Optimizing Electric Vehicle Charging Routes with Deep Reinforcement Learning Considering Driver Preferences

Shivendu Mishra, Anurag Choubey, Harshit Dhankhar, Sri Vaibhav Devarasetty, and Rajiv Misra

***Abstract*—Navigating electric vehicle (EV) charging can be challenging due to factors like limited batteries and uncertainties like traffic, user habits, costs, and charging station availability. This paper proposes a practical method for improving EV charging navigation by focusing on driver preferences, such as minimizing cost or distance travelled. Deep reinforcement learning (DRL), a widely used method, has successfully managed EV charging dynamics. We utilized three RL models: DQN, DDQN, and PPO. Our approach led to reduced overall costs, covering travel expenses, charging costs, and waiting costs. Moreover, we optimized the overall distance travelled by considering both the distance to reach charging stations and the distance travelled without visiting charging stations. Experiments conducted under normal and uniform distributions and various EV charging requests showed that the PPO model surpasses DQN and DDQN in minimizing cost (travel, waiting, and charging) and distance travelled, all while considering driver preferences.**

***Index Terms*—Charging navigation, Driver preferences, Deep reinforcement learning, Intelligent transport systems, Plug-in electric vehicles (PEVs), Proximal Policy Optimization.**

I. Introduction

PLUG-in Electric Vehicles (PEVs) have recently gained recognition as a promising and environmentally conscious mode of transportation, mitigating the environmental impacts of traditional petroleum-based energy sources [1]. While electric vehicles have a design that eliminates the need for petroleum and coal while emitting zero greenhouse gases, the increased use of EVs raises a potential issue. Depending on user behaviour, the energy demand for charging could become concentrated within specific time frames, leading to significant spikes in demand. This event could eventually cause increased power losses, voltage fluctuations, and pressure on the power grid, compromising the effectiveness of power plants and adversely affecting overall reliability and stability [2].

Establishing charging infrastructure for EVs, called charging stations, is critical. These stations obtain electricity from the grid at a low cost to generate revenue by reselling this electricity to EVs at a higher price point [3]. The optimization of charging schedules is a critical way to lower the cost of charging EVs. Many utility companies have made steps by implementing real-time power rates, which strategically encourage EV owners to recharge their vehicles during periods of low demand. The ability of an EV to generate revenue by feeding extra electricity back into the electrical grid while operating in vehicle-to-grid (V2G) mode is a fascinating feature in this arena. However, the environment is challenging because EV variables such as energy consumption, departure and arrival times, and electricity costs are dynamic and time-varying. The interaction of uncertainties caused by traffic conditions, user behaviour, and the pricing mechanisms of electricity providers adds to the complication. As a result, coordinating efficient control over EV charging schedules to reduce costs becomes an intricate task [4]–[6].

The management system steering EV charging is crucial to the above intricate task. Various strategies have emerged from an extensive number of studies aimed at optimizing EVs' scheduling and operational paradigms. In most of these solutions, dynamic electricity pricing takes centre stage, acting as a means to mitigate peaks in power demand. However, many of these studies focus on scenarios where EVs remain parked for extended periods at home or in parking lots, allowing for a comprehensive measurement of charging times and rates [7]–[13]. However, the environment of EV usage is not limited to stationary scenarios. Due to the inherent limitations of EV battery capacity, charging services are frequently required, even during short-distance travel. Creating navigation strategies for moving EVs presents a unique set of challenges and includes factors like the EV's current location, charging mode (rapid or gradual), and battery charge level. Additional factors include user behaviour's stochastic nature, traffic conditions' fluidity, charging station wait times, and associated costs. In light of this complex tapestry, numerous studies [14]–[19] propose innovative EV navigation and charging systems. Yet, they often overlook uncertainties like traffic conditions and charging station wait times, necessitating adaptive approaches. RL offers a dynamic solution in uncertain environments.

RL effectively addresses complexities in energy management and cost reduction domains for electric vehicles, charging

Shivendu Mishra is with the Department of Computer Science and Engineering, Indian Institute of Technology, Patna, 801106, Bihar, India and the Department of Information Technology, Rajkiya Engineering College Ambedkar Nagar, Ambedkar Nagar, 224122, Uttar Pradesh, India. (e-mail: shivendu_2021cs08@iitp.ac.in).

Anurag Choubey is with the Department of Computer Science and Engineering, Indian Institute of Technology, Patna, 801106, Bihar, India and School of Computer Science Engineering and Technology, Bennett University, Greater Noida, 201310, Uttar Pradesh, India (e-mail: anurag.pcs17@iitp.ac.in).

Harshit Dhankhar is with the Department of Mathematics, Indian Institute of Technology, Patna, 801106, Bihar, India (e-mail: harshit_2101mc20@iitp.ac.in).

Sri Vaibhav Devarasetty is with the Department of Computer Science and Engineering, Indian Institute of Technology, Patna, 801106, Bihar, India (e-mail: devarasetty_2101cs24@iitp.ac.in).

Rajiv Misra is with the Department of Computer Science and Engineering, Indian Institute of Technology, Patna, 801106, Bihar, India (e-mail: rajivm@iitp.ac.in).

stations, and intelligent buildings [8], [20]–[29]. However, charging station selection remains relatively unexplored, with only a few studies delving into this domain, as seen in studies like [30] and [31], though scalability remains a concern [31], [32]. Recent works like [33] address EV driver preference and scale RL for citywide charging management, while others [27], [34], [35] also address EV driver preferences.

However, these studies on electric vehicle charging planning primarily focus on driver preference for reducing charging time, minimizing travel distance, selecting preferred charging stations, and identifying optimal charging times. They often overlook the drivers' preferences for minimizing distance or reducing charging costs. Our study addresses this gap by considering both aspects, demonstrating how incorporating drivers' preferences for either cost savings or distance coverage can optimize charging schedules.

The proposed work offers several significant contributions, including:

i. Introduces a novel optimal EV charging scheduling and navigation method based on model-free deep reinforcement learning. The suggested strategy prioritizes driver preferences, such as cost or distance travelled, to provide a personalized navigation experience.
ii. The proposed approach reduces overall costs, including driving, charging, waiting, and distance costs, making EV usage more feasible for drivers/owners.
iii. Three RL models, namely DQN, DDQN, and PPO, were utilized to optimize the charging navigation process, providing flexibility and robustness to the proposed approach.
iv. Experiments conducted under various scenarios, including normal and uniform distributions with '80', '100', '120', and '140' EV charging requests, demonstrate the effectiveness of the suggested approach. The experiments reveal that the PPO model surpasses DQN and DDQN in terms of cost and distance travelled while considering driver preferences.
v. The proposed approach outperforms several benchmark strategies, including DQN with driver preference, DQN without driver preference, MDS, MTTS, and MWTS, with performance improvements of approximately 24.54 %, 62.58 %, 73.84 %, 70.30 %, and 65.14 %, respectively, in the uniform distribution case. Under normal distribution, it achieves even greater improvements of 76.28 %, 69.58 %, 78.33 %, 71.10 %, and 75.66 %, respectively.

Table I displays annotations and their meanings. The subsequent sections of this paper are structured as follows: Section II provides an overview of related works. Section III discusses the proposed DRL-based optimal charging scheduling and navigation. Section IV presents a case study to illustrate the efficacy of the proposed approach. Section V presents thorough discussions of experimental results, highlighting the efficiency of the suggested approach. Section VI presents the conclusions.

TABLE I: Annotations and their meanings.

Annotations	Meanings
EV, DQN, RL, CSNP	Electric vehicle, Deep Q-network, reinforcement learning, and charging scheduling & navigation problem, respectively
DRL, DDQN, PPO	Deep reinforcement learning, Double DQN, and Proximal Policy Optimization, respectively
t, SP, DL, CDSP	Current time, starting position, destination location, and charging-discharging scheduling problem, respectively
s^{t+1}, a^t, s^t, r^t	Next state, action, Current state, and reward, respectively
MDS, MTTS, MWTS	Minimum distance, minimum travel time, and minimum waiting time strategies, respectively
$CR^t, WT^t, AT^t, DT^t, DD^t$	EV charging request, expected waiting time, arrival time, driving time, and driving distance, respectively, at time t.
Pref., M, SOC, CSP	Driver preference, the set of all possible actions, state of charge, and charging scheduling problem, respectively
MDP, CSNS, RSU, ITS	Markov decision process, charging station navigation system, roadside unit, the intelligent transportation system (network), respectively
CSs, $C_{ch}, C_{drive}, C_{wait}$	Charging stations, Charging cost, driving cost, waiting cost, respectively
P, $\pi, \gamma, \hat{\alpha}$	parameter, policy, discount factor, daily driver wage per unit of time, respectively
$T_r, C_{dist.}$, TRPO	Charging request time, distance cost, and trust region policy optimization, respectively
Cr^t, D_{total}, λ	cumulative reward at time t, distance cost, energy consumption per distance unit, respectively

II. Related work

In [14], an integrated rapid charging navigation system is presented, combining a power system control centre, intelligent transportation system centre, EV terminals, and charging stations. This collaboration enhances EV technology by seamlessly integrating transportation and power domains. Similarly, [15] introduces an electric vehicle navigation system supported by a vehicle ad-hoc network and a hierarchical architecture, with a traffic information centre overseeing operations. Additionally, authors in [16] propose a charging navigation paradigm integrating real-time crowd sensing and EV route selection facilitated by a central control centre and charging stations. In the paper, [17], an integrated EV navigation system employs a hierarchical game model to optimize transportation and power infrastructures, with charging station dynamics modelled as a non-cooperative game. Finally, works in [18] suggest an IoT-based EV rapid charging strategy to mitigate power grid strain, dynamically setting charging prices based on power regulation and real-time traffic data.

A hybrid charging management system is a clever idea suitable for urban EV taxis using the knowledge provided in the work [19]. A combination of EVs, charging stations, battery-swapping facilities, and a stern global controller stands out in this context. The global controller plays a vital role in this architecture, managing real-time data from charging and swapping stations. This data-driven analysis is critical in determining the optimal charging or swapping station and optimizing the EV taxi fleet's efficiency. However, because the earlier methods operate in a deterministic framework, they ignore essential uncertainties like the constantly changing nature of traffic conditions and the fluctuating lengths of time drivers must wait at charging stations. These uncertainties significantly impact the effectiveness of both route and charging station selection strategies, necessitating a more adaptive approach to guarantee optimal performance.

In complex situations, relying solely on existing knowledge of uncertainty is insufficient. There has recently been

a greater emphasis on proposing learning-centred techniques for identifying optimal charging stations for electric vehicles in urban settings. To address this issue, Reinforcement Learning excels at tackling complex decision-making challenges while avoiding the need for advanced knowledge in uncertainty. In this context, MDP provides a versatile structure for depicting decision-related situations in unpredictable and uncertain environments. MDP is the foundation for reinforcement learning, providing a stable platform for making optimal decisions in the face of uncertainty. While the reinforcement Learning scenery has grown, its focus has frequently shown energy management and cost reduction domains for electric vehicles, charging stations, and intelligent buildings [8], [20]–[29]. Surprisingly, charging station selection still needs to be explored, with only a few studies stepping into this domain. The authors in the work [30] introduce an innovative approach to alleviate congestion in the CS allocation that has been presented, employing Q-learning. The study comprehensively considers both travel and queuing time within CSs, culminating in developing a cohesive joint-resource congestion game. This model effectively captures the dynamic interplay between EVs and available resources. By harnessing the power of the Q-learning algorithm, the authors have successfully tackled the challenge and resolved the problem at hand. Notably, a ground-breaking project described in [31] proposes a novel approach to EV charging navigation. This approach uses reinforcement learning to select charging tactics that reduce charging costs and time, aligning with optimizing the EV charging experience. This novel approach effectively addresses the challenge of route and charging station selection, operating independently without needing prior knowledge of variables such as traffic patterns, charging costs, or waiting durations.

