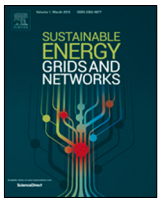
Sustainable Energy, Grids and Networks 36 (2023) 101174

Contents lists available at ScienceDirect

Sustainable Energy, Grids and Networks

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



A reinforcement learning based energy optimization approach for household fridges

Juan Pablo Giraldo-Pérez [a,*], Ricardo Mejía-Gutiérrez [a], Jose Aguilar [b,c,d]

[a] Design Engineering Research Group (GRID), Universidad EAFIT, Carrera 49 No 7 Sur-50, Medellín, 050022, Antioquia, Colombia
[b] GIDITIC, Universidad EAFIT, Carrera 49 No 7 Sur-50, Medellín, 050022, Antioquia, Colombia
[c] CEMISID, Facultad de Ingeniería, Universidad de Los Andes, Mérida, Venezuela
[d] Departamento Automática, Universidad de Alcalá, Alcalá, 28801, Madrid, Spain

ARTICLE INFO

Article history:
Received 12 February 2023
Received in revised form 21 May 2023
Accepted 13 September 2023
Available online 16 September 2023

Keywords:
Reinforcement learning
Renewable energy sources
Fridge control
Energy saving
Artificial Intelligence

ABSTRACT

The use of machine learning algorithms for the control of schedulable loads like Heating, ventilation, and air conditioning (HVAC), illumination, dryers and irrigation systems to optimize the use of RES and increase energy saving has obtained remarkable results in the last years. However, in the residential sector of tropical countries where HVAC systems are not necessary, these loads represent only a small percentage of the total energy consumption. In order to achieve a significant impact on energy savings and promote the use of RES, other residential loads must be taken into account in tropical countries. In the case of Colombia, for example, fridges account for 24% of residential energy consumption. This research proposes the use of RL for the development of a fridge energy management system capable of minimizing energy consumption and optimizing the use of RES for cooling. The fridge energy management system is based on an RL agent to control the fridge, and an artificial neural network to model the environment and assess the impact of its actions. Compared to the original fridge control, the RL-based control successfully reduced the total energy usage by 23% while also increasing the use of RES energy.

1. Introduction

RES are becoming the center of attention as the awareness for climate change increases and nations set new targets to reduce the emission of greenhouse gases in order to mitigate its consequences [1]. As part of this worldwide change not only new sources of energy are required, sectors such as industry and mobility that are currently using fossil fuels as their main source of energy are expected to continue to electrify in the next years. This transition will cause an increase in electricity demand, which paired with the increasing production of clean energy, will be key to achieve climatic goals [2].

In the specific case of Colombia, the government set a goal of reducing by 51% the generation of greenhouse gases by the year 2030. To do this, one of the main tools is to increase the installed capacity of RES such as solar and wind, which have been growing at a fast rate in the last years. In 2018, the total capacity of these sources was 50 MW, by the end of 2022 it is expected to be close to 2,5GW. Looking at the demand side of the energetic change, by the year 2018 residential energy consumption represented 20.1% of the total country's demand. In the same year, 33.8% of this energy was consumed in the form of electricity. The small participation of electricity in the residential sector is explained by the large amount of individuals who have no access to this source of energy, and therefore, use firewood as their principal cooking method. This scenario is predicted to change in the following years as the government aims to transition to less polluting sources of energy. The estimates calculate that by the year 2050 electricity will represent 55% of the total demand [3].

1.1. Previous works

Machine Learning (ML) algorithms are taking part of the solution both from the generation and consumer side. From the generation side, literature is found on the topic of solar and wind generation, where ML is being used for prediction [4,5], topology optimization [6] and material selection [7], among other topics. From the consumer side, ML is currently being used for the automation of schedule loads [8,9] and electricity consumption forecasting [10–12], among other applications. In both cases, supervised and unsupervised learning have been widely used and have obtained encouraging results.

However, the quality of these methods have a big dependency on the nature and size of the datasets used to create the models [13,14]. For this reason, RL algorithms have gained popularity

* Corresponding author.
E-mail address: jgiral95@eafit.edu.co (J.P. Giraldo-Pérez).

https://doi.org/10.1016/j.segan.2023.101174
2352-4677/

among researchers, as they do not depend on a preexisting model or dataset to train, and can learn the environment dynamics purely from experience. For example, in Heating, ventilation, and air conditioning (HVAC) systems have been successfully applied RL with the objective of minimizing costs and optimize energy usage [15–18]. In [19], a policy optimization method was used to control an HVAC system of a building taking into consideration the demand response. In [20], a Deep Reinforcement Learning (DRL) algorithm was implemented to control a multi-zone residential HVAC system, and Mason et al. [21] present a review of the literature relating to the application of RL in building energy management systems. Lee et al. [22] explored more complex environments by taking into consideration multiple home appliances, e.g., air conditioner, washing machine, and energy storage systems (ESS), to create a home energy management system. RL has also been used for more complex tasks, for example, [23] used a DRL algorithm to optimize the energy usage on a microgrid, where both the environment and the action space are multi-dimensional. Also, Ren et al. [24] proposed a forecasting based optimization method for real-time scheduling in a household energy management system (HEMS). The method is based on DRL mixed with a correntropy assisted long short-term memory.

