Energy Reports 10 (2023) 494–509

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Research paper

Electric vehicles charging management using deep reinforcement learning considering vehicle-to-grid operation and battery degradation

Mostafa M. Shibl [a,*], Loay S. Ismail [b], Ahmed M. Massoud [a]

[a] Department of Electrical Engineering, Qatar University, Doha, Qatar
[b] Department of Computer Science and Engineering, Qatar University, Doha, Qatar

ARTICLE INFO

Article history:
Received 16 January 2023
Received in revised form 2 May 2023
Accepted 4 July 2023
Available online 13 July 2023

Keywords:
Distribution grid
Optimization
Deep reinforcement learning
Electric vehicles charging
Vehicle-to-grid
Power system management

ABSTRACT

EVs are becoming more popular and widely used worldwide due to their environmentally friendliness as part of the world efforts to decrease the effects of climate change. Moreover, more users are buying EVs due to governmental incentives, development of charging technologies and cheaper maintenance costs. Thus, the increased electrical loads on the distribution grid caused by the charging of EVs can have negative impacts such as high voltage fluctuations, power losses and power overloads. Thus, a power system management solution is required to protect the distribution grid from the harmful effects of EVs charging through the regulation of the charging of EVs. In this paper, a deep RL-based EVs charging management solution is presented, while considering fast charging, conventional charging and V2G operation, in order to satisfy the requirements of the user and the utility. Deep RL is utilized to model the EV chargers and the EV users. The EV chargers are considered the RL environment and the EV users are considered the RL agent. Finally, the system was tested with a range of case studies using real-life EVs charging data, which proved the effectiveness and reliability of the system to protect the distribution grid and satisfy the EV user's charging requirements.

1. Introduction

Since the inception of EVs, they have been gaining immense popularity over conventional petrol and diesel vehicles. EVs are known for their environmentally friendliness, decreased maintenance costs, higher performance and cheaper charging compared to refuelling. As a result, the EVs market share in the global car market has seen around 170% increase in 2021, as well as the exponential increase in the yearly sales of EVs (Virta, 2021). Moreover, as a result of the environmentally friendliness of EVs, countries are starting to provide incentives for the users of EVs to encourage others to buy EVs to decrease the effects of climate change. For instance, the UK has passed legislation to ban all new diesel and petrol cars by 2030 (Harrabin, 2020).

Thus, undoubtedly, this has caused a change in the infrastructure of countries with a large increase in the number of EV chargers. It is expected that the number of public EV chargers will reach 2.9 million by 2030 (Virta, 2021). In addition, with the advancement of new technologies, fast chargers are gaining popularity, as well as V2G operation. Between 2020 and 2024, it is expected that the V2G market will experience growth of up to \$5 billion (Virta, 2021). The increase of EVs penetration in the distribution grid can have negative effects on the grid due to the increased loads. As a result, it is vital that such heavy electrical loads are properly regulated and controlled to protect the distribution grid from undesirable consequences such as voltage fluctuations, power losses and power overloads.

In this paper, a power system management solution considering EVs penetration is presented using deep RL. The penetration of EVs is studied with fast and conventional charging, as well as V2G, to model the EV chargers with the newest technologies. RL is utilized to model the EV chargers and the EV users to produce a system that can protect the distribution grid from the unwanted consequences of EVs penetration. The RL environment is modelled as the EV chargers and the RL agent is modelled as the EV users. The objective of the system is to coordinate the charging of EVs to protect the distribution grid by decreasing voltage fluctuations, power losses and power overloads through minimizing the charging cost for the user, while considering battery degradation. Fig. 1 provides an overview of the presented method.

The main contributions of the paper are twofold:

1. Introducing an RL-based solution for the management of EVs charging considering fast and conventional charging, as

* Corresponding author.
E-mail address: ma1902206@qu.edu.qa (M.M. Shibl).

https://doi.org/10.1016/j.egyr.2023.07.008

Nomenclature

Environment Constants

P_C	Conventional charging power
P_F	Fast charging power
Δ	Action space
σ	State space

Environment Variables

τ	Time period in hours
$C(t)$	Electricity cost at time t
$alpha$	Attenuation factor
κ	Discount factor

Agent Constants

$T_{c,max}$	Maximum charging time
Q	Maximum battery capacity
SoC_f	Final SoC

Agent Variables

$SoC(t)$	State of charge at time t
B	Battery degradation cost
$W_i(t)$	Idle waiting cost at time t
$W_c(t)$	Charging waiting cost at time t
ξ	Charging mode
h	Charging hour
m	Charging minute
ω	Total waiting time in hours

Acronyms

EV	Electric Vehicle
RL	Reinforcement Learning
V2G	Vehicle-to-Grid
SoC	State of Charge
MDP	Markov Decision Process
DNN	Deep Neural Network

well as V2G with the consideration of battery degradation, to satisfy the requirement of the utility and the EV user.

2. Testing the RL model using case studies with real-life EVs charging data, including the performance assessment under the impact of electricity price uncertainty, to ensure its effectiveness and reliability in managing the distribution system.

In the upcoming sections, Section 2 will review the literature. Next, Section 3 describes the RL model formulation. After that, Section 4 will present the results and discussion. Finally, Section 5 and VI will conclude the paper and examine the future work, respectively.

2. Related work

There have been lots of previous studies done in the domain of EV charging solutions, which included the use of machine learning, game theory, quadratic programming, etc. (Abid et al., 2022). Tan and Wang (2017) and Liu et al. (2018) proposed the use of game theory to control EVs charging. The use of support vector machine, random forest, XGBoost, deep neural networks, and recurrent neural networks for coordinating the charging of EVs was studied by Shahriar et al. (2021) and Van Kriekinge et al. (2021). Stojkovic (2019), Deilami et al. (2011), and Wei et al. (2018) proposed the coordination of EVs charging based on different optimization problems with different objective functions. Malekshah et al. (2021) discussed the use of an energy reserve operational scheduling method to prevent emergency situations and blackouts using energy storage systems, such as EVs, and shiftable loads.

Recently, the utilization of RL has been extensively studied by researchers for EVs charging, through different RL modelling techniques and methods. Hu and Li (2022) studied the use of offline RL to produce a computationally efficient energy management system for the charging of EVs. However, the performance of the system considering the uncertainty of EVs charging and the availability of offline data was not investigated.

The creation of an EV charging scheduling scheme utilizing Q-learning was examined by Dang et al. (2019). Time-of-use electricity pricing was utilized to produce the Q-learning tables for the reward function. However, due to the use of look-up tables, continuous states, such as the SoC, cannot be used. Thus, the model only considers the charging mode as the only state of the agent. Therefore, the model does not fully represent the EV charging problem and its stochastic nature.

Sadeghianpourhamami et al. (2020) studied the use of RL as a model-free approach to coordinate EV charging. The RL model controls a set of EVs and makes use of Q-iteration to find the optimal charging policy. The model is seen to be only 8.5% more expensive compared to the optimum solution. However, only a single charging speed is considered, which means that the model does not consider the possibility to use fast and conventional charging, and V2G operation.

Similarly, Zhao and Lee (2022) studied an RL model using dynamic pricing of charging cost, with the aim of maximizing the quality of service for the user. RL is utilized to solve the formulated dynamic pricing problem, which is defined as a finite discrete horizon Markov decision process with a mixed state space. The results show that the model successfully solves the problem. However, the main drawbacks of the system is the assumption that the arrival rates of EVs is known and the stochastic nature of EVs charging is not considered, which are not realistic for real-life EV chargers.

On the other hand, Ding et al. (2020) proposed an RL solution to find an optimum EV charging strategy, with the objective of maximizing the profits of the distribution grid operator. As a result, the model benefits the voltage profile of the distribution grid, increasing the power quality of the power system. However, the model does not consider the needs of the EV user such as decreased charging times and cheaper charging costs.

In addition, Dabbaghjamanesh et al. (2021) studied the use of Q-learning for load forecasting of an EVs charging station. The authors propose a method of predicting the expected EVs loads in a charging station to mitigate any possible harmful effects on the grid, such as increased power overloads and power losses. It was seen that the model predicts the future loading well in three charging scenarios, which are uncoordinated, coordinated, and smart charging.

Moreover, Liang et al. (2021) utilized deep RL for scheduling the charging of a fleet of mobility on demand EVs. The model consider charge scheduling, vehicle re-balancing and order dispatching, which is modelled using a Markov decision process. Deep RL and binary linear programming are applied to be able to find a near optimal solution to the model. The RL model chooses the action for the EV, which are taking an order, re-balancing to a position or charging. The actions are taken based on the state value at different locations, times, SoCs and electricity prices. The model was simulated in Haikou City, which demonstrated the

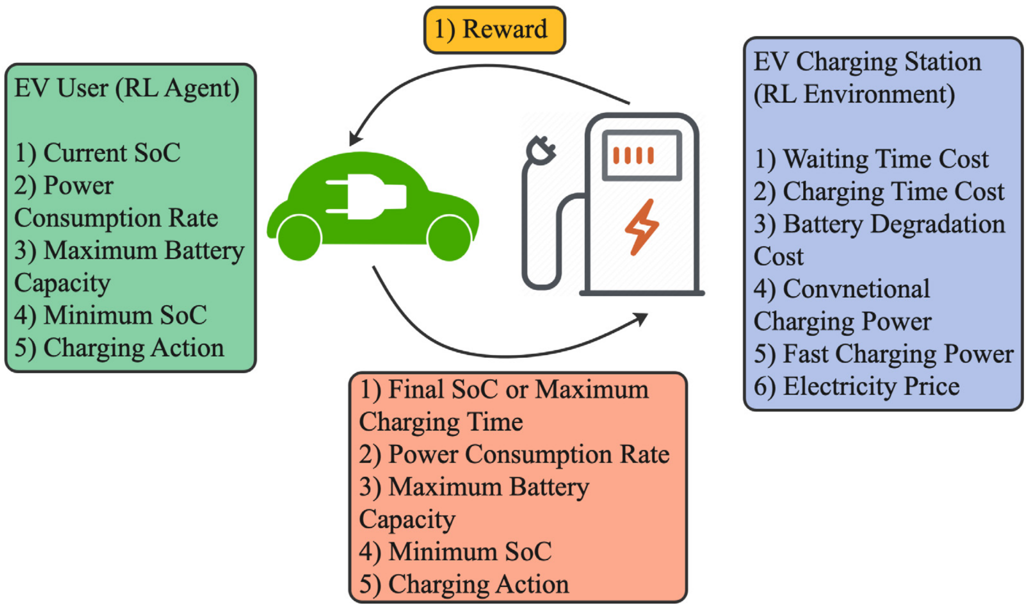

Fig. 1. EV charging station management solution overview.

viability of the model and increased the profits of the mobility on demand service provider.