Nonetheless, it is worth noting that the proposed strategy only considers the path from the origin to the chosen charging station. This narrow focus has a chance to increase network complexity due to the intricate feature extraction process necessitated by inter-node movements, which is frequently accomplished through optimisation methodologies. Furthermore, it is critical to recognise that the efficacy of the above-suggested approach depends on several variables, including the number of electric vehicles using the navigation system and the inherent uncertainty surrounding future EV charging demands. These dimensions, while critical, have yet to be considered in the approach, leaving room for further refinements and enhancements to ensure a practical solution that accounts for the complexities of real-world EV usage scenarios. The authors introduce a cutting-edge paradigm for route and charging station selection (RCS) guided by deep reinforcement learning, known for its model-free nature, in their ground-breaking work documented in [32]. Considering the inherent unpredictability of traffic conditions and the constantly shifting environment of arrival charging requests, this RCS algorithm emerges as a sign of efficiency, carefully minimizing the collective travel time, including charging duration, waiting, and driving times.

However, it is essential to note that while both solutions [31], [32] are remarkable in their efficacy, they have limitations in terms of scalability. The scope of the assessment was limited to relatively minor instances, with graphs consisting of only 39 nodes and a tiny 3-charging station. While these examples provide valuable insights into the procedure's capabilities, their limited scope indicates the potential for further development and exploration into more complex scenarios. Zhang et al. [33] use deep reinforcement learning to manage charging schedules on a larger scale, as demonstrated by their study of a large city with over 1,000 charging stations. Notably, their methodology includes the use of Dijkstra's algorithm [36] for route selection as well as a primitive energy consumption model that is primarily based on travel distance. While this configuration is suitable for city navigation, the demands of long-distance navigation necessitate more complex modelling. Factors such as the unpredictability of future traffic conditions, the dynamic nature of incoming charge requests, and real-time electricity pricing considerations must be included for a more comprehensive approach. These factors contribute to a thorough strategy that optimizes scheduling and increases drivers' earnings by allowing for more daily trips. By incorporating these elements into the scheduling framework, the resulting approach promises to lower charging costs and wait times, resulting in a holistic solution sensitive to the complexities of dynamic and evolving transportation scenarios.

None of those above studies consider EV driver/owner preferences. Recent literature, such as [27], [33]–[35], addresses this aspect. Authors in [33] used RL for charging navigation and considered driver preferences for reducing total charging time or minimizing the origin-to-destination distance. In their study [34], the authors present a comprehensive three-stage bi-objective model aimed at minimizing installation costs while maximizing coverage of EV charging stations. They consider driver risk-taking behaviour and route selection policies to determine the optimal locations and allocation of charging stations effectively. The authors in [27] aim to predict the best EV charging time slots a day in advance, reducing energy costs and waiting times at CS while ensuring EV batteries' fast and complete charging. It also considers EV driver choices for charging location, connector type, and preferred charging time of day. In [35], authors explore various EV driver charging preferences, including as fast as possible, as late as possible, peak-shaving, shared DCFC station use, and valley-filling. They assess how these preferences influence charging patterns, peak power, grid load, and driver flexibility, going beyond mere cost minimization.

TABLE II: Comparison of RL-based EV Charging Scheduling and Navigation Methods.

Reference	Year	Problem Type	Objective Type	Methods	Driver Preference	Network Topology	Station Location	Waiting Time	EV-Charging Navigation
[8]	2019	CDSP	Min. Charging Cost	DQN	NO	NA	NO	NO	NO
[23]	2019	CSP	Min. Charging Cost	Fitted Q-iteration	NO	NA	NO	NO	NO
[24]	2019	CSP	Max. Profit of EVCS	SARSA	NO	NA	NO	NO	NO
[31]	2019	CSNP	Min. Total Travel Time and Charging Cost	DQN	NO	Xi'an City, China	Yes	Normal Distribution	Navigate an EV to EVCS
[32]	2020	CSNP	Min. Total Travel Time of Multiple EVs	DQN	NO	Xi'an City, China	Yes	Estimated waiting time	Navigate Multiple EVs to destination via EVCS
[33]	2020	CSNP	Total Charging Time for EVs and Reduction in OD distance	DQN	Reducing Total Charging Time or OD Distance	Charging Stations in Beijing	Yes	Estimated waiting time	Navigate Multiple EVs to Destination via CS
[27]	2023	CSP	Max. EV profit (charging Cost Reduction, Waiting Time Minimization)	DQN	Preferred CS, Best Charging Time	NO	Yes	Estimated Waiting Time	NO
[29]	2024	CSP	Minimize Operational Costs of Charging Stations, Ensuring QoS	DQN, PPO	NO	NA	NO	NO	NO
[28]	2024	CDSP	Min. Charging Cost	DQN, DDPG	NO	NA	NO	NO	NO
[37]	2024	CSNP	EV Charging Destination Optimization, Charging Route	HEDQN	NO	34-Node, 108-Node Network	Yes	Normal Distribution	Navigate an EV to EVCS
[25]	2020	CSP	Maximize operator Profits	DDPG	NO	33-bus Network	NO	NO	NO
[26]	2023	CSP	EV Charging Target, Voltage Stability	DDPG, DQN, PPO, SAC	NO	IEEE 33-Node System	NO	NO	NO
Proposed	-	CSNP	Min. Total Cost (Travel, Waiting, Charging and Distance Cost)	PPO, DQN, DDPG	Reducing Cost or Distance	Xi'an City, China	Yes	Estimated Waiting Time	Navigate Multiple EVs to Destination via CS

The summary of the above RL-based methods is discussed in Table II. it is clear from the table II that the existing charging navigation methodologies overlook a crucial factor: the driver's or owner's preference when selecting routes and charging stations, whether prioritizing cost or distance. Minimizing travel distance is a key element in urban mobility, as it directly impacts travel time, fuel consumption, and overall efficiency, a common goal among drivers and widely examined in studies such as [33], [37]. Similarly, cost minimization is a primary concern for many drivers, affecting the total expenses associated with travel, including charging, driving, and waiting costs, as emphasized in works like [8], [23], [27], [28], [31]. Our model aligns with practical user concerns by reducing costs, promoting greater EV adoption and supporting sustainable practices. This perspective is essential in an EV-centric transportation framework. The impact of long wait times at charging stations extends beyond the individual driver, affecting the broader EV transportation system. Consequently, addressing user preferences for seamless driving experiences or maximizing drivers' earnings becomes vital in EV charging and scheduling. A comprehensive approach that recognizes and addresses these factors can meet user demands while enhancing the overall efficiency and sustainability of the EV transportation ecosystem.

III. DRL-BASED OPTIMAL CHARGING SCHEDULING AND NAVIGATION

A. Problem definition

The rise of EVs in urban areas has highlighted the importance of managing on-the-go charging effectively. Limited battery capacity poses a challenge for EVs during journeys, impacting the driver experience. Intermittent charging increases wait times at CS, affecting user satisfaction and driver earnings. Driver preferences for cost or distance during EV navigation influence the path, affecting trip costs and distance. Balancing driver/owner's preferences within EV battery constraints is challenging. To address this, we propose a DRL-based solution for optimal EV charging scheduling and navigation, catering to driver preferences. Our study aims to develop cost-effective and strategic charging schedules that meet user demands while maximizing drivers' earning potential through cost or distance minimization based on preference.

B. Navigation system model

Our novel and comprehensive navigation system model is depicted graphically in Figure 1. The charging scheduling navigation system (CSNS), the central component of our architecture, manages the ever-changing coordination of real-time traffic data and charging requests, with a preference for cost distance. Utilizing a range of communication technologies, from wired to wireless networks, the CSNS serves as a central nervous system that coordinates complex communication among numerous stakeholders, including EVs, CSs, and ITS. The CSNS manages the identification of optimal routing solutions and carefully selects appropriate charging stations to respond to charging requests via its complicated

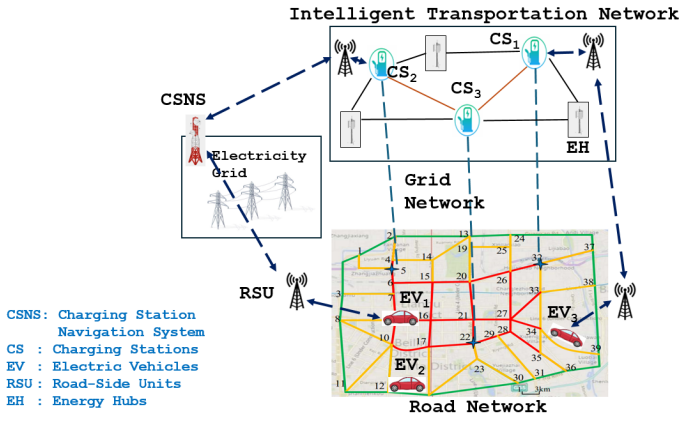


Fig. 1: Navigation system model.

web of connectivity. This decision-making is guided by real-time EVs, CSs, and ITS data, creating a responsive and intelligent decision-making process.

The integrated system is explained in the following sequence to obtain the most effective charging navigation strategies for electric vehicles:

(a) EV needing a charging service seeks assistance from the CSNS, which provides route guidance and charging station selection. This request from the EV is seamlessly transmitted to the CSNS via RSU through advanced wireless communication technologies.
(b) The CSNS receives a continuous stream of monitoring data from CSs, including data on the number of charging and waiting vehicles, among other things. It also collects real-time traffic insights from intelligent transportation networks, including road conditions and average velocities, to provide an up-to-date and comprehensive picture.
(c) The CSNS suggests the route and charging station most appropriate for the particular EV based on information collected from both CSs, intelligent transportation networks, and the driver's preferences (Pref.).
(d) After the EV accepts the suggested charging station, it sends a comprehensive confirmation message to the CSNS through the RSU, including all reservation details. Following this, the CSNS effectively catalogues and secures this valuable reservation data, strategically streamlining future charging requests.

C. Network model

A directed graph $\mathbb{G} = (V, E)$ indicates the network topology, where $V = \{1, 2, 3, ..., n\}$ denotes a set of vertices and $E = \{X_{i,j} \,|\, i, j = 1, 2, 3, ..., n\}$ represents a set of edge. Each node stands for an intersection or the end of a road, using a road connecting nodes i and j denoted as $X_{i,j} \in E$. The traffic network can be represented as a weighted directed graph, with each link assigned a specific weight. The weight of each link is defined as follows.

$$W_{i,j}^{t} = \frac{d_{i,j}}{v_{i,j}^{t}}, \tag{1}$$

here, $d_{i,j}$ represents the distance between nodes i to j, while $v_{i,j}^t$ denotes the velocity from node i to j at time t. Additionally, $W_{i,j}$ represents the weight value associated with the connection between nodes i and j. In other words, the weight value is the time required to traverse the $X_{i,j}$. The Dijkstra algorithm [36] discovers the shortest path by applying weights to all links. The ITS must update and maintain real-time traffic data.

D. Optimal charging navigation scheme

Consider an EV leaving its starting point and travelling through a complex traffic network to its destination. Throughout the journey, the EV's onboard system continuously assesses its SOC and determines whether a recharge is required. When the remaining SoC cannot complete the journey, the EV's terminal creates an ingenious charging schedule. This schedule is carefully planned, considering driver preferences to minimize cost (travel, waiting, and charging) or distance travelled.

The navigation path is dynamically updated throughout the journey based on real-time information the EV terminal receives. When the EV arrives at the CS, it obtains the necessary charging energy following the first come, first served principle. This operational principle prioritizes vehicles that arrive first, a strategy widely supported in previous studies addressing CS-selection issues [6], [17], [38]. Considering cost and distance factors as a preference is central to the proposed optimal charging navigation strategy. The cost includes the EV's waiting, charging, and driving costs at various CSs.

Furthermore, the distance cost encompasses the disparity between the distance from the origin to the destination when visiting a charging station and the distance from the origin to the destination when not visiting a charging station. This strategic framework yields a comprehensive strategy that balances the complex dynamics of EV charging, travel efficiency, cost optimization, and driver preferences. The details are as follows:

1) Cost (C_{cost}): It includes the following cost categories:

(a) **Charging cost** (C_{ch}) : We establish the definitions for energy consumption and charging time as follows:

$$E_c^l = \lambda \times d_l, \forall l \in E, \tag{2}$$

here, E_c^l denotes the energy consumption of link l, λ represents the energy consumption rate, and d_l stands for the distance of link l. Now, when the EV arrives at CS m, the state of charge at arrival (SOC_{arr}^m) is computed as follows:

$$SOC_{\text{arr}}^m = SOC_{\text{cur}} - \sum_{\forall l \in L_f^m} \frac{E_c^l}{E_{\text{max}}}, \quad \text{where } 0 < SOC_{\text{arr}}^m < SOC_{\text{req}}, \forall m \in M. \tag{3}$$

Here, SOC_{cur}, SOC_{req}, E_{max}, and L_f^m correspond to the current charge status at the point of the request, the required state of charge, the EV's maximum battery capacity, and the set of links from the origin to CS m, respectively.

