RESEARCH ARTICLE

INTERNATIONAL RESEARCH JOURNAL OF MULTIDISCIPLINARY TECHNOVATION

Asian Research Association

DOI: 10.54392/irjmt26214

Optimal Charging and Discharging Scheduling of Electric Vehicles using Reinforcement Learning for Efficient Operation

C. Srinivasan [a, *], C. Sheeba Joice [b]

[a] Department of EEE, Rajalakshmi Engineering College, Chennai, India
[b] Department of ECE, Saveetha Engineering College, Chennai, India.
* Corresponding Author Email: srinivasanc@ieee.org

DOI: https://doi.org/10.54392/irjmt26214
Received: 05-11-2025; Revised: 11-03-2026; Accepted: 23-03-20XX; Published: 28-03-2026

Abstract: The rapid growth of electric vehicles (EVs) has increased the demand for efficient charging and discharging strategies. This paper addresses EV scheduling in small cities to enhance charging efficiency and reduce user costs. We propose a hybrid framework that combines Deep Deterministic Policy Gradient (DDPG) with Hindsight Experience Replay (HER) and employs a Genetic Algorithm (GA) for hyperparameter optimization. The approach simultaneously targets three objectives: maximizing aggregator profit, minimizing EV charging cost, and maximizing the proportion of EVs departing with the target State of Charge (t-SOC). Simulations are performed under three scenarios with different aggregator capacities and EV counts: Scenario 1 (50 kW, 100 EVs), Scenario 2 (500 kW, 750 EVs), and Scenario 3 (50 kW, 150 EVs). Results indicate that the optimized Reinforcement Learning (RL) model outperforms conventional and other RL approaches like First Come First Serve (FCFS), Proximal Policy Optimization (PPO), Soft Actor-Critic (SAC). Aggregator profits increase from \$148–\$503 (FCFS), \$158–\$530 (PPO), \$162–\$545 (SAC) to \$181–\$571 (optimized RL); EV charging costs decrease from \$0.17–\$0.25/kWh (FCFS), \$0.12–\$0.15/kWh (PPO), \$0.11–\$0.14/kWh (SAC) to \$0.08–\$0.09/kWh (optimized RL); and the percentage of EVs achieving t-SOC rises from 55–71% (FCFS), 76–85% (PPO), 78–87% (SAC) to 87–94% (optimized RL). These findings demonstrate that the proposed DDPG + HER + GA framework effectively balances multiple objectives while adhering to operational constraints, providing a robust solution for real-time EV charge-discharge scheduling.

Keywords: Electric Vehicle, Reinforcement Learning, Aggregator, Charging cost, Electricity, Optimization.

1. Introduction

In view of growing worldwide environmental challenges such as air pollution and climate change, an increasing number of individuals regard EVs as a viable alternative to cars powered by fossil fuels [1]. As a result, the number of EVs and charging stations is fast increasing, fuelled by government funding and technological advancements. Recent research has found that the widespread adoption of EVs will significantly increase electricity demand and introduce new operational challenges to distribution networks between 2023 and 2025 [2]. Due to this growth in alternative vehicles, the grid may experience power outages and voltage fluctuations, which are expected to alter the demand profile [3]. Therefore, effective EV charging planning is considered essential to ensure stable grid operation. In addition, to enhance the resilience of smart grids, it is necessary to manage the intermittent and unpredictable nature of renewable generation, which is becoming increasingly significant as the share of renewable energy in total production rises [4]. Dynamic pricing, one of the most widely used demand-side mechanisms, facilitates load shifting between peak and off-peak hours by providing consumers with economic incentives to adjust their electricity usage [5]. Hourly pricing, also referred to as real-time pricing, is another approach in which power tariffs vary hourly. A notable example is the hourly rate applied by the power utility "ComEd" in Illinois, USA, to both residential clients and EVs [6]. Real-time pricing promotes economic efficiency by capturing supply and demand instantaneously and represents the future of electricity pricing schemes. Similarly, vehicle-to-grid (V2G) technology has enabled EV batteries to serve as a viable energy source that can be fed back to the grid. It has been demonstrated that coordinated V2G development can reduce peak load, improve grid flexibility, and provide economic benefits to EV aggregators and users [7]. Price signals can be used to schedule real-time charging and discharging to reduce EV charging costs and maintain grid reliability. This strategy involves the integration of V2G and dynamic pricing. However, monitoring real-time price signals to proactively control charging and discharging remains a major challenge for EV operators. Consequently, there

is a need for algorithms that can automatically interpret pricing cues and determine optimal charging and discharging actions.

Regarding the scheduling of EV charging and discharging, three primary approaches are commonly used: dynamic programming–based, day-ahead scheduling–based, and model-based. Recent studies have proposed optimization methods and heuristic algorithms to improve EV charging schedules and reduce operational costs [8, 9]. However, traditional algorithmic approaches alone struggle to optimize EV charging effectively [10]. This is because the high number of EVs leads to high-dimensional optimization variables, often referred to as the curse of dimensionality. Furthermore, uncertainties in algorithm control and overall performance—caused by fluctuations in the energy system and the inability to accurately predict EV demand—limit the development of precise models and complicate control. RL algorithms offer a potential solution to the complexity of EV charging schedules. These approaches are effective for sequential decision-making problems even when a complete model of the environment is unavailable. In recent years, deep RL models such as DDPG, PPO, and SAC have been applied in energy management and EV charging coordination, as they can handle continuous control tasks in dynamic environments. Some studies have shown that RL-based charging strategies can reduce operational costs and improve system efficiency compared to traditional scheduling approaches. However, RL algorithms can be slow to converge and unstable in high-dimensional settings, particularly when trained with sparse or delayed feedback.

This paper focuses on optimizing the timing of EV charging and discharging using RL. The proposed framework integrates DDPG with HER to improve learning efficiency, while the GA is used to optimize training procedures and model parameters. This combined strategy is expected to enhance scheduling performance and stability under varying operational conditions by leveraging both evolutionary optimization and RL.The key contributions of this study to EV charging and discharging scheduling are:

- The three objective functions used to assess EV systems are: increasing aggregator profitability, reducing EV charging costs, and maximizing departing EVs with a t-SOC.
- The study evaluates the efficacy of three models in achieving these objectives: FCFS, RL without optimization, and RL with optimization.
- To gain better insight into model dynamics under different conditions, experiments are conducted under three scenarios that vary EV volumes and power capacities.

2. Related Works

The current works on the scheduling of EV charging and discharging are as follows. The analysis [11] addresses the challenge of balancing the fundamental needs of the power grid with the charging requirements of EVs while considering the environmental consequences of power-generating resources. This study assumes a high-confidence wind power scenario. The structural development of a multi-objective optimal scheduling model is based on this scenario, taking into consideration the following aspects: V2G characteristics, generator operation cost, pollution rate, unused aeration, and the price charged to EV clients. The problem is solved using the CPLEX solver program. The proposed optimal scheduling technique enables joint optimization of EVs, wind energy, and thermal power stations, as evidenced by simulation results. The article [12] proposes a three-tiered infrastructure to manage and streamline EV operation modes in grid-to-vehicle (G2V) and V2G applications within a smart Micro-Grid (MG). The first tier aims to enhance efficiency and flexibility. Further effectiveness of the MG is achieved through a Demand Response Program (DRP) based on Time-Of-Use pricing. This paradigm assumes that the G2V and V2G modes of EVs are treated as commodities in the power market, with the V2G mode contributing to the efficiency and flexibility of the MG. The distribution system operator is responsible for managing transactions based on clearing prices. Among the various factors influencing EV owners' participation in demand response networks, range anxiety is a crucial one that the study considers. Simulation results validate the suggested three-level structure by demonstrating a significant increase in MG efficiency and flexibility when compared to the DRP case study.

