RESEARCH ARTICLE

Reinforcement Learning-Based E-Scooter Energy Minimization Using Optimized Speed-Route Selection

KARIM ABOELENEEN, NIZAR ZORBA, (Senior Member, IEEE), AND AHMED M. MASSOUD, (Senior Member, IEEE)
Electrical Engineering Department, College of Engineering, Qatar University, Doha, Qatar

Corresponding author: Nizar Zorba (nizarz@qu.edu.qa)

ABSTRACT In the evolving urban transportation, the emergence of Micro-Mobility (MM), symbolized by Electric Scooters (ESs), has become a pivotal response to private automobiles' environmental and logistical challenges. However, the limited battery capacity of ESs presents a challenge in realizing their full potential. This paper addresses the problem of optimizing energy consumption in ESs by jointly considering path and speed selection all while considering user dissatisfaction levels. Our approach considers two types of ESs, one with regenerative braking (i.e., able to recharge the battery from kinetic energy of movement) and the other without regenerative braking. In order to build a realistic environment, we considered dynamic factors such as time-varying road congestion, road conditions, and ambient temperature. We considered a comprehensive energy consumption model for the ES that includes parameters such as rolling resistance, air friction, road gradient, auxiliary power and ambient temperature influence. Moreover, we introduced a user dissatisfaction model that accounts for traffic conditions, congestion, and ambient temperature to enhance the user experience. The optimization problem was then formulated and solved with Deep Reinforcement Learning (DRL-DQN) approach considering the time-varying environment, road-specific parameters (i.e., road angle, road shading, road speed limit, and road condition), and user dissatisfaction levels. The DRL approach was designed to make timely and context-aware decisions the minimize the energy consumption of the ES. Rigorous validation and comprehensive testing demonstrate the effectiveness of our approach. We evaluated the proposed solution's performance against alternative methodologies used by fleet operators in different tests, including energy consumption, average user dissatisfaction, and average trip duration. The results showed that the proposed approach saved nearly 53-67% of energy for regenerative braking cases and 25-55% for non-regenerative braking cases when compared with other approaches and offers high adaptability to the environment and less complexity when compared with the exhaustive solution.

INDEX TERMS Electric scooters, reinforcement learning, energy minimization, user dissatisfaction, route and speed selection.

I. INTRODUCTION

Over the past few years, Micro-Mobility (MM) has surfaced as a promising approach to urban transportation that prioritizes the needs of passengers on short-distance journeys, such as those within a city's first and final

kilometre [1], [2], [3]. The appeal of MM lies in its on-demand transportation option that is flexible, sustainable, affordable, and easy to use. Moreover, MM represents one of the solutions to reduce reliance on private automobiles for transportation over short distances, ultimately mitigating traffic congestion problems [4]. Access to motorized private or public transportation for underprivileged communities can be improved through shared MM services. Therefore,

MM has become a top priority for countries worldwide to address transportation challenges and provide equitable mobility solutions for their citizens. MM refers to a range of compact vehicles or micro-vehicles that generally travel at speeds less than 45 kilometres per hour [4]. These vehicles include two- to four-wheelers and other equipment such as scooters, skateboards, bicycles, and tricycles [5]. They are typically shared or privately owned and can be either motorized or human-powered. Figure 1 provides an overview of the various types and sizes of MM vehicles, ranging from two-wheelers to four-wheelers. These vehicles are suitable for a wide range of users, including active individuals, those with partial mobility impairment, and even those who are fully disabled. This paper focuses on the two-wheel Electric Scooter (ES), highlighted in red.

In light of the favourable aspects of ESs as a subset of electric vehicles (EVs), it is crucial to address the persistent challenge of limited driving range, necessitating substantial enhancements. Numerous methodologies have been proposed in the literature to mitigate the energy consumption limitations associated with typical electric-type vehicles. One prominent approach involves the implementation of regenerative braking (i.e., converting the kinetic energy of movement into electrical energy to be stored in the battery) as illustrated by studies [6], [7], [8], [9], [10]. The primary objective here is to maximize the energy harvested from regenerative braking, consequently minimizing overall energy consumption. This is achieved through the application of optimal brake allocation and efficient braking control, employing advanced optimization and machine learning algorithms. Another way to minimize energy consumption centres around energy management strategies, leveraging sophisticated algorithms and control mechanisms to optimize power distribution among different energy sources (e.g., batteries) and components within the vehicle. For instance, [11] introduced strategies that optimize speed profiles and control parameters, resulting in reduced energy consumption. The work in [12] delves into battery management, incorporating strategy switching between power flow and thermostat control. Predictive control stands out as another effective method for minimizing energy consumption. This technique utilizes information about the road ahead, traffic conditions, and other pertinent data to anticipate power demand. Subsequently, it adjusts the energy flow and driving pattern (acceleration, deceleration) accordingly, as explored in [13], [14], [15], [16], [17], and [18]. Furthermore, energy minimization is attainable through the design of lightweight and low-drag vehicles, as shown in [19], and also by designing optimal multi-speed transmission systems as in [20]. Authors in [21] incorporated EV energy minimization by investigating the battery health and state of charge prediction techniques to enable a correct energy minimization. In addition, one of the common methods to minimize energy consumption is known as the energy-efficient routing problem, where it is required to optimally find the best route (or trajectory) that minimizes the energy of a vehicle considering environmental factors

such as road link speed limit, slope, and external factors such as the forces affecting on the vehicle [22], [23], [24], [25]. Some interesting papers adopted a scheme to control a fleet of agents through intermittent communication (rather than continuous communication) in case of a fault to save energy [26], [27]. Finally [28] discussed the different RL algorithms for optimal EV dispatch problems in which one of the applications (Grid to Vehicle) is to concerned with maintaining a healthy recharge of the EV from the grid and eventually conserving the battery state of health, which affects the state of charge of the battery.

To summarize, we have reviewed the different methods used for minimizing the consumption of a typical electric vehicle. In addition, we reviewed the routing optimization to minimize energy use for EVs. Our analysis has led us to observe the following research gaps:

- Most of the energy minimization methods for EVs/MM in the literature do not include the effect of the temperature on the battery state of charge.
- Most of the energy minimization techniques associated with routing consider only the energy-efficient path and lack the speed selection per each road segment.
- Limited contributions on energy consumption models for MM vehicles (especially ESs), Furthermore, there is a scarcity of research on energy-efficient routing strategies for these vehicles.
- Regenerative braking was not considered together with the speed and route selection in EVs and was not addressed in MM vehicles (especially two-wheel ES).
- Different from road slopes & angles, road type and quality were not considered for the routing problem in energy minimization for EVs in general and ESs in particular.

Therefore, this paper presents an approach for minimizing ES's energy consumption through optimized routing and optimal speed selection, using reinforcement learning, while considering user dissatisfaction as a constraint. The problem considers a graph composed of bi-directional roads and road segments, where each road segment has a different speed limit, road angle, road shading, traffic congestion, and road quality. To provide a more realistic problem, we modelled the system to be time dynamic, meaning that some parameters change with time, as in the case of temperature, road congestion and auxiliary power of ES. More details on the problem are discussed in the next chapter. The main contributions of this paper can be listed as follows:

- Modelling the problem of energy consumption optimization in ESs by jointly addressing path and speed selection. This comprehensive approach considers dynamic factors such as temperature fluctuations, time zones and road conditions. In addition, it considers both regenerative and non-regenerative braking in ES.
- Proposing an exhaustive energy consumption model that incorporates all energies affecting ESs and resulting from rolling resistance, air friction, road gradient, aux-

FIGURE 1. Micro mobility vehicles, with the focus on the two-wheel ES highlighted in red.

FIGURE 2. System grid overview, showing the different locations and bi-directional roads.

iliary and temperature/pressure. In addition, proposing a user dissatisfaction model that incorporates factors affecting the user's experience and includes traffic conditions, congestion and temperature.

- Proposing a DRL approach for addressing the optimization problem, taking into account time constraints, road constraints, and user dissatisfaction.

Considering how important ESs are as MM vehicles in shaping the world, their energy capacity is essential; thus, this paper aims to reduce the amount of energy ESs use, especially in commuting, through optimized routing and speed selection, considering user dissatisfaction and stochastic environment.

II. SYSTEM MODEL

In this section, we present the system model adopted for analysis. As depicted in Figure 2, the system encompasses an educational city graph, wherein autonomous ESs are deployed for student/faculty transportation. These ESs have been leased by an operator and strategically positioned across the grid. The primary objective of the operator is to minimize the energy consumption of ESs during each trip, all while considering user dissatisfaction. This optimization is realized through the appropriate selection of optimal paths and speeds. By proposing an energy-efficient approach to path-speed selection for ESs, the operator can enhance the ES's trip capacity, thereby increasing revenue generation and establishing a more sustainable business model.

A. SYSTEM DESCRIPTION

The educational city graph, denoted as G, comprises interconnected blocks linked by bidirectional roads. Furthermore, a set of L locations representing the most frequently visited campus buildings is considered. These locations are interconnected through I distinct paths, each comprising J road segments with lengths denoted as D_j. Each road segment is characterized by diverse parameters, including Road Speed Limit (R_{SL}), Road Slope (R_S), Road Condition
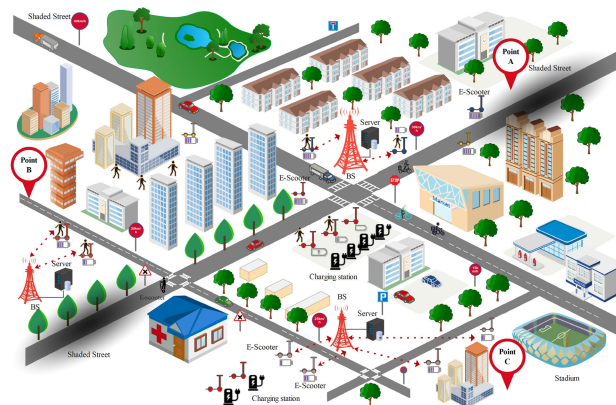
(R_C), and Road Shading (R_{SH}), as illustrated in 2. The system operates in a dynamic temporal context, wherein certain parameters, such as temperature and Traffic Congestion (T_{CO}), vary over time. The objective is to identify the most energy-efficient route and corresponding speed for travelling between locations within this dynamic grid while taking into account user dissatisfaction.

B. SYSTEM CONFIGURATION AND USER INTERACTION

Figure 3 provides a comprehensive representation of the system configuration. It is assumed that there exist L locations, for example, (A, B, C), interconnected by a network of paths (e.g., Path A-F2-K3-B connecting location A to B), with (A-F2), (F2-K3), and (K3-B) representing two-way road segments. For simplicity, vehicles traversing this grid are restricted from halting or altering their direction within a road segment after selecting a specific path between two locations. Additionally, it is assumed that speed remains constant along a road segment, and any deviations due to factors such as traffic congestion or road conditions are neglected, given the relatively low speeds of ESs. The adopted system in Figure 3 is referred to as the origin-destination graph, which shows all possible destinations from and to a set of locations. However, for generality, the adopted modelling and solution can be used on any type of graph or grid.

C. ES DEPLOYMENT AND OPERATION

The study takes into account a collection of scattered Electric Scooters (ESs), denoted as S, where S is a finite integer within the set of positive integers ($S \in \mathbb{Z}^+$). These ESs establish connectivity with a central server through base stations distributed across the grid. A scenario can be stated as follows: at a specific time instant t, a set of users U is defined as $U = \{1, 2, \ldots, S\}$, representing individuals requesting transportation from their current location (e.g., A, B, or C) to another predefined destination. ESs facilitate user requests through a mobile application that gathers essential user information. Subsequently, ESs transmit their current (source) location $\omega_t = (x_t, y_t)$, intended destination $\omega_d =$
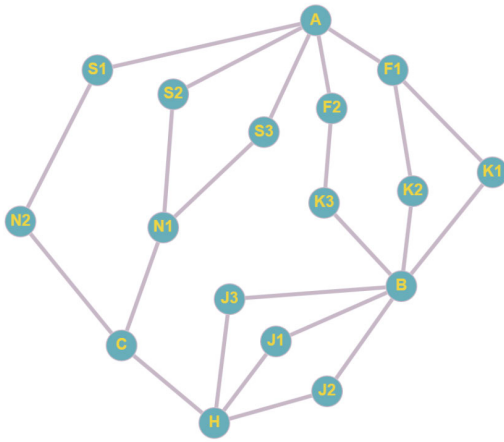
FIGURE 3. An example of a system graph showing *L* locations (e.g. *A*, *B*, and *C*) with all the intermediate road segments.