Utilizing equations 2 and 3, the calculation of the energy charging amount at CS m (E_{ch}^m) can be expressed as follows:

$$E_{ch}^m = [(SOC_{req} - SOC_{arr}^m) \times E_{max}], \forall m \in M. \tag{4}$$

Now, the estimation for the charging time at CS m is outlined as follows:

$$CT^m = \frac{E_{ch}^m}{\eta \cdot \gamma}, \forall m \in M, \tag{5}$$

here, γ denotes the charging efficiency, and η signifies the charging power of CS. It is important to mention that we assume that the charging stations charging poles have uniform power capabilities.

Therefore, the charging cost (C_{ch}) at CS m is estimated as follows:

$$C_{ch} = \hat{\alpha} \times CT^m + p \times E_{ch}^m, \tag{6}$$

here p and $\hat{\alpha}$ represent current prices and the driver's daily wage per unit of time. It is important to note that we have used the flat price value p=0.8395 [18]. Additionally, $alpha$ is set at 0.75$/h, based on China's average hourly wage in 2017 [31].

(b) **Driving cost:** Each EV has a different travel time to the charging station because different EVs have different distances from the same charging station. This variation leads to a variety of queuing patterns at the charging station. The associated travel time $T_{i,j}$ can be represented by the following equation when we assign an electric vehicle (EV_i) to a charging station ($m \in M$):

$$T_{i,j} = \frac{d_{i,j}}{V_i}, \tag{7}$$

here, $d_{i,j}$ signifies the distance between EV_i and CS m, whereas V_i represents the velocity of EV_i or road avg velocity. Analogous to equation 7, the time taken to traverse a link l at time step t is expressed as $T_{drive}^l = \frac{d_l}{V_l^t}$. Consequently, the aggregate driving time (T_{drive}^m) for route L^m from the origin to the destination via CS m can be formulated as depicted in equation 8.

$$T_{drive}^m = \sum_{\forall l \in L^m} T_{drive}^l . \forall m \in M. \tag{8}$$

Consequently, the assessment of the driving cost C_{drive} is undertaken in the subsequent manner:

$$C_{drive} = \hat{\alpha} \times T_{drive}^m . \forall m \in M. \tag{9}$$

(c) **Waiting cost:** Let us assume that each charging station has only one charging point. As a result, the waiting time for EVs is determined by the charging time ahead of them in the charging queue at that particular station. Furthermore, we assume that the order in which EVs arrive at the charging station is determined by their arrival times. This arrival time corresponds to the driving duration required for the EVs to travel from their current location to the selected charging station, i.e., the expected arrival time of the EV at CS m can be calculated as follows:

$$AT^m = T_r + T_{drive}^m, \forall m \in M, \tag{10}$$

here, T_r represents the time when a charging request is made, and T^m_{drive} denotes the duration required to drive to CS m.

Given this framework, we proceed to estimate the waiting time (W_{n+1}) for the $(n+1)^{th}$ EVs, denoted as EV_{n+1}, at the charging station (CS) labelled as m.

$$\begin{aligned} W_{n+1} =& AT_1 + (CT_1 + CT_2 + \cdots + CT_n) - AT_{n+1}, \\ & \text{for } n \geq 1 \\ W_{n+1} =& 0, \quad \text{for } n < 1, \end{aligned} \tag{11}$$

here, AT_1 represents the arrival time of the initial EV in the queue, while $CT_1 + CT_2 + \ldots + CT_n$ accounts for the cumulative charging time of EVs already in the queue. Additionally, AT_{n+1} signifies the time of arrival for the current EV (ev_{n+1}) at charging station m. Hence, in conclusion, the waiting cost (C_{wait}) is computed in the subsequent manner:

$$C_{wait} = \hat{\alpha} \times W_{n+1}, \tag{12}$$

here, the symbol $\hat{\alpha}$ denotes the daily driver wage per unit of time, which we have set to 0.75\$ per hour based on the average hourly wage in China, as cited in [32]. Additionally, it is important to note that when multiple charging points are available at a CS, the waiting time is determined by the minimum of W_{xn+1}, where x represents the number of charging points at each station. This rule applies when x is less than or equal to n. However, if x exceeds n, the waiting time is 0.

2) Distance cost (D_{total}): The calculation of the total distance D_{total} is executed using a shortest path-finding algorithm, like Dijkstra's algorithm. The formula for D_{total} is given by D_{total} = shortest origin-destination distance in case of visiting a charging station − shortest origin-destination distance without visiting the charging station. Finally, the cost associated $C_{dist.}$ is computed as follows:

$$C_{dist.} = \phi \times D_{total}, \tag{13}$$

here, ϕ represents the conversion factor that translates distance into cost.

3) Objective function and constraints: Charging navigation scheduling aims to reduce the overall synthetic cost. The objective function is stated as follows:

$$C_{Cost} = C_{drive} + C_{wait} + C_{Ch}. \tag{14}$$

$$min[C_{total} x^m_{i,j}] = C_{Cost} + \phi \times D_{total}. \tag{15}$$

Here, ϕ is the weight coefficient. The constraints on the objective function are:

$$E_{SOC_0} \geq \lambda \times D_{ocs} + E_{min}, \tag{16}$$

$$E_{SOC_0} - \lambda \times D_{ocs} + E_{ch} \geq \lambda \times D_{csd} + E_{min}, \tag{17}$$

$$E_{SOC_0} + E_{ch} - \lambda(D_{Ocs} + D_{csd}) = E, 0 < E_{ch} < E, \tag{18}$$

here, E denotes the rated battery capacity, while E_{min} represents the minimum battery capacity. The initial state of charge at the origin is denoted by SOC_0, and λ signifies the energy consumption per unit distance. Furthermore, D_{OCS} represents the driving distance from the origin to the selected Charging Station (CS), and D_{csd} corresponds to the driving distance from the chosen CS to the destination.

To elaborate further, the constraints can be understood as follows: Constraint 16 stipulates that the remaining battery energy of the electric vehicles at the origin must exceed the energy consumed from the origin to the chosen CS. In other words, the EV should have ample energy to cover the initial distance. Moving on to constraint 17, it requires that the battery energy after charging at the CS must surpass the energy needed from that CS to the destination. This ensures that the EV possesses adequate energy to complete the remaining journey after recharging. Assumptions continue with the consideration that, upon reaching the destination, the EV will be charged to its rated capacity E using the prescribed charging mode. This leads to equation 18, which governs the requisite battery capacity at the destination. Additionally, the constraint on E_{ch} dictates the amount of energy that can be charged, aligning with operational limitations. Finally, the task entails selecting both the optimal driving path and the CS yielding the lowest total cost, denoted as min $[C_{total} x^m_{i,j}]$.

The optimal EV charging navigation approach proposed is illustrated in Figure 2. It comprises three primary components: DRL MDP tuples representing state, action, transition, and rewards; the system model; and an agent. The agent, trained with reinforcement learning-based models (such as PPO, DQN, or DDQN), recommends actions, including charging station selection. Additionally, based on driver preferences, a path is determined to achieve the overall objective (O2: minimizing the total cost) while satisfying constraints C1, C2, and C3. The following subsections explain MDP tuple formulation and RL model training in depth.

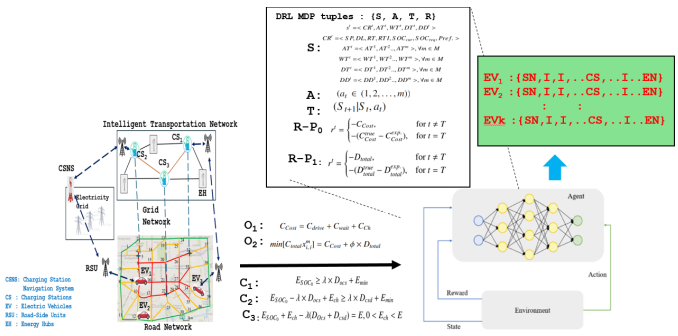


Fig. 2: Proposed EV charging scheduling and navigation approach.

E. Markov decision process modeling

We proposed a real-time EV charging scheduling and navigation problem from the standpoint of EV drivers and owners. To model the aforementioned problem, we employ a finite MDP operating within discrete time steps. The MDP framework offers a mathematical structure for representing decision-making scenarios where outcomes blend random events and decisions made by the decision-maker. An MDP is defined

by a five-tuple $(\mathbb{S}, \mathbb{A}, \mathbb{P}, \mathbb{R}, \gamma)$, where: $\mathbb{S}$ represents the system states, $\mathbb{A}$ is a finite set of available actions, $\mathbb{P}$ denotes the state transition probabilities, $\mathbb{R}$ signifies the immediate rewards of each state and action, and γ stands for the discount factor.

The details regarding the formulation of the MDP are elucidated as follows:

1) State: At time instance t, we perceive the system's condition, denoted by s^t. Which is defined as follows:

$$s^t =< CR^t, AT^t, WT^t, DT^t, DD^t >, \tag{19}$$

$$CR^t =< SP, DL, RT, RTI, SOC_{cur}, SOC_{req}, Pref. >, \tag{20}$$

$$AT^t =< AT^1, AT^2.., AT^m >, \forall m \in M, \tag{21}$$

$$WT^t =< WT^1, WT^2.., WT^m >, \forall m \in M, \tag{22}$$

$$DT^t =< DT^1, DT^2.., DT^m >, \forall m \in M, \tag{23}$$

$$DD^t =< DD^1, DD^2.., DD^m >, \forall m \in M, \tag{24}$$

where CR^t, AT^t, DT^t, WT^t and DD^t are the EV charge request, expected arrival time, driving time, waiting time, and driving distance, respectively. The EV charge request encompasses various details, including the starting position SP, destination location DL, request time RT, request time interval RTI, the current charge level of the battery SOC_{cur}, desired charge levelSOC_{req}, and driver's inclination denoted by $Pref.$, where $Pref.$ = 1 indicates a preference for distance optimization, while $Pref.$ = 0 signifies cost preference.

The decision regarding the charging action, specifying the charging station index where the electric vehicle battery will be charged, is contingent upon the state s^t. Following the execution of the charging action, the system state transitions to s^{t+1}, and subsequently, the next charging action a^{t+1} is determined for the succeeding time step $t + 1$.

2) Action: An action is taken for each state denoted as a^t. This action indicates a charging station index selection $(a^t \in (1, 2, \ldots, m))$, encompassing the planned route associated with CS m, facilitating travel from the starting position SP to the destination DL via CS m. The set of possible actions constitutes the action space, representing the collection of indices $M = (1, 2, \ldots, m)$. Consequently, it is a member of this action space, denoting a specific action undertaken within the system.

3) Transition probability: The function $(s^{t+1}|s^t, a^t)$ reflects the transition probability between the current state s^t and the next state s^{t+1}, given that the agent takes action a^t. However, accurately defining this transition probability becomes challenging without precise environmental models and a prior understanding of uncertainties. To address this issue comprehensively, this study adopts a model-free DRL approach to handle situations with unknown transition probabilities. Our approach aims to learn the transition probability through iterative learning to maximize cumulative rewards. This learning process revolves around refining the agent's policy over multiple iterations, guided by the outcomes of trial-and-error interactions with the environment.