The purpose of this research [13] is to create a hierarchical stochastic optimal scheduling framework for an Electric Thermal Hydrogen Integrated Energy System (ETH-IES) that integrates EVs' V2G function. The goal of the EV charging and discharging management layer is to make V2G customers happy by reducing load curve variance. In the paper, the MSCSO technique, which refers to multi-objective Sand Cat Swarm Optimization, is utilized to solve the suggested model. This basic study is then extended to look into ways to cut operating expenses at ETH-IES through daily stochastic economic scheduling. According to the simulation results, both MG operators and EV owners stand to benefit from the proposed technique. When compared to an uncontrolled charging and discharging approach, the proposed model and algorithm are effective since they reduce operation costs. The study's [14] purpose is to optimize the operation of an EV and an energy storage system to reduce power costs and peak demand in a smart house. The study gives multiple decision vectors for charging and discharging based on market prices, the power production of residential solar panels, and the electricity

demand of the smart house. The final stage is to identify the decision vectors by ranking the selection criteria using the Analytical Hierarchy Process technique. When selecting the EV and electrical storage system operation schedule, these final decision vectors take operational constraints and charging/discharging priority into account. The proposed approach is tested on a representative smart house with varying priorities for the selection criteria. The proposed method performs satisfactorily in all scenarios for determining the best time to run EVs and power storage systems. The research [15] proposes a Deep RL technique for scheduling EV charging and distribution network voltage management. The first layer of the DRL-based strategy is concerned with minimizing distributed generator operational expenditures and EV power consumption; the second layer regulates devices to ensure distribution network voltage stability. To account for charging process uncertainties induced by EV user behaviour, unpredictable load, dynamic power costs, and renewable energy output, they propose a Markov Decision Process (MDP) framework for the coordinating problem. The researchers trained a two-layer agent employing a DDPG architecture, which results in a policy network with outcomes of both continuous and discrete control actions. The simulation findings demonstrate that the recommended approach works effectively for voltage stability in distribution networks as well as collaborative scheduling of EV charging stations.

The paper [16] provides a hybrid technique termed Balancing Composite Motion Optimization (BCMO) and Political Optimizer (PO), abbreviated BCMPO, for the optimal scheduling of EV aggregators with market value uncertainty. The hybrid optimization system studies how EV aggregators might be programmed to maximize profitability in the face of unpredictable market pricing. Market price volatility is reported due to the high and low upstream grid rates used by the EV aggregator to increase revenues. Considering the unpredictability of upstream grid prices, the proposed algorithm offers alternative methods for EV aggregators to schedule charging and discharging effectively. The effectiveness of the proposed solution is compared with existing systems using the MATLAB platform. The results demonstrate that the proposed strategy is effective, as EV aggregators can earn more revenue than in previous approaches. The paper [17] outlines a power management approach for networked MGs and EV fleets, supported by a dedicated EV charging/discharging scheduling technique. The difference between photovoltaic (PV) power generation and local load demand is used to determine the maximum discharge power of parked EVs, aiming to eliminate imbalances. Multi-objective optimization is then applied to reduce the operating costs of the PV-based charging station and its reliance on the grid. The proposed technique enhances EV charging and discharging to satisfy the charging requirements. Moreover, a Gated Recurrent Unit (GRU) network is utilized to model the PV-based charging station, providing a cost-effective and energy-efficient solution by predicting future behaviors. Compared to existing algorithms, the proposed EV charging and discharging approach lowers the operating costs of PV-based charging stations. Overall, this strategy emphasizes efficient and cost-effective operation of PV-based charging stations through the proposed power management system and EV charging/discharging scheduling method.

3. Reinforcement Learning

Consider a typical RL scenario in which a learning agent interacts with its environment. State (S), action (A), initial state distribution $p(s_0)$, and transition probability $r: S * A \rightarrow R, p(s_{t+1}|s_t, a_t)$ are the variables that constitute the environment. $\gamma \in [0,1]$ also provides a discount factor [18]. In this setting, a policy is said to be deterministic if it maps states to actions: $\pi : S \rightarrow A$. Every episode starts with a snapshot of the initial state s_0. At each time step t, the agent reacts based on its current state: $a_t = \pi(s_t)$. The distribution $p(. s_t, a_t)$ aids in sampling the new environment state, and the action's reward $r_t = r(s_t, a_t)$ is a result of the action. The sum of all returns, R_t, is calculated by multiplying the sum of all returns, $\sum_{i=T}^{\infty} \gamma^{i-t} r_i$. An ideal policy, denoted by π^*, is one that maximizes the expected return $E[R_t|s_t, a_t]$ for every $s \in S$, $a \in A$ and any policy. The optimum course of action is connected to an ideal Q-function Q^* that shares the same Q function as the Bellman equation [19].

$$Q^*(s,a) = E_{s' p(.|s,a))}[r, (s,a) + \gamma max_{a' \in A} Q^*(s', a') \quad [1]$$

Problem Formulation:

State ($\boldsymbol{S}$) : $\left[SOC_i, t_{arrival,i}, t_{depart,i}, P_{i,max}, EV_{mode,i}, P_{up,t}, P_{down,t}, N_t, P_T\right]$

Action ($\boldsymbol{A}$) : Continuous Charging/discharging powers $a_t = \left[P_{c,i,t}, P_{d,i,t}\right]$, clipped by $0 \leq P_{c,i,t}, P_{d,i,t} \leq P_{i,max}$

Reward ($\boldsymbol{R}$): Multi-objective reward combining aggregator profit, EV cost reduction, and number of EVs departing at t-SOC:

$$r_t = w_1 . profit_t - w_2 . cost_{EV\ s,t} + w_3 . n_{target,t}$$

Constraints: SOC limits $SOC_{min} \leq SOC_{i,t} \leq SOC_{max}$, transformer power $\left|P_{up,t} + P_{down,t}\right| \leq P_T$, parking limit $N_t \leq n$. Violations are penalized in reward or corrected by action clipping.

Episode: Each episode is a 24-hour simulation with 30-minute steps. Ends when all EVs depart.

3.1 DDPG

Two Neural Networks (NN), one acting as an actor and the other as a critic, are involved in the DDPG paradigm [20]. The Critic NN approximates the action-value function $Q: S * A \rightarrow R$, whereas the Actor NN represents a target policy $\pi: S \rightarrow A$. The critic NN denoted as $Q(s,a|\theta^Q)$, and the actor network, denoted as $\mu(s|\theta^\mu)$, are both started with weights θ^Q and θ^μ that are randomly chosen. An episode-generating behavioral policy, $\pi_b(s) = \pi(s) + \aleph(0,1)$, is used to depict a noisy version of the target policy. Deep Q Networks' Q function is in sync with the critic NN's training. The equation $y_t = r_t + \gamma Q(s_{t+1}, \pi(s_{t+1}))$, is used to calculate the target $y_t = r_t$, with γ being the discounting factor. For training the actor NN, the loss function $\mathcal{L}_a = -\mathrm{E}_a Q\big(s, \pi(s)\big)$ is utilized.

