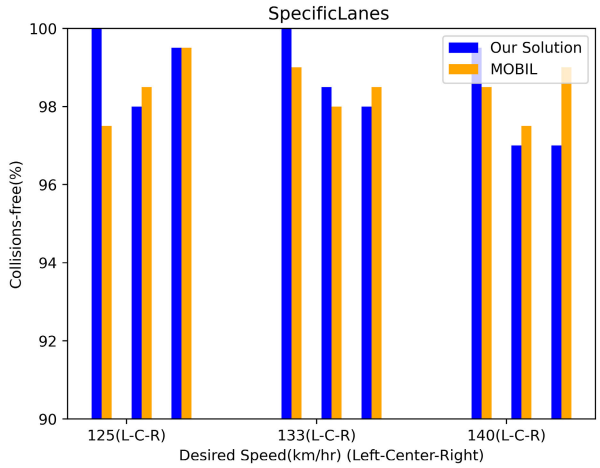

FIGURE 5. Specific lane collision-free.

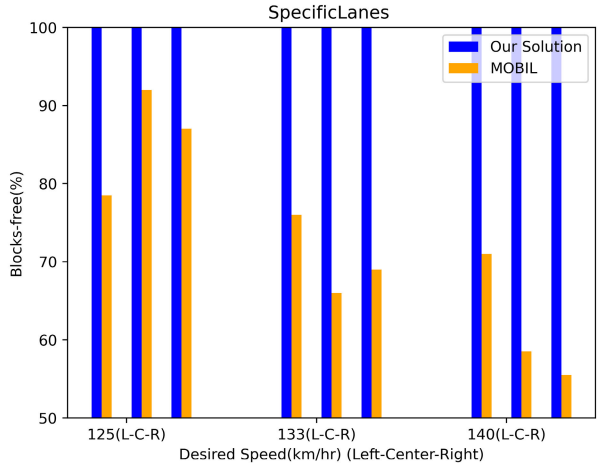

FIGURE 6. Specific lane blocks-free.

better when it gets approached by an EMV while being in a different lane when compared to MOBIL. This fact is also maintained for the proposed solution across all lanes, as evident in Fig. 6. In fact, the scores obtained by MOBIL were certainly expected, as the nature of MOBIL is inconsiderate to an approaching EMV. This bolsters the claim that we need to complement existing LCD models, which can intervene as the ego-vehicle gets approached by an EMV.

VII. CONCLUSION

In this paper, we proposed an EMV-LCD using techniques from the field of DRL. The objective was to develop an EMV-LCD that is deployed by an AV and used only when being approached by an EMV. To our knowledge, this is one of the pioneering works that has used DRL to specifically tackle the discussed scenarios, addressing potential limitations of existing work. The proposed EMV-LCD was compared against MOBIL. The results demonstrated that an ego-vehicle deploying the proposed EMV-LCD will be far more cooperative than using MOBIL as evident by the significantly less time it takes the AV to leave the lane for the EMV while achieving high safety levels. Additionally, the EMV-LCD avoids blocking the path of the EMV. This allows concluding that if the proposed EMV-LCD was to complement a MOBIL or any other well-known LCD, it will produce an AV that is capable of cooperating with an approaching EMV.

ACKNOWLEDGMENT

This is to acknowledge the value of NEP 3.0 in supporting the leadership of Dr. Ameena.

REFERENCES

[1] P. Mallozzi, P. Pelliccione, A. Knauss, C. Berger, and N. Mohammadiha, "Autonomous vehicles: State of the art, future trends, and challenges," in *Automotive Systems and Software Engineering*, 2019, pp. 347–367.
[2] E. Yurtsever, J. Lambert, A. Carballo, and K. Takeda, "A survey of autonomous driving: Common practices and emerging technologies," *IEEE Access*, vol. 8, pp. 58443–58469, 2020.
[3] L. Zhong and Y. Chen, "A novel real-time traffic signal control strategy for emergency vehicles," *IEEE Access*, vol. 10, pp. 19481–19492, 2022.
[4] (Nov. 22, 2022). *NHS Ambulance Services*. [Online]. Available: https://www.nao.org.uk/wp-content/uploads/2017/01/NHS-Ambulance-Services.pdf
[5] J. P. De Lone, "Emergency response time in Dubai improving," Gulf News, Sep. 2015. [Online]. Available: https://gulfnews.com/uae/
[6] P. G. Gipps, "A model for the structure of lane-changing decisions," *Transp. Res. B, Methodol.*, vol. 20, no. 5, pp. 403–414, Oct. 1986.
[7] E. Leurent, "A survey of state-action representations for autonomous driving," Tech. Rep. HAL-01908175, 2018.
[8] Y. Ali, Z. Zheng, M. M. Haque, M. Yildirimoglu, and S. Washington, "Understanding the discretionary lane-changing behaviour in the connected environment," *Accident Anal. Prevention*, vol. 137, Mar. 2020, Art. no. 105463.
[9] Z. Zheng, "Recent developments and research needs in modeling lane changing," *Transp. Res. B, Methodol.*, vol. 60, pp. 16–32, Feb. 2014.
[10] A. Kesting, M. Treiber, and D. Helbing, "General lane-changing model MOBIL for car-following models," *Transp. Res. Rec.*, vol. 1999, pp. 86–94, Jan. 2007.
[11] S. Aradi, "Survey of deep reinforcement learning for motion planning of autonomous vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 2, pp. 740–759, Feb. 2022.
[12] K. Arulkumaran, M. P. Deisenroth, M. Brundage, and A. A. Bharath, "Deep reinforcement learning: A brief survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26–38, Nov. 2017.
[13] V. Mnih et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
[14] T. Bonjour, M. Haliem, A. Alsalem, S. Thomas, H. Li, V. Aggarwal, M. Kejriwal, and B. Bhargava, "Decision making in monopoly using a hybrid deep reinforcement learning approach," *IEEE Trans. Emerg. Topics Comput. Intell.*, vol. 6, no. 6, pp. 1335–1344, Dec. 2022.
[15] H. Nguyen and H. La, "Review of deep reinforcement learning for robot manipulation," in *Proc. 3rd IEEE Int. Conf. Robot. Comput. (IRC)*, Feb. 2019, pp. 590–595.
[16] T. Théate and D. Ernst, "An application of deep reinforcement learning to algorithmic trading," *Expert Syst. Appl.*, vol. 173, Jul. 2021, Art. no. 114632.
[17] G. Wang, J. Hu, Z. Li, and L. Li, "Harmonious lane changing via deep reinforcement learning," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 5, pp. 4642–4650, May 2022.
[18] C.-J. Hoel, K. Wolff, and L. Laine, "Automated speed and lane change decision making using deep reinforcement learning," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2148–2155.
[19] S. Zhang, H. Peng, S. Nageshrao, and E. Tseng, "Discretionary lane change decision making using reinforcement learning with model-based exploration," in *Proc. 18th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2019, pp. 844–850.
[20] B. Mirchevska, C. Pek, M. Werling, M. Althoff, and J. Boedecker, "High-level decision making for safe and reasonable autonomous lane changing using reinforcement learning," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 2156–2162.