4) Reward: In our proposed scenario, the reward is calculated with consideration to the EV driver's perspective and the operational time T, reflecting the driver's preference denoted as $Pref.$, i.e.,

(a) When $Pref. = 0$, indicating a cost preference, the reward is evaluated subsequently:

$$r^t = \begin{cases} -C_{Cost}, & \text{for } t \neq T \\ -(C^{true}_{Cost} - C^{exp.}_{Cost}). & \text{for } t = T \end{cases} \tag{25}$$

(b) When $Pref. = 1$, indicating a distance preference, the reward is evaluated subsequently:

$$r^t = \begin{cases} -D_{total}, & \text{for } t \neq T \\ -(D^{true}_{total} - D^{exp.}_{total}). & \text{for } t = T \end{cases} \tag{26}$$

In both instances above, the reward function entails negative values, as the fundamental objective of RL is to optimize the accumulation of rewards. Hence, the function is assigned a negative value, which serves to minimize both costs and distances.

F. Training

We initially simplify the problem by employing MDP techniques, as discussed earlier, to tackle EV route charging and scheduling challenges. Subsequently, we utilize RL algorithms DQN, DDQN, and PPO to evaluate their efficacy in addressing our EV route charging scheduling and navigation issues. The training methodologies for these models are elaborated on in the subsequent paragraph.

1) Training of DQN: $\mathbb{Q}$-learning is a model-free RL approach to optimizing policy through environmental interaction. Introduced by Watkins in 1989, it is widely applied, as seen in Mhaisen et al.'s real-time EV charging schedules [39], [40]. However, $\mathbb{Q}$-learning faces high-dimensional problems, like EV charging, due to its reliance on lookup tables [41]. Combining RL and deep neural networks, deep reinforcement learning addresses this issue using DQN. DQN merges deep neural networks with $\mathbb{Q}$-learning, making RL applicable in complex environments [8], [42]. DQN operates effectively with discrete action spaces. The DQN approach to solving the proposed problem is discussed in Algorithm 1.

Algorithm 1 Training process of DQN

Randomly set the initial DQN parameters Θ
Set the initial target network parameters $\overline{\Theta} = \Theta$ Episode $= 1$ to E_m Node $= 1$ to $|V|$
Generate the starting state s^0 $EV \notin EN$ $EV \notin CS$
Choose a CS and a relevant route L_r through action a^t.
employing a ε-greedy approach
Proceed along route L_r, note reward r^t, obtain fresh information,
and generate new state s^{t+1}.
Maintain a tuple (s^{t+1}, a^t, r^t, s^t) within the replay buffer RM
Sample batch $\phi = \{(s^t, a^t, r^t, s^{t+1})\}_{t=1}^{\#\phi}$ from RM
Compute target $\mathbb{Q}$-value for each batch of transition:

$$y^t = \begin{cases} r^t & \text{if episode terminates at step } t+1 \\ r^t + \gamma \cdot \max_{a'} \mathbb{Q}(s^{t+1}, a'; \overline{\Theta}) & \text{otherwise} \end{cases}$$

Determine the loss function
$L(\Theta^t) = \sum_{t=1}^{\phi} \left[y^t - \mathbb{Q}(s^t; a^t; \Theta^t)\right]^2$
Update DQN parameters $\Theta^t \leftarrow \Theta^t - \lambda \nabla_{\Theta^t} L(\Theta^t)$ every P steps
Reset $\overline{\Theta} = \Theta$
Proceed along route L_r, note reward r^t, obtain fresh information,
and generate new state s^{t+1}.
Repeat Steps 10 to 16 until EN reached

2) Training of DDQN: Double Deep Q-Network enhances the original DQN algorithm. Unlike DQN, DDQN separates action selection and value estimation by employing two neural networks: one for selecting actions and another for estimating their values. This helps prevent overestimation biases and enhances training stability. The DDQN approach to solving the proposed problem is discussed in Algorithm 2.

Algorithm 2 Training process of DDQN

Randomly set the initial DDQN parameters Θ
Set the initial target network parameters $\overline{\Theta} = \Theta$ Episode $= 1$ to E_m Node $= 1$ to $|V|$
Generate the starting state s^0 $EV \notin EN$ $EV \notin CS$
Choose a CS and a relevant route L_r through action a^t,
employing a ε-greedy approach
Proceed along route L_r, note reward r^t, obtain fresh information,
and generate new state s^{t+1}.
Maintain a tuple (s^{t+1}, a^t, r^t, s^t) within the replay buffer RM
Sample batch $\phi = \{(s^t, a^t, r^t, s^{t+1})\}_{t=1}^{\#\phi}$ from RM
Compute target $\mathbb{Q}$-value for each batch of transition:

$$y^t = \begin{cases} r^t & \text{if episode terminates at step } t+1 \\ r^t + \gamma \cdot \mathbb{Q}(s^{t+1}, \text{argmax}_{a'} \mathbb{Q}(s^{t+1}, a'; \Theta); \overline{\Theta}) & \text{otherwise} \end{cases}$$

Determine the loss function
$L(\Theta^t) = \sum_{t=1}^{\phi} \left[y^t - \mathbb{Q}(s^t; a^t; \Theta^t)\right]^2$
Update DQN parameters $\Theta^t \leftarrow \Theta^t - \lambda \nabla_{\Theta^t} L(\Theta^t)$ every P steps
Reset $\overline{\Theta} = \Theta$
Proceed along route L_r, note reward r^t, obtain fresh information, and generate
new state s^{t+1}.
Repeat Steps 10 to 16 until EN reached

3) Training of PPO: PPO is a policy gradient method in reinforcement learning designed to combine the data efficiency and robust performance of TRPO while employing only first-order optimization techniques. The PPO approach to solving the proposed problem is discussed in Algorithm 3.

Algorithm 3 Training process of PPO

Randomly set initial policy π_θ with parameters Θ
Set initial target network parameters $\overline{\Theta} = \Theta$ Episode $= 1$ to E_m Node $= 1$ to $|V|$
Generate the initial state s^0 $EV \notin EN$ $EV \notin CS$
Use the $\hat{V}$ function to run π_θ in order to choose a CS
and associated route L_r through a^t action
Proceed along path L_r, note reward r^t, obtain fresh information,
and produce new state s^{t+1}.
maintain tuple (s^t, a^t, r^t, s^{t+1}) in replay memory RM
Sample batch $\phi = \{(s^t, a^t, r^t, s^{t+1})\}_{t=1}^{\#\phi}$ from RM
Determine advantage estimates $\hat{A}_i = r_i + \gamma \hat{V}_\phi(s_i') - \hat{V}_\phi(s_i)$
Apply gradient ascent to update the policy: $\theta \leftarrow \theta + \hat{\alpha} \nabla_\theta J^{\text{policy}}(\theta)$
Compute value loss: $L(\phi) = \frac{1}{|\mathcal{D}|} \sum_i \left(\hat{V}_\phi(s_i) - (r_i + \gamma \hat{V}_\phi(s_i'))\right)^2$
Revise the old PPO policy $\Theta_{\text{old}} \leftarrow \Theta$ $j = 1$ to N
Revise actor policy by policy gradient
$\mu \leftarrow \mu - \hat{\alpha} \nabla_\mu \min\left[\rho(\pi(\theta), \pi_{\text{old}}(s)) \hat{A}_t, \min(\rho(\pi(\theta), \pi_{\text{old}}(s)), 1) \hat{A}_t\right]$
Revise critic by:
$\hat{V} \leftarrow \hat{V} + \beta \nabla_{\hat{V}} \frac{1}{2} [(\hat{V}(s^t) - \hat{r}^t)^2]$ every P steps
Reset $\overline{\Theta} = \Theta$
Proceed along route L_r, note reward r^t, obtain fresh information,
and generate new state s^{t+1}.
Repeat Steps 10 to 16 until EN reached

IV. Case Study

This section uses an example to illustrate our proposed EV charging scheduling and navigation approach. We utilize Figure 3 as a model scenario to demonstrate how our approach functions. Specifically, we explain through the following three cases:

Case1- Suppose an EV at point C wants to go to node I, and all charging stations have the same waiting time: In this analysis, charging time has been excluded from consideration, and it is assumed that charging costs are identical at all charging stations. This case is further divided into two sub-categories, as outlined below:

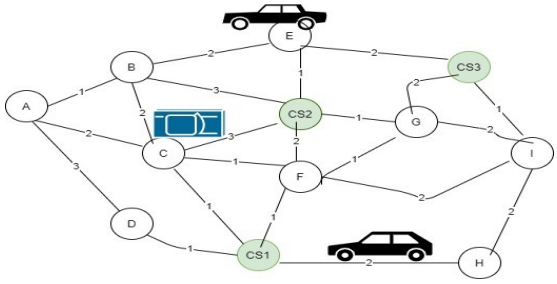


Fig. 3: Model network for case study.

a) **Drivers prioritize cost (C_{Cost}) reduction over distance (D_{total}):** EV has three paths in this case:
- Path 1: $C \xrightarrow{3} CS2 \xrightarrow{1} G \xrightarrow{2} I$ (Total time 6hrs)
- Path 2: $C \xrightarrow{1} CS1 \xrightarrow{1} F \xrightarrow{2} I$ (Total time 4hrs)
- Path 3: $C \xrightarrow{1} F \xrightarrow{1} G \xrightarrow{2} CS3 \xrightarrow{1} I$ (Total time 5hrs)

Hence, based on the criteria above, the CSNS selects one of the Paths—1, 2, or 3—by evaluating each path's total cost, denoted as C_{Cost}. In this scenario, Path 2 is determined to be the optimal choice.

b) **The driver prioritizes reducing the distance over cost**: If we assume that the velocity of the EV at each link is the same, then the shortest distance without visiting a charging station is as follows: $C \xrightarrow{1} F \xrightarrow{2} I$ (3 units). Furthermore, the EV has three possible paths to reach node I in this scenario:
- Path 1: $C \xrightarrow{3} CS2 \xrightarrow{1} G \xrightarrow{2} I$ (Total 6 unit) $D_{\text{total}} = 6 - 3 = 3$ unit.
- Path 2: $C \xrightarrow{1} CS1 \xrightarrow{1} F \xrightarrow{2} I$ (Total 4 unit) $D_{\text{total}} = 4 - 3 = 1$ unit.
- Path 3: $C \xrightarrow{1} F \xrightarrow{1} G \xrightarrow{2} CS3 \xrightarrow{1} I$ (Total 5 unit) $D_{\text{total}} = 5 - 3 = 2$ unit.

Thus, Path 2 is selected when the driver prefers to reduce the distance (D_{total}).

Case2- Suppose an EV at point C wants to go to node I, and all charging stations have different waiting times: Assume CS1 has a 3-hour wait time (Path time 7 hours), CS2 has a 30-minute wait time (Path time 6.30 hours), and CS3 has a 30-minute wait time (Path time 5.30 hours). This case is further divided into two sub-categories, as described below:

a) **Drivers prioritize cost reduction over distance**: In this case, EV again has three options:
- Path 1: $C \xrightarrow{3} CS2 \xrightarrow{0.30+1} G \xrightarrow{2} I$ (Total time 6.30hrs)
- Path 2: $C \xrightarrow{1} CS1 \xrightarrow{3+1} F \xrightarrow{2} I$ (Total time 7hrs)
- Path 3: $C \xrightarrow{1} F \xrightarrow{1} G \xrightarrow{2} CS3 \xrightarrow{0.30+1} I$ (Total time 5.30hrs)

CSNS chooses Path 3 for the preceding if the EV can travel with their remaining charge level up to CS3—otherwise, Path 1 is selected.

b) **The driver prioritizes reducing the distance over cost**: For the same EV velocity assumption at each Path and without visiting a charging station, the shortest distance is: $C \xrightarrow{1} F \xrightarrow{2} I$ (3 unit). Hence, based on distance, the Path options are:

TABLE III: Analysis of cost parameters in Case3.

