IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 1, 2022, accepted April 28, 2022, date of publication May 3, 2022, date of current version May 9, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3172327

An Advanced Satisfaction-Based Home Energy Management System Using Deep Reinforcement Learning

ALI FOROOTANI[1], MOHAMMAD RASTEGAR[1], (Member, IEEE), AND MOHAMMAD JOOSHAKI[2], (Senior Member, IEEE)

[1]Department of Power and Control, School of Electrical and Computer Engineering, Shiraz University, Shiraz 71348-14336, Iran
[2]Circular Economy Solutions Unit, Geologian Tutkimuskeskus (GTK), 02151 Espoo, Finland

Corresponding author: Mohammad Jooshaki (mohammad.jooshaki@gtk.fi)

ABSTRACT Home energy management (HEM) systems optimize electricity demand of appliances according to the price-based demand response (DR) programs. Undoubtedly, customer satisfaction is of such importance that if not taken into consideration, it prevents customers from participating in the DR. HEM systems suffer from high nonlinearity due to the variety of smart appliances and different criteria for customer satisfaction. In this paper, an advanced satisfaction-based HEM system using deep reinforcement learning is proposed to hourly schedule the controllable and time-shiftable appliances, including electric vehicle, air conditioner, and lighting system as controllable loads and washing machine, and dishwasher as time-shiftable loads. The proposed framework deploys a Deep Q-Network (DQN) method. Regarding customer dissatisfaction, this paper takes into consideration nonlinear precise functions. The Kano model for EV departure SoC, charging duration and lighting system satisfaction, desired temperature span for air conditioner, and the desirable operation period, waiting time, and consecutive mode of dishwasher and washing machine are taken into account. The proposed HEM system is applied to a smart home, and the results are compared with those of the Q-Learning algorithm. Numerical results prove the effectiveness of the proposed HEM system in reducing electricity cost and customer dissatisfaction, as well as the superiority of DQN over Q-Learning as well.

INDEX TERMS Deep reinforcement learning, demand response, home energy management, customer dissatisfaction.

I. INTRODUCTION

In modern societies, residential customers use advanced and technological appliances. Home appliances account for around 41% of the total residential energy consumption in the United States [1]. Development of the smart grids and significant advances in smart household appliances and the internet of things have paved the way for home energy management (HEM) to schedule controllable appliances. An optimal HEM strategy yields the optimum time and amount of energy consumption under a price-based demand response (DR) program [2].

An in-depth review of the relevant literature reveals the considerable efforts devoted to optimizing the HEM problem. In this respect, a wide range of classic optimization methods such as heuristic-based [3], fuzzy methods [4], MINLP [5] or commercial optimization solvers such as Scheduler [6] have been put forward. However, as the environment with which a HEM system interacts changes dynamically, solving the HEM problem with a fixed environment and set of scenarios via conventional optimization methods fails to yield a pragmatic solution [7].

In contrast to traditional methods, machine learning is able to tackle this handicap through a learning process. A machine-learning algorithm solves the problem by constructing a generalized description of the input data rather than memorizing the data. Reinforcement learning (RL) [8], one of the main subcategories of machine learning, has recently been used to implement energy management and accomplish the DR program. A review of the RL approaches for the HEM is provided in [9]. In [7] and [10], [11], researchers deal with residential energy management via

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

Q-Learning, a prevalent model-free algorithm. In [7], authors apply Q-Learning to solve the HEM problem, where user dissatisfaction is considered by calculating the deviation of energy consumption from the maximum power ratings of appliances. It should be mentioned that electric vehicle (EV) is not considered in [7]. The authors in [10] develop this field by applying a multi-agent Q-Learning to the DR for a smart home, where EV is also considered. However, the battery degradation and the customer dissatisfaction caused by waiting to reach the desired state of charge (SoC) are not taken into account. Authors in [10] take advantage of fuzzy reasoning to consider human preferences and make use of Q-Learning to implement DR. Researchers in [12] put forth an incentive-based DR in which Q-Learning is adopted, where dissatisfaction is formulated in the light of minimizing the load curtailment. In [13], the authors use the fitted-Q iteration algorithm to apply RL to an electric water heater. In contrast to previous works, thermal comfort is included in [13], precisely. More recently, authors in [14] made use of Q-Learning for an HVAC control system.

Despite all the advantages, Q-Learning suffers from a variety of shortcomings such as the curse of dimensionality and using Q-table with a fixed size. To tackle these downsides, the combination of RL with deep learning has recently proved promising [15]. Deep Q-Network (DQN) [16], which is the combination of a deep neural network (DNN) and Q-Learning, has solved complex problems such as playing Atari2600 games. In [17], [18], DQN is adopted for the optimal EV charging and navigation, respectively. Deep reinforcement learning (DRL) has also been used in studies that are more recent to optimize the indoor temperature [19], [20]. A HEM system based on the deep deterministic policy gradient is developed in [21], aiming to fulfill thermal comfort. Authors in [22] propose an optimization strategy for time-shiftable and controllable appliances where they suggest that customer dissatisfaction is only responsive to usage periods. Similar to [22], authors in [23] model customer dissatisfaction. Furthermore, some controllable loads such as the air conditioner are considered non-responsive in [23] or time-shiftable in [24], which are not realistic assumptions. The air conditioner is modeled precisely regarding thermal comfort in [25] via DRL, where the scheduling of other controllable and shiftable appliances is ignored. In summary, the following gaps are identified in the existing literature:

1) HEM is involved with an unstable environment. Hence, using conventional optimization methods is challenging.
2) Precise modeling of customer dissatisfaction has been considered only in the case of modeling an individual appliance (commonly air conditioner or EV). When it comes to considering various appliances, dissatisfaction is disregarded or, at best, is simply modeled by calculating the deviation from the maximum power rating of appliances.
3) Most of the previous works which made use of DRL have focused on scheduling one or a limited number of appliances owing to the hardship of deploying this algorithm.