Table 1 presents the objective and target household items of previous works. Other authors have also used ML with the objective of optimizing energy consumption and generation but were not included in Table 1 as they focused on studying loads on large buildings, districts and manufacturing facilities, and did not take into consideration small residential loads such as home appliances [25–28].

Even though previous research has been successful on applying ML and minimize energy consumption, their findings do not have a significant impact on the residential sector of countries like Colombia as they concentrate on loads that are not as common on a developing and tropical country.

Table 1
Related previous work.

Article	Method	Objective	Household item	Considers RES
[30]	Evolutionary algorithms	Minimize energy consumption	6 different appliances	No
[22]	RL	Cost reduction	Air conditioner Washing machine	Yes (PV)
[31]	Supervised learning algorithms	Reduce charging cost	Electric vehicle	No
[15]	RL	Minimize energy consumption	HVAC	No
[17]	RL	Cost reduction	HVAC	No
[32]	RL	Cost reduction	HVAC, washing machine EV and ESS	Yes (PV)
[33]	RL	Cost reduction	14 different appliances	Yes
Present	RL	Maximize the use of RES energy	Fridges	Yes

1.2. Scope and contributions

According to the National Planning Department of Colombia, the biggest consumers of residential energy in urban environments are kitchens and fridges, with 45% and 24%, respectively, followed by lighting, with 11% [29]. Having as baseline the previous works and with the objective of contributing to the residential energy transition of tropical developing countries, this research proposes a fridge energy management system capable of maximizing the use of RES energy and reducing the total energy usage. The proposed energy management system is intended to be implemented initially on Top-Freezer fridges, which are the most common and acquirable fridges for Colombian households. The system consists of three components. The first is a monitoring system installed on a fridge, which collects data about its use, internal temperature and energy consumption. The second component analyses and processes this data to generate a model to simulate the behavior of the fridge. Finally, a control system is developed and tested using the simulation model, an RL agent is used for this task. The control system uses information from a solar charging station located in Medellin, Colombia, as the RES energy generation input data. The last row of Table 1 presents the present study information in order to compare it with the existing literature.

Comparing our proposed method with the previous works, the main difference is that the majority has not focused on fridges but rather on schedulable loads such as HVAC, dishwashers and dryers. Four of the found studies include fridges in their target appliances, however, the implementation they used is different from the one presented in this work. Some treat fridges as non-controllable loads, and only use their current power consumption to decide whether to turn on or off traditional schedulable loads [30,33]. Finally, others have used ML algorithms for fridge energy usage prediction [34,35]. In contrast, our energy management system uses a fully dedicated RL agent to directly control a fridge. This allows fridges to go from simple data generators to an active part of the household energy environment, such that its utilization can be improved.

This work contributes in two areas: RL and RES optimization. From the RL perspective, this work presents the implementation of an agent to control a fridge, load that has been treated as non-schedulable by previous researches, and thus, no previous work is found on this topic. This agent uses an RL to control the fridge, and models the environment using an artificial neural network to assess the impact of its actions. From the RES perspective, the implementation of the RL agent maximizes the amount of RES energy that the fridge uses. This works is a complement to the already existing research on traditional schedulable household loads. The next sections are organized as follows: Section 2 gives insight into the three components that integrate the energy management system, Section 3 presents the case study that was selected to validate the system, Section 4 explains the training of the RL agent and the obtained results, and finally, Section 5 present the conclusions of the research.

2. Proposed system

This section describes the three components that integrate the fridge energy management system: monitoring, modeling and control.

2.1. Monitoring system

The monitoring system is in charge of obtaining data about the use and behavior of the fridge. Three variables were selected as targets for this system: temperature, power consumption and door state. The internal temperature of the fridge is essential, as the quality of the stored food directly depends on it [36]. Energy consumption is important as it is correlated with the temperature of the fridge and is the variable that is going to be optimized. Finally, the door state impacts the internal temperature and gives information about how the users interact with the fridge. This component is composed of a group of sensors, a data logger, and a database.

2.2. Modeling system

The second component of the energy management system is the model of the fridge. Here, the data gathered by the monitoring system is processed, analyzed and used to create a simulation of the fridge. The objective of this model is to create synthetic data that can be used to train a RL agent. Specifically, the model predicts the temperature of the fridge given some initial conditions. The model has as entry the power consumption, state of the fridge door, and previous temperature, and it outputs the next expected temperature. To create this model, a supervised learning approach is used, which is formulated as a regression model.

2.3. Control system

The final component of the system is the control. For this, a RL agent is used. The agent is in charge of deciding when to turn the fridge on or off, with the objective of maximizing the amount of RES energy that is used and minimize the total energy usage. This on/off control is used since the idea is to be able to implement the RL agent even on old hardware. Sections 2.3.1 and 2.3.2 present the general functioning of the RL agent, and the selected RL algorithm to build the control agent, respectively.

2.3.1. General description of RL algorithms

RL algorithms are composed by an environment, an agent, actions and rewards. The final objective of our RL agent is to learn a policy π that is able to map states into actions in order to maximize the long-term reward [37]. The agent is able to learn this mapping purely from experience.