3.2 Hindsight Experience Replay (HER)

By learning from mistakes, HER attempts to mimic human learning [21]. When compared to traditional experience replay, HER takes a more holistic view, drawing lessons from every episode—even the ones that fail to accomplish the initial objective [22]. Whatever the agent achieves, HER will consider it a goal that has been adjusted. Instead of storing transitions as $(s_t||g, a_t, r_t, s_{t+1}||g)$, HER uses the modified goal g' and stores them as $\big(s_t||g', a_t, r_t', s_{t+1}||g'\big)$. With its superior performance in cases of highly sparse incentives, HER surpasses conventional approaches and shows considerable improvements for sparse payouts as compared to formed ones.

3.3 GA

Genetic Algorithms (GAs) are designed to explore complex and poorly understood search spaces where exhaustive search is infeasible or conventional optimization methods may fail [23]. This paper employs GAs to optimize the parameters of the DDPG + HER agent, enhancing both the convergence rate and task performance.

- Encoding of Chromosomes: A chromosome corresponds to a candidate RL parameter set, encoded as the polyak-averaging coefficient (τ), discount factor (γ), actor learning rate (α_actor), critic learning rate (α_critic), random action probability (ϵ), and noise standard deviation (η).
- Population Settings: The GA begins with 50 randomly initialized chromosomes sampled over the specified parameter ranges.
- Genetic Operators: Parents are selected using tournament selection (size 3) to form the next generation. Simulated Binary Crossover (SBX) is applied with a probability of 0.9 to create new offspring. Each gene has a probability of 0.1 of undergoing mutation, introducing variation.
- Fitness Function: The fitness of a chromosome is calculated as a weighted sum of three objectives: maximizing aggregator profit, minimizing EV charging costs, and maximizing the percentage of EVs that reach the t-SOC. Improved parameter configurations are reflected in higher fitness values.
- Stopping Criteria: The GA terminates after 100 generations or 20 generations without improvement in the best fitness.
- Training/Evaluation Budget: Each chromosome is used to train the RL agent over 50 episodes. The most effective chromosome is further evaluated across more than 10 simulation runs to ensure stability and reliability.

This GA-based parameter tuning ensures systematic exploration of the RL hyperparameter space and enhances the overall effectiveness of the hybrid DDPG + HER framework.

3.4 Hybrid RL with GA

Our proposed model is detailed here: we use a GA to reduce the number of training epochs while enhancing task performance by exploring the parameter space of DDPG + HER. The below parameters are being analysed: the polyak-averaging coefficient [24], the discounting factor (γ), the learning rates for the α_{critic}, and α_{actor}, the frequency of random actions, and the standard deviation of the maximum absolute score of actions on various coordinates (η)added to them expressed as a percentage. The equations that follow in this section justify setting all variable parameter ranges to 0 and 1.

When we modified parameter values experimentally, we found no linear or evident pattern of increased or decreased learning for the agents. As a result, improving these parameters with a conventional hill climber strategy is unlikely. We employ our GA, which was designed for challenging and poorly understood circumstances, to optimize these parameter values. We demonstrate the nonlinear performance for various values of employing the polyak-averaging coefficient (τ). As indicated in Equation (2-3), the algorithm is based on.

$$\theta^{Q'} \leftarrow \tau\theta^Q + (1-\tau)\theta^{Q'} \quad [2]$$

$$\theta^{\mu'} \leftarrow \tau\theta^\mu + (1-\tau)\theta^{\mu'} \quad [3]$$

Equation (4) depicts the use of γ in the hybrid RL algorithm, Equation (5) denotes the Q-Learning

update. Network training is based on this update equation.

$$y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{t+1}|\theta^{\mu'})|\theta^{Q'} \quad [4]$$

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)] \quad [5]$$

Due to the presence of two different networks, we need two learning rates, one for the α_{actor} and one for the α_{critic}. Equation (6) details the function of the percentage of occurrences of a random action, denoted as ϵ.

$$a_t = \begin{cases} a_t^* & with\ probability\ 1-\epsilon, \\ random\ action & with\ probability\ \epsilon. \end{cases} \quad [6]$$

Pseudocode for Hybrid RL with GA for EV Charge–Discharge Scheduling

Input: State space S, action space A, reward function r_t, GA parameter ranges

Output: Optimal policy π^*

Initialize GA population P with chromosomes $C = (\tau, \gamma, \alpha_{actor}, \alpha_{critic}, \epsilon, \eta)$

For generation $g = 1$ to G do

For each chromosome C_i in population

Initialize Actor $\mu(s|\theta^{\mu})$ and Critic $Q(s, a|\theta^{Q})$.

Initialize target networks$\theta^{\mu'} \leftarrow \theta^{\mu}, \theta^{Q'} \leftarrow \theta^{Q}$.

Initialize replay buffer B.

For episode= 1 to E

Reset EV scheduling environment and observe state s_t.

Select action $a_t = \mu(s_t) + N(0, \eta)$ with random action probability ϵ.

Clip action within $0 \leq P_{c,i,t}, P_{d,i,t} \leq P_{i,max}$

Execute action, observe reward r_t and next state s_{t+1}.

Store transition (s_t, a_t, r_t, s_{t+1}) in B.

Apply **HER** to relabel goals and store modified transitions.

Sample mini-batch from B and compute target: $y_i = r_i + \gamma Q'\left(s_{i+1}, \mu'(s_{i+1})\right)$

Update Critic by minimizing $L = \left(y_i - Q(s_i, a_i)\right)^2$

Update Actor using policy gradient $\nabla_{\theta^{\mu}} J$.

Update target networks using Polyak averaging: $\theta' \leftarrow \tau\theta + (1-\tau)\theta'$

End episode loop

Compute chromosome fitness based on aggregator profit, EV cost reduction, and n_{target}.

End chromosome loop

Apply **tournament selection, SBX crossover,** and **polynomial mutation** to generate next population.

End generation loop

Select best chromosome and retrain the RL agent.

Validate the final policy over multiple simulation runs and outputπ^*.

4. Objective Functions

The charge-discharge scheduling challenge is structured as an optimization problem with numerous objectives and constraints, including:

4.1 Improve Profit for Aggregators

An aggregator is a profit company that links owners of EVs to the grid. As demonstrated in the calculation, the corporation makes money by selling and buying grid power and EVs.

$$max(profit = \sum_{t=0}^{t=T} Revenue_t - cost_t) \quad [7]$$

In the given term, $Revenue_t$ represents the total revenue the aggregator made during the current time slot t, and $cost_t$ represents the total cost the aggregator incurred during that time. Grid income (from energy sales to the grid) and EV revenue (from charging EVs) are revenue streams, according to equations (8)-(10).

$$Revenue_t == Revenue_{grid,t} += Revenue_{EVs,t} \quad [8]$$

$$Revenue_{grid,t} = p_{up,t} * P_{up,t} * \Delta t \quad [9]$$

$$Revenue_{EVs,t} = \Delta t * \sum_{i=1}^{i=N_t} M_{c,i} * w_{c,i} * p_{base} * P_{c,i,t} \quad [10]$$

Where

i → EV identification

t → Current timeslot

$Revenue_{grid,t}$ → Revenue generated by supplying V2G power from EV to the grid.

$Revenue_{EVs,t}$→ Revenue generated by EV.

$p_{up,t}$ → Electricity grid's upstream regulation price ($/kWh).