Source (C) to destination (I)	Path Cost	Total Charge required	Waiting time	Driving time	Charging time	Total time	r_pref_0	r_pref_1
C to I Path-1	$C \xrightarrow{1} CS1 \xrightarrow{1} F \xrightarrow{2} I$ (4 hr)	0.64- 0.48 = 0.16 kW	3	4	0.16 hr	7.16 hr	-7.16	-1
C to I Path-2	$C \xrightarrow{3} CS2 \xrightarrow{1} G \xrightarrow{2} I$ (6 hr)	0.96 - 0.48 = 0.48 kW	0.5	6	0.48 hr	6.98 hr	-6.98	-3
C to I Path-3	$C \xrightarrow{1} F \xrightarrow{1} G \xrightarrow{2} CS3 \xrightarrow{1} I$ (5 hr)	0.80-0.48=0.32 kW	0.5	5	0.32 hr	5.82 hr	-5.82	-2
C to I Path-4	$C \xrightarrow{1} CS1 \xrightarrow{2} H \xrightarrow{2} I$ (5 hr)	0.80-0.48=0.32 kW	3	5	0.32 hr	8.32 hr	-8.32	-2
C to I Path-5	$C \xrightarrow{3} CS2 \xrightarrow{1} G \xrightarrow{1} F \xrightarrow{2} I$ (7.5 hr)	1.28-0.48=0.80 kW	0.5	7	0.80 hr	8.30 hr	-8.30	-4

- Path 1: $C \xrightarrow{3} CS2 \xrightarrow{1} G \xrightarrow{2} I$ (Total distance 6 unit) (Total time 6.30 hrs) $D_{\text{total}} = 3$ unit
- Path 2: $C \xrightarrow{1} CS1 \xrightarrow{1} F \xrightarrow{2} I$ (Total distance 4 unit) (Total time 7 hrs) $D_{\text{total}} = 1$ unit
- Path 3: $C \xrightarrow{1} F \xrightarrow{1} G \xrightarrow{2} CS3 \xrightarrow{1} I$ (Total distance 5 unit) (Total time 5.30 hrs) $D_{\text{total}} = 2$ unit

Thus, Path 2 is selected.

Case3- Assume that EV at C desires to reach node I took into account the waiting, driving, and charging times and assumed that they differ from station to station: In this analysis, we assume a discharge rate of 0.16 kW/km, a charging efficiency of 1 kW/h, and a state of charge (SOC) of 0.48 kW. The shortest route from C to I is $C \rightarrow F \rightarrow I$, with a total travel time of 3 hours. Additionally, we have calculated the relevant cost parameters, which are presented in Table III. Now, let us consider the following scenarios:

a) **The driver's preference is to reduce cost over distance:** In this case, for $C \rightarrow I$, Path-2 is selected, although Path-3 is inexpensive, an EV with SOC = 0.48 cannot reach the charging station CS3 because 0.48 kW is required.

b) **The driver prefers reducing distance over cost:** In this case, for $C \rightarrow I$, Path-1 is selected. As the minimum value of D_{total} for Path-1 is 1 here.

V. Experimental Results

This section thoroughly examines the simulation environment, training parameters, convergence analysis, driver preference analysis, and path navigation analysis and discusses the proposed work's results. Specifically, the efficacy of the proposed work using the PPO model with DQN and DDQN is evaluated to indicate the usefulness of the suggested navigational approach.

A. Experimental setup

The efficacy of our suggested method is demonstrated within the real-scale area of Xi'an city, encompassing 39 nodes and 3 charging stations (4, 22, and 32), indicated as blue stars in Figure 4. Xi'an's roads are categorized into three distinct classes, visually represented by different colours in Figure 4. Green roads denote the ring highway encircling the city, yellow signifies the urban expressway, and red indicates inner ring roads [31]. Previous research [31], [32], [43], [44] suggests that road speeds follow truncated normal distributions. The parameters of these distributions vary proportionally with speed limits for various road categories, as shown in Table IV. Xi'an City's inner ring roads, urban motorways and ring highways have speed limits of 60 km/h, 80 km/h, and 120 km/h, respectively. EVs have a maximum battery capacity of

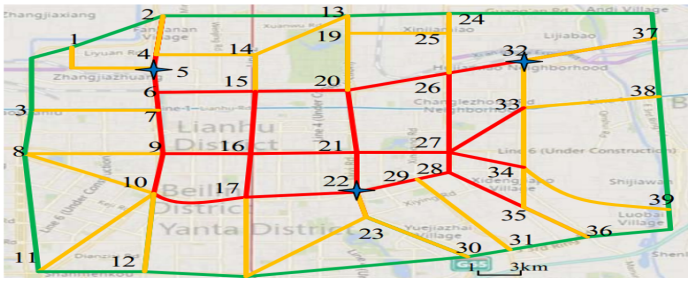


Fig. 4: Synthetic road network of Xi'an city China [31].

54.75 kWh [33]. EVs' initial and required SoC are uniformly distributed between 0.2 and 0.4. CS provides 60 kW charging power at an energy consumption rate of 0.16 kW/h. Three charging stations with two poles are available (see Table IV).

TABLE IV: Proposed work simulation parameters [31], [32].

S.No	Parameter	Values
1	Initial SOC	Uniform (0.2, 0.4)
2	Energy consumption rate	$0.16kW/km$
3	Charging power	$60kW$
4	Max. Battery capacity	$54.75kWh$
5	Charging efficiency	0.9
6	Required SOC	0.90
7	Number of CS	3
8	Number of nodes	39
9	Number of charging pole	2
10	Number of links	134
11	Green roads velocity (km/h)	$N(0.9^*120, (0.05^*120)^2)$
12	Yellow roads velocity (km/h)	$N(0.7^*80, (0.10^*80)^2)$
13	Red roads velocity (km/h)	$N(0.5^*60, (0.15^*60)^2)$
14	$\hat{\alpha}$, p, Φ	$0.75\$/h$, 0.8395$, and $0.5\$/km$, respectively

The simulation rigorously evaluates the proposed algorithm's effectiveness and adaptability across various scenarios, encompassing diverse numbers of EV charging requests and their arrival time distributions. These arrival times are stochastically generated using both uniform and normal distributions. Specifically, uniform and normal distributions generate arrival times within the $[20, 100]$ range. In the case of the normal distribution, the mean arrival time is 60, with a standard deviation of approximately 13.33. Furthermore, the simulation code is implemented in Python, leveraging the TensorFlow framework [45] and Stable Baselines3, implementations of reinforcement learning algorithms in PyTorch [46]. The computational setup entails an AMD Ryzen 7 5800H CPU operating at 3.20 GHz, an NVIDIA GeForce RTX 3050 GPU with 4GB of VRAM, and 16GB of RAM.

TABLE V: Model training parameters DQN and DDQN.

S.No	Parameter	Values
1	BATCH_SIZE	128 # minibatch size
2	LR	$8e-3$ # learning rate
3	BUFFER_SIZE	int(1e5) # replay buffer size
4	TAU	0.4 # for soft update of target parameters
5	n_episodes	2000
6	GAMMA	0.99 # discount factor
7	eps_end	0.01
8	UPDATE_EVERY	4 # how often to update the network
9	eps_start	1.0
10	max_grad_norm	10
11	eps_decay	0.7

TABLE VI: Model training parameters PPO.

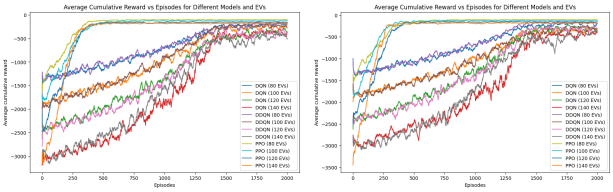
S.No	Parameter	Values
1	BATCH_SIZE	64 # minibatch size
2	LR	0.0003 # learning rate
3	BUFFER_SIZE	int(1e5) # replay buffer size
4	UPDATE_EVERY	4 # how often to update the network
5	GAMMA	0.99 # discount factor
6	eps_start	1.0
7	TAU	0.4 # for soft update of target parameters
8	vf_coef	0.50
9	n_episodes	2000
10	gae_lambda	0.95
11	eps_end	0.01
12	max_grad_norm	0.5
13	clip_range	0.20

B. Convergence analysis

Convergence analysis offers useful insights into the efficacy and reliability of RL algorithms, facilitating researchers and practitioners in understanding their behaviour and making informed decisions regarding algorithm selection and parameter tuning. In our study, we conducted convergence analysis using three RL models, DQN, DDQN, and PPO, to evaluate the proposed approach thoroughly. The training parameters utilized for these models are detailed in Tables V and VI. These parameters play an important role in shaping the learning dynamics of the models and directly impact their convergence behaviour. To delve deeper into the effectiveness of each model, we conducted extensive training sessions comprising 2000 episodes for DQN, DDQN, and PPO. We varied the number of electric vehicle (EV) charge requests between 80, 100, 120, and 140, utilizing uniform and normal distributions.

The convergence patterns of average cumulative rewards throughout the training process are graphically depicted in Figure 5. Figure 5 vividly illustrates the convergence behaviour of each model. Notably, DQN and DDQN exhibit a slower convergence rate than PPO, which consistently converges much faster, typically within 400 to 500 episodes across all EV cases. This swift convergence of PPO underscores its superior performance and adaptability relative to the other two models. Such findings shed light on the efficacy of the PPO model across diverse scenarios, highlighting its potential as a robust solution for real-world applications.

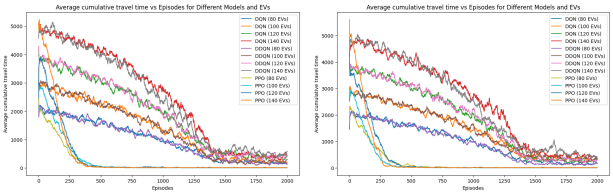
Furthermore, Figure 6 provides insight into the average cumulative travel time progress throughout the training process across different EVs and models. It is evident that as training advances, each model consistently reduces travel time,



(a) With uniform distribution (b) With normal distribution

Fig. 5: Average cumulative reward progress for different EVs and models during the training process.

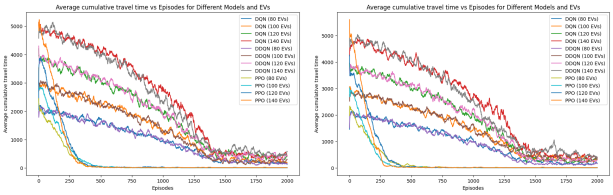
indicating an effective learning of the charging navigation strategy. Similar to previous observations, PPO demonstrates accelerated learning, achieving significant reductions in travel time within just 500 episodes, while DQN and DDQN require substantially more episodes, around 1500, to attain comparable performance. This stark contrast underscores the efficiency and rapid adaptability of the PPO model in optimizing travel time, further solidifying its prominence among the considered RL approaches.



(a) With uniform distribution (b) With normal distribution

Fig. 6: Average cumulative travel time progress for different EVs and models during the training process.

Moreover, Figure 7 showcases the average cumulative progress of waiting costs across various EVs and models during training. As training progresses, all models consistently improve in reducing waiting costs, indicating effective learning. Notably, PPO demonstrates rapid learning, achieving significant reductions within 500 episodes, whereas DQN and DDQN require around 1500 episodes to achieve comparable performance. This underscores the efficiency and rapid adaptability of the PPO model among the considered RL approaches.



(a) With uniform distribution (b) With normal distribution

Fig. 7: Average cumulative waiting cost progress for different EVs and models during the training process.

C. Driving preference analysis

The proposed work aims to accommodate driver preferences, which may prioritize either minimizing the overall cost, denoted as C_{cost}, or minimizing the distance travelled, denoted as D_{total}. C_{cost} encompasses various factors such as charging expenses, waiting times, and travel costs. On the other hand, D_{total} signifies the total distance covered by an electric vehicle (EV) within a specific scenario. D_{total} is computed as the disparity between the shortest distance between the origin and destination when the EV includes a stop at a charging station and the shortest distance when it does not. It quantifies the additional distance incurred from visiting a charging station during the journey.

We conducted a driver preference analysis for varying numbers of EVs using uniform and normal distributions across three RL algorithms: DQN, DDQN, and PPO. The results are presented in Tables VII, VIII, IX, and X. Analyzing Tables VII and VIII, it is clear that when prioritizing cost in Cumulative Costs (C_{Cost}) across both uniform and normal distribution scenarios, each model—DQN, DDQN, and PPO strategically select paths that minimize cost. Consistently lower cost values evidence this compared to scenarios where distance is preferred. Take, for instance, the uniform distribution scenario with 140 EV charge requests: the PPO model achieves a C_{Cost} of 51.59, whereas it registers 58.51 when distance is prioritized. This trend persists across all models, highlighting their adeptness at learning and adapting to preferences. Additionally, it is noteworthy that the PPO model consistently computes lower C_{Cost} values than both DQN and DDQN across both preferences.