In this paper, we propose an advanced satisfaction-based HEM system using DRL. The proposed model, aiming at reducing the electricity cost, takes into account controllable loads (EV, air conditioner, and lighting system), time-shiftable loads (dishwasher and washing machine), and non-responsive loads (TV and refrigerator). The proposed HEM system is equipped with the Kano model (a nonlinear model to quantify the dissatisfaction) to estimate and minimize the dissatisfaction caused by departure SoC and battery charging duration of EV. Furthermore, Kano model is deployed to quantify the lighting system satisfaction, as well. A nonlinear thermal comfort model based upon precise temperature calculating is employed for the air conditioner to preserve the temperature within the desired temperature span. Moreover, consecutive operation mode, waiting time dissatisfaction, and desirable operation period are considered for time-shiftable appliances. Deploying DRL is reasonable when the problem suffers from nonlinearity. Hence, it is imperative to model customer dissatisfaction precisely through nonlinear functions and solve this problem using DRL. To the best of the authors' knowledge, this paper, for the first time, proposes such an advanced satisfaction-based HEM system using DQN. Accordingly, we propose an advanced HEM system comprising the following contributions:

1) Putting forward an advanced hourly day-ahead HEM system equipped with DQN to reduce the electricity cost of a smart home possessing EV, air conditioner, and lighting system as controllable loads, and dishwasher and washing machine as time-shiftable loads.
2) Proposing a precise satisfaction-based framework including the Kano model for departure SoC and charging duration of EV and lighting system satisfaction. Furthermore, desirable temperature span for air conditioner, favorable operation time span, and consecutive operation mode for washing machine and dishwasher are taken into account.
3) Benchmarking the proposed DQN approach against the Q-Learning to prove the superiority of the developed HEM system in terms of reducing electricity cost and more importantly, improving customer satisfaction.

II. DEEP REINFORCEMENT LEARNING

In recent years, RL has shown remarkable progress and superhuman level performance in optimizing decision-making problems [16]. The fundamental elements of an RL algorithm are as follows: agent, environment, agent's action, reward, and state. The agent, as the decision-maker of RL, takes the actions. The environment is composed of appliances of the smart home and their relevant parameters. Each action executed by the agent leads to some changes in the environment. The information of the environment is monitored as state observation. In addition to the state, the agent receives a scaler reward corresponding to the action. RL methodology

can be modeled using the Markov decision process (MDP) [7]. MDP can solve a long-term optimal decision-making problem. Each MDP is defined by a tuple consisting of $<S, A, P, R, \gamma >$. In this tuple, S is the environment state. A stands for the action that is taken by the agent. P denotes the transition probability matrix, R is the reward signal, and γ is the discount factor. Q-Learning [26] is a model-free RL algorithm that solves nonlinear problems by estimating the maximum cumulative reward. The fundamental idea of this algorithm is to find the optimal state-action pair values in an iterative procedure. The Bellman equation describes this algorithm by:

$$Q_{new}(S_t, a_t) \leftarrow Q_{old}(S_t, a_t) + \theta \left(r + \gamma . \max_{a'_t} Q_{old}(S_{t+1}, a'_t) - Q_{old}(S_t, a_t) \right) \quad (1)$$

where $Q(S_t, a_t)$ is the state-action pair value, S_t stands for the state at time-step t, a_t and a'_t are actions taken by the agent, based on target policy and behavior policy, respectively. r is the current reward of the taken action, γ represents the discount factor, and θ denotes the learning rate.

DQN, which is a combination of the Q-Learning algorithm and a DNN, has been developed [16] to address the Q-Learning shortcomings. The idea of DQN is to use a DNN instead of a Q-table to estimate the state-action pair values. By doing so, deep sequential layers as processing units are deployed to perform a nonlinear transformation and abstract latent features from input data. The main advantage of utilizing a DNN for estimating the state-action pair value can be attributed to two main reasons. First, according to Cover's theorem, nonlinearly separable data can be transformed into linearly separable data with higher-dimensional space by means of a nonlinear transformation. Given that a neuron, with a nonlinear activation function, is a nonlinear transformation of its input, a DNN can be used to estimate a nonlinear Q-function. Second, using a DNN, rather than a Q-table with a fixed size, enables the algorithm to avoid discretizing the environment. Hence, any possible state which is not considered in a Q-table can be fed into the DNN. **Algorithm 1** explains the training of the agent with DNN.

Algorithm 1 Training Deep Q-Network

1. Initialization:
- 1.1 Setting hyperparameters of the algorithm (e.g., ϵ, batch-size, and experience-size)
- 1.2 initializing agent's memory and agent's experience
- 1.3 Initializing the DNN with random weights and biases
- 1.4 Initializing the environment (smart home)

2. Repeat for each 24-hour episode:
- 2.1 Start with an initial state S_{init}
- 2.2 Observe the state information
- 2.3 Predict the state-action pair value
- 2.4 implement the ϵ-greedy policy:
  - 2.4.1 Generate a random number
  - 2.4.2 If the generated number is less than ϵ: Select a random action and memorize the action index
  - 2.4.3 Else: Select the action with the maximum predicted value by DNN and memorize the action index
- 2.5. Calculate the current reward and obtain the next state
- 2.6. Append [state, action index, reward, next state] to the agent's memory
- 2.7 Denote the next state as the current state
- 2.8 If the length of the agent' experience is greater than the batch size:
  - 2.8.1 Randomly select a batch of experiences of a length equal to the batch size
  - 2.8.2 Train the DNN agent based on selected batch
- 2.9 Append the memory to the agent's experiences

3. Adopt the greedy policy at each time-step

III. PROPOSED HEM SYSTEM

As discussed above, this paper aims to provide an hourly day-ahead energy consumption strategy for a smart home. It is accomplished through determining the 24-hour ahead energy consumption of each appliance, aiming to reduce the electricity cost and user dissatisfaction. In this respect, it is assumed that the smart home is equipped with a HEM system consisting of an agent for each appliance. Also, smart meters are installed on appliances to monitor the situation and receive the command signals from the relevant agents regarding the electricity price at each hour. Appliances can be divided into three categories, non-responsive, time-shiftable, and controllable loads. In the remainder of this section, we explain the formulations for electricity consumption and customer dissatisfaction.

A. ENEGY CONSUMPTION MODELING

1) NON-RESPONSIVE APPLIANCES

Non-shiftable loads are appliances that cannot be turned off once they begin the operation, like a refrigerator [7], [10], [11]. Also, appliances such as TV are extremely reliant on user behavior and cannot be scheduled due to user priorities. Hence, the energy consumption of this kind of appliance at each hour is equal to their nominal energy consumption rate:

$$E_{N-R,t} = E_{Rated} \quad (2)$$

where $E_{N-R,t}$ is the amount of energy consumption of the non-responsive load. Also, E_{Rated} is the nominal electricity consumption of the appliance. Therefore, the electricity cost related to these appliances, $C_{N-R,t}$, is calculated by:

$$C_{N-R,\,t} = C_t . E_{N-R,\,t} \quad (3)$$

where C_t is the electricity price at hour t.