$P_{up,t}$ → Grid Power (kW)

Δt → Duration of each time slot (Hours),

N_t → Total parked EVs at t,

$M_{c,i}$→ Rate modifier for EV charging.

p_{base} → Charging/discharging base price ($/kWh)

$P_{c,i,t}$ → Power to charge EV (kW)

Equation (11-13) provides a clear sketch of the expression for the expenses that the aggregator bears. These expenses include grid costs (electricity purchased for EV charging) and EV costs (power discharged and collected from the EVs).

$$Cost_t = cost_{grid.t} + cost_{EVs,t} \quad [11]$$

$$cost_{grid,t} = p_{down,t} * P_{down,t} * \Delta t \quad [12]$$

$$Cost_{EVs,t} = \Delta t * \sum_{i=1}^{i=N_t} M_{d,i} * w_{d,i} * P_{base} * P_{d,i,t} \quad [13]$$

Where

$cost_{grid,t}$ → Cost associated with grid-supplied energy,

$Cost_{EVs,t}$ → Cost of using parked EVs for V2G,

$P_{down,t}$ → Power derived from the grid (kW),

$p_{down,t}$ → Cost of purchasing power from the grid ($/kWh),

$M_{d,i}$ → EV discharging price modifier,

$w_{d,i}$ → Weight for the priority level of discharge,

$P_{d,i,t}$ → EV power used to discharge (kW).

The underlying premise is that the modifier value increases with increasing charging power for $M_{c,i}$ and decreases with increasing discharging power for $M_{d,i}$.

4.2 Reduce Charging Cost for Users

When an EV's parking period is up, the owner is required by law to pay the aggregator the amount known as the EV charging cost. This cost can be expressed as follows, depending on the EV's charging and discharging schedule:

$$min(EV_i = \sum_{t=0}^{t=T} Revenue_{EVi,t} - Cost_{EVi,t}) \quad [14]$$

Equations (10) and (13) can be used to calculate the values of $Revenue_{EVi,t}$ and $Cost_{EVi,t}$ for a specific EV.

4.3 Improve the Evs Count Departing with the T-SOC

In order for all of the EVs to reach the t-SOC, the aggregator might not be able to fully charge them all. So to attain the t-SOC, the goal is to have as many EVs as possible.

$$max(\sum_{t=0}^{t=T} n_{target,t}) \quad [14]$$

Where, n_{target} → Total EVs depart with t-SOC, T → Simulation time (s)

4.4 Desired SOC (d-SOC)

The d-SOC for an EV is calculated using Equation 9. Assuming the aggregator has enough resources, this assures that the EV will attain the t-SOC at the highest possible charging/discharging rate chosen by the EV user during the parking time.

$$C_i \left(\frac{SOC_{i,target} - SOC_{i,arrival}}{t_{i,depart} - t_{i,arrival}} \right) \leq P_{i,max}, \forall i \quad [15]$$

Where

C_i → EV battery capacity

$P_{i,max}$ → User's maximum charging/discharging power

$t_{i,arrival}$ → Arrival time

$t_{i,depart}$ → Departure time

$SOC_{i,arrival}$ → SOC at arrival

$SOC_{i,target}$ → SOC at departure

i. SOC limits:

The aggregator can schedule the EV for discharge only when the $SOC_i > SOC_{min}$, the minimal safety threshold of the battery. Similarly, the aggregator is able to charge an EV when $SOC_i < SOC_{max}$, which represents the maximum safety threshold of the battery.

$$SOC_{min} \leq SOC_{i,t} \leq SOC_{max}, \forall t \quad [16]$$

In acceptance of Tesla's guidelines for typical usage to optimize an EV's battery life, the assumed values for SOC_{min} and SOC_{max} are 0.4 and 0.8, respectively.

4.5 Aggregator Transformer Limit

All net power inputs and outputs to the grid must be within the limits set by the transformer's maximum power rating.

$$|P_{down,t} + P_{up,t}| \leq P_T, \forall t \quad [17]$$

Where, P_T → Maximum power of transformer.

4.6 Limitation in Parking Capacity

The aggregator's maximum parking capacity also needs to be met by the amount of EVs present at all times.

$$N_t \leq n, \ \forall t \quad [18]$$

Where, n → the total EVs accommodated by the aggregator.

4.7 Constraint Enforcement

All operational constraints are strictly enforced during the scheduling process. The SOC constraint in Equation (16) ensures that EV batteries operate within safe limits by preventing charging when the SOC reaches SOC_{max} and preventing discharging when the SOC approaches SOC_{min}. The transformer capacity

constraint in Equation (17) guarantees that the total net power exchanged with the grid does not exceed the transformer ratingP_T, thereby avoiding feeder overload. The parking capacity constraint in Equation (18) ensures that the number of EVs simultaneously scheduled does not exceed the available charging ports$n$. In addition, the charging and discharging actions selected by the RL agent are clipped within the allowable power range$0 \leq P_{c,i,t}, P_{d,i,t} \leq P_{i,max}$. When multiple EVs request charging or discharging simultaneously, the scheduling allocation is performed based on the available transformer capacity and charging slots to ensure that all constraints are satisfied before executing the final action.

5. System Modeling

Due to the inherent problems in undertaking significant real-world testing for the EV scheduling approach, such as economic impracticality and time limits, a simulator has been built to enable the execution of the suggested charging strategy. As indicated in Figure 1, the simulator's development relies on three key components: a) EV, b) Aggregator, and c) Market. This section provides a high-level overview of the modelling of various components.

5.1 EV Model

EVs are treated in this study as dynamic energy storage systems that can be connected (arrive) or disconnected (depart) at any time. The mobility patterns outlined in [25] can be used to infer the EV arrival rate and parking length, where the latter is reliant on the former. The designers intended to show the absolute unpredictability of an EV owner's parking experience by emulating the layout of a conventional leisure parking lot. As a result, the EV model accounts for the uncertainty of parking time. Upon arrival, each EV offers precise information such as its current SOC, goal SOC for when it departs, projected departure time, the capacity of the battery, charging and discharging speeds, and how urgently it needs a charging or discharging port. To make the system more realistic, we additionally consider whether EV owners choose G2V, V2G, both, or none (Idle). For the aggregator to figure out when to charge and discharge the batteries, all the EVs in the parking area send this data. EVs use this technology, which is known as V2G and G2V bidirectional charging capabilities. Under normal conditions, the EV's power rating in V2G mode can supply -1.92 kW to the grid, while it remains at 0 kW while the car is idle. In G2V mode, the EV may draw power from the grid at 1.92 kW during ordinary charging, 7.68 kW during fast charging, and 19.2 kW during ultra-rapid charging. The power ratings reveal the EV's operational efficiency in various scenarios and emphasize its ability to connect to and disconnect from the grid.

Charging and discharging processes are assumed to have finite efficiencies to enhance the physical realism of the EV model. In particular, the charge and discharge efficiencies, (η_c) and (η_d), are assumed to be below unity to indicate power conversion losses between the EV battery and the grid. These efficiencies affect the effective energy stored in or delivered from the battery during any given time step. Additionally, operational constraints—including maximum charging/discharging power $P_{i,max}$, minimum and maximum SOC, and parking time restrictions—are strictly enforced during scheduling. To safeguard battery health and minimize excessive cycling, the SOC operating range is limited between SOC_{min} andSOC_{max}. While detailed modeling of electrochemical battery degradation is beyond the scope of this study, limiting the SOC window and avoiding unnecessary charge-discharge cycles provides a realistic estimation of battery operating conditions, which is commonly assumed in EV energy management systems.