[21] J. Liao, T. Liu, X. Tang, X. Mu, B. Huang, and D. Cao, "Decision-making strategy on highway for autonomous vehicles using deep reinforcement learning," *IEEE Access*, vol. 8, pp. 177804–177814, 2020.
[22] C.-J. Hoel, K. Driggs-Campbell, K. Wolff, L. Laine, and M. J. Kochenderfer, "Combining planning and deep reinforcement learning in tactical decision making for autonomous driving," *IEEE Trans. Intell. Vehicles*, vol. 5, no. 2, pp. 294–305, Jun. 2020.
[23] H. Shoaraee, L. Chen, and F. Jiang, "Decision-making of an autonomous vehicle when approached by an emergency vehicle using deep reinforcement learning," in *Proc. IEEE Int. Conf. Dependable, Autonomic Secure Comput., Int. Conf. Pervasive Intell. Comput., Int. Conf. Cloud Big Data Comput., Int. Conf. Cyber Sci. Technol. Congr.*, Oct. 2021, pp. 185–191.
[24] A. Alzubaidi, "Rain-aware lane change decision model for autonomous vehicles using deep reinforcement learning," M.S. thesis, Dept. Elect. Eng. Comput. Sci., Khalifa Univ., Abu Dhabi, United Arab Emirates, 2022.
[25] M. Treiber, A. Hennecke, and D. Helbing, "Congested traffic states in empirical observations and microscopic simulations," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 62, no. 2, p. 1805, 2000.
[26] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
[27] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, 1992.
[28] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing Atari with deep reinforcement learning," 2013, *arXiv:1312.5602*.
[29] E. Leurent. (2018). *An Environment for Autonomous Driving Decision-Making*. [Online]. Available: https://github.com/eleurent/highway-env
[30] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, "Stable-baselines3: Reliable reinforcement learning implementations," *J. Mach. Learn. Res.*, vol. 22, no. 268, pp. 1–8, 2021.

AHMED ALZUBAIDI received the B.Sc. degree in computer science from the University of Southampton, U.K., in 2020. He is currently pursuing the master's degree in computer science with Khalifa University, United Arab Emirates. His master's dissertation was on the use of reinforcement learning in autonomous vehicles decision making.

AMEENA SAAD AL SUMAITI (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from United Arab Emirates University, Abu Dhabi, United Arab Emirates, in 2008, and the M.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2010 and 2015, respectively. She was a Visiting Assistant Professor with the Massachusetts Institute of Technology, Cambridge, MA, USA, in 2017. She is currently an Associate Professor with the Advanced Power and Energy Center and the Department of Electrical Engineering and Computer Science, Khalifa University, Abu Dhabi. She is classed as one of the Top 2% scientists in the world. Her research interests include power systems, stochastic process, intelligent systems, intelligent transportation and autonomous vehicles, energy economics, and energy policy.

YOUNG-JI BYON was born in Seoul, South Korea, in 1979. He moved to Toronto, ON, Canada, in 1994. He received the B.A.Sc., M.A.Sc., and Ph.D. degrees from the University of Toronto, Toronto, in 2003, 2005, and 2011, respectively. From 2009 to 2010, he was a Visiting Researcher with the University of Chile, Santiago, Chile. From 2010 to 2011, he was a Postdoctoral Research Fellow with the University of Calgary, Calgary, AB, Canada. He is currently an Associate Chair with the Department of Civil Infrastructure and Environmental Engineering, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates.

KHALIFA AL HOSANI (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Notre Dame, Notre Dame, IN, USA, in 2005 and 2007, respectively, and the Ph.D. degree in electrical and computer engineering from The Ohio State University, Columbus, OH, USA, in 2011. He is currently an Associate Professor with the Department of Electrical and Computer Engineering, Khalifa University, Abu Dhabi, United Arab Emirates. He is also the Co-Founder of the Power Electronics and Advanced Sustainable Energy Center Laboratory, ADNOC Research and Innovation Center, Abu Dhabi. His research interests include a wide range of topics, including nonlinear control, sliding mode control, the control of power electronics, power systems stability and control, renewable energy systems modeling and control, smart grid, microgrid and distributed generation, and the application of control theory to oil and gas applications.