The environment is represented by all the variables that interact with the agent. The group of these variables in a specific time is called a state. The actions taken by the agent cause a change in the environment variables, and therefore, change the current state. The environment is also in charge of returning rewards to the agent.

The actions, rewards and states are the channels through which the RL agent and environment communicate. This problem can be formulated as a Markov Decision Process (MDP) [19,37], in short, this means that every state is independent, and present actions do not depend on past actions. In general, a MDP in the context of RL produces a trajectory of states, actions and rewards, as shown in expression (1):

$$S_0, A_0, R_1, S_1, A_1, R_2, S_2, A_2, R_3 \ldots \tag{1}$$

In the case of this study, the trajectory starts with the first state given by the environment (fridge temperature and available RES). Then, the agent takes an action (e.g. turn the fridge on or off) and gets a reward for taking that action in that state. The fridge temperature would then change, in part, due to the action taken by the agent, and this cycle continues indefinitely.

2.3.2. Selected RL algorithm

There are multiple families of RL algorithms, each designed to be implemented in different kinds of scenarios. In the case of this research, the SoftMax Actor–Critic algorithm was selected, as it can handle continuous tasks (the interaction between the agent and fridge is infinite) and has the capability of obtaining a policy π close to deterministic [37]. This means the policy will always have a random factor to it. However, after training the agent, the probability of taking the best action for a given state will be close to one. Eq. (2) shows how the SoftMax distribution is calculated.

$$\pi(a|S,\theta) \doteq \frac{e^{h(s,a,\theta)}}{\sum_b e^{h(s,b,\theta)}} \tag{2}$$

Where, a is the action, S is the current state and $h(s,a,\theta)$ corresponds to the action preferences, which is calculated using Eq. (3); θ is the policy's parameter vector and $x(s,a)$ is the feature vector.

$$h(s,a,\theta) = \theta^T x(s,a) \tag{3}$$

The output of Eq. (2) is an array with the same size as actions that can be performed, and each item of the array corresponds to the probability of choosing a certain action.

SoftMax actor–critic methods consist of two main parts. First, the actor is in charge of taking actions. Then, the critic tells the actor how good of a job it is doing depending on the rewards that the environment returns according to the action taken. The pseudo-code used to train the agent is shown in algorithm 1.

Algorithm 1 Actor–Critic with Softmax policy on continuous task

Input: Softmax Policy parameterization
Input: State-value parameterization
procedure
Initialize $\bar{R}$ *to 0*
Initialize w *and* θ *to 0*
Set $\alpha^w > 0$, $\alpha^\theta > 0$, $\alpha^{\bar{R}} > 0$
Get initial state S
Loop forever:
Get A from π
Take action A, observe S', R
$\delta \leftarrow R - \bar{R} + \hat{v}(S', w) - \hat{v}(S, w)$
$\bar{R} \leftarrow \bar{R} + \alpha^{\bar{R}}\delta$
$w \leftarrow w + \alpha^w \delta \nabla \hat{v}(S, w)$
$\theta \leftarrow \theta + \alpha^\theta \delta \nabla ln\pi(A|S,\theta)$
$S \leftarrow S'$

The algorithm works by using the SoftMax policy to select an action given an initial state. After the action is taken (turning the fridge on or off) the state changes, in this case the temperature increases or decreases. This change in temperature moves the environment into a new state S' and give the agent a positive or negative reward. For example a negative reward could be given to the agent if the temperature is too low or too high for the fridge. Then, depending on the reward given, the algorithm updates its parameters in order to make it less or more likely to select the previous action.

The variables used in the pseudo code are the following: $\bar{R}$ is the average reward, w is the state-value weights, α^w, α^θ and $\alpha^{\bar{R}}$ are the step-size parameters for the critic, actor and average reward respectively. The best way to determine these last two parameters is to perform a systematic sweep looping through different combinations, but in general, it is recommended to have $\alpha^w > \alpha^\theta$ to ensure the critic updates faster, and thus, is able to correctly evaluate the actor. w is the critic weight vector, it is updated along with θ on each time step to further improve the critic and actor approximations with the objective of maximizing the long term reward.

3. Case study

This section describes how the proposed system was implemented in a specific case study.

3.1. Monitoring system

3.1.1. Context

The fridge used in this study is a Samsung rt35k571js9, it is located in an apartment occupied by two people in Medellin,

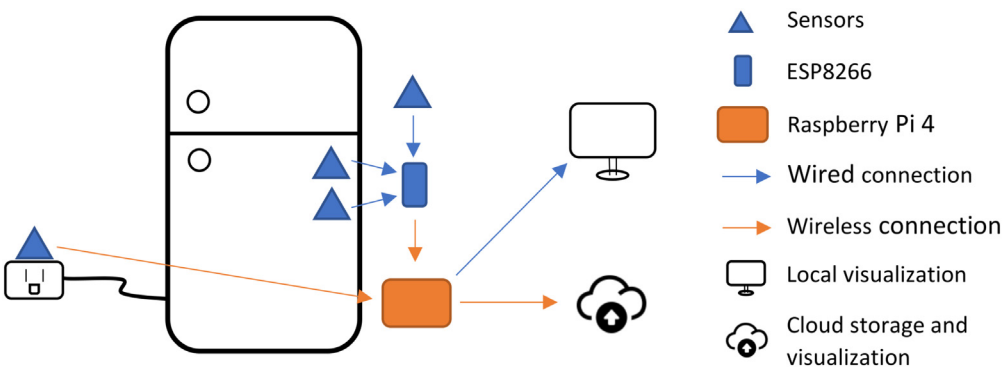


Fig. 1. Fridge variables monitoring system.