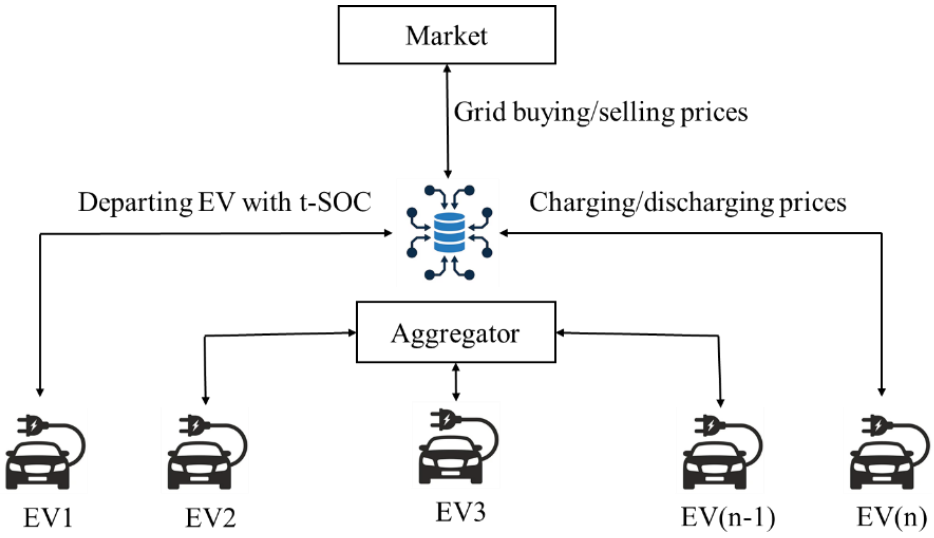

Figure 1. System Modeling

5.2 Aggregator Model

The aggregator in this model is a for-profit company that offers electric car charging and discharging services and has capacity for n EVs, where n is the number of charging (parking) outlets. It is expected that the aggregator has the ability to participate in the electrical market, specifically as a price taker. Section II explains how to calculate the aggregator's profit, which was indicated earlier in the model.

5.3 Market Model

Energy pricing is regulated by the market, which can be considered a decision-making authority. Grid pricing for this research is derived from the UK energy market, which includes both downstream purchasing prices and upstream regulation charges. The pricing signals used in this study consist of the upstream regulation price $p_{up,t}$ and the downstream electricity purchase price$p_{down,t}$, both expressed in \$/kWh. These price signals represent the electricity cost for buying energy from the grid and the revenue obtained when supplying energy back to the grid through vehicle-to-grid (V2G) operations. The price data are applied at a 30-minute time resolution, which matches the simulation time step and reflects the temporal granularity commonly used in electricity markets.

In this market model, the clearing price refers to the electricity price determined by the market at each time interval where supply and demand are balanced. The aggregator is assumed to act as a price-taker, meaning it cannot influence the market price and must accept the prevailing clearing price when buying or selling electricity.

Table 1. Notation used in the research

Symbol	Description	Unit
i	Index of EV	-
t	Time step index	-
T	Total simulation time	h
$SOC_{i,t}$	State of charge EV i at time t	(Fraction or %)
SOC_{min}	Minimum allowable SOC limit	-
SOC_{max}	Maximum allowable SOC limit	-
$SOC_{i,arrival}$	SOC of EV at arrival	-
$SOC_{i,target}$	t-SOC required at departure	-
C_i	Battery capacity of EV i	kWh
$P_{c,i,t}$	Charging power of EV i at time t	kW
$P_{d,i,t}$	Discharging power of EV i at time t	kW
$P_{i,max}$	Maximum charging/discharging power of EV i	kW
$P_{up,t}$	Power supplied from EVs to the grid (V2G)	kW
$P_{down,t}$	Power purchased from grid for EV charging	kW
P_T	Maximum transformer capacity	kW
N_t	Number of EVs present time at t	-
n	Maximum EVs that aggregator can accommodate	-
$p_{up,t}$	Electricity selling price to grid	\$/kWh
$p_{down,t}$	Electricity purchase price from grid	\$/kWh
p_{base}	Base price for EV charging/discharging	\$/kWh
Δ_t	Duration of each time slot	h
$Revenue_t$	Aggregator revenue at time t	\$
$Cost_t$	Aggregator cost at time t	\$
$M_{c,i}$	Charging price modifier for EV i	-
$M_{d,i}$	Discharging price modifier for EV i	-
$w_{c,i}$	Weight factor for charging priority	-
$w_{d,i}$	Weight factor for discharging priority	-
$n_{target,t}$	Number of EVs reaching t-SOC at time t	-
γ	Discount factor in RL	-
α_{actor}	Learning rate of actor network	-
α_{critic}	Learning rate of critic network	-
τ	Polyak averaging coefficient	-
ϵ	Probability of random exploration	-
η	Standard deviation of exploration noise	-

Consequently, the aggregator schedules EV charging during low-price periods and allows EV discharging during high-price periods to maximize operational profit while minimizing charging costs for EV users. The aggregator is supposed to be a middle-of-the-road non-domestic customer. The simulation runs for 24 hours at 30-minute intervals, which represents a typical daily operational cycle of the electricity market and ensures consistency between the pricing signals and the EV charging/discharging scheduling framework. The notation used in this research is given in Table 1.

6. Result and Discussion

6.1 Experimental Setup

The basic characteristics of the constructed simulator make it simple to put the suggested approach into action and compare it to different charge/discharge scheduling techniques via benchmarking. During the time an EV is parked, its charge and discharge schedule is laid out in 30-minute increments. The charge and discharge strategy of an EV is the sum of all the different schedules for charging and discharging existing EVs. The problem formulation is followed by a solution plan that considers the aggregator's and EV owners' perspectives. The cumulative profit is the profit produced by the aggregator from charge/discharge operations during the whole simulation period. EVs that leave before their specified departure time aren't included in the total amount of EVs with the t-SOC.

The transformer's power and the maximum accessible charging slots define an aggregator's capacity. These metrics can be used to calculate how many EVs an aggregator can handle. The goal is to complete three objective functions while staying within the constraints. As a result, we put the chosen framework through several scenarios. Table 2 summarizes the three alternative scenarios utilized in this study based on aggregator capacity and total EVs:

Table 2. Different Scenario of Aggregator Capacity and EV Count

Scenario	Aggregator Capacity (kW)	EV Count
Scenario 1	50	100
Scenario 2	500	750
Scenario 3	50	150

The heterogeneity of EV fleets is considered in the simulation to reflect more realistic charging scenarios. The participating EVs are assumed to have a battery capacity between 40 kWh and 80 kWh, which is typical for passenger EV models. EVs are assigned arrival state-of-charge (SOC) values randomly distributed between 20% and 60%, reflecting realistic driving dynamics prior to charging. Each EV has a t-SOC between 80% and 90%, which ideally should be reached before the vehicle departs. Moreover, the charging power of EVs is divided into three categories: slow charging (3.3 kW), standard charging (7.2 kW), and fast charging (11 kW), depending on the charger and infrastructure capacity. Only a finite number of charging ports are available, determined by the aggregator capacity and transformer rating, meaning the number of EVs in the system may exceed the number of charging ports.

Therefore, the scheduling algorithm must efficiently allocate charging and discharging operations among the EVs while respecting infrastructure constraints. Scenario 3 represents a case of high demand and comparatively low aggregator capacity, where the number of EVs is larger and the scheduling environment is more challenging for the proposed RL framework. Table 3 gives the configuration values used for the proposed optimized RL framework used in this research

6.2 Simulation Results

Here, we present the findings of our study, which looked at three primary objectives, under various settings. Three metrics are used to evaluate the performance of the proposed EV charging and discharging scheduling framework: aggregator profit, EV user charging cost, and t-SOC success rate.