Yan et al. (2021) approached the management of EVs charging using deep RL considering dynamic user behaviours and electricity price. The RL model aims to find the best sequence of charging events based on the charging cost and the EV user's experience. The EV user's experience is modelled to be the preference on charging times and locations. Thus, the model tries to balance between the price and the user's experience.

Furthermore, Liu et al. (2015) proposed an RL model as an energy management system for an EV. The model considers the state variables to be the SoC and generator speed. The driving schedule is utilized to calculate the transition probability and the model aims to increase fuel economy through the proper power split between the battery and the generator.

Another RL-based solution by Zhang et al. (2021) was built upon planning the charging of EVs. The objective of the RL model was to minimize the charging time, which was successfully achieved when tested with real-world data. In addition, the model considered the availability of two charging speeds, which are slow and fast, but V2G operation was not modelled.

A recent study by Wei et al. (2022) examined the use of deep RL for the fast charging of lithium-ion batteries. A model-based state observer and a deep RL optimizer model was utilized for the fast charging of lithium-ion batteries. The deep RL model penalized the over-temperature and degradation of the battery. To balance between the charging speed and the physical parameters of the battery, a novel environmental perceptive deep deterministic policy gradient algorithm is utilized. The results showed that the model is successful in providing a thermally safe charging solution for the battery and the increase in the lifetime of the battery. Nevertheless, the model did not undergo uncertainty analysis. In addition, the model was not being built for EVs, which means the utility requirements and V2G operation were not considered in the development of the model.

Furthermore, Shin et al. (2020) examined the utilization of a multiagent deep RL approach to manage EV chargers, while considering solar photovoltaic systems. The objective of the RL model was set to reduce the operation costs of the EV chargers. The model was designed to work with dynamically changing charging data and its use has shown that the operation of EV chargers were improved, while considering a single charging speed and V2G operation.

Similarly, Ye et al. (2022) and Mhaisen et al. (2020) researched the scheduling of EV charging using RL, while considering a single charging speed and V2G operation. The RL model provide a charging and discharging schedule for the user. However, the aim of the RL model proposed by Mhaisen et al. (2020) is the reduction of the charging costs of the user, while the aim of the RL model proposed by Ye et al. (2022) is the maximization of the profit of the charging station. The proposed RL method by Mhaisen et al. (2020) significantly reduces the charging cost, compared to heuristic and uncontrolled charging methods. Also, the proposed RL method by Ye et al. (2022) has a higher performance compared to the baseline model predictive control.

Additionally, Chu et al. (2022) examined the utilization of federated RL models for the minimization of the charging cost. Each user produces its own RL model, followed by the aggregation of the user models to produce a global model for all users. Li et al. (2022) studied the use of deep RL to minimize the charging cost for the user, considering the uncertainty in electricity pricing. Long short-term memory is utilized for the extraction of temporal features from the electricity price signal. The models successfully achieve their objectives and reduce the charging cost for the user.

The use of deep RL for the pricing of EVs charging by charging station operators was researched by Qiu et al. (2020), considering the discrete levels of charging and V2G operation. The model focuses on the requirements of the EVs charging station operators to increase their profits by adjusting the pricing of charging EVs. However, the model was not tested with real-life scenarios and EVs travelling data.

The advantage of RL over other methods such as quadratic programming and game theory is the adaptive nature of the RL method. Due to the stochastic nature of EVs charging, some models might not perform well in some situations due to predetermined assumptions and charging strategies. As a result, unlike other methods, RL can still perform well with the uncertainty in parameters such as electricity price. Additionally, RL is based on action and reward, which means that RL has an adaptable nature that allows it to perform well in all situations if the RL environment and RL agent are well defined. In addition, RL does not require time consuming problem formulation in comparison to optimization methods such as quadratic programming.

Table 1
Comparison of RL models.

Reference	Fast charging	V2G operation	User requirements (Cost, Final SoC and Time)	Utility requirements (Load Variance, Voltage Fluctuations and Overloads)	Uncertainty analysis
Sadeghianpourhamami et al. (2020) and Zhao and Lee (2022)	x	x	✓	x	x
Ding et al. (2020) and Dabbaghjamanesh et al. (2021)	x	x	x	✓	x
Liang et al. (2021)	x	x	✓	✓	x
Yan et al. (2021) and Liu et al. (2015)	x	x	✓	x	✓
Zhang et al. (2021) and Wei et al. (2022)	✓	x	✓	x	x
Shin et al. (2020) and Ye et al. (2022)	x	✓	x	✓	x
Mhaisen et al. (2020) and Chu et al. (2022)	x	✓	✓	x	x
Li et al. (2022)	x	✓	✓	x	✓
Qiu et al. (2020)	✓	✓	x	✓	x
Proposed	✓	✓	✓	✓	✓

The presented system in this paper manages the charging of EVs to decrease power losses, power overloads and voltage fluctuations. Unlike previous RL models created in the past literature, the presented system considers the utilization of conventional and fast charging, as well as V2G operation, and outputs the safest charging sequence for the grid, while taking into consideration the EV user's requirements, in terms of charging time, final SoC and charging cost. The RL model aims to decrease the charging cost for the user. Therefore, the system is able to satisfy the requirements of both the user and the utility.

Table 1 provides a brief comparison between the proposed RL model with RL models in the previous literature. The utilization of fast charging and V2G operation has become a requirement due to recent trends and the increase in the use of EVs. Thus, this requires the performance assessment of the model under the impact of uncertainty due to the highly uncertain nature of EVs charging, which becomes more critical with the use of fast charging and V2G operation. At the start of using deep RL for EVs charging, fast charging and V2G operation were not considered in the development of the RL models (Sadeghianpourhamami et al., 2020; Zhao and Lee, 2022; Ding et al., 2020; Dabbaghjamanesh et al., 2021; Liang et al., 2021; Yan et al., 2021; Liu et al., 2015). After that, either fast charging or V2G operation were defined in the RL models (Zhang et al., 2021; Wei et al., 2022; Shin et al., 2020; Ye et al., 2022; Mhaisen et al., 2020; Chu et al., 2022), with the performance of some models being tested under uncertain conditions (Li et al., 2022). Recently, models have been created that consider both fast charging and V2G operation (Qiu et al., 2020). Moreover, most models consider either the objectives of the user, or the objectives of the utility, and only a few models have undergone a performance assessment under the impact of uncertainties. In this paper, both fast charging and V2G operation were considered, as well as the requirements of the user and the utility. In addition, the RL model was tested under the uncertain conditions of electricity prices.

3. RL model formulation

3.1. MDP formulation

Due to the stochastic nature of EVs charging, EVs charging is formulated as a finite MDP with unknown transition probability. As previously mentioned, in the RL model, the EV charger is modelled as the RL environment, while the EV user is modelled as the RL agent. Firstly, the RL environment receives either a final, desired SoC, or a maximum charging time from the RL agent, as well as the minimum SoC and battery capacity. Then, the starting system states (s_t) are extracted, which are the SoC, charging hour, charging minute and charging mode. After that, the RL agent takes an action (a_t) and receives a reward (r_t). Next, the new updated data at time step $t+1$ is acquired by the RL agent to update the system state. The full details of the MDP, including the definitions of the system states, actions, transition function, rewards and action-value function, are shown below.

(1) States: The system state at time step t consists of the SoC at time step t, the hour h at time step t when the RL agent is taking the action, the minute m at time step t when the RL agent is taking the action, and the charging mode ξ at time step t, as shown in (1).

$$s_t = (SoC(t), h, m, \xi) \tag{1}$$

(2) Actions: The EV driver takes a specific action from the action space, which represents the charging action, including conventional charging, fast charging, V2G operation and idling, as seen in (2) and (3). When the RL agent chooses action 0, two consecutive time steps are utilized, which are idling at the first time step followed by one time step of conventional charging. Action 0 is defined in this way to prohibit the RL agent from always choosing idling, which provides a positive reward that can cause the EV to stay for hours without any charging action.

$$\Delta = \begin{cases} 0 & Idling \\ 1 & Conventional\,Charging \\ 2 & Fast\,Charging \\ 3 & V2G\,Operation \end{cases} \tag{2}$$

$$a_t = \xi \quad \xi \in \Delta \tag{3}$$

(3) Transition Function: The transition from the system state at time step t to the system state at time step $t+1$ is influenced by the action of the RL agent (a_t), as seen in (4). The SoC of the next state is the current SoC plus or minus (depending on charging or V2G operation) the capacity charged based on the charging power mode (P_x), where x represents the charging mode. The hour and

minute are updated based on the time period of each action and the current hour and minute.

$$s_{t+1} = \left(SoC(t) \pm \frac{\tau \cdot P_x}{Q}, h_{t+1}, m_{t+1}, \xi \right) \tag{4}$$

Due to having no prior knowledge of the EVs charging situation, the transition probability cannot be described. Thus, a deep RL method is utilized to implicitly find and calculate the transition probabilities.

(4) Rewards: The reward for action a_t is calculated from the perspective of the RL agent, as seen in (5) to (10). A range of values were utilized for the different weighting values in order to manually tune the values to obtain the required performance from the model. It should be noted that RL models are sensitive to such factors in reward functions, which means careful tuning should be performed.

$$r_t = \begin{cases} -\kappa(\tau \cdot P_C \cdot C(t) + \tau \cdot W_c(t) + \tau \cdot W_i(t)) & \xi = 0 \\ -\tau \cdot P_C \cdot C(t) - \tau \cdot W_c(t) & \xi = 1 \\ -\tau \cdot P_F \cdot C(t) - \tau \cdot W_c(t) & \xi = 2 \\ -\tau \cdot W_c(t) - B + \alpha \cdot \tau \cdot P_C \cdot C(t) & \xi = 3 \end{cases} \tag{5}$$

$$\kappa = \begin{cases} 1.0 & 1 - d \cdot \omega > 1 \\ p & 1 - d \cdot \omega < p \\ 1 - d \cdot \omega & others \end{cases} \tag{6}$$

$$\alpha = SoC^b(t) \tag{7}$$

$$B = \tau \cdot P_C \cdot \max(\mathcal{N}(\mu_1, z_1^2), L) \tag{8}$$

$$W_i(t) = \mathcal{N}(\mu_2, z_2^2) \tag{9}$$

$$W_c(t) = F \cdot W_i(t) \tag{10}$$

The discount factor κ is added to provide a set discount per hour with a set minimum and maximum discounts of 0% and p%, respectively, for idling in the charging station. Every hour corresponds to a set increase of d% in the discount. The discount is added in order to provide a reward for the user for leaving their EV in idle mode, which increases the chances of the user following the set charging scheme. In the proposed RL model, the maximum discount (p) was set to 50% and the discount per hour (d) was set to 0.5%. The use of higher values causes the model to overuse the idling mode to increase the discount, which is not suitable for real-life scenarios, where the time of charging should be controlled. On the contrary, the use of lower values causes the model to rarely utilize the idle mode, which is not preferred for the utility as the EV is constantly charging.