TABLE VII: Cumulative costs (C_{Cost}) for different numbers of EVs with different preferences following a uniform distribution.

Number of EVs	Algorithm	Cost Preference	Algorithm	Distance Preference
80	DQN	86.67	DQN	4450.83
	DDQN	96.72	DDQN	1045.33
	PPO	28.83	PPO	30.98
100	DQN	49.27	DQN	552.32
	DDQN	209.09	DDQN	214.74
	PPO	37.18	PPO	44.04
120	DQN	292.73	DQN	2544.72
	DDQN	439.11	DDQN	2478.08
	PPO	48.54	PPO	46.63
140	DQN	394.29	DQN	1822.32
	DDQN	371.79	DDQN	2507.90
	PPO	51.59	PPO	58.51

Turning to Tables IX and X, a similar trend emerges when prioritizing distance in D_{total} under both uniform and normal distribution scenarios. Each model DQN, DDQN, and PPO favourably selects paths that minimize D_{total} over C_{Cost}. For instance, prioritizing cost in the scenario of uniform distribution with 140 EV charge requests, the DQN, DDQN, and PPO models achieve D_{total} values of 154.5, 165.0, and 140.5, respectively. Conversely, when prioritizing distance, they record 91.5, 76.5, and 138.0 values for the same scenario. This underscores their proficiency in learning and adapting to distance preferences. Furthermore, it is noteworthy that these models exhibit mixed performance regarding D_{total}. One model may excel at one type of EV request, while another performs better on different requests across both preferences.

TABLE VIII: Cumulative Costs (C_{Cost}) for different numbers of EVs with different Preferences following a normal distribution.

Number of EVs	Algorithm	Cost Preference	Algorithm	Distance Preference
80	DQN	29.90	DQN	1132.08
	DDQN	32.50	DDQN	1433.63
	PPO	31.15	PPO	31.03
100	DQN	167.00	DQN	1517.74
	DDQN	46.44	DDQN	1019.91
	PPO	39.62	PPO	41.32
120	DQN	547.99	DQN	2070.97
	DDQN	796.08	DDQN	2002.99
	PPO	46.06	PPO	50.36
140	DQN	294.76	DQN	566.11
	DDQN	1318.66	DDQN	1378.84
	PPO	53.36	PPO	58.07

TABLE IX: Cumulative distance cost (D_{total}) for different numbers of EVs with different preferences following a uniform distribution.

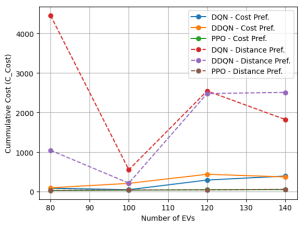
Number of EVs	Algorithm	Cost Preference	Algorithm	Distance Preference
80	DQN	84.0	DQN	59.5
	DDQN	73.0	DDQN	57.0
	PPO	84.5	PPO	79.5
100	DQN	99.0	DQN	82.5
	DDQN	116.0	DDQN	65.5
	PPO	102.5	PPO	98.0
120	DQN	127.5	DQN	61.0
	DDQN	142.5	DDQN	77.0
	PPO	142.5	PPO	129.5
140	DQN	154.5	DQN	91.5
	DDQN	165.0	DDQN	76.5
	PPO	140.5	PPO	138.0

TABLE X: Cumulative distance cost (D_{total}) for different numbers of EVs with different preferences following a normal distribution.

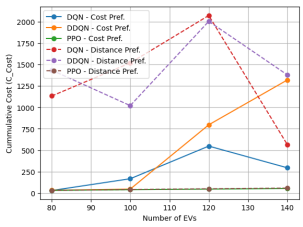
Number of EVs	Algorithm	Cost Preference	Algorithm	Distance Preference
80	DQN	71.5	DQN	33.0
	DDQN	87.0	DDQN	31.5
	PPO	87.5	PPO	64.5
100	DQN	82.0	DQN	49.0
	DDQN	91.0	DDQN	68.5
	PPO	135.5	PPO	107.5
120	DQN	101.5	DQN	61.0
	DDQN	120.5	DDQN	73.5
	PPO	134.0	PPO	130.5
140	DQN	146.0	DQN	125.0
	DDQN	125.0	DDQN	121.0
	PPO	166.0	PPO	149.5

Additionally, to visualize the cost and distance preferences for each model under both uniform and normal distributions, we present Figures 8 and 9, illustrating the progress of cumulative cost (C_{Cost}) and cumulative distance cost (D_{total}), respectively. Figure 8 illustrates that each model minimizes C_{Cost} when the cost is preferred, with PPO consistently producing the lowest cost compared to DQN and DDQN across all scenarios of uniform and normal distributions and varying numbers of EV charge requests. Moreover, in Figure 9, it is evident that each algorithm minimizes D_{total} more effectively when the preference is distance over cost. DDQN outperforms other models in uniform distributions in minimizing D_{total}

except for scenarios with 120 EVs, where DQN performs best. Conversely, in normal distribution scenarios, DQN performs best except for the 140 EV scenario, where DDQN is the most effective.

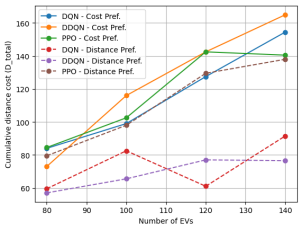


(a) With uniform distribution

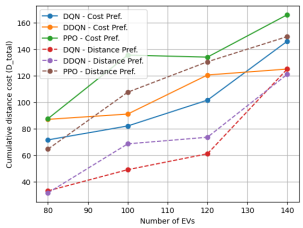


(b) With normal distribution

Fig. 8: Cumulative cost (C_{Cost}) progress.



(a) With uniform distribution



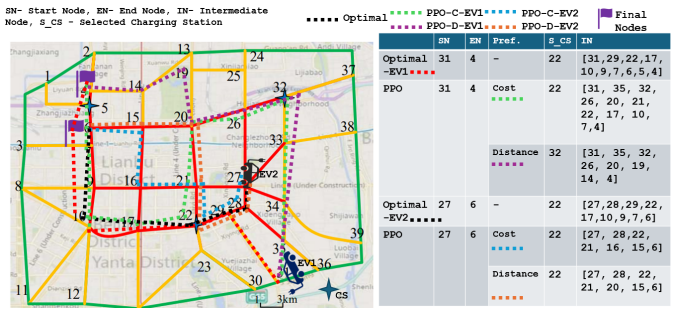
(b) With normal distribution

Fig. 9: Cumulative distance cost (D_{total}) progress.

D. Path navigation analysis

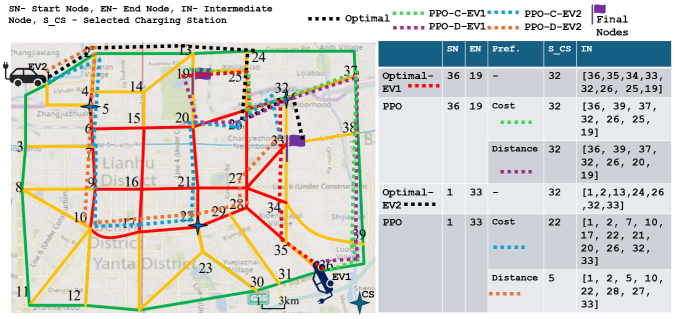
The driver preferences may lean towards minimizing the total cost (C_{cost}) or the distance traveled (D_{total}), which is the focal point of our proposed work. EV route selection is closely related to these preferences. We employed uniform and normal distributions in path navigation for our analysis, utilizing three distinct RL algorithms: DQN, DDQN, and PPO. The experimental findings of path navigation analysis are depicted in Figures 10, 11, and 12. Each of these figures illustrates the route from the source node (SN), where the EV initiates its journey, to the end node (EN), where it concludes, with S_{CS} denoting the selected charging station. The path taken, denoted as 'IN', includes SN, intermediate nodes, and EN. Additionally, 'Pref.' indicates whether the preference is for minimizing cost (C_{cost}) or distance (D_{total}).

On the left-hand side, a map displays the paths taken for both cost and distance preferences, distinguished by different colours as indicated in the figures. In Figure 10, we highlight the paths taken using the PPO model from nodes 31 to 4 and 27 to 6 in uniform distribution, as well as 36 to 19 and 1 to 33 in normal distributions, along with their corresponding optimal paths. Similarly, Figures 11 and 12 present path navigation using the DQN and DDQN models, respectively, in uniform and normal distributions. Notably, each navigation path includes at least one charging station (CS). These figures demonstrate how the models adapt to driver preferences and adjust their routes accordingly.



	SN	EN	Pref.	S_CS	IN
Optimal -EV1	31	4	-	22	[31,29,22,17, 10,9,7,6,5,4]
PPO	31	4	Cost	22	[31, 35, 32, 26, 20, 21, 22, 17, 10, 7,4]
			Distance	32	[31, 35, 32, 26, 20, 19, 14, 4]
Optimal -EV2	27	6	-	22	[27,28,29,22, 17,10,9,7,6]
PPO	27	6	Cost	22	[27, 28,22, 21, 16, 15,6]
			Distance	22	[27, 28, 22, 21, 20, 15,6]

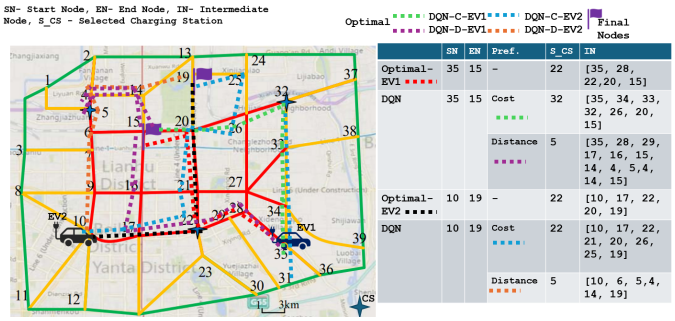
(a) With uniform distribution



	SN	EN	Pref.	S_CS	IN
Optimal-EV1	36	19	-	32	[36,35,34,33, 32,26, 25,19]
PPO	36	19	Cost	32	[36, 39, 37, 32, 26, 25, 19]
			Distance	32	[36, 39, 37, 32, 26, 20, 19]
Optimal-EV2	1	33	-	32	[1,2,13,24,26 ,32,33]
PPO	1	33	Cost	22	[1, 2, 7, 10, 17, 22, 21, 20, 26, 32, 33]
			Distance	5	[1, 2, 5, 10, 22, 28, 27, 33]

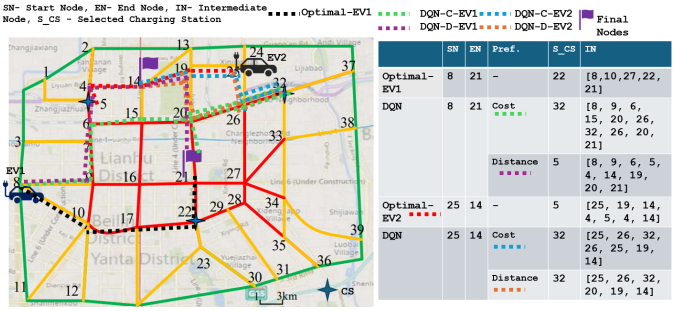
(b) With normal distribution

Fig. 10: Path navigation utilizing PPO models.