2) TIME-SHIFTABLE APPLIANCES

Time-shiftable loads have some flexibilities, which can be used to achieve a specific objective. For instance, they can be shifted to the off-peak hours with lower electricity prices to reduce the cost. In this paper, we develop a multiple decision model for time-shiftable appliances. Assuming that the nominal energy usage of a time-shiftable load in one hour is E_{Rated}, and it can normally finish its task in one hour [23], the multiple energy consumption modes can be derived as:

$$operation_t \in \{0, 1, 2, 3\} \quad (4)$$

where $operation_t = 0$ corresponds with the turned-off state at time step t. $operation_t = 1$ indicates that the appliance is turned on at t and operating with normal energy consumption. $operation_t = 2$ implies operating in two consecutive hours (t and $t+1$), consuming $E_{Rated}/2$ electricity power. In the same way, $operation_t = 3$ designates operating in three consecutive hours (t, $t+1$, and $t+2$), consuming $E_{Rated}/3$ energy at each hour. Regarding the above description, the electricity cost of time-shiftable loads can be derived as:

$$C_{Shiftable,\,t} = C_t\, E_{Rated}\,/operation_t \tag{5}$$

3) CONTROLLABLE APPLIANCES

In contrast to non-responsive and time-shiftable loads, controllable loads are able to operate flexibly in different levels of energy consumption. In this paper, a set of actions representing the different levels of EV charging is taken into consideration. Most of the previous works in the existing literature of HEM consider a binary-state model (i.e., charging mode and off mode) for the EV agent [22], [24]. But in this work, a quadruplet action level is taken into account:

$$action_{EV,\,t} \in \{0, E_1^{EV}, E_2^{EV}, E_3^{EV}\} \tag{6}$$

The arrival time, departure time, and SoC at arrival time adhere to normal distribution [17]. Accordingly, in this research, the agent will be trained based on various arrival and departure times and SoC at arrival time. Furthermore, EV discharging is not considered in this paper due to the damaging effect and shortening of the battery life [27].

A lighting system is another essential appliance that can be modeled as a controllable load [7]. Similar to the EV agent, a set of action levels is taken into account to formulate the lighting system as a controllable load.

$$action_{Lighting,t} \in \{E_1^L, E_2^L, E_3^L, E_4^L, E_5^L, E_6^L, E_7^L\} \tag{7}$$

Eventually, the action levels of the air conditioner are considered as below:

$$action_{AC,t} \in \{0, E_1^{AC}, E_2^{AC}, E_3^{AC}, E_4^{AC}\} \tag{8}$$

B. DISSATISFACTION MODELING

Although electricity cost reduction can make the price-based DR attractive for customers, user dissatisfaction is typically considered a significant barrier to participate in the DR programs. Thus, customers' dissatisfaction should be considered to pragmatically account for the customer participation in the DR programs [28]. In the following, elements of the proposed framework for dissatisfaction modeling are presented.

1) QUANTITATIVE KANO MODEL

Kano model is a helpful tool that seeks to give a map between customer satisfaction/dissatisfaction and requirement fulfillment [29]. Kano model characterizes the customer requirements (CRs) based on their impact on user satisfaction/dissatisfaction. Accordingly, CR is categorized into three main types, namely attractive, one-dimensional, and must-be attributes [29]. This categorizing is in line with how well different CRs can influence customer satisfaction. One-dimensional attributes are the general form of the relation between CR and customer satisfaction. These attributes lead to gratification when they are fulfilled and to displeasure when they are not. However, it should be noted that fulfilling the CRs more than expectation does not necessarily result in higher satisfaction. Attractive attributes, which follow exponential form, are the requirements whose absence does not result in dissatisfaction, whereas their presence leads to customer satisfaction. Must-be attributes are the ones whose shortage makes the user dissatisfied. Nonetheless, when these attributes are satisfied, the customer is neutral. Based upon the above description, a quantitative presentation of customer satisfaction/dissatisfaction can be provided.

In this paper, EV owner dissatisfaction is modeled in accordance with the above description. The deviation from desirable departure SoC and charging duration to achieve desirable SoC are foremost leading factors causing EV owner's dissatisfaction [30]. As an example, in the case of time limitation, when the optimal charging time is equal to the duration of charging the battery with maximum charging rate (uncontrolled manner), the customer is not dissatisfied. However, when the charging strategy lasts more than the uncontrolled manner, the customer will be dissatisfied. Therefore, EV owner dissatisfaction is a must-be attribute and is defined by:

$$Dissatisfaction_{EV,\,t} = a\; e^{-RF_{EV,\,t}} + b \tag{9}$$

The $RF_{EV,t}$ stands for requirement fulfillment and is in the range of [0-1]. In order to scale the dissatisfaction of each time step in the interval $[-1,\,0]$, the constants a and b are adjusted to -1.582 and 0.582, respectively, according to (10). **Algorithm 2** illustrates the $RF_{EV,t}$ calculation procedure.

$$a = \frac{\mathrm{e}(1)}{1-\mathrm{e}(1)}, \quad b = \frac{1}{\mathrm{e}(1)-1} \tag{10}$$

where t_{arr} and t_{dep} stand for arrival and departure times obeying normal distribution, ψ and λ_t denote the minimum charging duration and normalized deviation from the desired SoC, Cap_{batt} represents the maximum battery capacity, Ch_{max} and η_{ch} are maximum rating and efficiency of the charger, respectively.

Algorithm 2 EV Requirement Fulfillment

```
if t_arr < t < t_arr + ψ:
    RF_{EV,t} = 1 − λ_t
else:
    if t_arr + ψ ≤ t ≤ 23:
        t = t − (t_arr + ψ)
    else:
        t = (t + 24) − (t_arr + ψ)
    if SoC_t < desirable SoC:
        RF_{EV,t} = ((24 − t_arr) + (t_dep) − t) / ((24 − t_arr) + (t_dep)) × (1 − λ_t)
    else:
        RF_{EV,t} = 1
```

In addition to dissatisfaction caused by charging duration and deviation from desired departure SoC, battery degradation is also regarded in this work. According to [31], battery degradation is calculated by:

$$degradation_{EV,t} = \frac{cost_{batt}\, M_K\, action_{EV,t}}{Cap_{batt}} \tag{11}$$

where M_k stands for the slope of the linear approximation of battery life, $cost_{batt}$ represents the battery cost.