Similarly, the attenuation factor α is added to decrease the positive reward from V2G operation as the SoC approaches the minimum SoC. As a result, the value of b should be lower than 1 and greater than 0 to ensure that the positive reward of discharging is significantly decreased as the SoC approaches the minimum SoC. In the presented model, a value of b was set to 0.5, which provided suitable results for all case studies.

The battery degradation model seen in (8) was developed to penalize the RL agent for using V2G operation. This was done to avoid the RL agent from fully discharging and gaining the positive reward, which means the EV will never charge since charging has a negative reward only. In this model, the mean of the normal distribution was set to 0.1 and the standard deviation was set to 0.05. The maximum limit (L) for the battery degradation cost was set to 0.05. The values of μ_1, z_1 and L were chosen to be in proportion with the positive reward of V2G operation. The negative reward from battery degradation should not be too high as to discourage the RL agent from choosing V2G operation, and it should not be too low as to encourage the RL agent to always choose V2G operation.

Likewise, the idle waiting time was also chosen to be a normal distribution with a mean of 0.05 and a standard deviation of 0.0075. The values of μ_2 and z_2 were set to be in proportion with other rewards as well. However, the negative reward of waiting is lower compared to other costs, such as charging, since it is only added to encourage the RL agent to try and finish charging in the shortest time possible, in addition to charging with the lowest cost. In addition, the charging waiting time is considered to be a factor of the idle waiting time, since the RL agent will be benefiting during that time. In the presented model, the factor (F) was set to 0.5. It should be noted that if the value of k is too high, it can cause the overall waiting cost to increase significantly, which harms the performance of the RL model.

The definition of $C(t)$ is based on time-of-use pricing and is shown in Fig. 2 (Carlson, 2015). It should be noted that since the reward function is linear and dependent on the electricity cost, the issues of partial observability with electric vehicles charging is avoided. Moreover, the problem of dimensionality might arise with bigger networks, which increases the computational time for training the RL model. However, the use of deep neural networks helps in mitigating the curse of dimensionality, through function approximations. Characteristics from models, value functions, or policies are extracted to create a generalized representation of the entire function using deep neural networks.

(5) Action-Value Function: The influence of the RL agent taking an action a with a system state s following the policy Ψ is found from the expected discounted cumulative reward at time step t, as seen in (11) (Sutton and Barto, 2020). $Q^{\Psi}(s, a)$ represents the action-value function and γ represents the discount factor that expresses the balance between the short-term and long-term rewards.

$$Q^{\Psi}(s, a) = \mathbb{E}^{\Psi}\left[\sum_{k=0}^{K} \gamma^k r_{t+k} | s_t = s, a_t = a\right] \tag{11}$$

The purpose of the RL model is to calculate the optimal policy Ψ^* from all the possible policies that maximizes the reward, which is the minimization of the cost in this model, as seen in (12). $Q^*(s, a)$ denotes the optimal action-value function.

$$Q^*(s, a) = \max_{\Psi} Q^{\Psi}(s, a) \tag{12}$$

3.2. Deep RL approach

In order to find the optimal policy for the charging of EVs without having any knowledge of the uncertainties in the system, a deep RL approach is proposed. Deep RL is utilized as it is able to iteratively find the optimal policy through recursively updating the value of the action-value function, based on the Bellman equation, as seen in (13), where β denotes the learning rate. This process is called Q-learning (Sutton and Barto, 2020).

$$Q_{new}(s, a) = Q_{old}(s, a) + \beta\left(r + \gamma \max_{a'} Q(s', a') - Q_{old}(s, a)\right) \tag{13}$$

As the Bellman equation converges to the value of the optimal action-value function, the optimal policy can be denoted by a greedy strategy, as seen in (14) (Sutton and Barto, 2020).

$$a* = \arg\max_{a} Q^*(s, a) \tag{14}$$

The conventional Q-learning method consists of the use of look-up tables to approximate the action-value function. Nevertheless, in the EVs charging problem, the states are continuous, meaning that look-up tables will be extremely large and very impractical to deal with. Thus, a DNN is utilized to be able to approximate the action-value function.

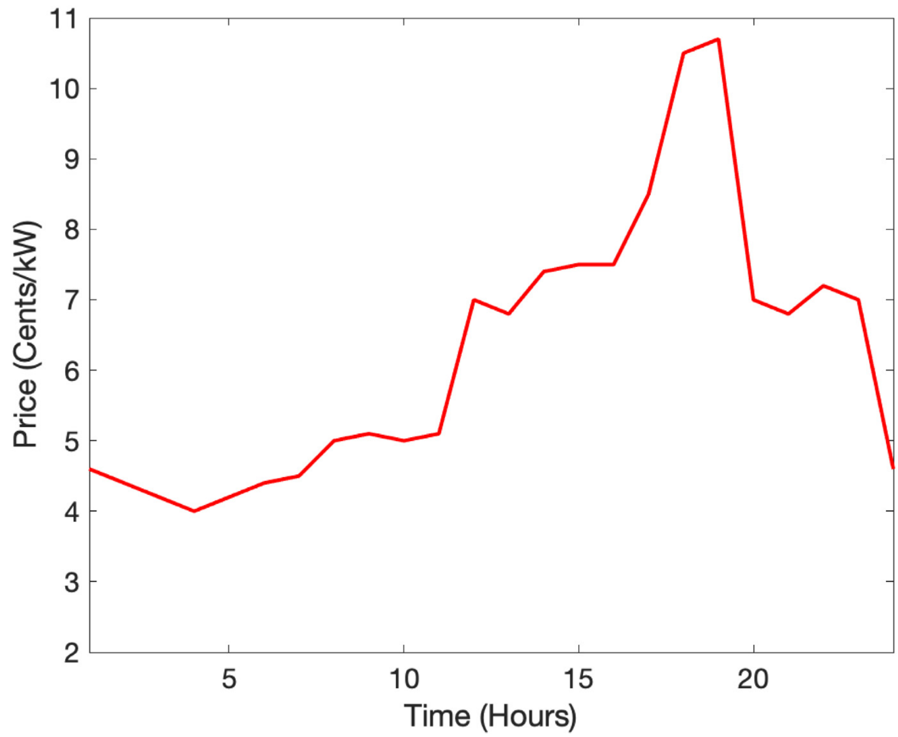

Fig. 2. Electricity cost function (Carlson, 2015).

As previously mentioned, deep RL allows for the state space to be in the infinite continuous space, instead of the conventional finite discrete space. The data produced from the interactivity of the agent with the environment is stored in the replay buffer. This is given to the main network, as well as the target network. After that, the policy is updated through the gradient descent method, which minimizes the loss function and updates the main network parameters. Finally, the parameters of the target network are updated based on the parameters of the main network after a certain interval of time (Pan et al., 2019).

Therefore, in the RL model, the first step is the extraction of the features of the EV from the RL agent by the RL environment. After that, the state of the RL agent is fed into the DNN. The action with the highest action-value function is chosen as the next action. Next, the action is taken, the RL agent receives the reward for the chosen action and the state of the user is updated. Finally, the process is repeated until either the final SoC or the maximum charging time is reached.

The training of the DNN is done based on Algorithm 1 using Python and a 2016 15-inch MacBook Pro. The 2016 15-inch MacBook Pro has a 2.7 GHz quad-core Intel Core i7, with 8 MB shared L3 cache, and a 16 GB of 2133 MHz LPDDR3 onboard memory. For N number of episodes, the state is initialized to the starting state. After that, based on the ϵ - greedy strategy, an action is chosen. Next, the action is done, and the reward and the new state are observed. Then, the values of the action-value function are updated based on the Bellman equation. Finally, the new state is set as the current state. This process is repeated until either the final SoC or the maximum charging time is reached. Finally, the exploration rate ϵ and the learning rate are decayed.

Consecutive states are extremely interlinked and actions have a significant effect on the following states in RL. Thus, running the algorithm using consecutive states eventually results in the divergence of training, which occurs when impractical increased values are given for the state–action pairs, decreasing the quality of the greedy control policy (Mnih et al., 2015).

Algorithm 1 DNN Training for RL Model

Input: State Space σ, Action Space Δ, Reward Function r, Number of Episodes N

Output: Estimation of Optimal Action-Value Function

```
for episodes=1:N do
    initialize state s = SoC(t), h, m, ξ
    while T < T_{c,max} or SoC(t) < SoC_f do
        with probability 1 − ϵ, set a_t as arg max_a Q*(s, a)
        or otherwise set a_t randomly
        Take action a_t and observe reward r_t and new state s_{t+1}
        Update Q(s, a) using Bellman equation
        s = s_{t+1}
    end while
    Decay exploration rate ϵ and learning rate α
end for
```

Moreover, Fig. 3 illustrates the flowchart for the EV charging algorithm. As the EV arrives, the EV user chooses to set a desired final SoC or a maximum time for charging. After that, charging actions are chosen until either the final SoC or the maximum time for charging is reached.