	SN	EN	Pref.	S_CS	IN
Optimal-EV1	35	15	-	22	[35, 28, 22,20, 15]
DQN	35	15	Cost	32	[35, 34, 33, 32, 26, 20, 15]
			Distance	5	[35, 28, 29, 17, 16, 15, 14, 4, 5,4, 14, 15]
Optimal-EV2	10	19	-	22	[10, 17, 22, 20, 19]
DQN	10	19	Cost	22	[10, 17, 22, 21, 20, 26, 25, 19]
			Distance	5	[10, 6, 5,4, 14, 19]

(a) With uniform distribution

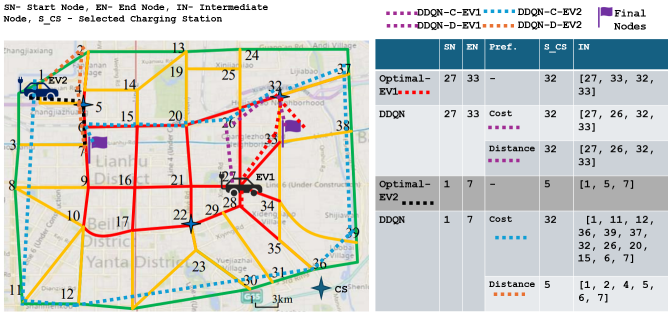


	SN	EN	Pref.	S_CS	IN
Optimal-EV1	8	21	-	22	[8,10,27,22, 21]
DQN	8	21	Cost	32	[8, 9, 6, 15, 20, 26, 32, 26, 20, 21]
			Distance	5	[8, 9, 6, 5, 4, 14, 19, 20, 21]
Optimal-EV2	25	14	-	5	[25, 19, 14, 4, 5, 4, 14]
DQN	25	14	Cost	32	[25, 26, 32, 26, 25, 19, 14]
			Distance	32	[25, 26, 32, 20, 19, 14]

(b) With normal distribution

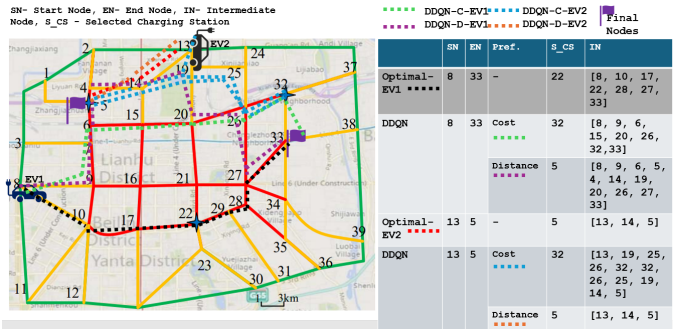
Fig. 11: Path navigation utilizing DQN models.

Moreover, in Figure 13, we also present the comparative path navigation utilizing PPO, DQN, and DDQN models, which shows that all the used models, PPO, DQN, and DDQN, learn effectively and choose the path as per the preference of the driver, i.e., cost or distance. Furthermore, the figure clearly shows that the PPO model will select a navigation path closer to optimal than the DQN and DDQN models.



	SN	EN	Pref.	S_CS	IN
Optimal-EV1	27	33	-	32	[27, 33, 32, 33]
DDQN	27	33	Cost	32	[27, 26, 32, 33]
			Distance	32	[27, 26, 32, 33]
Optimal-EV2	1	7	-	5	[1, 5, 7]
DDQN	1	7	Cost	32	[1, 11, 12, 36, 39, 37, 32, 26, 20, 15, 6, 7]
			Distance	5	[1, 2, 4, 5, 6, 7]

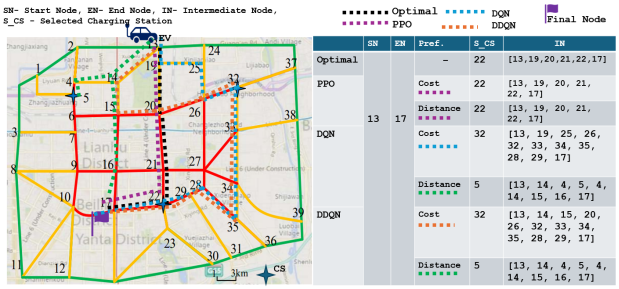
(a) With uniform distribution



	SN	EN	Pref.	S_CS	IN
Optimal-EV1	8	33	-	22	[8, 10, 17, 22, 28, 27, 33]
DDQN	8	33	Cost	32	[8, 9, 6, 15, 20, 26, 32,33]
			Distance	5	[8, 9, 6, 5, 4, 14, 19, 20, 26, 27, 33]
Optimal-EV2	13	5	-	5	[13, 14, 5]
DDQN	13	5	Cost	32	[13, 19, 25, 26, 32, 32, 26, 25, 19, 14, 5]
			Distance	5	[13, 14, 5]

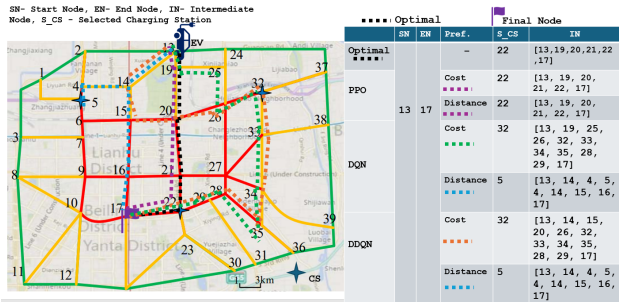
(b) With normal distribution

Fig. 12: Path navigation utilizing DDQN models.



	SN	EN	Pref.	S_CS	IN
Optimal	13	17	-	22	[13,19,20,21,22,17]
PPO			Cost	22	[13, 19, 20, 21, 22, 17]
			Distance	22	[13, 19, 20, 21, 22, 17]
DQN			Cost	32	[13, 19, 25, 26, 32, 33, 34, 35, 28, 29, 17]
			Distance	5	[13, 14, 4, 5, 4, 14, 15, 16, 17]
DDQN			Cost	32	[13, 14, 15, 20, 26, 32, 33, 34, 35, 28, 29, 17]
			Distance	5	[13, 14, 4, 5, 4, 14, 15, 16, 17]

(a) With uniform distribution



	SN	EN	Pref.	S_CS	IN
Optimal	13	17	-	22	[13,19,20,21,22 ,17]
PPO			Cost	22	[13, 19, 20, 21, 22, 17]
			Distance	22	[13, 19, 20, 21, 22, 17]
DQN			Cost	32	[13, 19, 25, 26, 32, 33, 34, 35, 28, 29, 17]
			Distance	5	[13, 14, 4, 5, 4, 14, 15, 16, 17]
DDQN			Cost	32	[13, 14, 15, 20, 26, 32, 33, 34, 35, 28, 29, 17]
			Distance	5	[13, 14, 4, 5, 4, 14, 15, 16, 17]

(b) With normal distribution

Fig. 13: Comparative path navigation utilizing PPO, DQN, and DDQN models.

E. Comparative Analysis of Cumulative Cost

The proposed PPO-based charging navigation algorithm is evaluated and compared with five benchmark strategies. These strategies are summarized as follows:

i. **DQN with driver preference:** This strategy uses a DQN algorithm that incorporates driver preferences for cost or distance.
ii. **DQN-based without driver preference** [31]: In this strategy, the DQN-based route and charging station selection algorithm optimizes route and charging station selection to minimize total travel time, addressing traffic uncertainties and dynamic EV charging requests, without factoring in user preferences.
iii. **Minimum distance strategy (MDS)** [18], [31], [47], [48]: This strategy minimizes the total distance travelled, ignoring charging and waiting times at the EVCS.
iv. **Minimum Travel Time Strategy (MTTS)** [31], [49]: This strategy aims to minimize total travel time, including driving, waiting, and charging times. The MTTS selects the EVCS and route, resulting in the shortest total travel time.
v. **Minimum waiting time strategy (MWTS)** [31], [49], [50]: This strategy prioritizes minimizing waiting time at the EVCS while also selecting a route with minimal driving time.

Table XI presents the comparative analysis of cumulative costs for 100 EVs under both uniform and normal distributions. As Table XI demonstrates, the proposed (PPO + driver preference) approach consistently achieves a lower cumulative cost than all five benchmark strategies in both distributions. Furthermore, in the uniform distribution case, the proposed approach outperforms the DQN with driver preference, DQN-based without preference, MDS, MTTS, and MWTS strategies with performance improvements of approximately 24.54 %, 62.58 %, 73.84 %, 70.30 %, and 65.14 %, respectively. In the normal distribution case, the proposed approach shows improvements of about 76.28 %, 69.58 %, 78.33 %, 71.10 %, and 75.66 %, respectively. Moreover, Figure 14 effectively illustrates the performance improvements of the proposed approach in terms of cumulative cost.

TABLE XI: Comparative Analysis of Cumulative Cost for 100 EVs.

Strategies	Uniform Distribution	Normal Distribution
Proposed (PPO + driver preference)	37.18	39.62
DQN with driver preference	49.27	167.00
DQN-based without driver preference	99.36	130.23
MDS	142.14	182.80
MTTS	103.40	137.08
MWTS	125.19	162.76

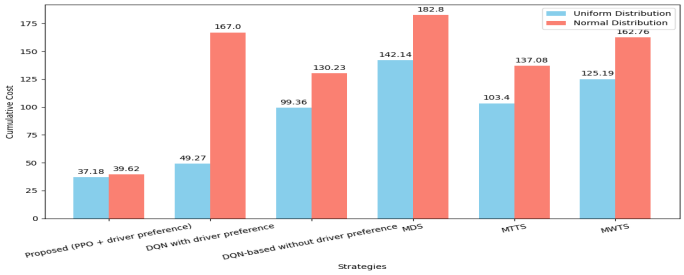


Fig. 14: Comparative analysis of cumulative cost with benchmark strategies.

F. Discussion

The machine learning models, particularly reinforcement learning, offer advantages over traditional mathematical programming, including adaptability to dynamic environments like fluctuating electricity prices and varying charging demands, scalability for handling larger, evolving datasets, the ability to capture complex non-linear relationships, and real-time decision-making, which enables responsive EV charging scheduling and management based on current conditions. Therefore, this paper introduces a novel method for scheduling EV charging navigation, considering driver preferences based on DRL. Our main goal is to create affordable charging schedules that increase drivers' profitability by considering driver preferences in terms of cost and distance. To demonstrate the effectiveness and adaptability of our EV charging navigation technique, We carried out experiments with varying numbers of EV charging requests arriving in a single day: "80", "100", "120", and "140", utilizing both uniform and normal distributions. We conducted a comprehensive performance analysis of our approach utilizing RL models: DQN, DDQN, and PPO.

We selected DQN, DDQN, and PPO for EV charging scheduling and navigation due to their unique strengths in reinforcement learning. DQN and DDQN excel in decision-making within discrete environments, effectively addressing overestimation issues, while PPO offers a stable and robust solution for both discrete and continuous control tasks. This combination enables the suggested approach to explore the advantages of value-based and policy-based learning strategies, ensuring optimized performance in EV charging navigation.

In Figure 5, PPO consistently outperforms DQN and DDQN in average cumulative reward, achieving stability within 400 to 500 episodes compared to DQN and DDQN, which stabilize after 1500 episodes. All models exhibit an apparent convergence trend, with cumulative rewards gradually increasing over time under uniform and normal distributions. However, DQN and DDQN display fluctuations and instability in their learning curves. Furthermore, our approach selects the EV charging station for each EV to minimize travel time, as illustrated in Figure 6. Figure 6 highlights that the PPO algorithm outperforms DQN and DDQN, sharply reducing cumulative travel time as training progresses and computing the minimum travel time compared to the others under both uniform and normal distributions across all numbers of EV charging requests arriving in a single day. Additionally, as shown in Figure 13, the comparative path navigation using PPO, DQN, and DDQN demonstrates that all models effectively learn and select routes based on the driver's preferences. However, the figure highlights that the PPO model consistently selects a navigation path closer to the optimal solution than DQN and DDQN.

Moreover, a comparative analysis of cumulative costs with benchmark strategies shows that the proposed approach outperforms several benchmark strategies, including DQN with driver preference, DQN without preference, MDS, MTTS, and MWTS, with performance improvements of approximately 24.54 %, 62.58 %, 73.84 %, 70.30 %, and 65.14 %, respectively, in uniform distribution. In the normal distribution case, it achieves even more significant improvements of 76.28 %, 69.58 %, 78.33 %, 71.10 %, and 75.66 %, respectively.

The findings from the preceding discussion indicate the success of our suggested EV charging navigation strategy, particularly in accommodating driver preferences. Furthermore, the PPO algorithm outperforms DQN and DDQN, two established reinforcement learning algorithms, demonstrating the efficacy of the suggested approach.