In addition to EV, the lighting system obeys the same modeling. In the existing literature, lighting system dissatisfaction is ignored or formulated based on simple deviation from maximum energy consumption. Similar to EV, the requirement fulfillment of lighting system is a fundamental expectation of the customer. When it is fulfilled, the user is neutral, whereas the user will be dissatisfied when it is not provided. Consistent with the Kano model, the lighting system belongs to the must-be category through the following nonlinear equation:

$$Dissatisfaction_{Lighting,t} = a\, e^{-RF_{L,t}} + b \tag{12}$$

where $RF_{L,t}$ stands for requirement fulfillment of the lighting system. As done for $Dissatisfaction_{EV,t}$, the constants a and b are adjusted to -1.582 and 0.582, respectively, in order to normalize the dissatisfaction. The $RF_{L,t}$ in (12) is derived by:

$$RF_{L,t} = \frac{action_{Lighting,t}}{E_{Max}} \tag{13}$$

where E_{Max} represents the maximum possible energy consumption, i.e., maximum brightness.

2) THERMAL COMFORT

Reducing the electricity cost for an air conditioner without considering thermal comfort might not be convincing for the customers due to the thermal dissatisfaction. As stated in [32], it is possible to reduce energy consumption and electricity cost and preserve user satisfaction at a satisfactory balance, concurrently. In this work, we aim to maintain the smart home temperature in the desired interval $[T_1, T_2]$, according to:

$$\begin{aligned} &\text{if } temp_t > temp_{max} \text{ or } tempt < temp_{min}: \\ &\quad TD_t = \min\{|temp_t - temp_{max}|, |temp_t - temp_{min}|\} \\ &\text{else}: TD_t = 0 \end{aligned} \tag{14}$$

where temperature at hour t is derived by [33]:

$$\begin{aligned} temp_t &= \varepsilon_{air}.temp_{t-1} \\ &+ (1-\varepsilon_{air}).(temp_{outdoor,t-1} + \eta_{ac}.action_{AC,t}/K_{air}) \end{aligned} \tag{15}$$

where $temp_t$ denotes the current indoor temperature, and $temp_{max}$ and $temp_{min}$ are the maximum and minimum admissible temperature, respectively. TD_t is the thermal discomfort, ε_{air} represents air inertia factor, $temp_{outdoor,t}$ stands for current outdoor temperature, η_{ac} is the coefficient performance, and K_{air} is thermal conductivity.

3) WAITING TIME

Operating in consecutive hours can lead to a lower electricity cost, but it brings about more dissatisfaction than operating in one hour (normal mode). Regarding desirable starting time, time-shiftable load dissatisfaction can be derived as:

$$dissatisfaction_{time-shiftable} = \sum_{i=1}^{n} |set\ time_i - T_{desirable}| \tag{16}$$

where *set time*, $T_{desirable}$ *and* n stand for current operating time (on mode), the customer desirable starting time, and the number of consecutive hours, respectively. The intention of using the absolute value operator is to appropriately calculate the dissatisfaction for *a set time* before $T_{desirable}$.

C. DRL IMPLEMENTATION

After determining the electricity cost and user dissatisfaction, the reward signal related to each agent can be modeled as follows:

$$R = -(B_1 C_{appliance} + B_2 Dissatisfaction_{appliance}) \tag{17}$$

where $C_{appliance}$ and $Dissatisfaction_{appliance}$ stand for electricity cost and dissatisfaction associated with an appliance. Also, B_1 and B_2 denote the weighting factors for electricity cost and dissatisfaction, respectively. It should be noted that weighting factors might vary for each smart home, owing to the fact that they depend on user preference [7]. As discussed in the previous subsections, we have considered several constraints such as desirable operation time for shiftable loads, desired SoC, arrival time and departure time for EV, favorable temperature span for air conditioner, and requirement fulfillment for the lighting system. Hence, the weighting factors are determined through trial and error [24] in such a manner that constraints are satisfied, and cost and dissatisfaction are minimized, as well. Moreover, the effect of manipulating them is investigated by designing a case study in Section IV.C.4.

It is should be noted that maximum energy consumption at each time step cannot exceed a threshold value due to the practical aspects of HEM system implementation. In the light of consecutive operating modes for time-shiftable loads, an additional constraint is formulated based on time-shiftable appliances to ensure that energy consumption does not exceed irrationally:

$$\sum_{i=1}^{N_a} E_{appliance,i} + E_{time-shiftable\,load} \leq E_{threshold} \tag{18}$$

where N_a represents the number of non-responsive and controllable appliances.

Afterward, each agent learns the optimal policy through maximizing the cumulative reward, separately. Consistent with the inherent nature of RL, agents pursue the optimal policy out of dynamic interaction with the environment. Due to the lack of experience of the agents at the beginning, the learning process commences with trial and error. Taking various actions and estimating the cumulative reward in cooperation with DNN which facilitates the learning process.

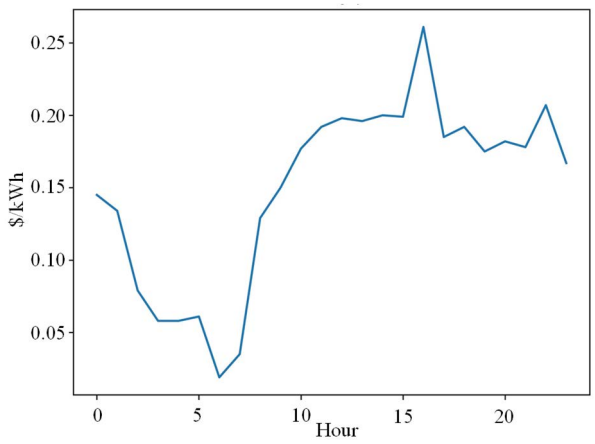

FIGURE 1. Hourly electricity price.

Gradually, agents learn to take the posterior actions, aiming to gather higher reward.

IV. SIMULATION RESULTS AND DISCUSSION

In this section, the performance of the proposed DRL approach is validated by applying it to smart home and comparing it with different scenarios.

A. SIMULATION SETTINGS

The electricity price for the price-based DR program is taken from [34]. Fig. 1 shows the hourly price for 24-hours. The simulated smart home consists of a TV and refrigerator as non-responsive loads, washing machine, and dishwasher as time-shiftable loads, and EV, lighting system, and air conditioner as controllable loads.