4. Case studies

A range of different case studies have been conducted in order to test and prove the effectiveness and reliability of the presented RL model to manage the charging of EVs. In each case study, the details of the charging situation are presented, alongside with the results of the RL model. In addition, the load demand and voltage fluctuations for the high charging price case studies are presented as the load demand is highest during such times, to prove and validate the effectiveness of the RL model in achieving the objectives of the utility. These are based on the load data of multiple residential households and using the IEEE 33-bus

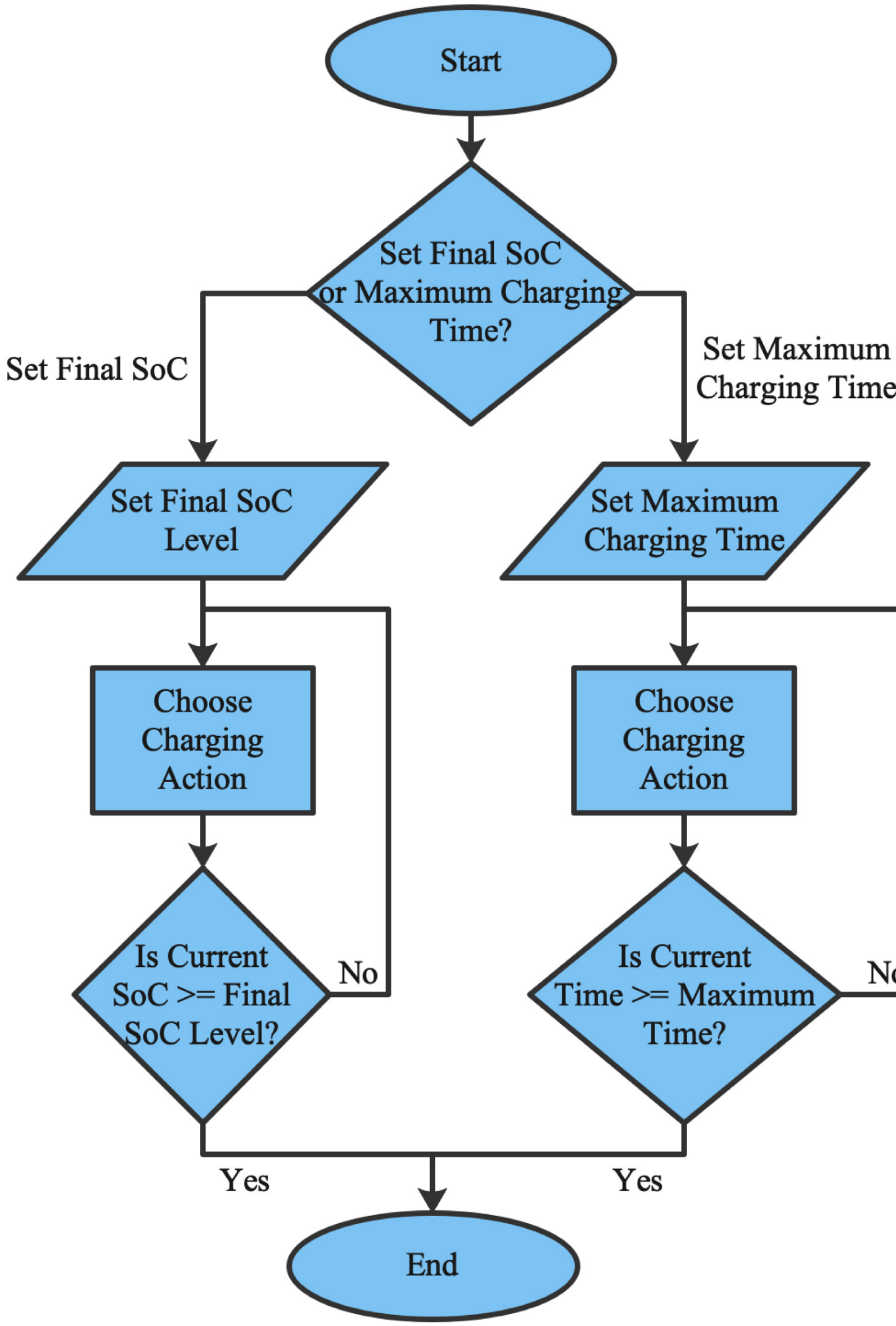

Fig. 3. Flowchart for EV charging algorithm.

distribution system model, which has a transformer rating of 3.5 kW.

Table 2 provides the list of all the values of the RL model parameters. The case studies done at high prices assume charging starts at 1500 h, while case studies done at low prices assume charging starts at 0000 h.

Fig. 4 provides a tree diagram for the illustration of the different case study scenarios, as well as the algorithm of the system. After the arrival of the EV, the RL model receives whether the user wants a set SoC level, or wants a specific time for charging.

Table 2
Values of RL model parameters.

Parameter	Value
Minimum SoC	20%
Battery Capacity	22 kWh
Time Period	15 min
Conventional Charging Power	7 kW
Fast Charging Power	22 kW

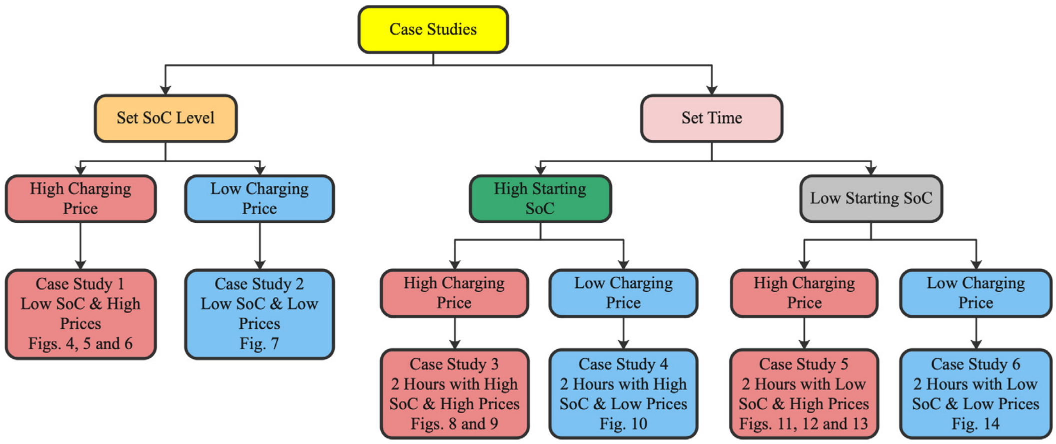

Fig. 4. Case study scenarios tree diagram.

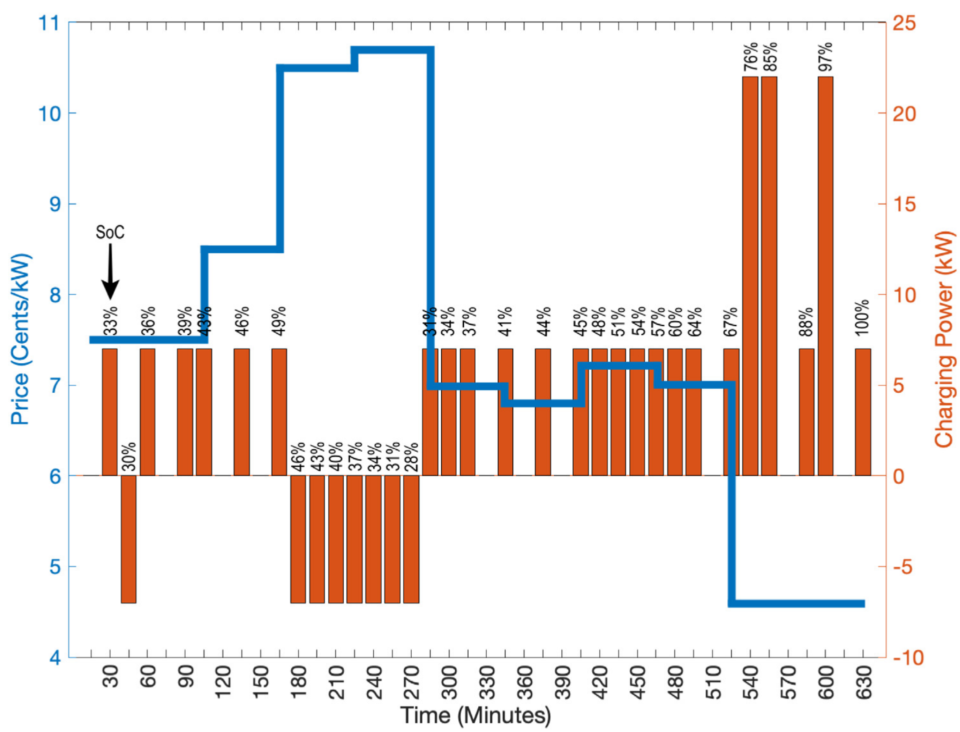

Fig. 5. Actions for full charge starting at high prices with low initial SoC.

Then, the RL model initiates the charging of the EV. Case study 1 and 5 were tested with amplitude and time uncertainty for the electricity price, since those case studies can be considered to be highly sensitive to the electricity price due to the low starting SoC. As a result, the uncertainty analysis was done on those case studies to test the robustness of the system.

4.1. Case study 1: Low SoC at high prices

The first case study involves an EV wanting to fully charge with an initial SoC of 30%, and the RL agent started to charge at a time where electricity prices are high. The full charge of the battery can be set based on the type of battery to maximize its life span. In this paper, linear charging is assumed up to 100%. However, this can be reduced based on the type of battery for practicality. Fig. 5 displays the charging actions taken by the RL agent, as well as the electricity price. The SoC of the EV after each charging action is placed above the bar. As seen in the figure, the RL agent makes use of the mid-price range to idle and charge conventionally at the start. After that, as the electricity price peaks, the RL agent starts to utilize V2G operation to make use of the high electricity prices. Next, when the electricity price starts to settle back down, the RL agent begins conventional charging once more and starts fast charging when electricity price is at its lowest point. However, some actions do not follow the general trend such as the V2G operation action at minute 45 and the idling action at minute 570. These actions can be attributed to the RL agent trying to increase its reward through discharging during mid-range electricity prices, as well as making use of the set discount from idling during low prices. It is noteworthy to mention that such actions do not significantly affect the performance of the RL model as the overall general trend is still the same.

Fig. 6 shows the results of the same case study with the electricity prices advanced by three hours and with an amplitude uncertainty of ± 20%. As seen in the figure, despite the advancement of the electricity prices by three hours and 20% amplitude

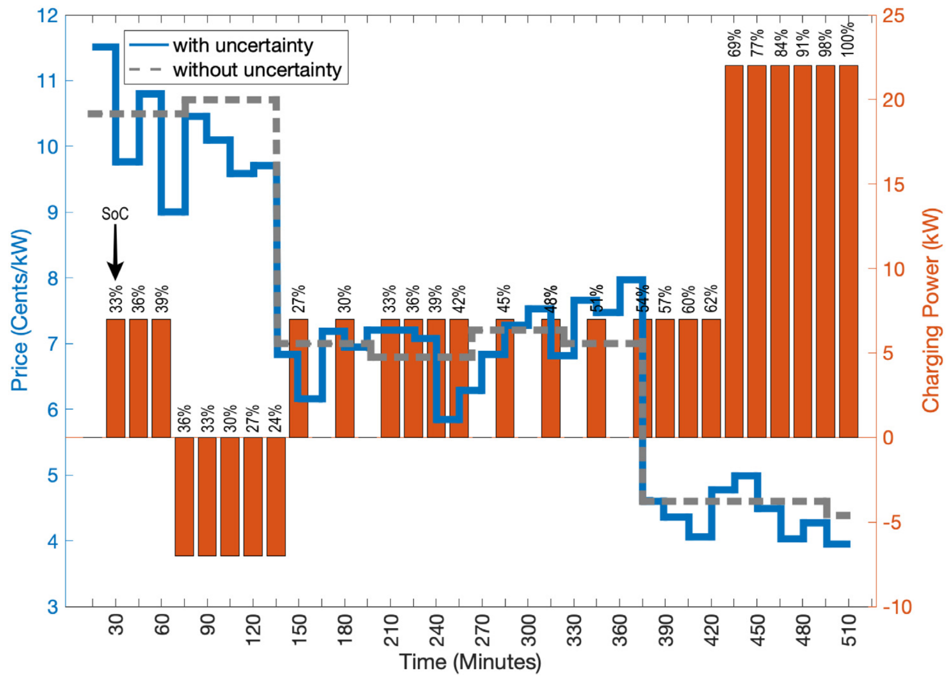

Fig. 6. Actions for full charge starting at high prices with low initial SoC (with a 3 h advance and 20% amplitude uncertainty).

uncertainty, the RL model is still able to achieve the goal of fully charging the EV. Additionally, the charging sequence generated by the RL model is able to adapt to the uncertainty in the electricity prices. It is seen that a similar charging pattern is seen even in the presence of an electricity price uncertainty, and the electricity price peak happening at an earlier time. In addition, during the time with low electricity prices, fast charging is always utilized, which decreased the charging time for the user, as the minimum of the electricity price happened at an earlier time.