VI. Conclusion

This paper addresses the challenges of EV charging navigation by proposing a practical method based on driver preferences, such as cost or distance minimization, using model-free DRL. The proposed approach, which utilizes three RL models, DQN, DDQN, and PPO, effectively reduces overall costs, encompassing travel expenses, charging, and waiting costs. Furthermore, the method optimizes the overall distance travelled by considering both the distance to reach charging stations and the distance travelled without charging stations. The experimental results, spanning different numbers of EV charge requests and distribution types, highlight the superiority of the PPO model over DQN and DDQN in minimizing both cost and distance travelled while considering driver preferences. Notably, the PPO model demonstrates fast convergence, with average cumulative waiting and travel costs decreasing more rapidly than DQN and DDQN.

Additionally, when evaluating cumulative costs across different numbers of EVs and distribution types, PPO consistently outperforms other models. However, the comparison of distance costs produces mixed results in identical scenarios. A comparative analysis of cumulative costs against benchmark strategies reveals that the proposed approach surpasses several benchmarks, including DQN with and without driver preference, MDS, MTTS, and MWTS. These findings highlight the importance of personalized navigation systems in fostering the widespread adoption of EVs and advancing sustainable transportation solutions.

Acknowledgement

Authors acknowledge Visvesvaraya PhD Scheme, MeitY, Govt. of India MEITY-PHD-2525 for supporting this research.

References

[1] A. Ghosh, "Possibilities and challenges for the inclusion of the electric vehicle (ev) to reduce the carbon footprint in the transport sector: A review," *Energies*, vol. 13, no. 10, p. 2602, 2020.
[2] J. Zhang, J. Yan, Y. Liu, H. Zhang, and G. Lv, "Daily electric vehicle charging load profiles considering demographics of vehicle users," *Applied Energy*, vol. 274, p. 115063, 2020.
[3] W. Lee, R. Schober, and V. W. Wong, "An analysis of price competition in heterogeneous electric vehicle charging stations," *IEEE Transactions on Smart Grid*, vol. 10, no. 4, pp. 3990–4002, 2018.
[4] C. Liu, K. Chau, D. Wu, and S. Gao, "Opportunities and challenges of vehicle-to-home, vehicle-to-vehicle, and vehicle-to-grid technologies," *Proceedings of the IEEE*, vol. 101, no. 11, pp. 2409–2427, 2013.
[5] F. C. Silva, M. A. Ahmed, J. M. Martínez, and Y.-C. Kim, "Design and implementation of a blockchain-based energy trading platform for electric vehicles in smart campus parking lots," *Energies*, vol. 12, no. 24, p. 4814, 2019.
[6] J. Tan and L. Wang, "Real-time charging navigation of electric vehicles to fast charging stations: A hierarchical game approach," *IEEE transactions on smart grid*, vol. 8, no. 2, pp. 846–856, 2015.

[7] S. Kim and H. Lim, "Reinforcement learning based energy management algorithm for smart energy buildings," *Energies*, vol. 11, no. 8, p. 2010, 2018.
[8] Z. Wan, H. Li, H. He, and D. Prokhorov, "Model-free real-time ev charging scheduling based on deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 10, no. 5, pp. 5246–5257, 2018.
[9] F. Wang, J. Gao, M. Li, and L. Zhao, "Autonomous pev charging scheduling using dyna-q reinforcement learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 11, pp. 12 609–12 620, 2020.
[10] H. Li, Z. Wan, and H. He, "Constrained ev charging scheduling based on safe deep reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2427–2439, 2019.
[11] F. Zhang, Q. Yang, and D. An, "Cddpg: A deep-reinforcement-learning-based approach for electric vehicle charging control," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3075–3087, 2020.
[12] L. Yan, X. Chen, J. Zhou, Y. Chen, and J. Wen, "Deep reinforcement learning for continuous electric vehicles charging control with dynamic user behaviors," *IEEE Transactions on Smart Grid*, vol. 12, no. 6, pp. 5124–5134, 2021.
[13] S. Li, W. Hu, D. Cao, T. Dragičević, Q. Huang, Z. Chen, and F. Blaabjerg, "Electric vehicle charging management based on deep reinforcement learning," *Journal of Modern Power Systems and Clean Energy*, vol. 10, no. 3, pp. 719–730, 2021.
[14] C. Jin, J. Tang, and P. Ghosh, "Optimizing electric vehicle charging: A customer's perspective," *IEEE Transactions on vehicular technology*, vol. 62, no. 7, pp. 2919–2927, 2013.
[15] Q. Guo, S. Xin, H. Sun, Z. Li, and B. Zhang, "Rapid-charging navigation of electric vehicles based on real-time power systems and traffic data," *IEEE Transactions on smart grid*, vol. 5, no. 4, pp. 1969–1979, 2014.
[16] J.-Y. Yang, L.-D. Chou, and Y.-J. Chang, "Electric-vehicle navigation system based on power consumption," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 8, pp. 5930–5943, 2015.
[17] H. Yang, Y. Deng, J. Qiu, M. Li, M. Lai, and Z. Y. Dong, "Electric vehicle route selection and charging navigation strategy based on crowd sensing," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 5, pp. 2214–2226, 2017.
[18] W. Mo, C. Yang, X. Chen, K. Lin, and S. Duan, "Optimal charging navigation strategy design for rapid charging electric vehicles," *Energies*, vol. 12, no. 6, p. 962, 2019.
[19] X. Zhang, L. Peng, Y. Cao, S. Liu, H. Zhou, and K. Huang, "Towards holistic charging management for urban electric taxi via a hybrid deployment of battery charging and swap stations," *Renewable Energy*, vol. 155, pp. 703–716, 2020.
[20] K. Mason and S. Grijalva, "A review of reinforcement learning for autonomous building energy management," *Computers & Electrical Engineering*, vol. 78, pp. 300–312, 2019.
[21] S. Lee and D.-H. Choi, "Reinforcement learning-based energy management of smart home with rooftop solar photovoltaic system, energy storage system, and home appliances," *Sensors*, vol. 19, no. 18, p. 3937, 2019.
[22] S. Kim and H. Lim, "Reinforcement learning based energy management algorithm for smart energy buildings," *Energies*, vol. 11, no. 8, p. 2010, 2018.
[23] N. Sadeghianpourhamami, J. Deleu, and C. Develder, "Definition and evaluation of model-free coordination of electrical vehicle charging with reinforcement learning," *IEEE Transactions on Smart Grid*, vol. 11, no. 1, pp. 203–214, 2019.
[24] S. Wang, S. Bi, and Y. A. Zhang, "Reinforcement learning for real-time pricing and scheduling control in ev charging stations," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 2, pp. 849–859, 2019.
[25] T. Ding, Z. Zeng, J. Bai, B. Qin, Y. Yang, and M. Shahidehpour, "Optimal electric vehicle charging strategy with markov decision process and reinforcement learning technique," *IEEE Transactions on Industry Applications*, vol. 56, no. 5, pp. 5811–5823, 2020.
[26] D. Liu, P. Zeng, S. Cui, and C. Song, "Deep reinforcement learning for charging scheduling of electric vehicles considering distribution network voltage stability," *Sensors*, vol. 23, no. 3, p. 1618, 2023.
[27] I. Azzouz and W. Fekih Hassen, "Optimization of electric vehicles charging scheduling based on deep reinforcement learning: A decentralized approach," *Energies*, vol. 16, no. 24, p. 8102, 2023.
[28] S. Mishra, A. Choubey, S. V. Devarasetty, N. Sharma, and R. Misra, "An innovative multi-head attention model with bimgru for real-time electric vehicle charging management through deep reinforcement learning," *Cluster Computing*, pp. 1–31, 2024.
[29] A. Zhang, Q. Liu, J. Liu, and L. Cheng, "Casa: cost-effective ev charging scheduling based on deep reinforcement learning," *Neural Computing and Applications*, pp. 1–16, 2024.
[30] L. Zhang, K. Gong, and M. Xu, "Congestion control in charging stations allocation with q-learning," *Sustainability*, vol. 11, no. 14, p. 3900, 2019.
[31] T. Qian, C. Shao, X. Wang, and M. Shahidehpour, "Deep reinforcement learning for ev charging navigation by coordinating smart grid and intelligent transportation system," *IEEE transactions on smart grid*, vol. 11, no. 2, pp. 1714–1723, 2019.
[32] K.-B. Lee, M. A. Ahmed, D.-K. Kang, and Y.-C. Kim, "Deep reinforcement learning based optimal route and charging station selection," *Energies*, vol. 13, no. 23, p. 6255, 2020.
[33] C. Zhang, Y. Liu, F. Wu, B. Tang, and W. Fan, "Effective charging planning based on deep reinforcement learning for electric vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 1, pp. 542–554, 2020.
[34] A. Yazdekhasti, M. A. Jazi, and J. Ma, "Electric vehicle charging station location determination with consideration of routing selection policies and driver's risk preference," *Computers & Industrial Engineering*, vol. 162, p. 107674, 2021.
[35] T. Ye, S. Liu, and E. Kontou, "Managed residential electric vehicle charging minimizes electricity bills while meeting driver and community preferences," *Transport Policy*, vol. 149, pp. 122–138, 2024.
[36] P. W. Eklund, S. Kirkby, and S. Pollitt, "A dynamic multi-source dijkstra's algorithm for vehicle routing," in *1996 Australian New Zealand Conference on Intelligent Information Systems. Proceedings. ANZIIS 96*. IEEE, 1996, pp. 329–333.
[37] C. Jiang, L. Zhou, J. Zheng, and Z. Shao, "Electric vehicle charging navigation strategy in coupled smart grid and transportation network: A hierarchical reinforcement learning approach," *International Journal of Electrical Power & Energy Systems*, vol. 157, p. 109823, 2024.
[38] F. C. Silva, M. A. Ahmed, J. M. Martínez, and Y.-C. Kim, "Design and implementation of a blockchain-based energy trading platform for electric vehicles in smart campus parking lots," *Energies*, vol. 12, no. 24, p. 4814, 2019.
[39] C. J. Watkins and P. Dayan, "Q-learning," *Machine learning*, vol. 8, pp. 279–292, 1992.
[40] N. Mhaisen, N. Fetais, and A. Massoud, "Real-time scheduling for electric vehicles charging/discharging using reinforcement learning," in *2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies (ICIoT)*. IEEE, 2020, pp. 1–6.
[41] S. Lee and D.-H. Choi, "Reinforcement learning-based energy management of smart home with rooftop solar photovoltaic system, energy storage system, and home appliances," *Sensors*, vol. 19, no. 18, p. 3937, 2019.
[42] J. Lee, E. Lee, and J. Kim, "Electric vehicle charging and discharging algorithm based on reinforcement learning with data-driven approach in dynamic pricing scheme," *Energies*, vol. 13, no. 8, p. 1950, 2020.
[43] D. M. Miranda and S. V. Conceição, "The vehicle routing problem with hard time windows and stochastic travel and service time," *Expert Systems with Applications*, vol. 64, pp. 104–116, 2016.
[44] M. Çimen and M. Soysal, "Time-dependent green vehicle routing problem with stochastic vehicle speeds: An approximate dynamic programming algorithm," *Transportation Research Part D: Transport and Environment*, vol. 54, pp. 82–98, 2017.
[45] "TensorFlow Framework," https://www.tensorflow.org/, accessed on 12 December 2023.
[46] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, "Stable-baselines3: Reliable reinforcement learning implementations," *Journal of Machine Learning Research*, vol. 22, no. 268, pp. 1–8, 2021. [Online]. Available: http://jmlr.org/papers/v22/20-1364.html
[47] J. Tan and L. Wang, "Real-time charging navigation of electric vehicles to fast charging stations: A hierarchical game approach," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 846–856, 2017.
[48] F. Xia, H. Chen, L. Chen, and X. Qin, "A hierarchical navigation strategy of ev fast charging based on dynamic scene," *IEEE Access*, vol. 7, pp. 29 173–29 184, 2019.
[49] Y. Cao, X. Zhang, R. Wang, L. Peng, N. Aslam, and X. Chen, "Applying dtn routing for reservation-driven ev charging management in smart cities," in *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2017, pp. 1471–1476.
[50] Y. Cao, S. Liu, Z. He, X. Dai, X. Xie, R. Wang, and S. Yu, "Electric vehicle charging reservation under preemptive service," in *2019 1st International Conference on Industrial Artificial Intelligence (IAI)*, 2019, pp. 1–6.