Table 1 lists the appliances considered in this research, where specification are taken from [7], [10], [17]. The non-responsive appliances affect (18) related to maximum power usage. Operation time for refrigerator and TV are 24-hour and three random hours, respectively. The desired operating period for the washing machine is assumed to be [13:00–19:00]. Similarly, for the dishwasher, an interval of [20:00–23:00] is taken into account as a desirable operating period. The power rating of the washing machine and dishwasher are 1.5 and 1.6 kWh, and these appliances can operate in consecutive hours, consuming a fraction of the power rating. In addition, the $E_{threshold}$ to adjust the time-shiftable loads is assumed 8 kWh. The desired indoor temperature interval for the air conditioner is assumed to be 20-22 degrees Celsius. Air conditioner corresponding parameters, namely inertia factor, coefficient performance, and thermal conductivity, are 0.7, 2.5, and 0.14, respectively [35]. The lighting system starts operating from 06:00 until 23:00 [10], [11]. Regarding EV parameters, a Nissan Leaf battery with 24 kWh capacity and 6 kW as maximum charging rate is considered. Charger efficiency is assumed to be 93% [30], and the desired SoC is 90%. As discussed before, the arrival time, departure time, and the initial SoC adhere to normal distribution [17].

TABLE 1. Appliances specifications.

Appliance	*Type*	*Power rating (kWh)*
Air Conditioner	Controllable	[0, 0.3, 0.6, 0.9, 1.2]
Lighting System	Controllable	[0.2, 0.3, 0.4, 0.5, 0.6, 0.7 ,0.8]
EV	Controllable	[0, 2, 4, 6]
Washing Machine	Time-shiftable	1.5
Dishwasher	Time-shiftable	1.6
TV	Non-Responsive	0.1
Refrigerator	Non-Responsive	0.2

In this research, a normal probability distribution function with mean and standard deviation equal to 20% and 10%, i.e., N (20%, 10%), is considered for the initial SoC. Arrival time obeys a normal distribution function with mean and standard deviation equal to 16:00 and 1 hour, i.e., N (16:00, 1 hour). Similarly, for departure time, 8:00 and 1 hour are the mean and standard deviation, respectively. Accordingly, random episodes are created to train the agents. An episode represents a whole day in the HEM problem. After the training phase, a new random episode is created to test the agents. The simulation is implemented in Python 3.6 programming language. Regarding the hyperparameters, the DNN of agents is composed of 3 hidden layers. The First, second, and third hidden layers are composed of 64, 128, and 64 neurons, respectively. The batch size is 64, and the experience-size is 200. To execute ϵ-greedy policy as behavior policy, ϵ is set to 0.05.

B. SCENARIO DEFINITION

1) SCENARIO 1: APPLYING Q-LEARNING TO THE HEM SYSTEM

In this scenario, a Q-Learning-based HEM is deployed, aiming to minimize electricity cost and customer dissatisfaction. The Kano model for lighting system and EV, thermal comfort through nonlinear thermal comfort model, and time-shiftable dissatisfaction function are applied to the agents to achieve the optimal policy.

2) SCENARIO 2: PROPOSED HEM SYSTEM BASED ON DQN

This scenario is to test the effectiveness of the proposed approach. It is similar to the previous scenario, except that it makes use of DQN rather than conventional RL. As discussed before, RL algorithms can solve nonlinear problems. However, implementing a DNN rather than a fixed size Q-Table enables the HEM system to reach better policy. Therefore, DRL is expected to outperform Q-Learning.

C. RESUTLS AND DISCUSSION

1) LEARNING PROCESS

The cumulative negative reward gathered at each episode is shown in Fig. 2 to visualize the convergence of agents. Agents learn the optimal policy through dynamic interactions with the environment. As they are not equipped with prior knowledge, the learning process starts with trial and error rather than experience. As illustrated in Fig. 2, in the beginning,

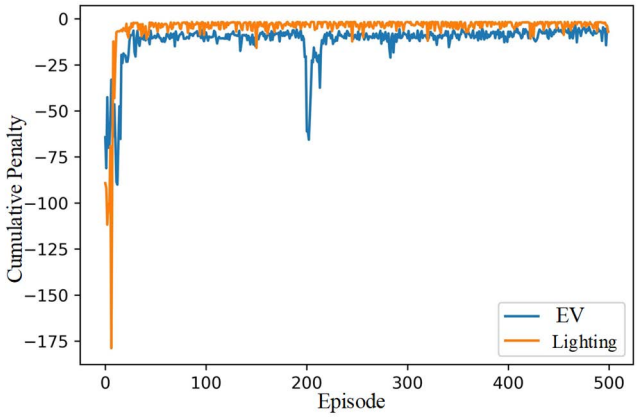

FIGURE 2. Agents' convergence in scenario 2.

TABLE 2. Electricity cost results.

Appliance	*Applying Q-Learning to the HEM system*		*Proposed HEM system based on DQN*	
	Consumption (kWh)	Cost ($)	Consumption (kWh)	Cost ($)
Refrigerator	4.8	0.715	4.8	0.715
TV	0.3	0.053	0.3	0.053
Washing Machine	0.5+0.5+0.5	0.274	0.5+0.5+0.5	0.274
Dishwasher	1.6	0.285	1.6	0.285
Air conditioner	10.8	1.644	10.5	1.576
Lighting system	8.4	1.434	8.5	1.453
EV	10	1.872	10	1.843
Total (daily)	37.4	6.277	37.2	6.199

agents are not aware of the environment and the consequences of their actions. Gradually, during the learning process, agents acquire foresight vision and learn to minimize the subsequent penalties.

2) ELECTRICITY COST

The results obtained from the described scenarios are listed in Table 2, where the energy consumption of appliances and their share in the electricity cost are provided. Comparing the electricity cost reduction in Scenarios 1 (Q-Learning) and scenario 2 (DQN) implies that DRL outperforms regular RL due to solving the problem continuously rather than discretely.