(x_d, y_d), and their current SoC ζ_t to the central server via base stations strategically positioned within the grid. The central server then undertakes the task of determining the optimal route P_i and segment-specific speeds $V_{i,j}$ to minimize the total energy consumption E_T while adhering to the User Dissatisfaction *UDS* constraint and environment fluctuations. To accomplish this, we propose a comprehensive energy consumption model for ESs that accounts for all relevant energy factors across the terrain. Furthermore, we formulate a *UDS* model that encompasses the factors contributing to user discomfort. Both models are presented later in this paper.

D. SYSTEM DYNAMICS

In this sub-section, we delve into the dynamic aspects of the system, highlighting the dynamic and uncertain nature of temperature, traffic congestion, and auxiliary energy for electric scooters (ESs) over time. Our analysis considers a dynamic temperature model for the entire grid, where temperature fluctuations occur both stochastically and location-wise. Temperature variations are influenced by several factors, including shading effects within distinct areas. In addition to temperature fluctuations, the system experiences dynamic changes in traffic congestion over time. These changes result from varying user demand and travel patterns throughout the day. We consider these temporal fluctuations in traffic congestion as a critical factor influencing the selection of optimal paths and speeds for ESs. Furthermore, the auxiliary energy required by ESs is subject to time-dependent variations. This auxiliary energy accounts for factors such as environmental lighting, which differ during different time periods.

To facilitate a simplified analysis, we focused on a predefined time frame of 12 hours. This time frame corresponds to the operational hours of the operator, spanning from 8:00 AM to 8:00 PM. Within this time window, we further partition the system dynamics into three equal time periods, each encompassing 4 hours, as outlined below in Equation (1),

where we consider a number of *K* trips done per time zone.

$$Time\ Zones = \begin{cases} Zone_1, 8:00 \leqslant t < 12:00 \\ Zone_2, 12:00 \leqslant t < 16:00 \\ Zone_3, 16:00 \leqslant t < 20:00 \end{cases} \quad (1)$$

E. EXHAUSTIVE ENERGY CONSUMPTION MODEL

This section discusses the exhaustive energy model, which includes mechanical and electrical energies affecting ESs. It is defined as follows:

$$E_T = E_{roli} + E_{af} + E_{gr} + E_{aux} + E_{temp} + E_{add} \quad (2)$$

where E_T is the total energy consumption, E_{roli} is the energy consumed to overcome rolling resistance, E_{af} is the energy consumed to overcome air friction, E_{gr} is the energy consumed or regenerated due to road slope angle, E_{aux} is the energy consumed due to the powering of auxiliary loads, E_{temp} is the energy loss due to influence of temperature and, E_{add} is the additional energy due to losses of moving parts. All these energies are to be calculated per 1-kilometre distance.

1) ENERGY DUE TO ROLLING RESISTANCE

Rolling resistance is defined as a loss of energy per unit of travelled distance [29]. It comprises mechanical energy loss owing to aerodynamic drag, friction between the terrain and tire, and contact between tire and rim, generating heat energy in the tire's material and leading to high energy consumption [30]. Rolling resistance may account for up to 30% of the vehicle's energy usage under specific traffic situations and fixed travelling velocity [31], [32]. The energy consumption due to rolling resistance E_{rol} can be calculated by multiplying rolling force F_{rol} in line with the distance D as follows:

$$E_{rol} = F_{rol}\ D \quad (3)$$

rolling resistance's force can be formulated as follows [33] and [34]:

$$F_{rol} = \tau\ m\ g\ cos(\theta) \quad (4)$$

where τ depicts the coefficient of rolling resistance, m is the vehicle mass loaded (with driver), g represents the gravitational acceleration constant, and θ denotes the road slope angle. τ can be represented by the following formula [35]

$$\tau = \frac{C_i}{r} + \frac{1}{p_t}(0.01 + 0.095(\frac{V}{100})^2) \quad (5)$$

in which C_i represents the rolling resistance constant, *r* is the radius of the vehicle's wheel, p_t symbolises the tire pressure, and *V* is the travelling velocity. Therefore, E_{rol} can be written as follows:

$$E_{rol} = \left[\frac{C_i}{r} + \frac{1}{p_t} \left(0.01 + 0.095 \left(\frac{V}{100} \right)^2 \right) \right] m\ g\ D\ cos(\theta) \quad (6)$$

2) ENERGY DUE TO AIR FRICTION

Air friction energy consumption E_{af} can be calculated by multiplying force F_{af} in line with the distance D as follows:

$$E_{af} = F_{af} D \tag{7}$$

air friction force can be represented by the following equation [33], [34], [36].

$$F_{af} = \frac{1}{2} q A_f C_d V^2 \tag{8}$$

such that q is the air density, A_f is the frontal area of the vehicles, including the riding passenger, C_d is the air drag coefficient. Air density q can be determined by the subsequent formula [37]

$$q = \frac{p_d}{R_d T_k} + \frac{p_v}{R_v T_k} \tag{9}$$

in which p_d is the pressure of dry air, p_v is the pressure of water vapour present in the atmosphere, R_d is the gas constant of dry air, R_v is the gas constant of water vapour, and T_k is the atmospheric temperature in Kelvin. p_v can be formulated in equation (10) [37], such that p_x represents the saturation vapour's pressure, and r_h is the relative humidity present in the atmosphere. p_x can be estimated by equation (11) [37], where T_c is the atmospheric temperature in Celsius.

$$p_v = p_x r_h \tag{10}$$

$$p_x = 6.1078(10)^{\frac{7.5\, T_c}{T_c + 237.3}} \tag{11}$$

Therefore, E_{af} can be represented as follows:

$$E_{af} = \frac{1}{2} \left[\frac{p_d}{R_d T_k} + \frac{6.1078(10)^{\frac{7.5\, T_c}{T_c + 237.3}} r_h}{R_v T_k} \right] A_f C_d V^2 D \tag{12}$$

3) ENERGY DUE TO ROAD GRADIENT

E_{gr} is the energy consumed or regenerated due to the slope of the terrain and formulated as follows:

$$E_{gr} = F_{gr} D \tag{13}$$

where F_{gr} is the gravitational force experienced when driving on non-horizontal roads, or in other words, the additional force experienced while driving uphill or downhill [34], and is calculated in equation (14) [36]. This equation consists of three functions: When the vehicle goes uphill, energy will be consumed regardless of the slope and speed. However, if the vehicle is travelling downhill (meaning the slope of the road is negative) and with a speed exceeding $20\,\frac{km}{h}$, the kinetic energy will be harvested and sent back as regenerated energy to the battery, but with an assumed efficiency of only 60% due to the light weight of the vehicle (considering ES) and the low speed, which contributes to the lower energy regeneration, as discussed by [36]. F_{gr} will equal zero for full horizontal terrain, i.e., $\theta = 0$. Furthermore, in this paper,
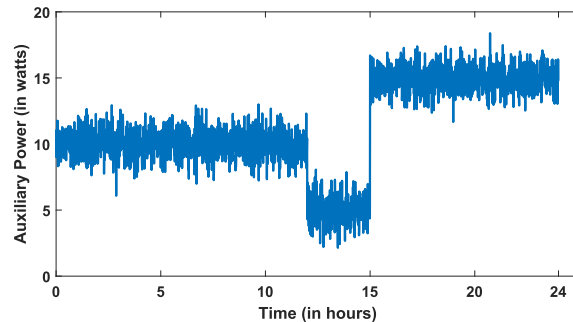

FIGURE 4. Auxiliary power of e-scooter throughout the day.

we assume the road angle R_{ag} varies between $-3°$ and $+3°$, i.e., $\{-3° \leq R_{ag} \leq +3°\}$

$$F_{gr} = \begin{cases} mg\sin(\theta), & \theta > 0 \ and \ v_{min} > 0 \\ 0.6mg\sin(\theta), & \theta < 0 \ and \ v_{min} \geqslant 20\frac{km}{h} \\ 0, & \theta = 0 \end{cases} \tag{14}$$

therefore E_{gr} can be represented as follows

$$E_{gr} = \begin{cases} mg\sin(\theta)D, & \theta > 0 \ and \ v_{min} > 0 \\ 0.6mg\sin(\theta)D, & \theta < 0 \ and \ v_{min} \geqslant 20\frac{km}{h} \\ 0, & \theta = 0 \end{cases} \tag{15}$$

4) ENERGY DUE TO AUXILIARY DEVICES

Auxiliary Energy E_{aux} is the energy required to run the vehicle's electronics such as the additional power needed for lighting, heating/cooling (in closed MM vehicles), radio, and other functions based on the surrounding conditions (ambient light, temperature, etc.) [36], [38], [39]. As we consider a light 2-wheel ES, E_{aux} accounts only for the lighting and is given by the following equation:

$$E_{aux} = P_e(t)\, t = P_e(t)\, \frac{D}{V} \tag{16}$$

such that $P_e(t)$ is the auxiliary power mainly consumed by lighting and is variable with time, traveling time t represents the trip time determined by the traveled distance D and the ES's traveling velocity V. Figure 4 illustrates an averaged $P_e(t)$ of a scooter throughout a week. $P_e(t)$ is influenced by two primary factors: Time of day and environmental light intensity. The variation in power throughout the day can be segmented into three distinct periods.

- During the early morning period, which spans from 00:00 to 12:00, $P_e(t)$ fluctuates around $10 \pm 2W$. This is attributed to the moderate sunlight intensity experienced during these hours.
- In the afternoon, from 12:00 to 15:00, the lighting conditions are at their brightest. Consequently, $P_e(t)$ fluctuates at a minimum value of $5 \pm 2W$, reflecting the reduced electrical power required during these hours.
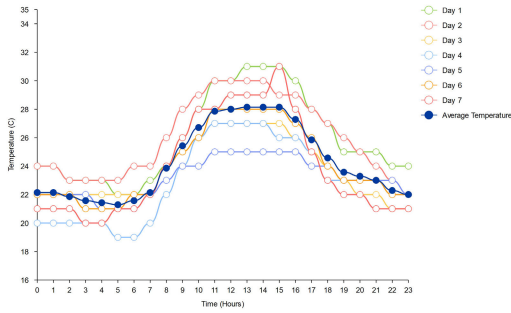
FIGURE 5. Temperature fluctuations during the day for shaded road segments for several days and their average.

TABLE 1. Traffic congestion.

Traffic Congestion Level	Description
0	Clear Road (No congestion, smooth traffic flow)
1	Light Congestion (Minor slowdowns, but traffic is still moving).
2	Moderate Congestion (Noticeable slowdowns, some stop-and-go traffic).
3	Heavy Congestion (Significant slowdowns, long delays, and slower overall traffic speed).
4	Severe Congestion (Almost standstill traffic, long delays).
5	Stand-Still Congestion (Complete gridlock, no movement at all).

- The evening period, extending from 15:00 to 24:00, witnesses the highest light intensity of the day. Consequently, the power consumption of $P_e(t)$ rises to $15W \pm 2W$ during this time.

The reason behind the fluctuations of $P_e(t)$ values is related to the environmental light intensity, since the lighting at the same spot, at the same time, varies throughout a week.