Colombia. Medellin is a city that is located near the equator line, so its temperature remains stable throughout the year. To obtain information about the behavior of the selected three variables (temperature, power consumption and door state), a network of sensors and a gateway was used. A Raspberry Pi 4 was used as the gateway to centralize all the gathered data and store it on a MongoDB database.

To acquire power consumption data, a Shelly Plug US smart plug was used, which has the capability of sensing the power drawn by any device up to 1650 W. It also has MQTT capabilities that enable a wireless connection to the Raspberry Pi. The power sensor sends an update every 30 s, or when there is a change in the power consumption.

For the temperature, two AHT21B digital sensors were used, one inside and one outside the fridge. These sensors have a resolution of 0.01 °C and work with the I2C protocol. The sensors were configured to send data when a delta of 0.05 °C with respect to the last update was detected. The internal temperature sensor was positioned and secured to the fridge's wall half way between the door and the back of the fridge. The final variable is the state of the door, which is sensed through a LDR (light-dependent resistor) that detects the moment the internal light bulb turns on when the fridge is opened. These three sensors connect to an ESP8266 module, which gathers the data and sends it to the Raspberry Pi through MQTT.

The flow of the data through the monitoring system is shown in Fig. 1. All the sensors communicate with the Raspberry Pi using MQTT, then a python script interprets this information and stores it in the MongoDB database. Finally, a backup of the database was made daily and uploaded to the cloud.

A total of 2'624.320 data entries were taken beginning on November 26 of 2021 and finishing on April 13 of 2022. The resulting database was separated and processed by variables.

3.1.2. Temperature

A total of 899.709 internal temperature measurements were made. The data is plotted to see the behavior of the variable and possible outliers. A communication problem between the ESP8266 and the temperature sensors was found to affect 10 days of the tests causing large outliers. These days were deleted from the dataset. Then, the IsolationForest algorithm was used to delete the remaining outliers [38]. The result of this process can be seen in Fig. 2 where an entire day of data is shown. It is important to notice that the temperature peaks shown in the figure are caused by the fridge being opened by the users, letting outside air to come into contact with the temperature sensor which is located inside fridge. The mean temperature the fridge had was calculated after the data cleaning process, obtaining 5.5 °C. The procedure done for the external temperature was identical to the one performed for the internal temperature. Fig. 2 also shows how the external temperature varies during the same day. The external temperature sensor was located next to the fridge, this explains its oscillating behavior that matches the on/off state of the fridge.

3.1.3. Fridge power consumption

In the case of power consumption, it is important to notice that the rate at which this sensor delivers data is different from the temperature sensors. But it is known that the power consumption between each new data point is equal. For example, if the sensor sends a power draw of 2 W and 30 s later sends 30 W, then this means that for those 29 s the power draw was 2 W. Taking this into consideration, the power consumption data was shifted to the same timestamps of the temperature so that the relation between the variables could be later analyzed.

3.1.4. Fridge state

This variable has to go through more processing, since the only information it gives directly is when the door is opened or closed. The first step is to calculate the probability of the fridge being open. To do this, the original dataset, which has 5239 entries, is transformed in a way that we get the total number of seconds the fridge is in either state. The total number of entries is now 11'192.381, one per second of the test. The results show that the fridge is open 0, 2% of the time. To further extract data, the probability per hour is calculated, the result can be seen in Fig. 3.

The results highlight the times of the day that the fridge is used the most, 6am, 8pm and 7pm being the top three. This results match the usage the two users give to the fridge since they work from 8 am to 6 pm, usually have breakfast between 6 am and 7 am, most days they do not have lunch at home, and usually, have dinner between 7:30 p.m. and 8:30 p.m.

Another important factor in this variable is the length of the openings. Fig. 4 shows a histogram of the opening times. The most frequent time is 4 s, followed by 5 and 6 s. The longest opened time was 148 s.

3.1.5. Photovoltaic production

A photovoltaic charging station located in Medellin, Colombia, was used to get data on the amount of available RES energy. The station sends information about power generation every 5 min to an online cloud service. It was considered that the power generation was equal throughout the 5 min intervals. This dataset did not need any further processing since no outliers were found.