Fig. 3 shows the disaggregated energy consumption of all appliances during 24 hours. Regarding Fig. 1, the electricity price at 06:00 and 07:00 is low, hence, controllable loads consume more energy at these hours compared to the other hours of daylight. After daylight, the electricity price increases and peaks twice at 16:00 and 22:00. Therefore, agents have learned to consume energy within 16:00 and 22:00, rather than these two peaks, to decrease the electricity cost. Turning off the appliances in this period leads to high dissatisfaction, which is discussed in the next section. It should be pointed

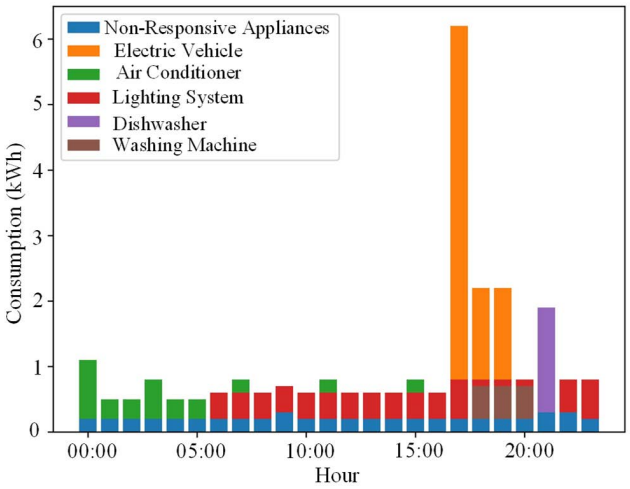

FIGURE 3. Disaggregated presentation of energy consumption in scenario 2.

TABLE 3. EV and lighting quantitative dissatisfaction based on Kano model.

Appliance	*Dissatisfaction*	
	DQN	*Q-Learning*
EV	*0.175*	*0.309*
Lighting System	*1.256*	*1.355*
Cumulative	*1.431*	*1.664*

that the peak of consumption at 17:00 is due to EV arrival time.

In both Scenarios 1 and 2, the agent of the washing machine decided to operate consecutively at 18:00, 19:00, and 20:00 to reduce the electricity cost and dissatisfaction. If the maximum energy consumption constraint (18) is disregarded, the time-shiftable load scheduling may change. Therefore, the proposed approach in Scenario 2 was executed once again, ignoring equation (18). In this case, the agent decided to turn on the washing machine one hour earlier, at 17:00, due to a lower electricity price and more closeness to the desired starting time. This decision led to exceeding the constraint (18) by 0.9 kWh.

3) DISSATISFACTION

Besides the electricity cost, customer dissatisfaction reduction is an objective of the agents. Table 3 shows the quantitative dissatisfaction related to EV and lighting systems based on the Kano model.

Considering Table 3, DQN outperforms Q-Learning by a 14% reduction in customer dissatisfaction. Although according to the inherent nature of RL, Q-Learning is capable of solving nonlinear problems, the superiority of DQN over Q-Learning is due to the extreme nonlinearity of the Kano model and the ability of DNN to solve nonlinear problems. Consequently, DQN has achieved a better policy to satisfy customer comfort concurrent with decreasing the electricity cost.

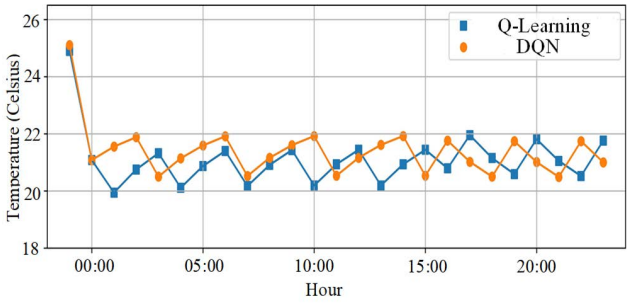

FIGURE 4. Thermal comfort in scenarios 1 and 2.

Air conditioner operation affecting the thermal comfort is illustrated in Fig. 4. In this figure, the hourly temperature of the smart home resulting from the implementation of DQN and Q-Learning on the air conditioner is presented.

To test the agent of the air conditioner, the initial temperature of the house is assumed to be 25 degrees Celsius, which is 3 degrees higher than the maximum acceptable customer temperature. As can be seen, DQN and Q-Learning have tried to decrease the temperature very quickly within the admissible temperature interval. Discussing Fig. 4, the temperature pattern in both algorithms is approximately similar, and none of them exceed the acceptable temperature interval, i.e., 20-22 degrees Celsius. However, with regard to DQN, the agent has identified the hours that can reduce energy consumption without exceeding the highest acceptable temperature. Consequently, thermal comfort is satisfied in both Q-Learning and DQN, but DQN has succeeded in satisfying thermal comfort simultaneous with leading to less electricity cost.

Discussing the acceptable operating period, the agent of the washing machine has decided to operate in three consecutive hours, starting at 18:00. By doing so, not only the electricity cost is reduced, but also customer satisfaction is met. Explaining dishwasher operation, the agent has decided to operate regularly at 21:00. It is worth mentioning that operating at 22:00 in the form of 3 consecutive hours was one of the attractive alternative decisions found by the agent. But this policy led to less reward and was overlooked by the agent.

4) EVALUATING CUSTOMER DISSATISFACTION

As this paper takes into account both electricity cost and customer dissatisfaction, it is evident that there are tradeoff solutions for HEM, depending on user sensitivity to comfort [7], [25]. A plausible case in point is where the customer has more inclination towards plunging the electricity cost rather than being comfortable. In this case, the electricity cost is expected to witness a decrease, whereas customer dissatisfaction is expected to experience a soar. Hence, a new case is studied to investigate the effect of customer comfort on the electricity cost in which the dissatisfaction factor is decreased. The results obtained for the new case (*decreased sensitivity to the user comfort*) are presented in the following.

TABLE 4. Numerical results – impact of customer dissatisfaction.

Appliance	*Decreased sensitivity to user comfort* case		Scenario 2	
	Consumption (kWh)	Cost ($)	Consumption (kWh)	Cost ($)
Refrigerator	4.8	0.715	4.8	0.715
TV	0.3	0.053	0.3	0.053
Washing Machine	1.5	0.028	0.5+0.5+0.5	0.274
Dishwasher	1.6	0.030	1.6	0.285
Air conditioner	9.6	1.388	10.5	1.576
Lighting system	7.7	1.302	8.5	1.453
EV	10.0	1.362	10.0	1.843
Total (daily)	35.5	4.878	37.2	6.199

a: ELECTRICITY COST

Table 4 shows the performance of the agents in the case of *decreased sensitivity to user comfort*. The results for Scenario 2 are also listed to facilitate the comparison. According to Table 4, in the *decreased sensitivity to user comfort* case, the electricity cost has decreased by 21.30%, compared to Scenario 2. As expected, the electricity cost has significantly plunged in the new case. However, the user dissatisfaction has been notably jeopardized, which is discussed in the following.

b: DISSATISFACTION

Given the decreased sensitivity of the agents to the user dissatisfaction, both the dishwasher and washing machine are turned on at 06:00. This policy does not satisfy the customer due to the high deviation from the desired starting times but reduces the electricity cost notably due to low electricity price at 06:00.