Moreover, Fig. 7 illustrates the load demand curves for the first case study. The green curve is the original load demand with no EVs penetration, and the red curve is the load demand with uncontrolled fast charging only. The orange curve is the controlled EVs charging using the RL model. As seen from the graph, the RL model successfully flattens the load curve through peak shaving and valley filling. This leads to have a stable power distribution system and avoids the overload of the transformer. To add to that, the maximum voltage fluctuation for the three situations are 0.12, 0.17 and 0.09, respectively. Thus, the power losses are reduced since the voltage profile of the distribution system is within the required levels.

4.2. Case study 2: Low SoC at low prices

Similarly, the second case study has the same conditions as the first case study, however, the RL agent started to charge at a time where electricity prices are low. Fig. 8 shows the charging actions taken by the RL agent, alongside with the electricity price. The SoC of the EV after each charging action is placed above the bar. The figure shows that the RL agent starts with conventional charging and as the electricity price drops continues charging with fast charging. In addition, the RL agent also occasionally chooses to idle in order to make use of the set discount on the charging cost. Finally, at the time where the electricity price starts to increase, the RL agent utilizes the increased prices to discharge for a little while.

As a result, it is seen that the RL model successfully learns useful charging strategies in these situations (case studies 1 and 2), where time is not a constraint and the user would like to fully charge his EV. The RL model is able to minimize the charging cost and learn the electricity price pattern and provide a suitable charging scheme for the EV.

4.3. Case study 3: 2 h with high SoC at high prices

Next, the third case study involves an RL agent with a maximum charging time of 2 h, with an initial SoC of 90% and the time of charging is a time with high electricity prices. Fig. 9 portrays the charging actions taken by the RL agent and the electricity price. The SoC of the EV after each charging action is placed above the bar. As seen from the figure, during the charging time, the electricity price is constant at 7.5 cents/kW. As a result of the high SoC and high prices, the RL model chooses mainly to discharge. This can be considered as a suitable strategy as the user would want to make money from the high SoC of the battery.

In addition, Fig. 10 displays the different load demand curves for the third case study. The controlled charging of the EVs using the RL model provides an improved, flattened load curve. Peak shaving is achieved through V2G operation, while valley filling is achieved through controlled conventional and fast charging. As seen from the orange curve, power overload only occurs for a short time, which is a high improvement compared to the load curve without EVs penetration. The maximum voltage fluctuation for the three situations are 0.1, 0.14 and 0.08, respectively.

4.4. Case study 4: 2 h with high SoC at low prices

The fourth case study involves an RL agent with a maximum charging time of 2 h, with an initial SoC of 90% and the time of charging is a time with low electricity prices. Fig. 11 conveys the charging actions chosen by the RL agent, as well as the electricity price. The SoC of the EV after each charging action is placed above the bar. The figure shows that the RL agent chooses to be idle for most of the time and fast charges at the end when the price drops further. Due to the low prices, the RL model chooses to fully charge the EV despite the high initial SoC. Nevertheless, the model starts with idling in order to make use of the set discount and further decrease the charging cost.

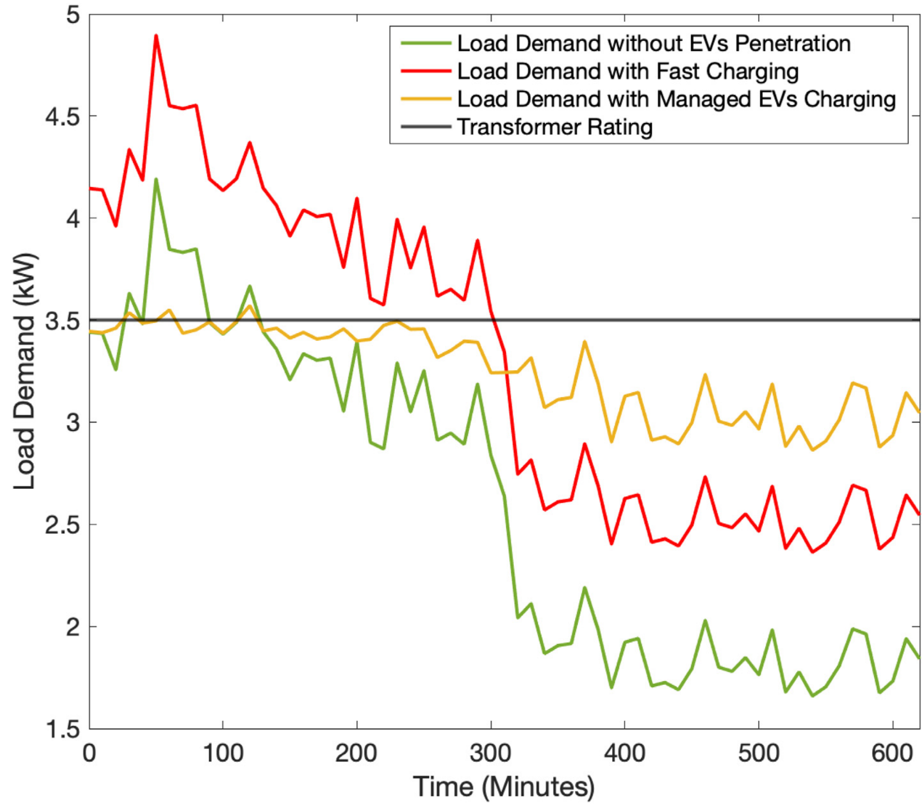

Fig. 7. Load demand for case study 1. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

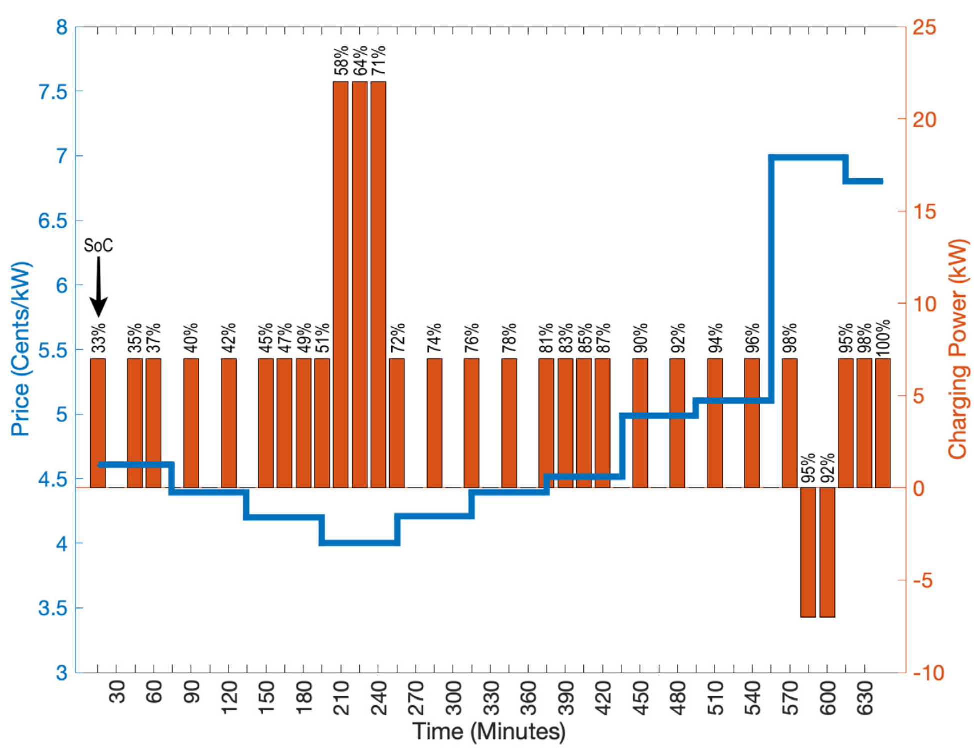

Fig. 8. Actions for full charge starting at low prices with low initial SoC.

Thus, it is seen that the RL model finds suitable charging schemes for such situations that time is a constraint and the starting SoC is high (case studies 3 and 4). The model can predict whether it is suitable to charge or discharge the high SoC of the battery, in terms of the electricity price and the possible reward from each option.

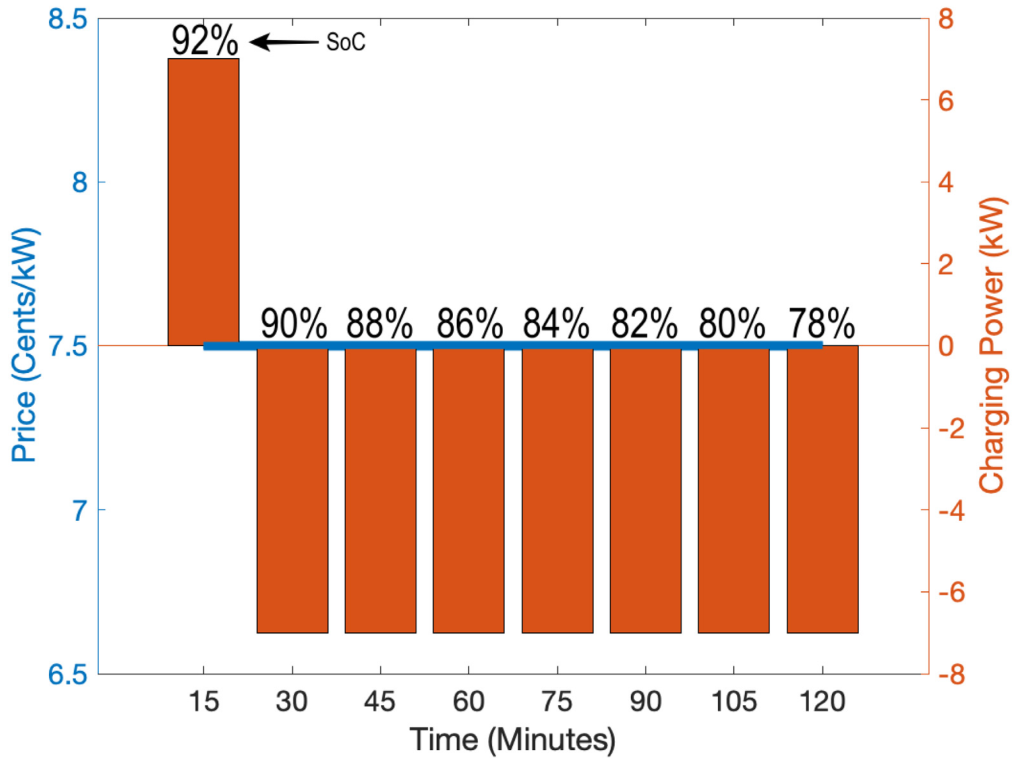

Fig. 9. Actions for 2 h starting at high prices with high initial SoC.