3.1.6. Variables visualization

Fig. 2 shows the variables behavior on a randomly selected day (Saturday, November 27, 2021). The vertical lines in the first graphic represent the opening of the door, the second graphic is the internal temperature, the third is the external temperature,

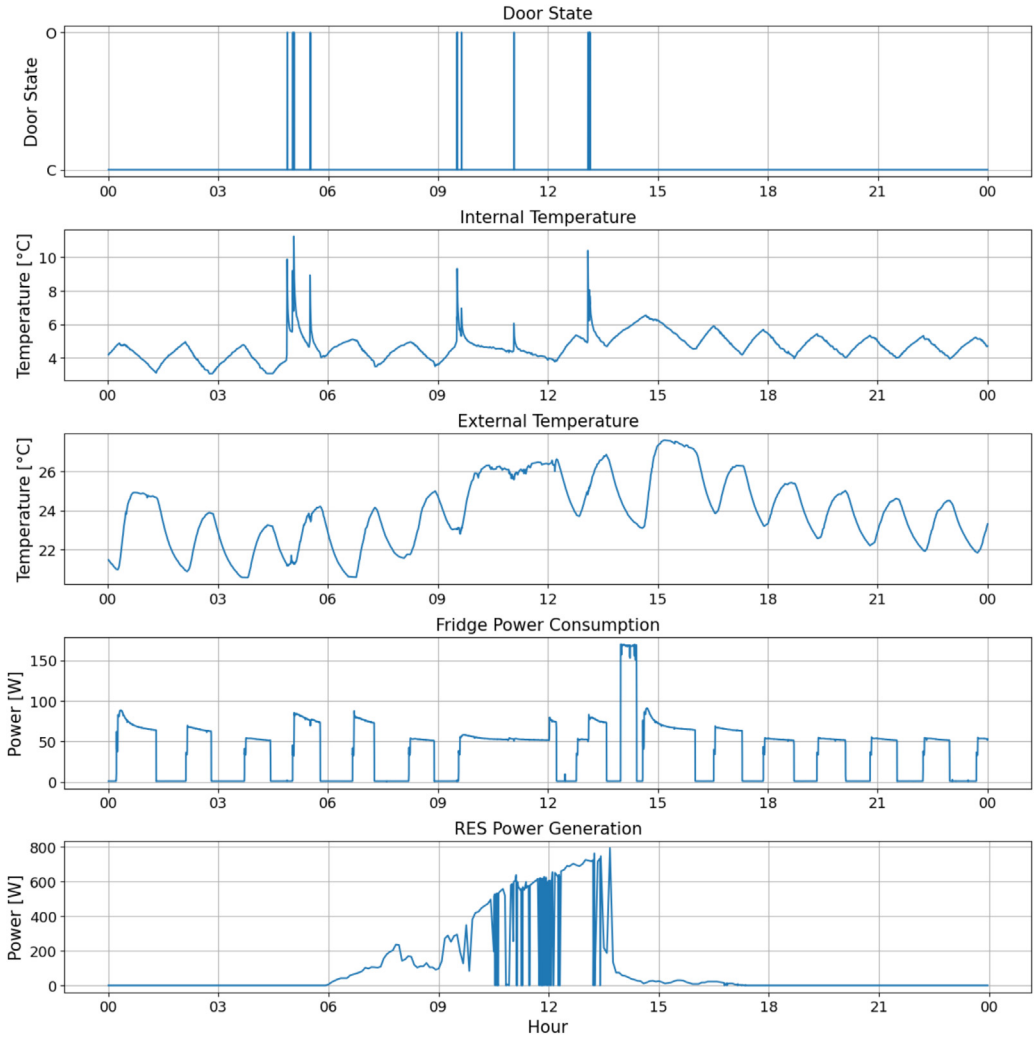


Fig. 2. Visualization of one day of data.

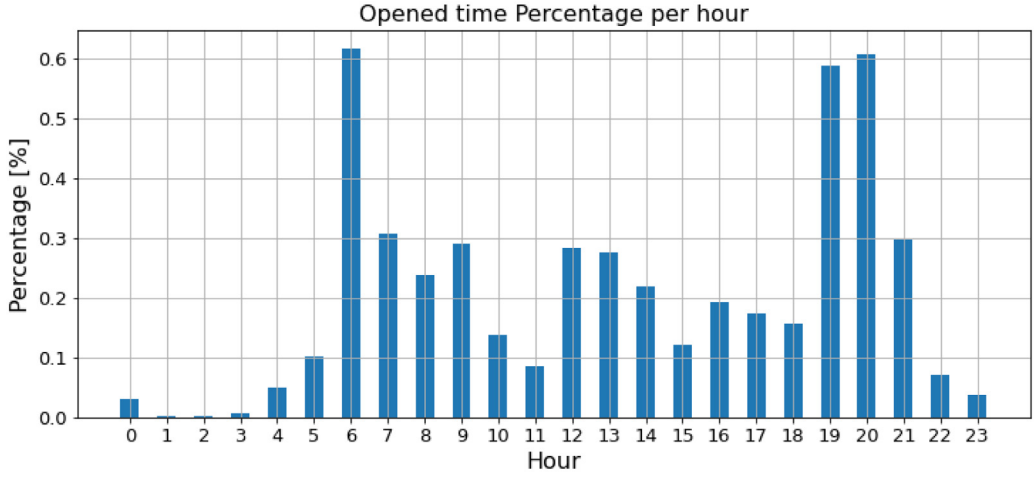


Fig. 3. Time percentage of the fridge being open per hour.

the fourth is the power consumption, and the last is the available RES power. The figure shows how the variables interact with each other. For example, how opening the door affects the internal temperature, and how the internal temperature gets colder when the fridge is on (power is being consumed).