5) ENERGY DUE TO TEMPERATURE

E_{temp} refers to the energy loss that occurs due to experiencing high ambient temperature. Earlier research mentioned that E_{temp} is the energy loss brought on by the electric motor's power output fluctuating, owing to the impact of the surrounding temperature [33], while in [40], E_{temp} was defined as the increase in the battery's energy intensity that occurs due to the presence of a heated environment which causes a speed-up in the battery chemical reactions and eventually causing losses. In addition, both authors have modelled E_{temp} as a third-order polynomial function, based on sedan electric vehicles travelling on different terrains and at different temperatures. Typical MM vehicles such as e-scooters are usually left in open areas with direct sunlight, and thus, it is crucial to include E_{temp} effect on the energy consumption of these vehicles. Furthermore, E_{temp} in this paper was formulated using [40], and scaled-down, taking into consideration the battery capacity between EVs and ESs, and is shown as follows:

$$E_{temp} = \frac{287 - 16T_c + 0.502T_c^2 - 0.00439T_c^3}{218} \quad (17)$$

To simulate a realistic system, we used realistic temperature data in Doha-Qatar for March 2023 [41], in which the temperature varies throughout the day, with the lowest temperature occurring during the early morning hours and peaking at noon before dropping again at night. In addition, we assumed that some roads are shaded (i.e. roads that are shaded by buildings or trees next to them), while some roads are in open areas with no shade. Temperature varies in different areas, with shaded roads having lower temperatures than open areas by 15%. Figure 5 shows the temperature fluctuation over 1 day period in shaded roads for 1 week and the average temperature for the same. The same

pattern applies to non-shaded areas with a 15% increase in temperature as stated previously.

6) ADDITIONAL ENERGY

E_{add} refers to the additional loss of energy that is lost during a process and cannot be attributed to any specific cause [34], [36]. E_{add} can be assumed as follows:

$$E_{add} = 1\frac{Wh}{km} \quad (18)$$

F. USER DISSATISFACTION MODEL

User Dissatisfaction UDS is defined as a combination of factors that could cause discomfort to the users while driving. These factors include high ambient temperature, road congestion, and poor road conditions. High temperatures can lead to discomfort and heat stress, which can affect the drivers' mortality or mental health [42], [43], while road congestion can increase travel time and eventually frustrate the user causing aggressive driving pattern [44]. Poor road conditions, such as potholes and uneven surfaces, can cause an increase in the risk of accidents to the users [45]. Considering these factors, the proposed approach seeks to optimize routing and speed selection to minimize energy consumption while keeping user dissatisfaction within an acceptable range.

It is always preferable for ES users to operate their vehicles in a shaded area to avoid heat strokes and exhaustion. Specifically, excessive temperatures & hot weather have a substantial negative impact on human health, raising the likelihood of respiratory death and morbidity [46]; thus, it is well considered. In addition, navigating the vehicle through roads with better conditions other than unpaved roads that can cause injuries is desirable. According to [47], the growth of scooter-sharing firms has led to an increase in recent reports of injuries involving motorized scooters, with more than 50% injuries affecting the lower arm, lower leg, and wrist in their study, which involved nearly 8400 electric scooter driver's injuries.

UDS is defined in equation (19), where T_{cn} is the normalized temperature, T_{con} is the normalized traffic congestion, and R_{cn} designates the normalized road condition.
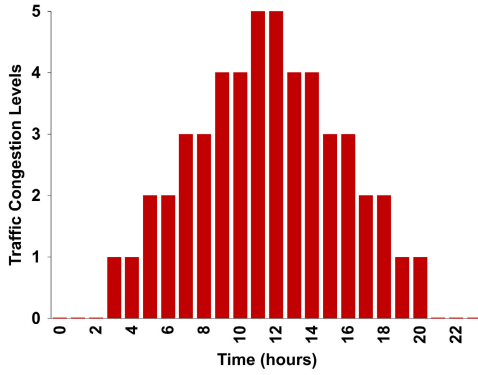
FIGURE 6. Traffic congestion variation throughout the day per 1 road segment.

Normalization was applied since the *UDS* adds different factors with different units.

$$UDS = T_{cn} + R_{cn} + T_{con} = \frac{T_C}{T_{Cmax}} + \frac{R_c}{R_{cmax}} + \frac{T_{CO}}{T_{comax}} \quad (19)$$

T_{cn} represents the percentage of the road segment's ambient temperature T_C to the grid's maximum temperature T_{Cmax}. R_{cn} designates the normalized road condition, which is basically the percentage of the road condition R_c per segment to the maximum road condition in the grid R_{cmax}. R_c describes the quality factor of the road and is mainly divided into four levels which are excellent, good, fair, and poor conditioned roads, that are ranging from the paved road (excellent for ESs to travel such as asphalt roads) to unpaved road that is not barely suitable for travelling. R_c ranges from a value of 1 to 4, with excellent roads having a value of 1 and worst roads with a R_c value of 4. T_{con} refers to the percentage of traffic congestion of a road segment to the maximum traffic congestion in the grid T_{comax}. T_{co} is related to car density per road segment. We consider the roads to have six different levels of traffic congestion, ranging from 0 (Clear road) to 5 (stand-still congestion). Traffic congestion can be seen in table 1. In order to simulate a realistic environment, we considered that T_{co} changes with time and location. As seen in 6 where the traffic congestion varies throughout the day of a single road segment while other road segments have a time-shifted replica of the same.

III. PROBLEM FORMULATION

As mentioned previously, we consider a number of scattered ESs denoted as S, where S is a finite integer, i.e., $S \in \mathbb{Z}^+$. These ESs are connected to a central server via base stations scattered throughout the grid. At a specific time moment t, a number of users U is defined as $U = \{1, 2, \ldots, S\}$ request to be transported from their current location to any other saved location. All ESs then send their current location, intended destination, and their current SoC to the central server through the base stations located in the grid. Afterwards, the central server will choose the optimal path P_i

and speed per each segment $V_{i,j}$ to minimize the trip's Total Energy E_T while abiding by the *UDS* constraint. We consider two types of ESs, where the first type has a regenerative braking feature (can recover energy E_{gr}), while the other type does not. The central server will solve the formulated optimization problem P as follows:

$$\mathbf{P:} \quad \min_{P_{i,t}, V_{i,j,t}} \sum_{i=1}^{I} \sum_{t=1}^{T} (E_{i,t} \cdot P_{i,t}) \quad (20)$$

$$E_{i,t} = \sum_{j=1}^{J(i)} (E_{roli,j,t} + E_{af,j,t} + E_{gr,j,t} +$$

$$E_{aux,j,t} + E_{temp,j,t} + E_{add,j,t}) \quad (21)$$

such that:

$$P_{i,t} \in \{0, 1\} \quad (22)$$

$$\sum_{i=1}^{I} P_{i,t} = 1, \forall t \quad (23)$$

$$UDS_{i,t} = \sum_{j=1}^{J} \frac{1}{J(i)} \left(\frac{T_{C,j,t}}{T_{Cmax}} + \frac{T_{CO,j,t}}{T_{comax}} + \frac{R_{c,j,t}}{R_{cmax}} \right) \quad (24)$$

$$UDS_{i,t} \le UDS_{max}, \forall t \quad (25)$$

$$t_{i,t} \le t_{max}, \forall t \quad (26)$$

$$T_{C,i,j,t} \le T_{Cmax}, \forall j \in J, \forall t \quad (27)$$

$$V_{i,j,t} \le V(j)_{max}, \forall j \in J, \forall t \quad (28)$$

$$P_{i,t} = 1 \Rightarrow V_{i,1,t} > 0, \forall t \quad (29)$$

$$V_{i,j,t} > 0 \Rightarrow V_{i,j-1,t} > 0, \forall j > 1, \forall t \quad (30)$$

where equation (20) is the optimization problem that corresponds to minimization of energy $E_{i,t}$ for all paths $i \in I$, for all time steps $t \in T$, considering each path i contains number of road segments $j \in J$. $E_{i,t}$ is the total energy per path i and is defined in (21).

We introduce a path selector $P_{i,t}$ which serves as a decision variable for selecting a path per time step, where $P_{i,t} = 1$ for selecting a path and zero otherwise as seen in (22). Since we wish to find the only optimal path per time step, constraint (23) guarantees the selection of one path. Constraints (24) and (25) ensure that the average user dissatisfaction per path per time step does not exceed a maximum value UDS_{max}. To avoid long trip times, constraint (26) was added to ensure the trip time along the path does not exceed a specific maximum time t_{max}. In addition, Constraint (27) ensures that the central server shall not choose any road segment j that exceeds a certain maximum temperature T_{Cmax}. Furthermore, constraint (28) ensures that the selected speed per each road segment does not exceed the speed limit for the road segment for all time steps. Constraint (29) and (30) are flow rule constraints, where (29) is the starting segment constraint that ensures the first road segment is guaranteed to be connected to the starting location for all times; that is for each chosen path i, the first road segment $j = 1$ must be connected to the starting location for each time a trip starts; this is to ensure that the vehicle starts its journey from the chosen path. On the other hand, (30) is

the intermediate segment connectivity constraint that ensures there are no gaps between road segments on the chosen path, and they are connected in a continuous manner, that is; for each chosen path i and each road segment j along the path (excluding the first segment $j = 1$), if the road segment j is used (i.e., $V_{i,j,t} > 0$), the previous road segment $j - 1$ must also be used (i.e., $V_{i,j-1,t} > 0$).

IV. DEEP REINFORCEMENT LEARNING (DRL)

This section is dedicated to examining the proposed solution. This approach considers both environmental dynamics and the requirements of the problem, aiming to provide an efficient resolution. First, we will highlight the advantages of employing Deep Reinforcement Learning as our optimization method, comparing it with other approaches. Subsequently, we will delve deeper into the process of transforming the problem into a Markov Decision Process (MDP) and utilizing the DQN algorithm for its solution. The rationale behind incorporating DRL in this problem stems from its capacity to address sequential decision-making. Actions taken during a specific time instant t have a lasting impact on subsequent trips at $t + 1$ due to the cumulative nature of energy consumption. Moreover, DRL excels in handling uncertainties and stochastic parameters, particularly temperature and road congestion in our context. Notably, DRL's adaptability to dynamic environmental changes and its ability to facilitate real-time decision-making distinguish it from other optimization techniques.

A. ADVANTAGES OF DRL

DRL can offer several advantages over MINLP, LP, and GA for solving our optimization problem, particularly when dealing with complex, dynamic, and high-dimensional scenarios like route optimization with discrete decision variables (path selection and discrete speed combination). The following reasons illustrate the advantages of DRL over other optimization approaches in the case of our problem:

- Dynamic Adaptation: DRL can adapt to changing conditions in real time. As conditions on the road change (e.g., due to traffic or weather), DRL agents can adjust their strategies, whereas LP and GA require re-running, which may not be feasible in real-time scenarios.
- Exploration and Exploitation: DRL naturally balances exploration (trying different paths and speeds) and exploitation (choosing known good strategies), allowing it to discover alternative newly-made routes when applicable.
- Handling Discrete Actions: DRL can naturally handle discrete action spaces, such as selecting paths and discrete speed levels. In contrast, LP and traditional optimization techniques are typically designed for continuous decision variables.
- Transfer Learning: DRL models can potentially be transferred or fine-tuned for different road networks or

locations, saving computational resources compared to re-optimizing LP or GA models.
- Scalability: DRL algorithms are scalable and can handle problems with high-dimensional state and action spaces. This is important for large-scale route optimization problems.
- Real-Time Decision Making: Since DRL at the testing phase can have low complexity, it is well-suited for real-time decision-making, such as route planning, where decisions need to be made quickly based on current information. LP and GA may not provide timely solutions in such scenarios.