- Aggregator Profit (Objective-1) represents the total revenue earned by the aggregator from charging and discharging operations during the simulation period and is expressed in US dollars ($).
- EV User Charging Cost (Objective-2) represents the total cost incurred by EV owners for charging their vehicles based on the amount of energy consumed and the real-time electricity price. This metric is also expressed in US dollars ($).
- The t-SOC Success Rate (Objective-3) represents the percentage of EVs that successfully reach their t-SOC before departure, calculated as the ratio of EVs achieving t-SOC to the total EVs remaining until departure. The value is expressed as a percentage (%).

All models—First Come First Serve (FCFS) technique, Proximal Policy Optimization (PPO) technique, Soft Actor-Critic (SAC) technique, and optimized RL technique are listed in Table 4, which summarizes the final results. To completely understand how effectively these models performed, they were evaluated in three different scenarios as mentioned in Table 4.

Table 3. Configuration of the Proposed optimized RL Framework

Parameter	Value / Description
RL Algorithm	DDPG
Experience Replay	HER
Optimization Method	GA
State Representation	SOC, arrival/departure time, EV mode, grid power, EV count
Action Space	Charging/discharging power
Reward Function	Profit – cost + t-SOC reward
Actor Network	Policy network
Critic Network	Q-value network
Exploration Policy	Noisy policy $\pi(s) + N(0,1)$
Discount Factor (γ)	RL discount factor
Target Update (τ)	Polyak averaging
Actor Learning Rate	α_{actor}
Critic Learning Rate	α_{critic}
Random Action Probability	$\in$
Noise Deviation	η
GA Chromosome	$T, \gamma, \alpha_{actor}, \alpha_{critic}, \in, \eta$
GA Population Size	50
GA Selection	Tournament
GA Crossover	SBX (0.9)
GA Mutation	Polynomial (0.1)
GA Generations	100
Training Episodes	50 per chromosome
Validation Runs	10 simulations
Simulation Horizon	24 hours
Time Resolution	30 minutes

Table 4. Performance of optimized-RL model on various objective functions under different scenarios

MODEL	OBJECTIVE-1 ($)			OBJECTIVE-2 ($)			OBJECTIVE-3 (%)		
	Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3
FCFS	148 ± 1.2	225 ± 3.8	503 ± 1.2	0.17 ± 0.004	0.25 ± 0.002	0.19 ± 0.003	67 ± 1.1	55 ± 0.9	71 ± 0.8
PPO	158 ± 2.6	242 ± 3.4	530 ± 4.1	0.12 ± 0.006	0.18 ± 0.008	0.15 ± 0.007	82 ± 1.7	76 ± 1.9	85 ± 1.6
SAC	162 ± 2.2	250 ± 3.0	545 ± 3.8	0.11 ± 0.005	0.16 ± 0.007	0.14 ± 0.006	84 ± 1.5	78 ± 1.7	87 ± 1.5
Optimized RL	181 ± 1.8	270 ± 2.5	571 ± 3.2	0.08 ± 0.004	0.11 ± 0.005	0.09 ± 0.004	91 ± 1.2	87 ± 1.3	94 ± 1.1

For Objective-1, which focuses on maximizing aggregator profit, the FCFS technique achieved profits of 148 ± 1.2, 225 ± 3.8, and 503 ± 1.2 dollars in the three scenarios, as shown in Table 4. When RL techniques such as PPO and SAC were applied, the profit increased to 158 ± 2.6, 242 ± 3.4, 530 ± 4.1 and 162 ± 2.2, 250 ± 3.0, 545 ± 3.8, respectively. The proposed optimized RL model achieved the highest profits of 181 ± 1.8, 270 ± 2.5, and 571 ± 3.2, demonstrating that the optimized RL framework effectively schedules EV charging and discharging operations to improve aggregator revenue. The comparison of the models for aggregator profit is illustrated in Figure 2.

For Objective-2, which aims to reduce the EV user charging cost, the FCFS approach resulted in charging costs of 0.17 ± 0.004, 0.25 ± 0.002, and 0.19 ± 0.003 across the three scenarios (Table 4). The PPO

and SAC models reduced these costs to 0.12 ± 0.006, 0.18 ± 0.008, 0.15 ± 0.007 and 0.11 ± 0.005, 0.16 ± 0.007, 0.14 ± 0.006, respectively. The optimized RL model achieved the lowest charging costs of 0.08 ± 0.004, 0.11 ± 0.005, and 0.09 ± 0.004, indicating that the proposed framework effectively schedules charging during lower electricity price periods. The comparison of EV charging cost among the models is shown in Figure 3.

The third objective was to maximize the number of EVs leaving with a t-SOC. FCFS was successful for 67% of For Objective-3, which maximizes the number of EVs reaching the t-SOC before departure, the FCFS scheduling method achieved success rates of 67 ± 1.1%, 55 ± 0.9%, and 71 ± 0.8% in the three scenarios (Table 4). The PPO and SAC models improved the success rates to 82 ± 1.7%, 76 ± 1.9%, 85 ± 1.6% and 84 ± 1.5%, 78 ± 1.7%, 87 ± 1.5%, respectively. The optimized RL framework achieved the highest success rates of 91 ± 1.2%, 87 ± 1.3%, and 94 ± 1.1%, indicating that the proposed approach effectively prioritizes EV charging schedules so that more vehicles reach the desired SOC before departure. The comparison of the models for t-SOC success rate is illustrated in Figure 4.

To further analyze the contribution of each component in the proposed framework, an ablation study was conducted using four model variants: DDPG only, DDPG + HER, DDPG + GA, and DDPG + HER + GA. The results are summarized in Table 5.