3.2. Modeling system

As previously mentioned, the modeling component of the system is defined as a regression problem where the output is the next expected internal temperature of the fridge given some

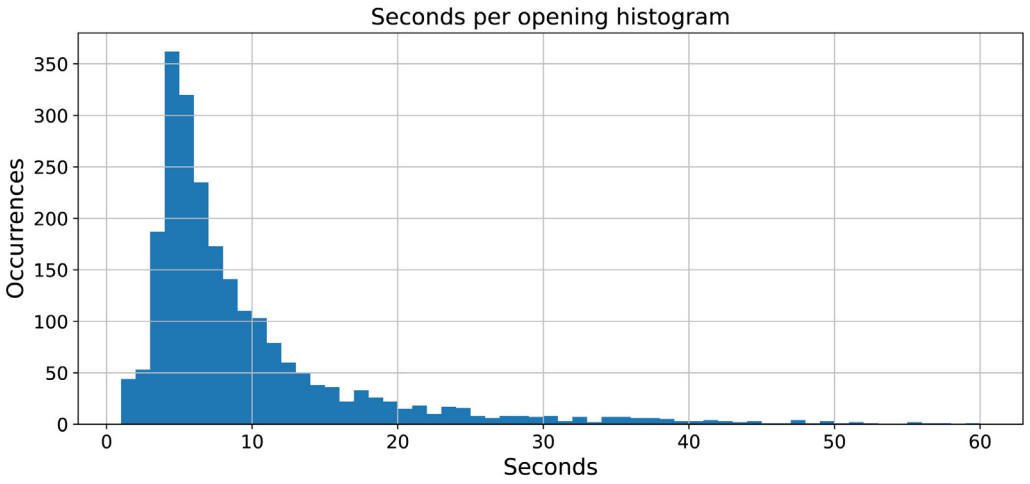


Fig. 4. Histogram of seconds the fridge is kept open.

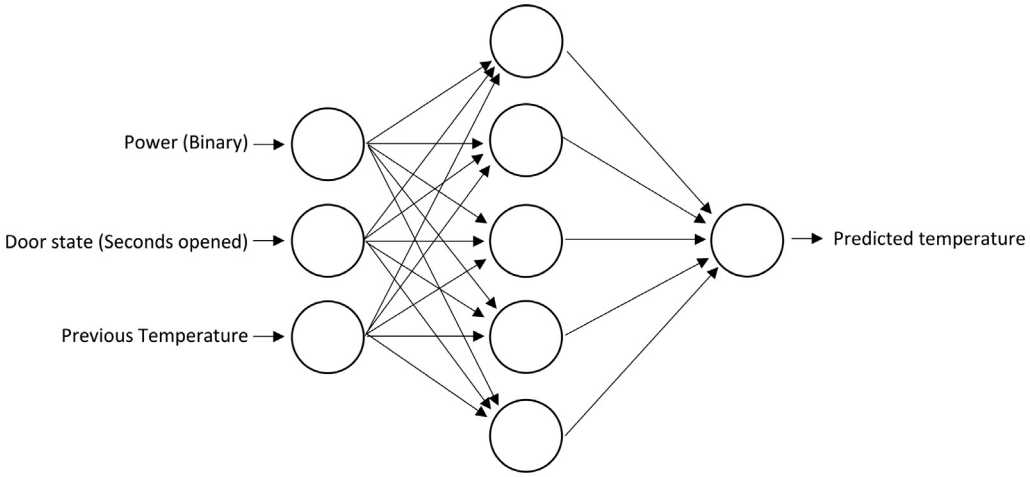


Fig. 5. Used ANN topology.

initial conditions. To solve this regression problem, the data acquired from the fridge's monitoring system and the solar charging station were used to train a supervised learning algorithm. A linear regression model was the first tested method. However, this model was not able to correctly fit the temperature when the fridge door was opened. An Artificial Neural Network (ANN) was then tested. Different combinations of inputs nodes were used in order to find the one with the best results. The first training was made using the variables gathered by the monitoring system without any previous treatment. The ANN obtained a R-squared score of 0.1528 in the testing phase.

While training the first batch of algorithms, multiple improvement opportunities were identified. The original data was organized by seconds, which meant large data sets and long training periods. As the RL agent would not be intended to turn the power of the fridge on and off on one second intervals, the dataset was grouped by minutes. The temperature and the power consumption were averaged and the door state was turned into a binary value (1 for open and 0 for close), and was added to get the amount of seconds the fridge was opened each minute. The power was then turned into a binary variable, where 1 means that the fridge was on during that minute and 0 that it was off. Additionally, the external temperature was found to have a weak correlation with the internal, for this reason, it was discarded from the model inputs.

The new dataset was used to train an ANN. The performance of the network improved and reached a 0.986 R-squared score in the testing phase. Fig. 5 presents the used ANN topology. An iterative process was used to find the best topology and activation function for the network. Only one hidden layer with five units and ReLU activation function was used as larger networks only increased the computation time while having no significant accuracy improvements.

However, even though the network had a better performance, when it was tested and compared with real data, it became clear that the temperature profile is different from the real one. As shown in Fig. 6, temperature rises and falls at a faster rate and the temperature peaks caused by the opening of the door do not fall as fast as the real ones. The mean absolute error metric was used to quantify the fit of the model, obtaining 1.56 °C. Thus, the results obtained show that on average, the model has an error of 1.56 °C on its temperature predictions.