Table 5 shows the quantified dissatisfaction in the *decreased sensitivity to user comfort* case. To ease the comparison, Scenario 2 is also listed in this Table. To have a fair comparison, EV arrival and departure times and initial SoC are the same. As can be seen in Table 5, dissatisfaction is extremely increased. It is worth noting that the electricity consumption of EV in Table 4 for both Scenarios 2 and the new case is 10 kWh. But according to Table 5, dissatisfaction has increased remarkably. The reason for paying less electricity cost with equal energy consumption in the *decreased sensitivity to user comfort* case is that the agent of EV has decided to postpone the charging to the late hours in the midnight when the electricity price is low. This policy decreased the electricity cost, but dissatisfaction increased dramatically.

Fig. 5 shows the performance of the air conditioner in the *decreased sensitivity to user comfort* case. Similar to the two discussed scenarios, the initial temperature is assumed to be 25 degrees Celsius, whereas the maximum customer admissible temperature is considered 22 degrees Celsius. According to Fig. 5, in the *decreased sensitivity to user comfort* case, the agent of the air conditioner responds to this discomfort (high initial temperature) slowly, aiming to reduce

TABLE 5. EV and lighting dissatisfaction for *decreased sensitivity to user comfort case.*

Appliance	Dissatisfaction	
	Decreased sensitivity to user comfort case	Scenario 2
EV	*2.362*	*0.175*
Lighting System	*2.107*	*1.256*
Cumulative	*4.469*	*1.431*

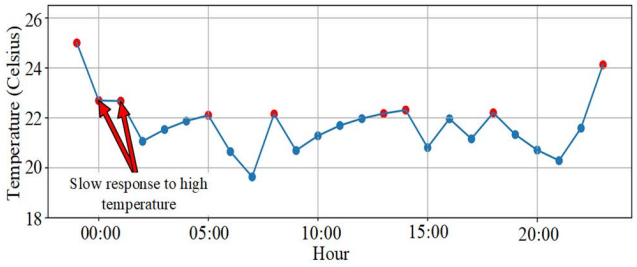

FIGURE 5. Thermal comfort discomfort case.

the electricity cost. Furthermore, the agent failed to maintain the temperature within the acceptable temperature range in several time-steps, indicated in red points.

V. CONCLUSION

This paper proposed an advanced satisfaction-based HEM system using DQN, in which a smart home including EV, air conditioner, lighting system, washing machine, dishwasher, refrigerator, and TV was simulated to test the proposed HEM system. Customer dissatisfaction was modeled precisely through the quantified Kano model, nonlinear thermal comfort, desirable operation period, waiting time, and consecutive operation mode. The proposed HEM succeeded in lowering the electricity cost, where customer dissatisfaction was also satisfied. In addition, the superiority of a DQN-based HEM system over a Q-Learning-based HEM was shown in this research. The results demonstrated that the proposed advanced satisfaction-based HEM approach outperformed the Q-Learning, especially in terms of customer dissatisfaction.

For future works, the authors plan to equip the proposed framework with recurrent neural networks such as gated recurrent unit model to forecast the EV owner behavior. Additionally, developing a satisfaction-based approach using DQN to investigate a smart grid including a number of smart homes to participate in the electricity market, is a further step to expand this field.