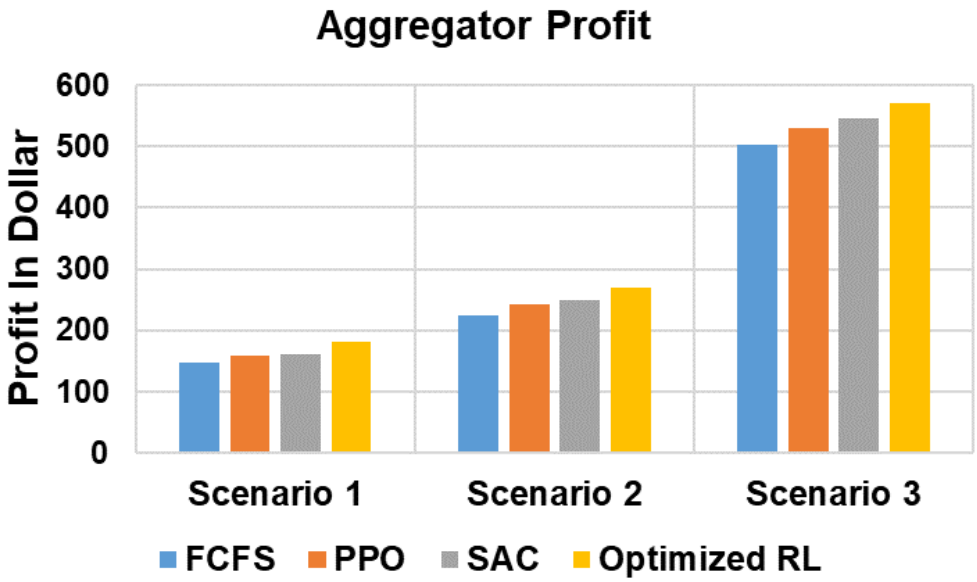

Figure 2. Comparison of the model outcome on aggregator profit

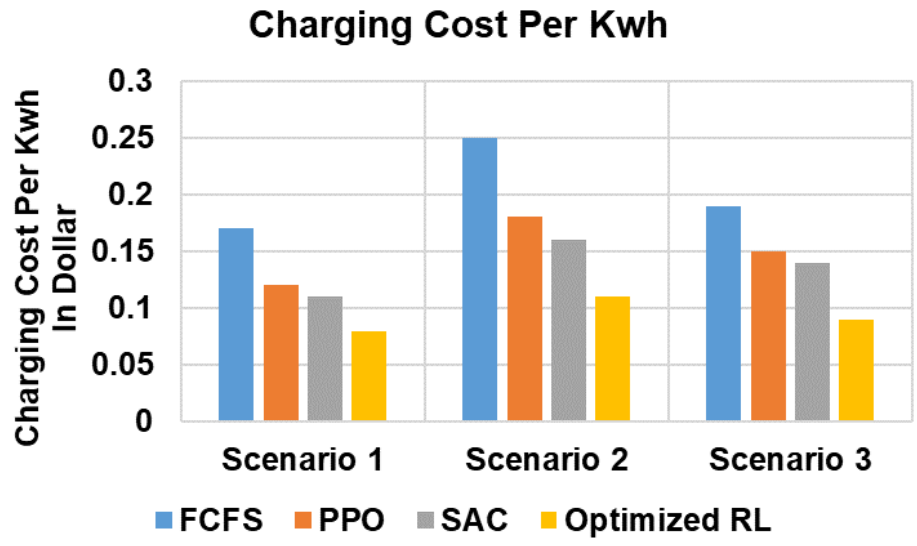

Figure 3. Comparison of the model outcome on EV user's charging cost

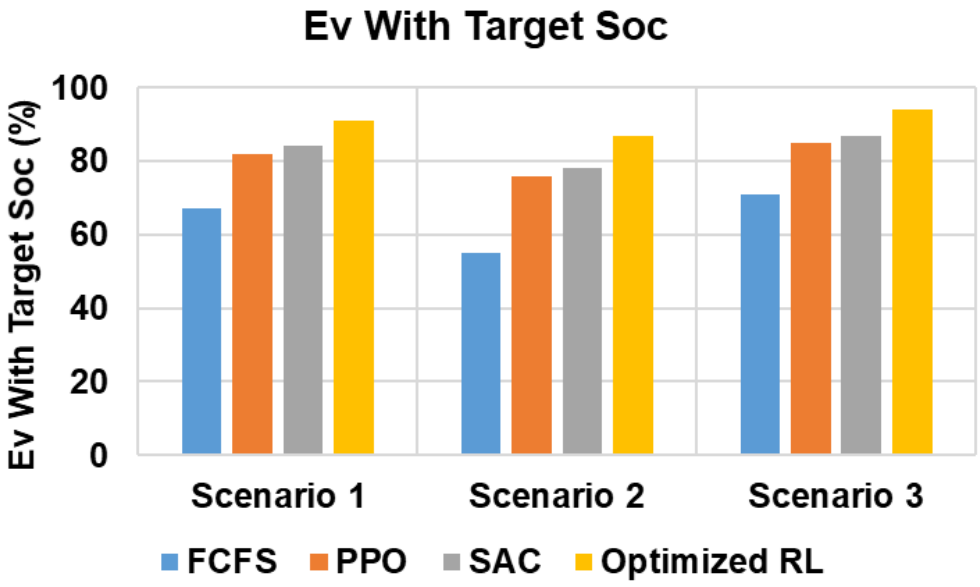

Figure 4. Comparison of model outcome on EV with t-SOC

Table 5. Performance of optimized-RL model and its ablation variants

MODEL	OBJECTIVE-1 ($)			OBJECTIVE-2 ($)			OBJECTIVE-3 (%)		
	Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3	Scenario 1	Scenario 2	Scenario 3
DDPG only	153 ± 2.9	232 ± 3.5	519 ± 4.3	0.11 ± 0.006	0.16 ± 0.007	0.14 ± 0.006	85 ± 1.8	79 ± 2.0	88 ± 1.7
DDPG + HER	164 ± 2.5	246 ± 3.2	543 ± 3.9	0.10 ± 0.005	0.14 ± 0.006	0.12 ± 0.005	88 ± 1.6	82 ± 1.8	90 ± 1.5
DDPG + GA	173 ± 2.1	261 ± 2.8	557 ± 3.4	0.09 ± 0.004	0.13 ± 0.005	0.11 ± 0.004	89 ± 1.4	85 ± 1.6	91 ± 1.3
DDPG + HER + GA	181 ± 1.8	270 ± 2.5	571 ± 3.2	0.08 ± 0.004	0.11 ± 0.005	0.09 ± 0.004	91 ± 1.2	87 ± 1.3	94 ± 1.1

The basic DDPG model achieved profits of 153 ± 2.9, 232 ± 3.5, and 519 ± 4.3 dollars across the three scenarios, with corresponding charging costs of 0.11 ± 0.006, 0.16 ± 0.007, and 0.14 ± 0.006, and t-SOC success rates of 85 ± 1.8%, 79 ± 2.0%, and 88 ± 1.7%. When HER was incorporated, the performance improved due to better learning from past experiences, increasing profits to 164 ± 2.5, 246 ± 3.2, and 543 ± 3.9, while also improving t-SOC success rates. Similarly, the introduction of the GA to hyperparameter optimization further improved performance, achieving 173 ± 2.1, 261 ± 2.8, and 557 ± 3.4 in profits. It also led to greater reductions in charging costs as well as higher SOC success rates.

The final DDPG + HER + GA model performed optimally, with profits of 181 ± 1.8, 270 ± 2.5, and 571 ± 3.2, while simultaneously achieving the lowest charging costs and the highest t-SOC success rates. These results support the claim that the HER and GA components are essential for enhancing both the learning efficiency and the effectiveness of the RL–based scheduling framework.

7. Conclusion

Since EVs are becoming mainstream, charging stations and the control of charging operations are increasingly important for system efficiency and grid stability. This paper explores how EV charging and discharging can be optimally scheduled through an optimized RL model, combining DDPG and HER, with hyperparameters optimized using a GA. The framework is designed to maximize aggregator profit, reduce EV charging costs, and increase the number of EVs leaving with a t-SOC. Simulations were performed under three conditions with different aggregator capacities and numbers of EVs: Scenario 1 (50 kW, 100 EVs), Scenario 2 (500 kW, 750 EVs), and Scenario 3 (50 kW, 150 EVs). Compared to FCFS, PPO, and SAC strategies, the optimized RL achieved the highest aggregator profits ($181–$571 vs. $148–$503, $158–$530, and $162–$545), lowest EV charging costs ($0.08–$0.11/kWh vs. $0.17–$0.25, $0.12–$0.15, and $0.11–$0.14), and the largest proportion of EVs leaving with t-SOC (87–94% vs. 55–71%, 76–85%, and 78–87%). These findings indicate that the proposed optimized RL model is

effective in balancing objectives while respecting operational constraints, providing a feasible and scalable solution for real-time EV charging and discharging scheduling, improving user experience, and increasing aggregator profitability.