B. MARKOV'S DECISION PROCESS (MDP)

MDP is represented as a tuple of 5 elements as $(\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma)$, such that $\mathcal{S}$ depicts a list of potential states, whereas $\mathcal{A}$ represents the possible action list at specific state $\mathcal{S}_t$, and $\mathcal{R}$ represents the reduced reward by a reduction/discounted factor $\gamma \in [0, 1]$. The training process of DRL initiates with a DRL agent obtaining environment state representation $s \in \mathcal{S}$ at each time increment t. Then, by employing a policy $\pi(a|s)$, the agent performs an action $a \in \mathcal{A}$ and, in return, obtains a reward $r_t \in \mathcal{R}$ then transits with a probability of $P(s'|s, a)$ to new/next state s'. The accumulated sum of discounted rewards up to some horizon X is given as $\mathcal{R}_t = \sum_{t'=t}^{X} \gamma^{t'-t} r_t$ [48], [49].

For the agent to determine how good it is to execute a specific action in a state s following some policy π, it follows a state-action value function (alternatively named as Q function) which is represented as $Q^{\pi}(s, a) = \mathbb{E}_{\pi}[R_t|S_t = s, a_t = a]$. This essentially pertains to the rewards accumulated by performing an action a in a state s and then obeying a policy π. In order to evaluate a state's quality when following a policy π, a state value function is applied. The state value function is formulated as $V^{\pi}(s) = \mathbb{E}_{\pi}[Q^{\pi}(s, a)]$.

C. ADOPTED ALGORITHM

Among different DRL algorithms, we opt to use Deep Q-Network (DQN) [50], [51], which is a variant of the classic Q-Learning algorithm. Q-Learning and Deep Q-Network (DQN) are two popular reinforcement learning techniques used to solve complex decision-making problems. Q-learning is a model-free, off-policy algorithm that learns the optimal action-value function Q for a given policy. It works by estimating the expected future rewards for each action in a given state and updating the action-value function based on the observed rewards. The action-value function Q represents the expected reward for taking a specific action in a given state. It can be calculated as follows:

$$Q(s, a) = \mathbb{E}[r + \gamma \max_{a'} Q(s', a')|s, a]$$

where r is the immediate reward, s' is the next state, and a' is the next action. γ is the discount factor that determines the importance of future rewards. Q-Learning updates the action-value function iteratively by employing

the Bellman equation [52]. Q-Learning continues to update the action-value function until convergence is reached. Unlike Q-Learning, DQN uses a deep neural network to approximate the Q-function, thus handling high-dimensional state spaces and is more stable than traditional Q-Learning. There have been many extensions and variations of Q-Learning and DQN. Double Q-Learning [53] and Dueling DQN are two examples of improved algorithms that have been shown to outperform the original Q-Learning and DQN. Distributional reinforcement learning and Rainbow are other recent developments that combine multiple techniques to achieve state-of-the-art results in various domains [54], [55], [56]. In our specific case, the choice of DQN stems from its favourable attributes, including ease of implementation owing to its simplicity, its proven stability and robustness in training, and alignment with the discrete nature of our action space.

V. SYSTEM'S STATE, ACTION AND REWARD

After modelling the system and all its time dynamic changes, we intend to solve the problem presented in this paper using DRL. As discussed previously, to solve our optimization problem P using DRL, we are required to translate our problem to MDP where we introduce our states, actions, and rewards as follows:

A. ENVIRONMENT STATE

As DRL algorithms require awareness of the environmental state and the ability to respond to it, it was imperative to include sufficient information for the agent to make decisions at each time step, denoted as t. Firstly, because determining the scooter's path and speed relies on its current position, we incorporated the coordinates of the electric scooter's location at time t, denoted as ω_t, such that $\omega_t = (x_t, y_t)$. This information enables the agent to deduce its next destination. Secondly, to enable the agent to consider the temporal aspect of its actions and assess the effectiveness of its chosen path and speed under varying environmental conditions (e.g., weather, shading, etc.), we introduced a time clock variable denoted as $\mathbb{C}_t$. Lastly, we provided information about the remaining battery capacity in Watt-hours (Wh) at time t, denoted as $\mathbb{B}_t$. This variable assists the agent in making decisions that stay within the battery's limits and understanding the relationship between its actions and the battery's status. Furthermore, it prompts the agent to adjust its strategy when the battery is running low to maximize the number of trips the electric scooter can complete. Compiling the above, the environmental state at each time step can be defined as in eq.(31)

$$s_t = (\omega_t, \mathbb{C}_t, \mathbb{B}_t) \tag{31}$$

B. ACTION

After evaluating the environmental state, denoted as s_t, the agent takes action a_t to induce a transition in the environmental state. In this context, each action corresponds to a specific trip, such as moving from location A to B. As our primary objective revolves around selecting the optimal route and the appropriate speeds for that route, the action must encompass both these critical elements. Thus, the action is defined as in eq. (32):

$$a_t = (P_t, V_t(i, j)) \tag{32}$$

where P_t represents the chosen path at time t and $V_t(i, j)$ signifies the selected speeds for each road segment j along path i at time t. In this formulation, P_t is an integer variable indicating the path ID leading to the destination. Simultaneously, $V_t(i, j)$ is a vector containing the speeds applied to different street segments within the chosen path P_t. As an example, if a user requests to be transported from location A to location C at time 9:00; then if there are three possible paths from location A to location C (i.e., path $A - S_1 - N_2 - C$, path $A - S_2 - N_1 - C$, and path $A - S_3 - N_1 - C$), then the DRL should choose the best combination of path and segment speeds that will result in lowest energy consumption, taking into consideration the time-varying environment and user dissatisfaction levels. An example of the chosen path and speed can be path $A - S_2 - N_1 - C$ with speeds of 20km/h for segment $A - S_2$, 25 km/h for segment $S_2 - N_1$, and a speed of 30 km/h for segment $N_1 - C$.

It's important to note that the choice of path impacts the energy consumption of the battery and the user's level of dissatisfaction, while the speed directly influences energy consumption. Moreover, the dynamic nature of the grid introduces variability into the decision-making process, as different times may necessitate distinct optimal path-speed combinations. Additionally, the road segments exhibit varying angles, $\{-3° \leq R_{ag} \leq +3°\}$, meaning that the optimal path from one location to another may not be the same in the reverse direction. Furthermore, different paths come with specific speed limits that must not be exceeded, further complicating the problem's complexity.

C. REWARD

After applying an action a_t per trip, the agent is rewarded with a scalar value r_t that identifies the goodness of the action taken. To better judge the performance of the agent, our reward function is divided into an immediate reward, given after each trip in the different time zones, and a final reward, provided after completing all trips. While the former influences the agent to abide by the constraints and search for a better low-energy configuration, the latter focuses on the long-term goal of conserving energy for the ES battery. For immediate rewards, we keep track of the lowest energy configuration achieved by the agent for each trip and zone, denoted as $E(T, Z)_{min}$ because there can be multiple configurations per trip and zone. This helps us measure how close the agent is to finding the best configuration for a specific zone and trip. In our reward system, the agent receives the maximum reward if it discovers a new minimum-energy configuration that meets user dissatisfaction and time constraints, indicated as $\mathbb{U} = 1$

TABLE 2. Road segment parameters.

From Node	To Node	Shading	Speed Lim. (km/h)	Road Angle	Road Cond.
A	F1	T	35	+1	1
A	F2	F	35	+3	2
F1	K1	F	20	+3	2
F1	K2	F	35	+2	3
F2	K3	F	35	-1	1
K1	B	T	35	+2	1
K2	B	F	35	+1	2
K3	B	F	20	0	3
B	J1	F	35	+3	1
B	J2	T	35	-2	2
B	J3	T	35	-1	3
J1	H	F	35	+3	1
J2	H	F	20	-2	2
J3	H	T	35	+1	3
H	C	T	35	-1	3
C	N1	T	35	+1	1
C	N2	F	20	-2	2
N1	S2	F	35	-2	1
N1	S3	T	35	+1	2
N2	S1	T	20	+2	3
S1	A	F	35	-2	1
S2	A	F	35	-1	2
S3	A	F	20	0	3

TABLE 3. System parameters.

Parameter	Value
Max Battery capacity	275 Wh
Mass m	75 kg
Rolling resistance constant C_i	0.005
Radius of ES wheel r	15 cm
Tyre pressure p_t	3.10 bar
Gravitational constant g	9.8 m/s^2
ES's Frontal area A_f	0.65 m^2
Air drag coefficient C_d	0.8
dry air gas contatn R_d	287.058 $J/kg.k$
Gas constant for water vapour R_v	461.495 $J/kg.k$
relative humidity r_h	50%
Discount rate γ	0.99
Training episodes	$2*10^5$
Replay buffer size $\mathcal{D}$	$2*10^6$
Batch size	1024
Learning rate α	0.0003
Neural Network	[128,128]
Exploration Decay	0.995
Final Epsilon Value	0.05

& $\mathbb{T} = 1$ (a value of 1 is given when the UDS and Trip time constraints are within boundaries). If the agent explores an action that is not better than the current minimum energy configuration, it incurs a penalty equal to the difference between the current obtained energy consumption and the minimum energy consumption, multiplied by a scaling factor λ, represented as $-\lambda|E(T, Z) - E(T, Z)_{min}|$. Additionally, if the agent does not satisfy any of the constraints for a trip, it receives no reward. Furthermore, if the agent doesn't complete the required trips because of battery depletion, it receives a significant penalty of -100, which will encourage the agent to complete as many trips as possible to reduce or avoid this penalty. Now, regarding the final rewards: Since we have a specific number of trips the agent must complete, it's crucial to ensure that all of them meet the requirements and reward the agent when they are successfully fulfilled, especially if there's remaining battery capacity. Therefore, we assign a penalty ranging from 0 to -10 based on the number of satisfied trips F_T. Satisfying all trips results in no penalty, achieving none results in a -10 penalty, and other cases receive penalties proportionally. Moreover, if the agent meets all trips with some battery capacity left, it gets this remaining battery as a reward. This encourages the agent to adopt a policy that conserves energy effectively. Compiling all of the above, creates the reward function r_t as follows:

$$
r_t = \begin{cases}
1 & \text{if } E(T, Z) \leq E(T, Z)_{\min} \ \& \ \mathbb{U} = 1, \mathbb{T} = 1 \\
-\lambda|E(T, Z) - E(T, Z)_{\min}| & \text{if } E(T, Z) > E(T, Z)_{\min} \ \& \ \mathbb{U} = 1, \mathbb{T} = 1 \\
-1 & \text{if } \mathbb{T} = 0 \text{ or } \mathbb{U} = 0 \\
-100 & \text{if } \mathbb{B} = 0 \\
(0, \cdots, -10) & \text{if } 18 > F_T > 0 \\
\mathbb{B}_{\text{remain}} * 10 & \text{Episode over}
\end{cases}
$$

$$(33)$$

VI. EVALUATION AND TESTING

In order to evaluate the proposed solution, we consider a system graph illustrated in Figure 3, our system graph consists of three central locations (A, B, and C) with three distinct paths from and to each location. Each path consists of three segments, each with different road parameters, speed limit, road slope, road condition, traffic congestion and road shading, as seen in Table 2. In addition, the grid is time dynamic, meaning that some parameters change with time, as in the case of temperature (see Figure 5) and road congestion.

In this section, we present our testing experiments along with different evaluation and comparison methods to ensure the correctness of the obtained results. We first ran convergence tests, followed by three distinct tests: the energy consumption test, user dissatisfaction test, and total trip time test. In each test, we compared our solution with four distinct solutions (more on that later). Lastly, we justified the use of an DRL-based solution by introducing a sudden disturbance in the grid and showing how quickly the DRL agent converged to the optimal solution. Our simulation parameters and simulation constraints can be found in Table 3.