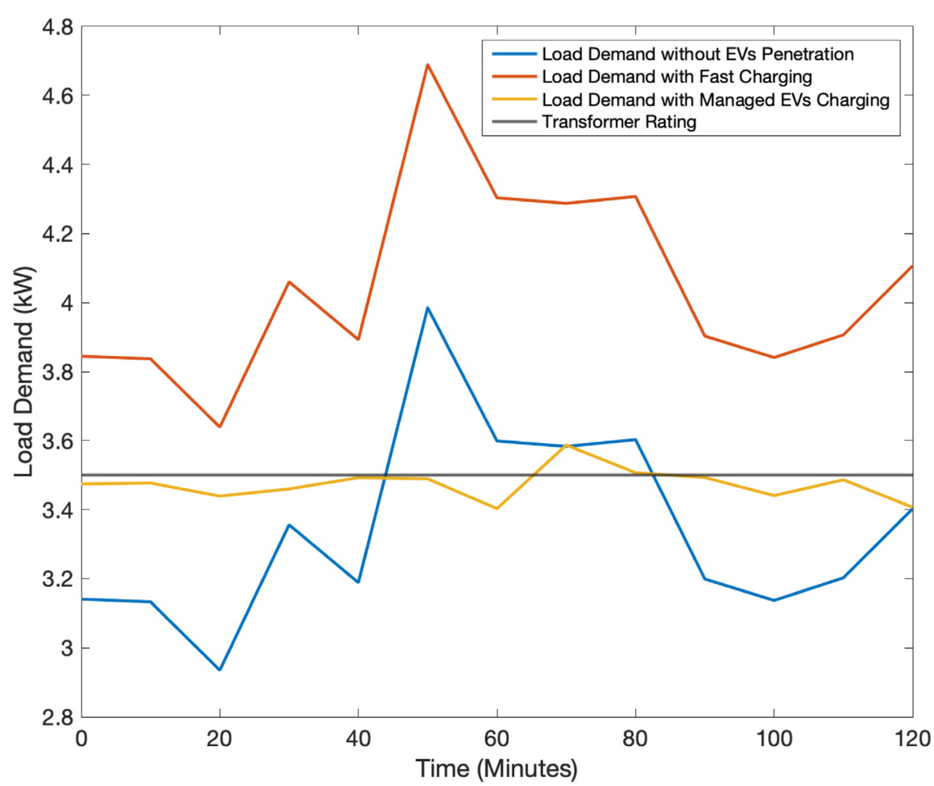

Fig. 10. Load demand for case study 3. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.5. Case study 5: 2 h with low SoC at high prices

After that, the fifth case study has the same conditions as the third case study, however, the initial SoC is set as 30%. Fig. 12 displays the charging actions chosen by the RL agent, alongside with the electricity price. The SoC of the EV after each charging action is placed above the bar. As seen in the figure, the RL agent chooses to conventionally charge and discharges for a single time period. This can be attributed to the low SoC. Despite the high electricity prices, the RL model realizes that it cannot use V2G

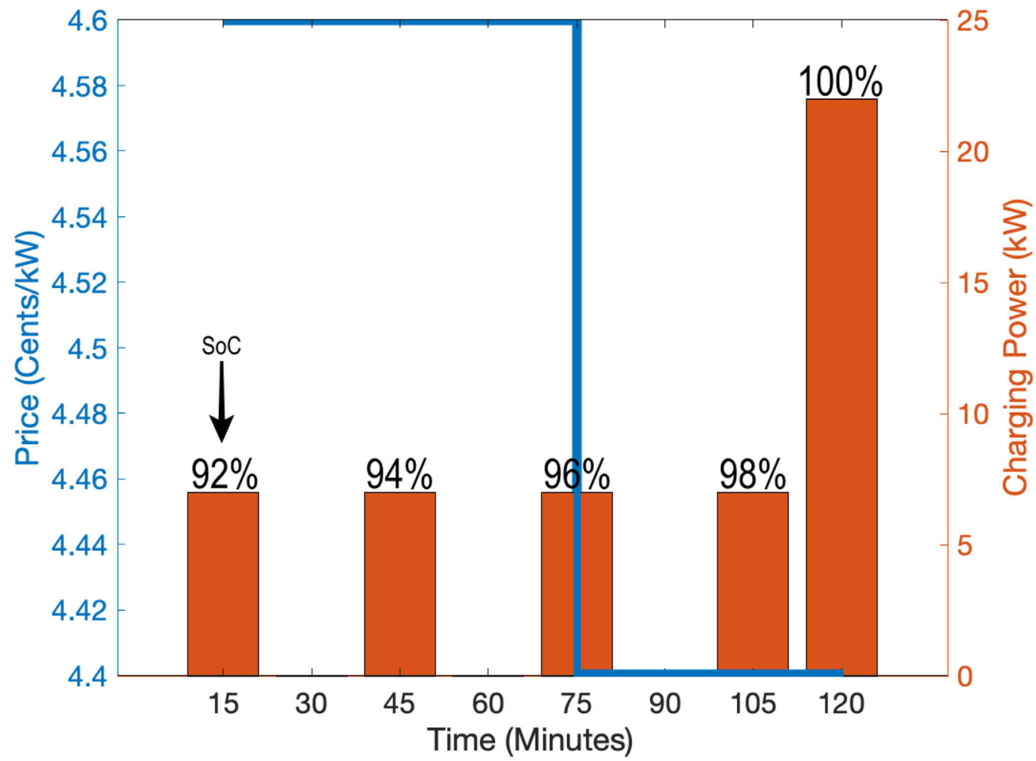

Fig. 11. Actions for 2 h starting at low prices with high initial SoC.

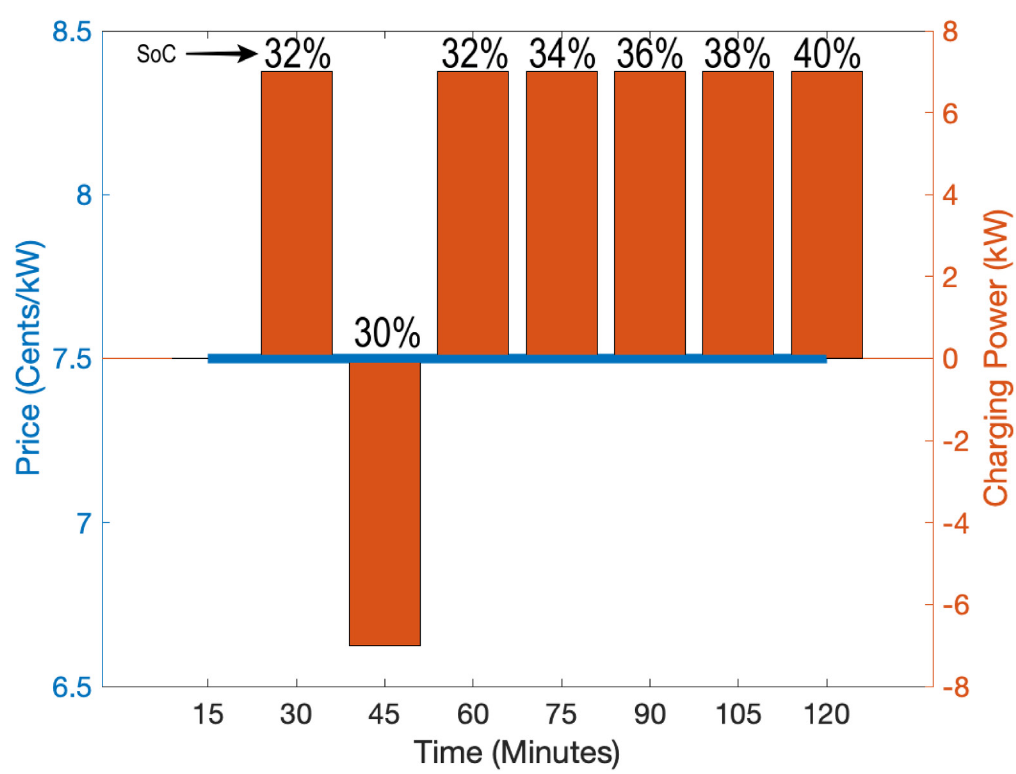

Fig. 12. Actions for 2 h starting at high prices with low initial SoC.

operation as the SoC is very low. Further, the RL model chooses to use conventional charging, instead of fast charging, in order to minimize charging cost.