References

[1] F. Alanazi, Electric vehicles: Benefits, Challenges, and Potential Solutions for Widespread Adaptation. Applied Sciences, 13(10), (2023) 6016. https://doi.org/10.3390/app13106016

[2] N. Panossian, M. Muratori, B. Palmintier, A. Meintz, T. Lipman, K. Moffat, Challenges and Opportunities of Integrating Electric Vehicles in Electricity Distribution Systems. Current Sustainable/Renewable Energy Reports, 9(2), (2022) 27–40. https://doi.org/10.1007/s40518-022-00201-2

[3] M. İnci, Ö. Çelik, A. Lashab, K. Ç. Bayındır, J. C. Vasquez, J. M. Guerrero, Power System Integration of Electric Vehicles: A review on Impacts and Contributions to the Smart Grid. Applied Sciences, 14(6), (2024) 2246. https://doi.org/10.3390/app14062246

[4] F. Z. Peng, C. Liu, Y. Li, A. K. Jain, D. Vinnikov, Envisioning the Future Renewable and Resilient Energy grids—A Power Grid Revolution enabled by Renewables, Energy Storage, and Energy Electronics. IEEE Journal of Emerging and Selected Topics in Industrial Electronics, 5(1), (2023)8–26. https://doi.org/10.1109/jestie.2023.3343291

[5] K. Kessels, C. Kraan, L. Karg, S. Maggiore, P. Valkering, E. Laes, Fostering Residential Demand Response through Dynamic Pricing Schemes: A Behavioural Review of Smart Grid Pilots in Europe. Sustainability, 8(9), (2016) 929. https://doi.org/10.3390/su8090929

[6] G. Pereira, J. Deason, A. Sandonato, (2025).Unlocking Load Growth at the Grid Edge: Practices for Managing, Recovering, and Allocating Distribution System Investments.

[7] F. Giordano, C. Diaz-Londono, G. Gruosso, Comprehensive Aggregator Methodology for EVs in V2G Operations and Electricity Markets. IEEE Open Journal of Vehicular Technology, 4, (2023) 809–819. https://doi.org/10.1109/ojvt.2023.3323087

[8] H. I. Shaheen, G. I. Rashed, B. Yang, J. Yang, Optimal Electric Vehicle Charging and Discharging Scheduling using Metaheuristic Algorithms: V2G Approach for Cost Reduction and Grid Support. Journal of Energy Storage, 90, (2024) 111816. https://doi.org/10.1016/j.est.2024.111816

[9] P.K. Shanmugam, P. Thomas, Optimization Strategies for Electric Vehicle Charging and Routing: A Comprehensive Review. Gazi University Journal of Science, 37(3), (2024) 1256–1285. https://doi.org/10.35378/gujs.1321572

[10] A.K. Kalakanti, S. Rao, Computational Challenges and Approaches for Electric Vehicles. ACM Computing Surveys, 55(14s), (2023) 1–35. https://doi.org/10.1145/3582076

[11] W. Yin, X. Qin, Cooperative Optimization Strategy for Large-Scale Electric Vehicle Charging and Discharging. Energy, 258, (2022) 124969. https://doi.org/10.1016/j.energy.2022.124969

[12] E. Shokouhmand, A. Ghasemi, Stochastic Optimal Scheduling of Electric Vehicles Charge/Discharge Modes of Operation with the Aim of Microgrid Flexibility and Efficiency Enhancement. Sustainable Energy Grids and Networks, 32, (2022) 100929. https://doi.org/10.1016/j.segan.2022.100929

[13] S. Jia, X. Kang, J. Cui, B. Tian, S. Xiao, Hierarchical Stochastic Optimal Scheduling of Electric Thermal Hydrogen Integrated Energy System considering Electric Vehicles. Energies, 15(15), (2022) 5509. https://doi.org/10.3390/en15155509

[14] M. Alilou, G. B. Gharehpetian, R. Ahmadiahangar, A. Rosin, A. Anvari-Moghaddam, Day-Ahead Scheduling of Electric Vehicles and electrical Storage Systems in Smart Homes using a Novel Decision Vector and AHP Method. Sustainability, 14(18), (2022) 11773. https://doi.org/10.3390/su141811773

[15] D. Liu, P. Zeng, S. Cui, C. Song, Deep Reinforcement Learning for Charging Scheduling of Electric Vehicles considering distribution network voltage stability. Sensors, 23(3), (2023) 1618. https://doi.org/10.3390/s23031618

[16] B. Sukumar, S. Aslam, N. Karthikeyan, P. Rajesh, A hybrid BCMPO technique for Optimal Scheduling of Electric Vehicle Aggregators under Market Price Uncertainty. IETE Journal of Research, 70(3), (2023) 2974–2988. https://doi.org/10.1080/03772063.2023.2177756

[17] H. Jin, S. Lee, S. H. Nengroo, D. Har, Development of Charging/Discharging Scheduling Algorithm for Economical and Energy-Efficient Operation of Multi-EV Charging

Station. Applied Sciences, 12(9), (2022) 4786. https://doi.org/10.3390/app12094786

[18] E. Muškardin, M. Tappler, B. K. Aichernig, I. Pill, Reinforcement Learning under Partial Observability guided by Learned Environment Models. In Lecture Notes in Computer Science, (2023) 257–276. https://doi.org/10.1007/978-3-031-47705-8_14

[19] H. Bojun, (2022). Lagrangian method for q-function learning (with applications to machine translation). In International Conference on Machine Learning, PMLR, 2129-2159. https://proceedings.mlr.press/v162/bojun22a.html

[20] M. Wu, Y. Gao, A. Jung, Q. Zhang, S. Du, The Actor-Dueling-Critic Method for Reinforcement Learning. Sensors, 19(7), (2019) 1547. https://doi.org/10.3390/s19071547

[21] H. Nguyen, H. M. La, M. Deans, (2019) Hindsight Experience Replay with Experience Ranking. In Proc. IEEE 9th Int. Conf. Development and Learning and Epigenetic Robotics (ICDL-EpiRob), Oslo, Norway. https://doi.org/10.1109/DEVLRN.2019.8850705

[22] B. Manela, A. Biess, Bias-Reduced Hindsight Experience Replay with Virtual Goal Prioritization. Neurocomputing, 451, (2021) 305–315. https://doi.org/10.1016/j.neucom.2021.02.090

[23] P. Dasha, A comparative review of Approaches for the Evolutionary Search to Prevent Premature Convergence in GA. Applied Soft Computing, 25, (2023) 1047–1077.

[24] B. T. Polyak, A. B. Juditsky, Acceleration of Stochastic Approximation by Averaging. SIAM Journal on Control and Optimization, 30(4), (1992) 838–855. https://doi.org/10.1137/0330046

[25] M. Huber, A. Trippe, P. Kuhn, T. Hamacher, (2012) Effects of large scale EV and PV Integration on Power Supply Systems in the Context of Singapore. In Proc. IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe),IEEE, Berlin, Germany. https://doi.org/10.1109/ISGTEurope.2012.6465831

Funding
The authors declare that no funds, grants or any other support were received during the preparation of this manuscript.

Competing Interests
The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

Authors Contribution Statement
C. Srinivasan: Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. C. Sheeba Joice: Supervision, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. Both authors have read and agreed to the published version of the manuscript.

Data Availability
The data supporting the findings of this study can be obtained from the corresponding author upon reasonable request.

Has this article screened for similarity?
Yes

About the License