A. CONVERGENCE TESTS

Considering the state, action, and reward mentioned previously, we ran the convergence tests to ensure that our DRL-based solution converges correctly to a specific answer every time, thus ensuring the correctness of the employed DRL training technique. We ran two different convergent tests. The first considers the Non-Regenerative Braking (NRB) (see Equation 14). The second test considers the regeneration braking (RB) factor, where the ES will have the capability to regenerate some of the lost energy by means of Regenerative braking. As previously mentioned, the system runs from 8:00 to 20:00, where we assume that the day is divided into three zones as depicted in equation 1. Therefore,
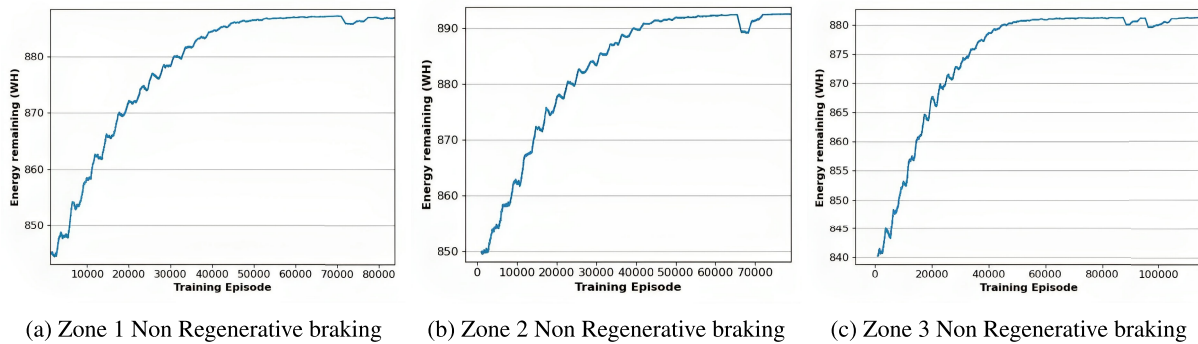
(a) Zone 1 Non Regenerative braking   (b) Zone 2 Non Regenerative braking   (c) Zone 3 Non Regenerative braking

FIGURE 7. Convergence of DRL for the case of non-regenerative Braking.

we train our DRL agent based on the three zones with and without RB and plot the agent's cumulative reward. We ran the agent for 70,000 to 100,000 episodes, each containing six trips. For DRL training we considered a battery of 1000 Wh capacity, while in testing we used a typical battery capacity of 275 Wh

1) CONVERGENCE TEST FOR NRB CASE

Figure 7 shows how our agent converges successfully for all zones in the cases of NRB, where these graphs are smoothed over a window of 500 episodes. From the same graph, it is observable that the amount of energy consumption is different in each zone due to the dynamic nature of the system (i.e. temperature, traffic congestion, and auxiliary power change with time). All sub-graphs have a small energy consumption dip (about 3-5 Wh only) near the final convergence, and this is due to the random exploration of the DRL agent as it continuously explores new solutions due to the fixed final epsilon value of 5%. Nevertheless, the outcomes are quite similar between the zones, such that zone 2 had the lowest energy consumption of around 105 WH, followed by zone 1 with an energy consumption of 110 WH and zone 3 with the highest energy consumption of 117.5 WH. In addition, we observed that the energy consumption pattern between zones matches exactly the auxiliary energy pattern rather than the temperature pattern (i.e., the lowest auxiliary energy consumption in zone 2 is the lowest followed by zone 1 and zone 3), meaning that the auxiliary energy had a larger impact on the energy consumption than temperature effect despite the huge temperature difference between the zones especially in zone 1 and 2, reaching about 15^oC in un-shaded areas and 10^oC in shaded areas.

2) CONVERGENCE TEST FOR RB CASE

Figure 8 illustrates the convergence with the RB in action. We remind that there is a specific condition for RB, such as negative road gradient and speed threshold (see equation 14). Similar to the NRB case, we notice the exact energy consumption pattern. However, the energy consumption is reduced significantly due to the regeneration action in which zone 2 had the lowest energy consumption of nearly 72 WH,

followed by zone 1 with a total consumption of 77 WH and then zone 3 with a consumption of 82 WH. In addition, it should be noted that there were relatively small energy consumption dips (as in 7). However, it did not appear in this case due to the smoothing of the graph.

B. SIMULATION RESULTS

After ensuring the DRL agent had been properly trained, we ran three tests to validate the results. Then, we compare our solution to other solutions in terms of 1) Energy consumption, 2) Average user dissatisfaction and 3) Average time for all trips. We compare our solution to four different realistic solutions, namely: *optimal*, *random*, *Fixed Lowest Speed (FLS)* and *aggressive*. Firstly, the optimal solution is obtained by an exhaustive method (i.e., trying all possible paths and speeds at all locations); this method is not always applicable but was considered a baseline. Secondly, the fully random solution, which means random path and random speed selection at each location, this solution simulates no given control policy by the ES operator, where a commuter can choose any path and velocities, which is common. Thirdly, the FLS solution, which is an energy-conservative solution that considers an operator policy limit on the maximum running speed to be the lowest speed but offers no guidance about routes. Finally, the aggressive solution is an energy-greedy solution that considers ES running on the maximum speed without guidance to the routes. The tests were conducted across all three zones, utilizing ES speed limits of 15km/h and up to 25km/h in a discrete form.

The tests were done on a PC with an Intel processor (I7-9700k), graphics card (RTX-2080 TI), and 32 GB of RAM. The computational time of the whole simulation to find the best paths/speeds for a total number of 18 Trips (i.e., the best path/speeds from and to each location, for each time zone) was 30-35 minutes, followed by a DRL inference of less than 0.1 seconds, which means that upon training the model, it can find the best paths and speeds instantly.

C. ENERGY CONSUMPTION TEST

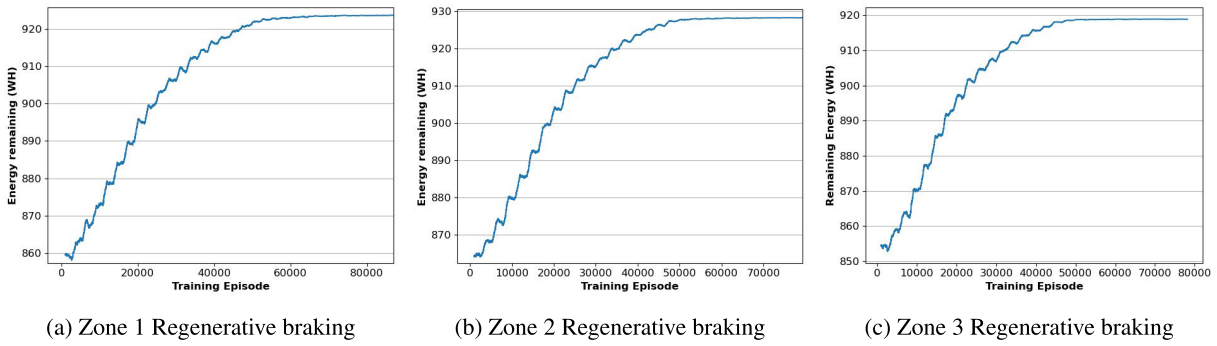First, we conducted an Energy consumption test, where we examined the least and most energy-efficient approaches. For

(a) Zone 1 Regenerative braking   (b) Zone 2 Regenerative braking   (c) Zone 3 Regenerative braking

FIGURE 8. Convergence of DRL for the case of regenerative braking in different zones.
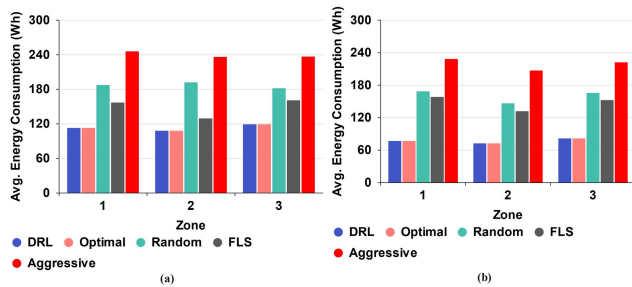

FIGURE 9. (a) Energy Consumption for a different solution (non-Regenerative case) (b) Energy Consumption for a different solution (Regenerative case).

FIGURE 10. (a) Average user dissatisfaction per zone for different solutions (non-Regenerative case). (b) Average user dissatisfaction per zone for different solutions (Regenerative case).

each of the three zones, we carried out two distinct energy consumption tests, with each zone comprising six trips. The first test involved RB, while the second did not. The results presented here represent the average energy consumption across the six trips in each respective zone. As shown in 9(a), which illustrates the scenario without RB, our DRL solution consistently matched the exhaustive/optimal solution for all three zones, indicating its correctness. Additionally, we observed a consistent energy consumption pattern, with Zone 2 exhibiting the lowest consumption, followed by Zone 1 and then Zone 3, as previously noted during the convergence test. Furthermore, our observations revealed that the second-lowest energy consumption was associated with the Fixed lowest Speed (FLS) approach, which aligns with our expectations, considering that speed significantly contributes to energy consumption. In contrast, aggressive driving behaviour consistently led to higher energy consumption, regardless of the route selected. On the other hand, the random solution fell somewhere in between the FLS approach and the aggressive approach in terms of energy consumption. Figure 9 (b) illustrates the RB scenario where all the approaches show less energy consumption and thus verify the regeneration. In addition, our DRL-based solution matched the exhaustive solution with an energy consumption of only 76 WH rather than 113 WH, with an overall efficiency increase of approximately 33% in all zones. Moreover, the same pattern of the NRB case appears here,
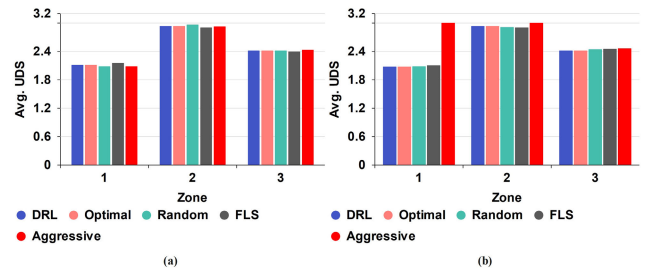
where the second minimum energy is obtained by the FLS approach, followed by the random and aggressive approach.

D. AVERAGE USER DISSATISFACTION TEST

Second, the average user dissatisfaction comparison. We compared the average user dissatisfaction of 6 trips for each zone. Figure 10 (a) depicts the average UDS for each zone for the NRB case, while Figure 10 (b) shows the same but for the RB case. In Figure 10 (a) we noticed that despite our DRL-based solution matching the optimal solution with $UDS = 2.11$ in zone 1, $UDS = 2.93$ in zone 2 and $UDS = 2.41$ in zone 3, it did not yield the minimum UDS. However, it is within the boundaries (i.e., $1 \leqslant UDS \leqslant 3$). In addition, we found a fluctuating trend of UDS where sometimes the random approach results in the minimum value while other times, the aggressive solution and FLS solution have the minimum value, which is expected since these approaches include randomness. Figure 10 (b) showed a similar trend compared to Figure 10 (a), with nearly the same values of UDS in zones 2 and 3, and a slight improvement in zone 1 with $UDS = 2.07$. This result is due to the DRL agent choosing alternative paths for zone 1 to enable the RB, and these paths happen to have low congestion and more shading.

E. AVERAGE TIME TEST

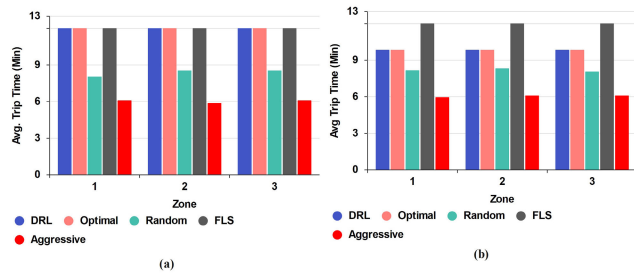Thirdly, Average time comparison. We compared the average time taken for all zones using the different

FIGURE 11. (a) Average time comparison (Non Regenerative braking Scenario). (b) Average time comparison (Regenerative braking Scenario).