Fig. 13 shows the results of the same case study with the electricity prices advanced by three hours and with an amplitude uncertainty of ± 20%. Once again, due to the low SoC,

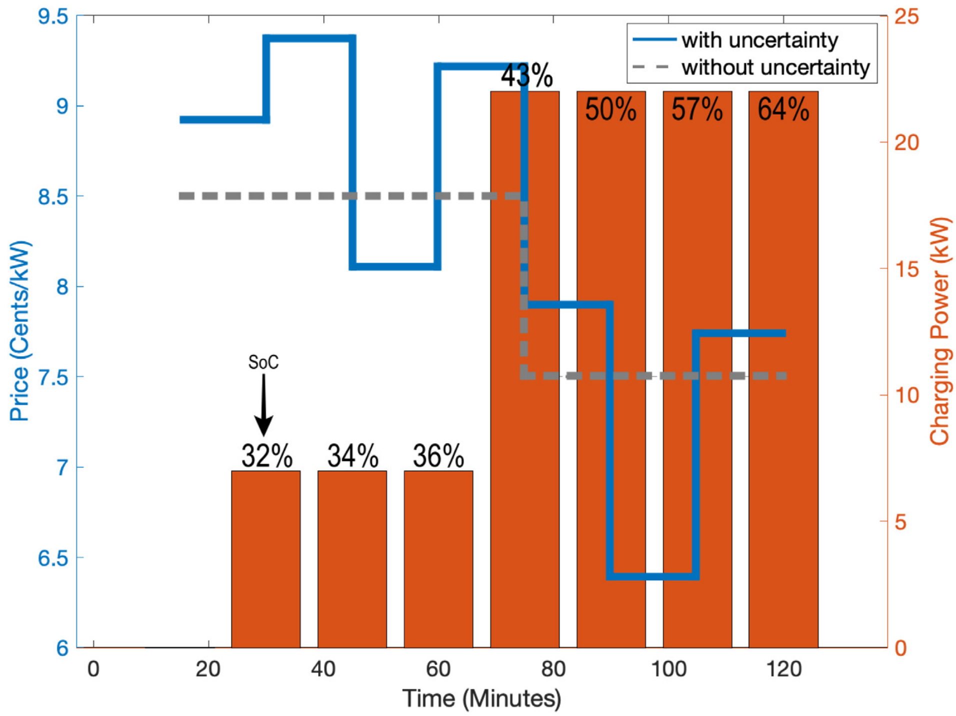

Fig. 13. Actions for 2 h starting at high prices with low initial SoC (with a 3 h advance and 20% amplitude uncertainty).

conventional and fast charging actions are utilized to quickly charge the EV to avoid reaching the minimum SoC. Thus, it is seen that in spite of introducing an advancement of 3 h in the electricity price and 20% amplitude uncertainty, the RL model is still able to find a suitable charging sequence.

Additionally, Fig. 14 presents the load demand curves for the fifth case study. Similar to case study 3, the load curve for the controlled EVs charging has been flattened and avoids significant power overloads. It can be deduced that the RL model successfully improves the operation of the power system and achieves the objectives of the utility. Moreover, the maximum voltage fluctuation for the three situations are 0.1, 0.14 and 0.07, respectively.

4.6. Case study 6: 2 h with low SoC at low prices

Finally, the last case study has the same conditions as the fourth case study, however, the initial SoC is set as 30%. Fig. 15 portrays the charging actions chosen by the RL agent and the electricity price. The SoC of the EV after each charging action is placed above the bar. As seen in the figure, the RL agent chooses a single conventional charging action, a single idle action and then continues with fast charging. The fast charging starts when the electricity price slightly drops, in order to quickly charge the EV and to exploit the low electricity prices. The use of the idle action at the start can be attributed to the RL model trying to further decrease the charging cost through the set discount.

Therefore, for the situations where time is a constraint, as well as the SoC of the EV nearly reaching a critical level, the RL model is able to effectively choose proper charging actions to decrease the charging cost for the user. When the electricity price is high, conventional charging is used and V2G operation is avoided. On the contrary, when the electricity price is low, fast charging is utilized to quickly charge the EV using decreased electricity prices.

In addition, the effect of the RL model on the power system parameters is similar to other more computationally intensive optimization techniques such as quadratic programming, as seen in Zhang et al. (2012). Also, the RL model also surpasses other optimization techniques, as seen in Sortomme et al. (2011), where quadratic programming was not able to successfully flatten the load curve, and machine learning algorithms, as shown in López et al. (2019). It is also noteworthy to mention that such techniques suffer greatly from the introduction of the uncertain behaviour of EVs charging, as seen in Zhang et al. (2012), where the flattening of the load curve is affected based on the EVs penetration level. Moreover, the performance of such optimization techniques might also suffer with the addition of V2G operation, due to the increased uncertainty in that case.

Thus, it can be deduced that the RL model successfully manages the charging of EVs, and the results yielded by the presented model is close to other studies, such as Mhaisen et al. (2020) and Qiu et al. (2020), in terms of the success of achieving its objective with similar efficiency. Nevertheless, as previously mentioned, unlike previous studies, the presented model takes into consideration both the user and utility requirements. In addition, both fast charging and V2G operation are defined in the RL model to allow the model to be up-to-date with the most recent technologies in EVs charging. Moreover, since the RL model is based on the reward function, which depends mainly on the electricity price, the size of the electrical distribution grid does not affect the performance of the RL model. Thus, the RL model can be successfully scaled to bigger networks. The drawback of scaling such a method would be the increased computational time for training the RL model. However, this can be mitigated by sharing the computational load over multiple parallel nodes.

5. Conclusion

In conclusion, this paper investigated the utilization of an adaptive RL method to manage and coordinate the charging of EVs, while taking into consideration the use of conventional charging, fast charging and V2G operation, to help protect the

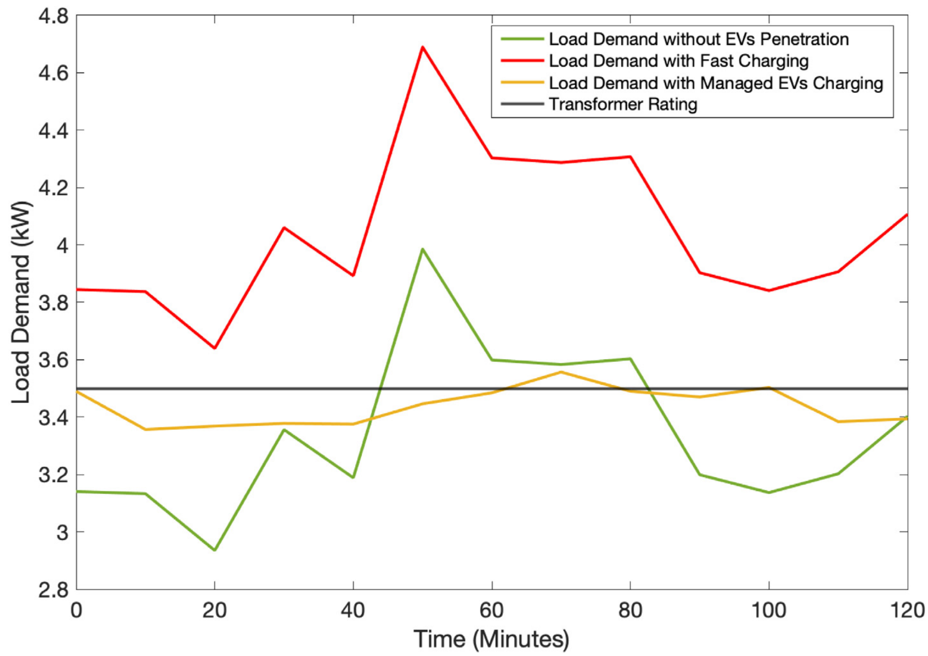

Fig. 14. Load demand for case study 5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

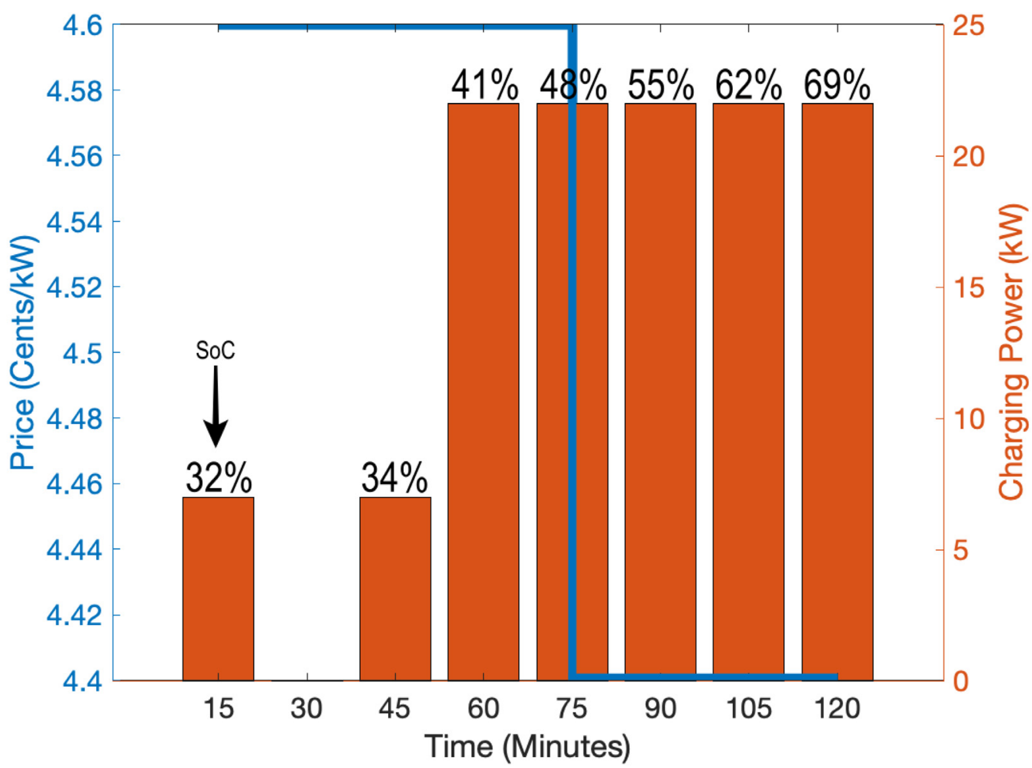

Fig. 15. Actions for 2 h starting at low prices with low initial SoC.

distribution system and satisfy the charging requirements of the EV user. Moreover, the RL model was tested with real-life charging data, and the uncertainty analysis showed that the model can work in the changing environment of EV charging, in terms of its robustness to time and magnitude uncertainty in electricity prices. The use of the RL model is highly adaptable with the environment and has the ability to constantly learn new charging strategies based on the status of the real-life electricity cost data. Thus, the model provides a coordinated charging method that is flexible in its usage and not localized to certain areas or to specific power systems.

In the RL model, the RL environment was modelled as the EV charger and the RL agents was modelled as the EV user. The possible actions that the RL agent can take are idling, conventional charging, fast charging and V2G operation. The reward function mainly consisted of the charging cost for the user, as well as some elements that contribute to achieving the utility's requirements. Thus, the objective of the RL model was to minimize the charging cost of the user.

Additionally, the RL model was tested using a range of different charging scenarios to investigate its effectiveness and reliability in managing the charging of EVs. It was seen that the model is able to successfully find charging strategies that obliges with the user's requirements, as well as protecting the distribution grid, even in the presence of uncertainties in the electricity price. Such actions include using V2G operation and idling at peak cost times, while charging during low electricity cost times.