The cause of prediction error with respect to the real data was found to be caused by the response the temperature has when the fridge is opened. The used ANN tried to generalize for all cases (open and close door) but in reality, the temperature behavior in these two cases is very different. To solve this issue, two separate ANNs were trained, one to simulate the temperature when the door is closed and has not been opened for several minutes, and

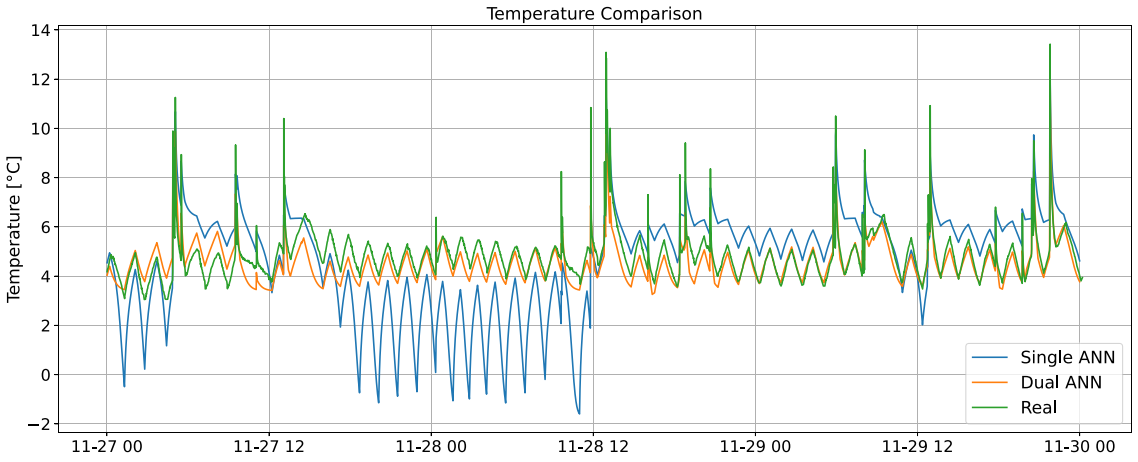


Fig. 6. Temperature prediction using ANNs.

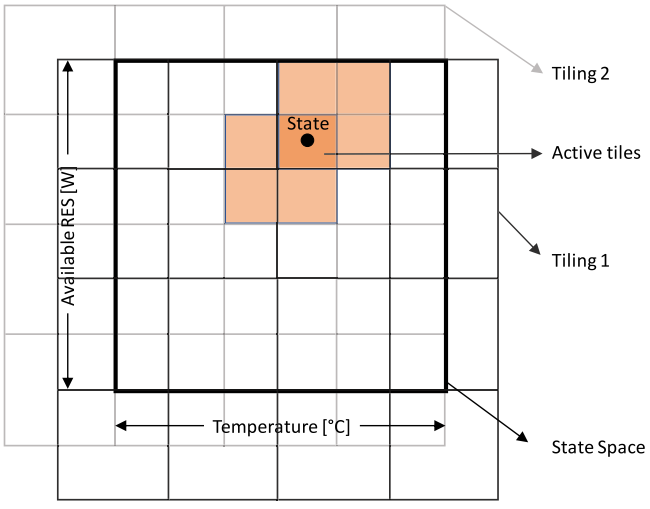


Fig. 7. Tile coding representation of the state space of the environment.

one for when the door is open. Fig. 6 shows the results obtained when using this method for a three day period prediction. The model was now able to correctly follow the tendency of the temperature change, both when the fridge is open or close. The model obtained a 0.45 °C mean absolute error, equivalent to a 9.2% mean absolute percentage error, meaning a 71.1% error reduction with respect to the single ANN approach. This approach was selected to be used for temperature prediction.

To evaluate the stability of the ANN a sensitivity analysis was performed. It consisted on introducing small noise to the inputs and observing the impact on the network's predictions. Two random numbers were generated, the first using a normal distribution with a mean of 0 and a standard deviation of 0.1 to add noise to the previous temperature, the second using a normal distribution with a mean of 0 and a standard deviation of 2 for the door state input. A total of 1000 cycles where run with different generated random numbers. A statistical analysis of the results was made in order to evaluate the effect of the introduced noise. First the original predicted temperature without noise was subtracted from the noisy results, then the mean and standard deviation was calculated. A mean of −0.011 and a standard deviation of 0.088 were obtained, showing that the noise introduced to the inputs did not have a large effect on the predicted temperature.

Finally, to simulate the duration and frequency of the fridge openings the post processed data from Section 3.1.4 was used alongside a random number generator to statistically match the amount of openings and their duration to the real ones in function of the hour.