FIGURE 12. Effect of reducing trip time threshold on energy consumption for RB and NRB cases.

approaches/solutions in two scenarios which are RB and NRB scenarios. Figure 11 (a) depicts the average time of 6 trips per zone for all solutions. Considering the non-regeneration, the DRL tends to fix the speed to the lowest value while looking for the best path in each time zone to avoid wasting energy, which is the reason why DRL based solution matched the exhaustive and FLS solutions with 12 minutes per trip (i.e., 4 min per segment, corresponding to a speed of $15\frac{km}{h}$). In addition, the aggressive approach showed nearly 50% less time due to high speeds while commuting. On the other hand, Figure 11 (b) illustrates the average time when considering the RB scenario. Regarding this case, the DRL chose the best path while taking advantage of the regeneration to increase the speed in some road segments without losing much energy, which justifies a nearly 18% time improvement.

F. ADDITIONAL ANALYSIS

In this subsection, three additional tests were done, which are limiting trip time, limiting *UDS* and adaptability test. We illustrate how limiting both the time and dissatisfaction thresholds affects energy consumption. Additionally, we show the adaptability and flexibility of the DRL solution when a sudden change occurs in the environment (graph).

1) EFFECT OF LIMITING TRIP TIME ON ENERGY CONSUMPTION

We conducted an experiment to assess the impact of limiting trip duration on energy consumption. We began by gradually reducing the trip time constraint, going from 12 minutes to 9 minutes, which represented a 25% reduction in trip duration. Figure 12 provides a visual representation of this relationship for both scenarios, RB and NRB. From the graph, a clear trend emerges: as the trip time constraint decreases, energy consumption tends to increase. This trend can be attributed to the shorter trip duration, which compels the DRL (Deep Reinforcement Learning) agent to opt for higher speeds. Consequently, the agent selects paths with higher speed limits and pushes the electric scooter to travel faster, even on roads with suboptimal conditions. While this behaviour is aimed at achieving shorter trip times, it is less energy-efficient overall. Specifically, we observed that energy consumption increased by nearly 5Wh (7%) when the trip duration was reduced for RB scenarios. However,
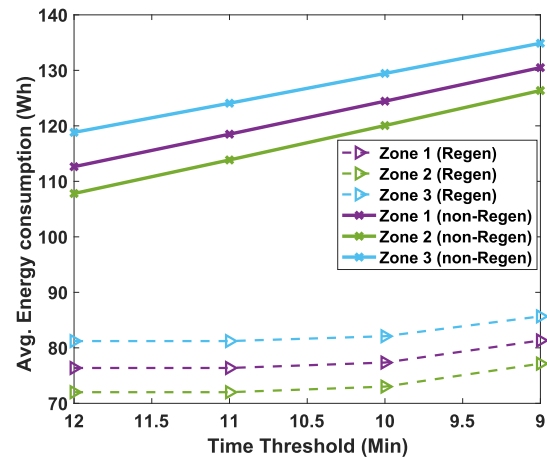
for electric scooters with NRB, the energy consumption increased by about 16Wh (15%). This difference can be explained by the fact that when the trip time was shortened, the DRL agent tended to choose paths with more negative slopes, allowing for regenerative braking. This analysis highlights the trade-off between trip duration and energy consumption in our study and shows the DRL's adaptability in choosing paths and speeds.

2) EFFECT OF LIMITING USER DISSATISFACTION ON ENERGY CONSUMPTION

In this experiment, we systematically reduced the user dissatisfaction threshold (UDS) from 2.93 to 2.43, and the results unveiled a noteworthy relationship between user dissatisfaction and energy consumption. As we lowered the UDS threshold, user dissatisfaction decreased, but we observed that this also led to a higher energy draw from the battery. This outcome can be attributed to the DRL agent's decision-making process, which seeks to minimize user dissatisfaction by avoiding congested roads, low-quality road conditions, and high-temperature areas. Consequently, the DRL agent tends to select paths that, while potentially of higher quality, also entail greater energy consumption, such as inclined roads (as depicted in Figure 13). We specifically chose to showcase this effect in Zone 2 because it was the most impacted by the gradual reduction of UDS. This zone experiences the highest temperatures (resulting in higher UDS values), requiring the DRL agent to expend more effort to find shaded paths. However, this quest for relief from high temperatures may lead to the selection of paths that are not energy-efficient. In our observations, decreasing UDS by 0.5 (a 17% reduction) had a substantial impact on energy consumption. In the RB case, energy consumption increased from 72Wh to 108Wh (a 50% rise), and in the NRB case, it increased from 107Wh to 128Wh (a 19.6% increase). This increase can be attributed to the fact that the chosen paths often have a positive slope that doesn't allow for regenerative
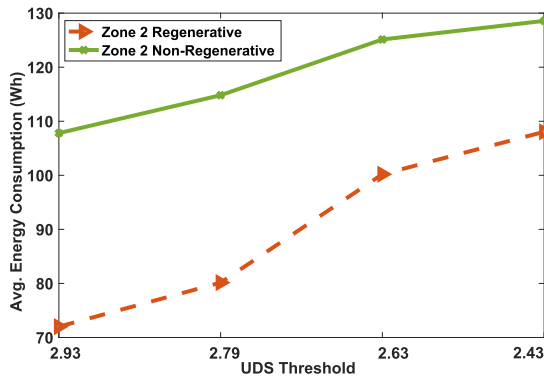
FIGURE 13. Effect of reducing user dissatisfaction on energy consumption.

FIGURE 14. Adaptability test: Introducing traffic congestion.

action to take place, resulting in a significant impact on energy consumption.

3) ENVIRONMENT ADAPTABILITY

In this experiment, we aimed to showcase the versatility of our DRL-based solution by simulating a scenario that reflects real-world challenges. Initially, we allowed DRL algorithm to converge to the optimal route-speed solution for energy efficiency using a training battery with a capacity of 1000 Wh. The initial DRL convergence resulted in an energy consumption of 113 Wh. Once the initial convergence was achieved, we introduced sudden and significant changes in traffic congestion levels on various paths. This abrupt shift in traffic congestion was depicted in Figure 14, illustrating the adaptability test. Our findings demonstrated the remarkable adaptability of the DRL algorithm to these sudden environmental changes. The algorithm swiftly re-converged and adjusted its route choices and speed selection in response to the new conditions, with the DRL convergence after the sudden change resulting in an energy consumption of 123 Wh. This reaffirmed the robustness and adaptability of our solution in the face of dynamic environmental challenges. An interesting observation from the same graph was that the newly adapted solution, while highly adaptive, was less energy-efficient compared to the initial configuration. This is due to the changes made to the environment,

Our previous observations have demonstrated that our DRL solution consistently achieves an optimal solution, just like the exhaustive solution. However, DRL offers several advantages over the exhaustive solution. One significant benefit of our solution is its ability to work under the uncertainty of temperature traffic congestion and take sub-optimal solutions achieving the best overall solution. Moreover, its ability to adapt to changes in the environment and learn the dynamics of the environment without prior knowledge. This contrasts the exhaustive solution, which lacks memory and must evaluate all possible scenarios at each time step. DRL's relative simplicity makes it a more efficient and scalable solution than the exhaustive approach, which can become increasingly complex as the system expands.
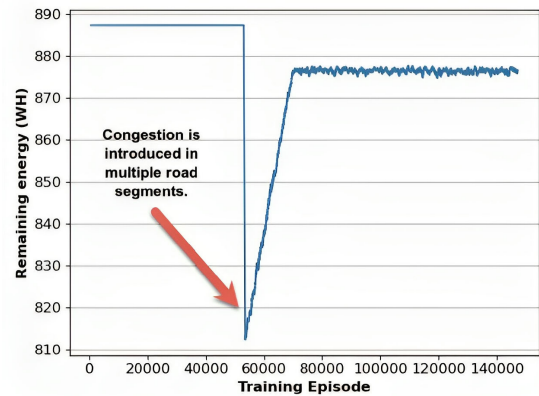
G. SENSITIVITY ANALYSIS TESTS

In this sub-section, we conduct a series of sensitivity analysis tests with the primary objective of understanding the effects of uncertainties in road parameters on both energy consumption and user dissatisfaction. These uncertainties may arise from factors such as sensor inaccuracies or potential exaggerations in user-reported data. A total of four sensitivity tests were performed, where two tests considered the DRL solution, while the other two tests considered the aggressive solution. All these tests were carried out in time zone 1 on the NRB case, specifically for trips from location A to B. The tests are explained in detail as follows.

1) ENERGY CONSUMPTION SENSITIVITY TESTS

For these tests, we specifically examined the influence of errors in road angle and temperature on energy consumption for both our DRL solution and aggressive solution. We fixed the chosen path and speeds for both solutions from location A to location B. For the first test, road angle errors were systematically increased from 10% to 50% within one of the road segments along this path. For the second test, road temperature was increased gradually from 10% to 50% and for the third test, both parameters were increased simultaneously from 10% to 50%, indicating a more severe scenario of having two sensors with failures. The results of both tests are illustrated in 17 and 18, where the former illustrates the energy consumption test for the DRL solution while the latter illustrates the energy consumption test for the aggressive solution. For both tests, introducing an error in the road angle seems to have a slightly more negative impact on energy consumption when compared with the impact of temperature errors. For the DRL case, varying the road angle by 50% has resulted in an energy consumption increase from 27 Wh to 32 Wh (about 18.5% or 5 Wh increase). This is due to the presence of road angle in the gradient energy, which is one of the dominant energies affecting vehicles moving in inclined streets. Varying the temperature by 50% resulted in an energy increase from 27 Wh to 27.5 Wh (about 1.6% or 0.5 Wh increase). This is due to three reasons, one being
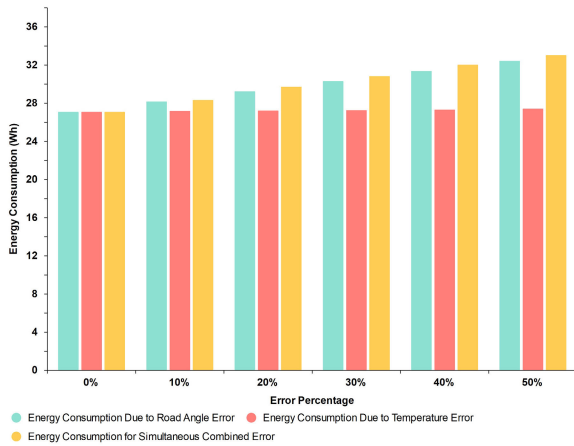
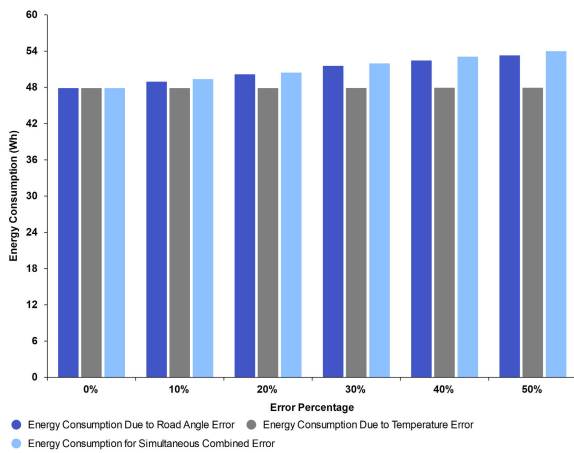**FIGURE 15.** Sensitivity test for energy consumption (DRL case).

FIGURE 16. Sensitivity test for energy consumption (aggressive case).