REFERENCES

[1] U.S. Energy Information Administration. *Use of Energy Explained Report*. Accessed: Aug. 4, 2020. [Online]. Available: https://www.eia.gov/energyexplained/
[2] P. Palensky and D. Dietrich, "Demand side management: Demand response, intelligent energy systems, and smart loads," *IEEE Trans. Ind. Informat.*, vol. 7, no. 3, pp. 381–388, Aug. 2011.
[3] K. Parvin, M. A. Hannan, A. Q. Al-Shetwi, P. J. Ker, M. F. Roslan, and T. M. I. Mahlia, "Fuzzy based particle swarm optimization for modeling home appliances towards energy saving and cost reduction under demand response consideration," *IEEE Access*, vol. 8, pp. 210784–210799, 2020.
[4] V. Hosseinnezhad, M. Shafie-Khah, P. Siano, and J. P. S. Catalao, "An optimal home energy management paradigm with an adaptive neuro-fuzzy regulation," *IEEE Access*, vol. 8, pp. 19614–19628, 2020.
[5] H. T. Dinh and D. Kim, "An optimal energy-saving home energy management supporting user comfort and electricity selling with different prices," *IEEE Access*, vol. 9, pp. 9235–9249, 2021.
[6] H. R. Gholinejad, J. Adabi, and M. Marzband, "An energy management system structure for neighborhood networks," *J. Building Eng.*, vol. 41, Sep. 2021, Art. no. 102376.
[7] R. Lu, S. H. Hong, and M. Yu, "Demand response for home energy management using reinforcement learning and artificial neural network," *IEEE Trans. Smart Grid*, vol. 10, no. 6, pp. 6629–6639, Nov. 2019.
[8] R. S. Sutton and A. G. Barto, "Introduction," in *Reinforcement Learning: An Introduction*, 1st ed. Cambridge, MA, USA: MIT Press, 1998.
[9] J. R. Vázquez-Canteli and Z. Nagy, "Reinforcement learning for demand response: A review of algorithms and modeling techniques," *Appl. Energy*, vol. 235, pp. 1072–1089, Feb. 2019.
[10] X. Xu, Y. Jia, Y. Xu, Z. Xu, S. Chai, and C. S. Lai, "A multi-agent reinforcement learning-based data-driven method for home energy management," *IEEE Trans. Smart Grid*, vol. 11, no. 4, pp. 3201–3211, Jul. 2020.
[11] F. Alfaverh, M. Denaï, and Y. Sun, "Demand response strategy based on reinforcement learning and fuzzy reasoning for home energy management," *IEEE Access*, vol. 8, pp. 39310–39321, 2020.
[12] R. Lu and S. H. Hong, "Incentive-based demand response for smart grid with reinforcement learning and deep neural network," *Appl. Energy*, vol. 236, pp. 937–949, Feb. 2019.
[13] F. Ruelens, B. J. Claessens, S. Quaiyum, B. De Schutter, R. Babuška, and R. Belmans, "Reinforcement learning applied to an electric water heater: From theory to practice," *IEEE Trans. Smart Grid*, vol. 9, no. 4, pp. 3792–3800, Jul. 2018.
[14] Z. Deng and Q. Chen, "Reinforcement learning of occupant behavior model for cross-building transfer learning to various HVAC control systems," *Energy Buildings*, vol. 238, May 2021, Art. no. 110860.
[15] M. Mohammadi, A. Al-Fuqaha, M. Guizani, and J. Oh, "Semisupervised deep reinforcement learning in support of IoT and smart city services," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 624–635, Apr. 2018.
[16] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, S. Petersen, C. Beattie, A. Sadik, I. Antonoglou, H. King, D. Kumaran, D. Wierstra, S. Legg, and D. Hassabis, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, Feb. 2015.
[17] Z. Wan, H. Li, H. He, and D. Prokhorov, "Model-free real-time EV charging scheduling based on deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 10, no. 5, pp. 5246–5257, Sep. 2019.
[18] T. Qian, C. Shao, X. Wang, and M. Shahidehpour, "Deep reinforcement learning for EV charging navigation by coordinating smart grid and intelligent transportation system," *IEEE Trans. Smart Grid*, vol. 11, no. 2, pp. 1714–1723, Mar. 2020.
[19] A. Gupta, Y. Badr, A. Negahban, and R. G. Qiu, "Energy-efficient heating control for smart buildings with deep reinforcement learning," *J. Building Eng.*, vol. 34, Feb. 2021, Art. no. 101739.
[20] Z. Jiang, M. J. Risbeck, V. Ramamurti, S. Murugesan, J. Amores, C. Zhang, Y. M. Lee, and K. H. Drees, "Building HVAC control with reinforcement learning for reduction of energy cost and demand charge," *Energy Buildings*, vol. 239, May 2021, Art. no. 110833.
[21] N. Kodama, T. Harada, and K. Miyazaki, "Home energy management algorithm based on deep reinforcement learning using multistep prediction," *IEEE Access*, vol. 9, pp. 153108–153115, 2021.
[22] Y. Liu, D. Zhang, and H. B. Gooi, "Optimization strategy based on deep reinforcement learning for home energy management," *CSEE J. Power Energy Syst.*, vol. 6, no. 3, pp. 572–582, Sep. 2020.
[23] A. Mathew, A. Roy, and J. Mathew, "Intelligent residential energy management system using deep reinforcement learning," *IEEE Syst. J.*, vol. 14, no. 4, pp. 5362–5372, Dec. 2020.
[24] E. Mocanu, D. C. Mocanu, P. H. Nguyen, A. Liotta, M. E. Webber, M. Gibescu, and J. G. Slootweg, "On-line building energy optimization using deep reinforcement learning," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3698–3708, Jul. 2019.
[25] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, and T. Jiang, "Deep reinforcement learning for smart home energy management," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 2751–2762, Apr. 2020.
[26] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Mach. Learn.*, vol. 8, nos. 3–4, pp. 279–292, 1992.

[27] S. M. Rezvanizaniani, Z. Liu, Y. Chen, and J. Lee, "Review and recent advances in battery health monitoring and prognostics technologies for electric vehicle (EV) safety and mobility," *J. Power Sources*, vol. 256, pp. 110–124, Jun. 2014.
[28] M. Rastegar, "Impacts of residential energy management on reliability of distribution systems considering a customer satisfaction model," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 6062–6073, Nov. 2018.
[29] T. Wang and P. Ji, "Understanding customer needs through quantitative analysis of Kano's model," *Int. J. Quality Rel. Manage.*, vol. 27, no. 2, pp. 173–184, Jan. 2010.
[30] M. Ebrahimi, M. Rastegar, M. Mohammadi, A. Palomino, and M. Parvania, "Stochastic charging optimization of V2G-capable PEVs: A comprehensive model for battery aging and customer service quality," *IEEE Trans. Transport. Electrific.*, vol. 6, no. 3, pp. 1026–1034, Sep. 2020.
[31] M. A. Ortega-Vazquez, "Optimal scheduling of electric vehicle charging and vehicle-to-grid services at household level including battery degradation and price uncertainty," *IET Gener., Transmiss. Distrib.*, vol. 8, no. 6, pp. 1007–1016, Jun. 2014.
[32] W.-T. Li, S. R. Gubba, W. Tushar, C. Yuen, N. U. Hassan, H. V. Poor, K. L. Wood, and C.-K. Wen, "Data driven electricity management for residential air conditioning systems: An experimental approach," *IEEE Trans. Emerg. Topics Comput.*, vol. 7, no. 3, pp. 380–391, Jul. 2019.
[33] Y.-Y. Hong, J.-K. Lin, C.-P. Wu, and C.-C. Chuang, "Multi-objective air-conditioning control considering fuzzy parameters using immune clonal selection programming," *IEEE Trans. Smart Grid*, vol. 3, no. 4, pp. 1603–1610, Dec. 2012.
[34] *IESO*. Accessed: Apr. 2021. [Online]. Available: http://reports.ieso.ca/public/PriceHOEPPredispOR/
[35] R. Deng, Z. Zhang, J. Ren, and H. Liang, "Indoor temperature control of cost-effective smart buildings via real-time smart grid communications," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.

ALI FOROOTANI received the B.Sc. degree in electrical power engineering from Shiraz University, Shiraz, Iran, in 2019, where he is currently pursuing the M.Sc. degree. His research interests include artificial intelligence, deep learning, reinforcement learning, energy management, and load forecasting.

MOHAMMAD RASTEGAR (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2009, 2011, and 2015, respectively. He joined the School of Electrical and Computer Engineering, Shiraz University, in 2016. His current research interests include modeling home energy management systems, plug-in hybrid electric vehicle operation, and power system reliability and resiliency studies.

MOHAMMAD JOOSHAKI (Senior Member, IEEE) received the M.Sc. degree in electrical engineering from the Sharif University of Technology, Tehran, Iran, in 2014, and the Ph.D. degree in power systems from Aalto University, Espoo, Finland, and the Sharif University of Technology, in 2020.

He is currently a Postdoctoral Researcher with the Circular Economy Solutions Unit, GTK, Espoo. His research interests include power system modeling and optimization, distribution system reliability, performance-based regulations, and machine learning.