At high charging prices, the RL model prefers V2G operation, with some idling to benefit from the set discount. However, when a maximum charging time is set with a low starting SoC, the model prefers to use conventional charging to avoid reaching the minimum SoC, and charge the EV at a decreased cost. Furthermore, when a maximum charging time is set with a high starting SoC, the model starts to discharge in order for the user to gain money. At low charging prices, the RL model starts fast charging, with some idling actions during specific case studies to benefit from the set discount. Nevertheless, when a maximum charging time is set with a low starting SoC, the model prefers to use fast charging to charge quickly and benefit from the low electricity prices.

6. Future work

Future work can encompass the testing of the RL model with other electricity pricing strategies to evaluate its performance when time-of-use pricing is not utilized. Also, the effect of a range of different load specific characteristics can be studied, including the quality factor of the load, the rating of the load, and the resistance and inductance of the load. Finally, the RL model can also be refined to fully take into consideration all of the utility's requirements to analyse such a model in real-life scenarios.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgement

This publication was made possible by IRCC grant IRCC-2022-609 from Qatar University, Qatar. The statements made herein are solely the responsibility of the authors.

References

Abid, M., Tabaa, M., Chakir, A., Hachimi, H., 2022. Routing and charging of electric vehicles: Literature review. Energy Rep. 8, 556–578. http://dx.doi.org/10.1016/j.egyr.2022.07.089, Technologies and Materials for Renewable Energy, Environment and Sustainability.

Carlson, M., 2015. "Now is the time for time-of-use pricing". URL https://cubillinois.wordpress.com/2015/02/13/now-is-the-time-for-time-of-use-pricing/.

Chu, Y., Wei, Z., Fang, X., Chen, S., Zhou, Y., 2022. A multiagent federated reinforcement learning approach for plug-in electric vehicle fleet charging coordination in a residential community. IEEE Access 10, 98535–98548. http://dx.doi.org/10.1109/ACCESS.2022.3206020.

Dabbaghjamanesh, M., Moeini, A., Kavousi-Fard, A., 2021. Reinforcement learning-based load forecasting of electric vehicle charging station using Q-learning technique. IEEE Trans. Ind. Inform. 17 (6), 4229–4237. http://dx.doi.org/10.1109/TII.2020.2990397.

Dang, Q., Wu, D., Boulet, B., 2019. A Q-learning based charging scheduling scheme for electric vehicles. In: 2019 IEEE Transportation Electrification Conference and Expo. ITEC, pp. 1–5. http://dx.doi.org/10.1109/ITEC.2019.8790603.

Deilami, S., Masoum, A.S., Moses, P.S., Masoum, M.A.S., 2011. Real-time coordination of plug-in electric vehicle charging in smart grids to minimize power losses and improve voltage profile. IEEE Trans. Smart Grid 2 (3), 456–467. http://dx.doi.org/10.1109/TSG.2011.2159816.

Ding, T., Zeng, Z., Bai, J., Qin, B., Yang, Y., Shahidehpour, M., 2020. Optimal electric vehicle charging strategy with Markov decision process and reinforcement learning technique. IEEE Trans. Ind. Appl. 56 (5), 5811–5823. http://dx.doi.org/10.1109/TIA.2020.2990096.

Harrabin, R., 2020. Ban on new petrol and diesel cars in UK from 2030 under PM's green plan. URL https://www.bbc.com/news/science-environment-54981425.

Hu, B., Li, J., 2022. A deployment-efficient energy management strategy for connected hybrid electric vehicle based on offline reinforcement learning. IEEE Trans. Ind. Electron. 69 (9), 9644–9654. http://dx.doi.org/10.1109/TIE.2021.3116581.

Li, S., Hu, W., Cao, D., Dragičević, T., Huang, Q., Chen, Z., Blaabjerg, F., 2022. Electric vehicle charging management based on deep reinforcement learning. J. Mod. Power Syst. Clean Energy 10 (3), 719–730. http://dx.doi.org/10.35833/MPCE.2020.000460.

Liang, Y., Ding, Z., Ding, T., Lee, W.-J., 2021. Mobility-aware charging scheduling for shared on-demand electric vehicle fleet using deep reinforcement learning. IEEE Trans. Smart Grid 12 (2), 1380–1393. http://dx.doi.org/10.1109/TSG.2020.3025082.

Liu, Z., Wu, Q., Huang, S., Wang, L., Shahidehpour, M., Xue, Y., 2018. Optimal day-ahead charging scheduling of electric vehicles through an aggregative game model. IEEE Trans. Smart Grid 9 (5), 5173–5184. http://dx.doi.org/10.1109/TSG.2017.2682340.

Liu, T., Zou, Y., Liu, D., Sun, F., 2015. Reinforcement learning of adaptive energy management with transition probability for a hybrid electric tracked vehicle. IEEE Trans. Ind. Electron. 62 (12), 7837–7846. http://dx.doi.org/10.1109/TIE.2015.2475419.

López, K.L., Gagné, C., Gardner, M.-A., 2019. Demand-side management using deep learning for smart charging of electric vehicles. IEEE Trans. Smart Grid 10 (3), 2683–2691. http://dx.doi.org/10.1109/TSG.2018.2808247.

Malekshah, S., Alhelou, H.H., Siano, P., 2021. An optimal probabilistic spinning reserve quantification scheme considering frequency dynamic response in smart power environment. Int. Trans. Electr. Energy Syst. 31 (11), http://dx.doi.org/10.1002/2050-7038.13052.

Mhaisen, N., Fetais, N., Massoud, A., 2020. Real-time scheduling for electric vehicles charging/discharging using reinforcement learning. In: 2020 IEEE International Conference on Informatics, IoT, and Enabling Technologies. ICIoT, pp. 1–6. http://dx.doi.org/10.1109/ICIoT48696.2020.9089471.

Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al., 2015. Human-level control through deep reinforcement learning. Nature 518 (7540), 529–533. http://dx.doi.org/10.1038/nature14236.

Pan, Y., Zhang, J., Yuan, C., Yang, H., 2019. Supervised reinforcement learning via value function. Symmetry 11 (4), 590–600. http://dx.doi.org/10.3390/sym11040590.

Qiu, D., Ye, Y., Papadaskalopoulos, D., Strbac, G., 2020. A deep reinforcement learning method for pricing electric vehicles with discrete charging levels. IEEE Trans. Ind. Appl. 56 (5), 5901–5912. http://dx.doi.org/10.1109/TIA.2020.2984614.

Sadeghianpourhamami, N., Deleu, J., Develder, C., 2020. Definition and evaluation of model-free coordination of electrical vehicle charging with reinforcement learning. IEEE Trans. Smart Grid 11 (1), 203–214. http://dx.doi.org/10.1109/TSG.2019.2920320.

Shahriar, S., Al-Ali, A.R., Osman, A.H., Dhou, S., Nijim, M., 2021. Prediction of EV charging behavior using machine learning. IEEE Access 9, 111576–111586. http://dx.doi.org/10.1109/ACCESS.2021.3103119.

Shin, M., Choi, D.-H., Kim, J., 2020. Cooperative management for PV/ESS-Enabled electric vehicle charging stations: A multiagent deep reinforcement learning approach. IEEE Trans. Ind. Inform. 16 (5), 3493–3503. http://dx.doi.org/10.1109/TII.2019.2944183.

Sortomme, E., Hindi, M.M., MacPherson, S.D.J., Venkata, S.S., 2011. Coordinated charging of plug-in hybrid electric vehicles to minimize distribution system losses. IEEE Trans. Smart Grid 2 (1), 198–205. http://dx.doi.org/10.1109/TSG.2010.2090913.

Stojkovic, J., 2019. Multi-objective optimal charging control of electric vehicles in PV charging station. In: 2019 16th International Conference on the European Energy Market. EEM, pp. 1–5. http://dx.doi.org/10.1109/EEM.2019.8916281.

Sutton, R.S., Barto, A.G., 2020. Reinforcement Learning: An Introduction. The MIT Press.

Tan, J., Wang, L., 2017. Real-time charging navigation of electric vehicles to fast charging stations: A hierarchical game approach. IEEE Trans. Smart Grid 8 (2), 846–856. http://dx.doi.org/10.1109/TSG.2015.2458863.

Van Kriekinge, G., De Cauwer, C., Sapountzoglou, N., Coosemans, T., Messagie, M., 2021. Peak shaving and cost minimization using model predictive control for uni- and bi-directional charging of electric vehicles. Energy Rep. 7, 8760–8771. http://dx.doi.org/10.1016/j.egyr.2021.11.207.

Virta, 2021. The global electric vehicle market in 2022. URL https://www.virta.global/global-electric-vehicle-market,

Wei, Z., Li, Y., Cai, L., 2018. Electric vehicle charging scheme for a park-and-charge system considering battery degradation costs. IEEE Trans. Intell. Veh. 3 (3), 361–373. http://dx.doi.org/10.1109/TIV.2018.2843126.

Wei, Z., Quan, Z., Wu, J., Li, Y., Pou, J., Zhong, H., 2022. Deep deterministic policy gradient-DRL enabled multiphysics-constrained fast charging of lithium-ion battery. IEEE Trans. Ind. Electron. 69 (3), 2588–2598. http://dx.doi.org/10.1109/TIE.2021.3070514.

Yan, L., Chen, X., Zhou, J., Chen, Y., Wen, J., 2021. Deep reinforcement learning for continuous electric vehicles charging control with dynamic user behaviors. IEEE Trans. Smart Grid 12 (6), 5124–5134. http://dx.doi.org/10.1109/TSG.2021.3098298.

Ye, Z., Gao, Y., Yu, N., 2022. Learning to operate an electric vehicle charging station considering vehicle-grid integration. IEEE Trans. Smart Grid 13 (4), 3038–3048. http://dx.doi.org/10.1109/TSG.2022.3165479.

Zhang, C., Liu, Y., Wu, F., Tang, B., Fan, W., 2021. Effective charging planning based on deep reinforcement learning for electric vehicles. IEEE Trans. Intell. Transp. Syst. 22 (1), 542–554. http://dx.doi.org/10.1109/TITS.2020.3002271.

Zhang, P., Qian, K., Zhou, C., Stewart, B.G., Hepburn, D.M., 2012. A methodology for optimization of power systems demand due to electric vehicle charging load. IEEE Trans. Power Syst. 27 (3), 1628–1636. http://dx.doi.org/10.1109/TPWRS.2012.2186595.

Zhao, Z., Lee, C.K.M., 2022. Dynamic pricing for EV charging stations: A deep reinforcement learning approach. IEEE Trans. Transp. Electrif. 8 (2), 2456–2468. http://dx.doi.org/10.1109/TTE.2021.3139674.