FIGURE 17. Sensitivity test for user dissatisfaction (DRL case).
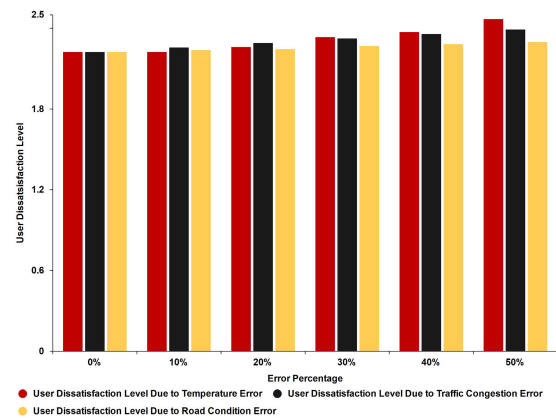
FIGURE 18. Sensitivity test for user dissatisfaction (aggressive case).

that the ES has a low battery capacity and is, therefore, less responsive to temperature fluctuation, and the second reason is that the temperature equation is a polynomial of third order with the lowest energy intensity due to temperatures at the range of $15°C - 35°C$ [40], which is the case in this test and finally, the temperature variation was introduced to only one road segment. However, this effect could become much larger for the case of larger battery-capacity scooters, high-temperature areas and variations in more road segments. Varying the road angle from 10% to 50% resulted in an energy increase from 47.8 Wh to 53.2 Wh (about 11.28% or 5.4 Wh increase). This increase can be seen as small. However, it should be noted that the aggressive solution consumes nearly 77% more energy than the DRL approach. Finally, for the third test, varying both road angle and temperature simultaneously in both approaches (DRL-Aggressive) has resulted in only 1%-5% energy increase compared with varying road angle alone; this is because the road angle effect is dominant in our case, and the temperature has a slight contribution to the energy consumption as explained previously.
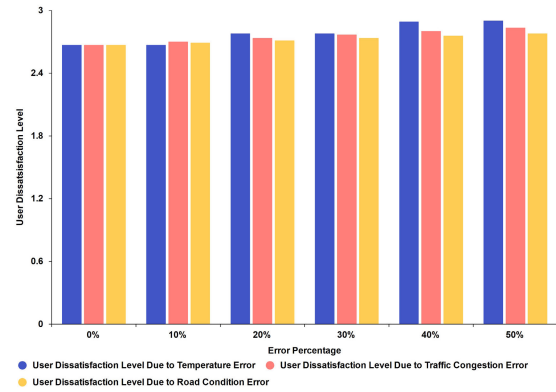
2) USER DISSATISFACTION SENSITIVITY TESTS

In contrast to the previous tests, which primarily examined energy consumption, this test focuses on the impact of errors in temperature, traffic congestion, and road conditions on user dissatisfaction in both the DRL solution and the aggressive approach. As mentioned earlier, we kept the chosen paths and speeds fixed, while introducing an increasing error ranging from 10% to 50% in temperature, traffic congestion, and road condition for one of the road segments. Figures 17 and 18 depict the variations in both the DRL solution and the aggressive approach. From these graphs, it becomes evident that temperature plays the most significant role in influencing user dissatisfaction, followed by traffic congestion and road conditions, in that order. In the case of the DRL solution, user dissatisfaction increased only slightly, from 2.67 to nearly 2.9. Conversely, for the aggressive approach, the user dissatisfaction level rose from approximately 2.2 to 2.46. Comparing the two approaches, the aggressive one exhibited a slight improvement in user dissatisfaction, with a range of approximately 15% to 18% in all tests. It is essential to note that this improvement in user dissatisfaction occurred primarily because the aggressive approach selected routes that were shaded and had better road quality, albeit at the cost of higher energy consumption.
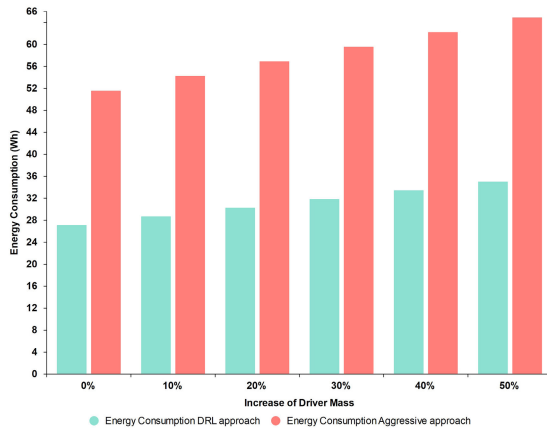
FIGURE 19. Effect of driver's mass on energy consumption for DRL and aggressive approaches.

3) DRIVER MASS AGAINST ENERGY CONSUMPTION

Unlike the previous sensitivity tests that investigated the effect of road parameters on energy consumption and user dissatisfaction, this additional test focuses on the variation of one user parameter affecting ES energy consumption: the driver's mass. In this test, we varied the driver's mass from 10% up to 50%, i.e., from 75kg to 112.5kg. This test was conducted for both the DRL and aggressive approaches, as depicted in Figure 19. The results revealed a substantial impact of the driver's mass on energy consumption, particularly on inclined road segments. When comparing the DRL approach to the aggressive approach, we observed a 29% increase in energy (8 Wh) for the DRL approach, primarily due to its preference for negatively inclined roads. Conversely, the aggressive approach exhibited a 25% increase in energy consumption (13.3 Wh). ESs, with their small size and light weight, are highly sensitive to both user parameters and ES-specific factors, such as mass, tire pressure, frontal area, and more

VII. CONCLUSION

This paper focused on minimizing ES energy consumption by optimally choosing energy-efficient paths and speeds, all while accommodating the inherent variability in time-dynamic environmental conditions and user dissatisfaction levels. Many parameters were considered, such as road condition, road angle, speed limit, road shading, traffic congestion, and temperature fluctuations. In addition, two types of ESs were considered RB and NRB. Three comprehensive validation tests were executed to evaluate our approach, such as energy consumption, average user dissatisfaction, and average trip duration. For each test, we assessed our DRL-based approach against four distinct operational policies closely mirroring strategies commonly employed by ES fleet operators: Optimal(brute force), Random, FLS, and Aggressive solutions. The results of the proposed DRL approach matched the brute force solution for both RB and NRB cases. In the RB case, the energy

consumption was averaged 76 Wh, while in the NRB case it was averaged 113 Wh. Notably, the DRL approach exhibited a significant energy consumption reduction of 53-67% in RB case and 25-55% in NRB case when compared to the other solutions.

Furthermore, we conducted three additional tests to thoroughly assess the flexibility of the DRL solution. The first two tests examined the impact of limiting trip time and user dissatisfaction on energy consumption. The third test introduced abrupt traffic congestion, spotlighting the solution's adeptness in adapting to dynamic environmental changes. Following that, we implemented three sensitivity tests to evaluate the consequences of inaccuracies in reading specific road parameters on both energy consumption and user dissatisfaction, considering our DRL and aggressive approaches. The initial test explored the implications of introducing errors in road angle and temperature on energy consumption, while the second test analyzed the effects of errors in road temperature, traffic congestion, and road conditions on user dissatisfaction. The third test appraised the influence of introducing errors in the driver's mass on energy consumption. All tests confirm the robustness of our solution against environmental changes.

A. POTENTIAL LIMITATIONS AND FUTURE WORK

When extending the solution methodology to other types of MM, there are several critical considerations:

1) The vehicle's frontal area, whether the driver is standing, sitting, or in a closed MM vehicle. Each MM vehicle has a distinct frontal area, impacting the energy required to overcome air friction and overall energy consumption. The solution should account for these variations to optimize energy usage accurately.

2) The auxiliary energy of the vehicle, as some MM vehicles are closed and require heating/cooling, eventually affecting the battery range. In addition, historical data of their energy consumption should be obtained for higher accuracy.

3) Energy consumption due to temperature influence shall not pose a problem, as the battery capacity of such vehicles is considerably small (i.e., less than 1KWh); therefore, the temperature effect shall be minimal and eventually neglected.

4) User dissatisfaction model can be customized based on the user preference of each MM vehicle, as example, closed MM vehicles are not much affected by the temperature factor as they might have an air conditioning system, therefore the temperature factor can be omitted or has a less weight compared to other factors, and the same for other MM vehicles.

As for future work, many enhancements can be considered as follows:

1) Considering and investigating of multi-agent deep reinforcement learning in managing fleets of ES and

optimizing their communication and sharing network updates instead of relying on a global SDN control.
2) Including the constraints that are related to the battery life cycle and depth of discharge.
3) Collaborating with e-mobility companies to obtain energy consumption data of MM and build a more detailed energy consumption model for MM that combines motion dynamic consumption, traffic, and environmental impact.
4) Surveying regular users of e-mobility services to identify the causes of their dissatisfaction and factoring those findings into the aforementioned constraints.

REFERENCES

[1] B. Fong, A. C. M. Fong, and G. Y. Hong, "Sustainable micromobility management in smart cities," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 12, pp. 15890–15896, Dec. 2023.

[2] D. Goswami, A. Riggins, and D. A. Paley, "Data-driven prediction of urban micromobility: A study of dockless electric scooters [applications of control]," *IEEE Control Syst. Mag.*, vol. 42, no. 5, pp. 18–31, Oct. 2022.

[3] S. He and K. G. Shin, "Distribution prediction for reconfiguring urban dockless e-scooter sharing systems," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 5722–5740, Dec. 2022.

[4] A. Jaber, J. Hamadneh, and B. Csonka, "The preferences of shared micro-mobility users in urban areas," *IEEE Access*, vol. 11, pp. 74458–74472, 2023, doi: 10.1109/ACCESS.2023.3297083.

[5] S. S. Sayed and A. M. Massoud, "Review on state-of-the-art unidirectional non-isolated power factor correction converters for short-/long-distance electric vehicles," *IEEE Access*, vol. 10, pp. 11308–11340, 2022.

[6] G. M. Scheepmaker and R. M. P. Goverde, "Energy-efficient train control using nonlinear bounded regenerative braking," *Transp. Res. C, Emerg. Technol.*, vol. 121, Dec. 2020, Art. no. 102852.

[7] S. Heydari, P. Fajri, R. Sabzehgar, and A. Asrari, "Optimal brake allocation in electric vehicles for maximizing energy harvesting during braking," *IEEE Trans. Energy Convers.*, vol. 35, no. 4, pp. 1806–1814, Dec. 2020.

[8] S. Heydari, P. Fajri, R. Sabzehgar, and M. Rasouli, "A novel approach for maximizing regenerative braking energy extraction of electric vehicles using motor performance lookup table," in *Proc. IEEE Transp. Electrific. Conf. Expo. (ITEC)*, Jun. 2019, pp. 1–5.

[9] G. Sandrini, D. Chindamo, and M. Gadola, "Regenerative braking logic that maximizes energy recovery ensuring the vehicle stability," *Energies*, vol. 15, no. 16, p. 5846, Aug. 2022.

[10] B. Prasanth, R. Paul, D. Kaliyaperumal, R. Kannan, Y. V. P. Kumar, M. K. Chakravarthi, and N. Venkatesan, "Maximizing regenerative braking energy harnessing in electric vehicles using machine learning techniques," *Electronics*, vol. 12, no. 5, p. 1119, Feb. 2023.

[11] Y. Farajpour, H. Chaoui, M. Khayamy, S. Kelouwani, and M. Alzayed, "Novel energy management strategy for electric vehicles to improve driving range," *IEEE Trans. Veh. Technol.*, vol. 72, no. 2, pp. 1735–1747, Feb. 2023.

[12] X. Meng, D. Hao, R. Wang, Y. Xu, Z. Wei, and L. Zhang, "Research on the energy management strategy of range extended fuel cell electric vehicle," in *Proc. IEEE 4th Inf. Technol., Netw., Electron. Autom. Control Conf. (ITNEC)*, Jun. 2020, pp. 503–507.

[13] S. Zhang, Y. Luo, K. Li, and V. Li, "Real-time energy-efficient control for fully electric vehicles based on an explicit model predictive control method," *IEEE Trans. Veh. Technol.*, vol. 67, no. 6, pp. 4693–4701, Jun. 2018.

[14] H. Chu, L. Guo, B. Gao, H. Chen, N. Bian, and J. Zhou, "Predictive cruise control using high-definition map and real vehicle implementation," *IEEE Trans. Veh. Technol.*, vol. 67, no. 12, pp. 11377–11389, Dec. 2018.

[15] J. Zhang, S. Dong, Z. Li, B. Ran, R. Li, and H. Wang, "An eco-driving signal control model for divisible electric platoons in cooperative vehicle-infrastructure systems," *IEEE Access*, vol. 7, pp. 83277–83285, 2019.

[16] Y. Zhang, Y. Zhang, Z. Ai, Y. L. Murphey, and J. Zhang, "Energy optimal control of motor drive system for extending ranges of electric vehicles," *IEEE Trans. Ind. Electron.*, vol. 68, no. 2, pp. 1728–1738, Feb. 2021.

[17] H. Wu, J. Zhang, Z. Cai, F. Liu, Y. Li, and A. Liu, "Toward energy-aware caching for intelligent connected vehicles," *IEEE Internet Things J.*, vol. 7, no. 9, pp. 8157–8166, Sep. 2020.

[18] K. Vatanparvar, S. Faezi, I. Burago, M. Levorato, and M. A. Al Faruque, "Extended range electric vehicle with driving behavior estimation in energy management," *IEEE Trans. Smart Grid*, vol. 10, no. 3, pp. 2959–2968, May 2019.

[19] K. Kivekäs, A. Lajunen, F. Baldi, J. Vepsäläinen, and K. Tammi, "Reducing the energy consumption of electric buses with design choices and predictive driving," *IEEE Trans. Veh. Technol.*, vol. 68, no. 12, pp. 11409–11419, Dec. 2019.

[20] M. Janulin, O. Vrublevskyi, and A. Prokhorenko, "Energy minimization in city electric vehicle using optimized multi-speed transmission," *Int. J. Automot. Mech. Eng.*, vol. 19, no. 2, pp. 9721–9733, Jun. 2022.

[21] M. Elmahallawy, T. Elfouly, A. Alouani, and A. M. Massoud, "A comprehensive review of lithium-ion batteries modeling, and state of health and remaining useful lifetime prediction," *IEEE Access*, vol. 10, pp. 119040–119070, 2022.

[22] M. Tang, B. Li, H. Liu, W. Zhuang, Z. Li, and J. Peng, "Energy-oriented routing strategy of electric vehicle: An end-to-end reinforcement learning approach," in *Proc. 6th CAA Int. Conf. Veh. Control Intell. (CVCI)*, Oct. 2022, pp. 1–7.

[23] O. Trigui, E. Mejri, Y. Dube, S. Kelouwani, and K. Agbossou, "Energy efficient routing estimation in electric vehicle with online rolling resistance estimation," in *Proc. IEEE Vehicle Power Propuls. Conf. (VPPC)*, Dec. 2017, pp. 1–6.

[24] A. M. Bozorgi, M. Farasat, and A. Mahmoud, "A time and energy efficient routing algorithm for electric vehicles based on historical driving data," *IEEE Trans. Intell. Vehicles*, vol. 2, no. 4, pp. 308–320, Dec. 2017.

[25] M. T. Dabiri, M. Hasna, N. Zorba, and T. Khattab, "Optimal trajectory and positioning of UAVs for small cell HetNets: Geometrical analysis and reinforcement learning approach," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 2667–2683, 2023.

[26] Q. Hou and J. Dong, "Distributed dynamic event-triggered consensus control for multiagent systems with guaranteed L_2 performance and positive inter-event times," *IEEE Trans. Autom. Sci. Eng.*, vol. 21, no. 1, pp. 746–757, Dec. 2022.

[27] C. Deng, "Cooperative fault-tolerant output regulation of linear heterogeneous multiagent systems under directed network topology," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 51, no. 8, pp. 4773–4781, Aug. 2021.

[28] D. Qiu, Y. Wang, W. Hua, and G. Strbac, "Reinforcement learning for electric vehicle applications in power systems: A critical review," *Renew. Sustain. Energy Rev.*, vol. 173, Mar. 2023, Art. no. 113052.

[29] *Passenger Car, Truck, Bus and Motorcycle Tyres—Methods of Measuring Rolling Resistance*, Standard ISO 18164:2005, 2005. Accessed: Aug. 16, 2022. [Online]. Available: https://www.iso.org/obp/ui/

[30] C. Beckers, I. Besselink, and H. Nijmeijer, "Combined rolling resistance and road grade estimation based on EV powertrain data," *IEEE Trans. Veh. Technol.*, vol. 72, no. 3, pp. 3201–3213, Mar. 2023, doi: 10.1109/TVT.2022.3220157.

[31] B. Świeczko-Żurek, J. Ejsmont, and G. Ronowski, "Reduction of tire rolling resistance by optimization of road surfaces and tires," in *Proc. Troisième Congrès Tunisien Mécanique*, 2014, pp. 24–26.

[32] J. Ejsmont, S. Taryma, G. Ronowski, and B. Swieczko-Zurek, "Influence of load and inflation pressure on the tyre rolling resistance," *Int. J. Automot. Technol.*, vol. 17, no. 2, pp. 237–244, Apr. 2016.

[33] J.-B. Wang, K. Liu, T. Yamamoto, and T. Morikawa, "Improving estimation accuracy for electric vehicle energy consumption considering the effects of ambient temperature," *Energy Proc.*, vol. 105, pp. 2904–2909, May 2017.

[34] L. Guzzella and A. Sciarretta, *Vehicle Propulsion Systems*, vol. 1, 2nd ed. Berlin, Germany: Springer, 2007.

[35] D. G. Wilson and T. Schmidt, *Bicycling Science*. Cambridge, MA, USA: MIT Press, 2020.

[36] J. Asamer, A. Graser, B. Heilmann, and M. Ruthmair, "Sensitivity analysis for energy demand estimation of electric vehicles," *Transp. Res. D, Transp. Environ.*, vol. 46, pp. 182–199, Jul. 2016.

[37] F. Cavallini and F. Crisciani, *Quasi-Geostrophic Theory of Oceans and Atmosphere: Topics in the Dynamics and Thermodynamics of the Fluid Earth*, vol. 45. Dordrecht, The Netherlands: Springer, 2012.

[38] B. Benders, K. Winther, M. Holst, and C. Kolf, "Performance validation—Results from EV measurements," *Green Emotion*, vol. 1, no. 1, pp. 124–127, 2014.

[39] A. Lahlou, F. Ossart, E. Boudard, F. Roy, and M. Bakhouya, "Optimal management of thermal comfort and driving range in electric vehicles," *Energies*, vol. 13, no. 17, p. 4471, Aug. 2020.

[40] H. Hamwi, T. Rushby, M. Mahdy, and A. S. Bahaj, "Effects of high ambient temperature on electric vehicle efficiency and range: Case study of Kuwait," *Energies*, vol. 15, no. 9, p. 3178, Apr. 2022.

[41] AccuWeather. (Mar. 17, 2023). *Doha, Doha, Qatar Current Weather*. [Online]. Available: https://www.accuweather.com/en/qa/doha/271669/current-weather/271669

[42] J. L. Kephart, B. N. Sánchez, J. Moore, L. H. Schinasi, M. Bakhtsiyarava, Y. Ju, N. Gouveia, W. T. Caiaffa, I. Dronova, S. Arunachalam, A. V. D. Roux, and D. A. Rodríguez, "City-level impact of extreme temperatures and mortality in Latin America," *Nature Med.*, vol. 28, no. 8, pp. 1700–1705, Aug. 2022.

[43] Y. Hua, Y. Qiu, and X. Tan, "The effects of temperature on mental health: Evidence from China," *J. Population Econ.*, vol. 36, no. 3, pp. 1293–1332, Jul. 2023.

[44] G. Li, W. Lai, X. Sui, X. Li, X. Qu, T. Zhang, and Y. Li, "Influence of traffic congestion on driver behavior in post-congestion driving," *Accident Anal. Prevention*, vol. 141, Oct. 2020, Art. no. 105508.

[45] A. S. El-Wakeel, A. Noureldin, H. S. Hassanein, and N. Zorba, "IDriveSense: Dynamic route planning involving roads quality information," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–6.

[46] E. Grigorieva and A. Lukyanets, "Combined effect of hot weather and outdoor air pollution on respiratory health: Literature review," *Atmosphere*, vol. 12, no. 6, p. 790, Jun. 2021.

[47] M. Aizpuru, K. X. Farley, J. C. Rojas, R. S. Crawford, T. J. Moore, and E. R. Wagner, "Motorized scooter injuries in the era of scooter-shares: A review of the national electronic surveillance system," *Amer. J. Emergency Med.*, vol. 37, no. 6, pp. 1133–1138, Jun. 2019.

[48] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.

[49] H. M. Abdullah, A. Gastli, and L. Ben-Brahim, "Reinforcement learning based EV charging management systems—A review," *IEEE Access*, vol. 9, pp. 41506–41531, 2021, doi: 10.1109/ACCESS.2021.3064354.

[50] D. T. Hoang, N. V. Huynh, D. N. Nguyen, E. Hossain, and D. Niyato, *Deep Reinforcement Learning Models and Techniques*. Hoboken, NJ, USA: Wiley, 2023, pp. 37–82, doi: 10.1002/9781119873747.

[51] A. Santra, S. Hazra, L. Servadei, T. Stadelmayer, M. Stephan, and A. Dubey, *Deep Reinforcement Learning*. Hoboken, NJ, USA: Wiley, 2023, pp. 115–150, doi: 10.1002/9781119910695.

[52] S. Singh, T. Jaakkola, and M. Jordan, "Reinforcement learning with soft state aggregation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 7, 1994, pp. 361–368.

[53] H. Hasselt, "Double Q-learning," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 23, 2010, pp. 2613–2621.

[54] J. Wang and L. Sun, "Robust dynamic bus control: A distributional multi-agent reinforcement learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 4, pp. 4075–4088, Apr. 2023, doi: 10.1109/TITS.2022.3229527.

[55] J. Xie and W. Sun, "Distributional deep reinforcement learning-based emergency frequency control," *IEEE Trans. Power Syst.*, vol. 37, no. 4, pp. 2720–2730, Jul. 2022, doi: 10.1109/TPWRS.2021.3130413.

[56] K. Min, H. Kim, and K. Huh, "Deep distributional reinforcement learning based high-level driving policy determination," *IEEE Trans. Intell. Vehicles*, vol. 4, no. 3, pp. 416–424, Sep. 2019, doi: 10.1109/TIV.2019.2919467.

KARIM ABOELENEEN received the B.Sc. degree in electrical engineering from Qatar University, Qatar, in 2018, where he is currently pursuing the M.Sc. degree. His research interests include machine learning, optimization, smart grids, and power systems.

NIZAR ZORBA (Senior Member, IEEE) received the B.Sc. degree in electrical engineering from JUST University, Jordan, in 2002, and the Ph.D. degree in signal processing for communications from UPC Barcelona, Spain, in 2007. He is currently a Professor with the Electrical Engineering Department, Qatar University, Doha, Qatar. He has authored six international patents, two books, contributed to seven book chapters, and coauthored over 170 papers in peer-reviewed journals and international conferences. He is a Former Chair of the IEEE Communication Society Technical Committee on Communications Systems Integration and Modeling. He is the Associate Editor-in-Chief of *IEEE Communications Magazine*, an Area Editor of IEEE Communications Letters, and an Associate Editor of IEEE Transactions on Cognitive Communications and Networking. He is also an IEEE ComSoc Distinguished Lecturer, during 2024–2025.

AHMED M. MASSOUD (Senior Member, IEEE) received the B.Sc. (Hons.) and M.Sc. degrees in electrical engineering from Alexandria University, Egypt, in 1997 and 2000, respectively, and the Ph.D. degree in electrical engineering from Heriot-Watt University, Edinburgh, U.K., in 2004. He is currently the Associate Dean for Research and Graduate Studies with the College of Engineering, Qatar University, Qatar, and a Professor with the Department of Electrical Engineering, College of Engineering, Qatar University. He holds 12 U.S. patents. He published more than 130 journal articles in the fields of power electronics, energy conversion, and power quality. He supervised several M.Sc. and Ph.D. students at Qatar University. He has been awarded several research grants addressing research areas, such as energy storage systems, renewable energy sources, HVDC systems, electric vehicles, pulsed power applications, and power electronics for aerospace applications. His research interests include power electronics, energy conversion, renewable energy, and power quality.

• • •