3.3. Control system

3.3.1. Reward

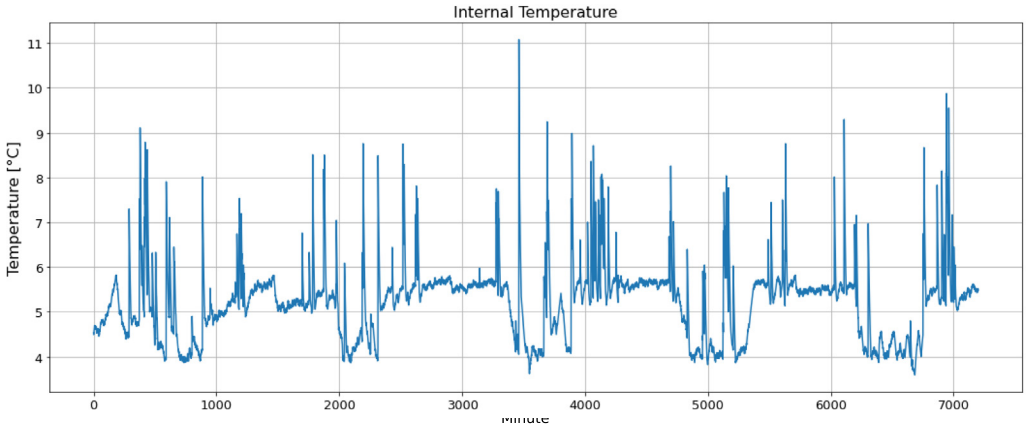
As previously mentioned, the environment in this research is composed of the temperature, power consumption, door state (opened or closed), and power generated by RES. The most important variable to control is the internal temperature, the agent must be able to keep it within a reasonable margin in order to preserve food. 6 °C was selected as the upper temperature limit taking into consideration the World Health Organization suggestion of not exceeding this temperature for residential fridges [39]. The lower limit was selected by looking at the data gathered from the monitoring system, 3.8 °C was selected. The environment was programmed to give the agent a −100 reward in case it went off those limits; and in case it was between the limits, the reward would be 0.

The second variable that has an impact on the reward is the RES generated power. In this case, the idea is to favor turning on the fridge when RES power is available. The environment gives a positive reward proportional to the generated power if the agent turns the fridge on and power is available. In this way, the agent should learn to turn the fridge on when RES power is being generated.

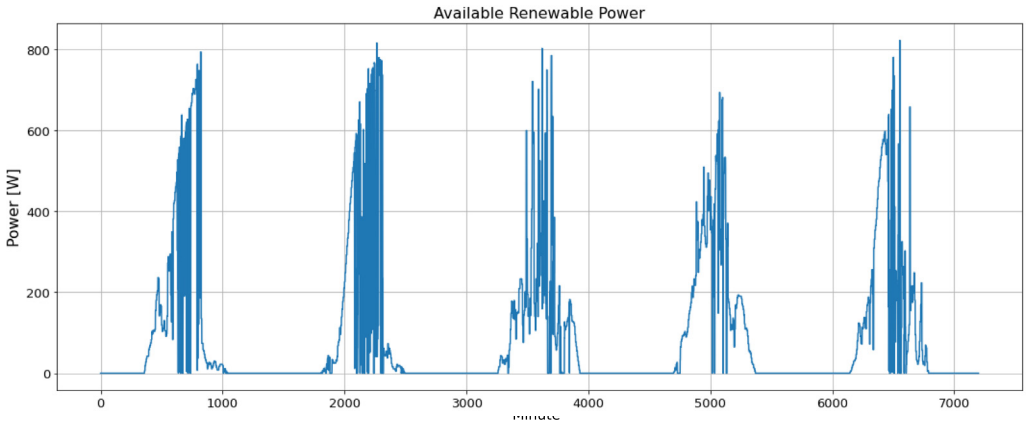
Finally, the agent must be encouraged to not turn on the fridge when it is not necessary. This is, in order to avoid it keeping a low temperature all the time, which would increase the total energy consumption. To do this, a −1 reward is given to the agent every time it decides to turn the fridge on.

After the first agent was trained a problem was found. When the simulation opened the door and the temperature had a spike the agent would get a negative reward even though the temperature spike had nothing to do with the actions it was taking. For this reason, the reward function was further improved by giving a reward of 0 for the next 5 actions after the fridge was opened. Eq. (4) shows the complete reward function, in case multiple conditions are fulfilled the total reward would be the sum of all corresponding rewards.

$$R = \begin{cases} 0, & 3.8 < temp < 6 \\ -100, & temp > 6 \\ -100, & temp < 3.8 \\ RES_{power}/100, & RES_{power} > 50 \ \& \ action = on \\ RES_{power}/50, & RES_{power} > 200 \ \& \ action = on \end{cases} \quad (4)$$



(a) Temperature [°C].



(b) Available RES Power [W].

Fig. 8. Results of the first five days of training.

3.3.2. The RL agent

The first step to implement the actor–critic algorithm is to choose how to parameterize the state-values. The fridge temperature could range from lower than 3.5 °C to higher than 15 °C, and the power being generated by the solar charging station from 0 W to 2 kW. This means that the state space is both continuous and large. Tile coding was selected to construct the feature vector since it is easy to implement on a two dimensional space, and has good generalization and discrimination between states [37]. Tile coding works by overlapping tilings on the state space, then each tiling is divided into smaller tiles. Fig. 7 presents an example where two tilings are divided into 16 tiles, dividing the state space into 32. The figure also illustrates the two active tiles, one per tiling. In the case of this study, a total of 32 tilings, each with 100 tiles were used.

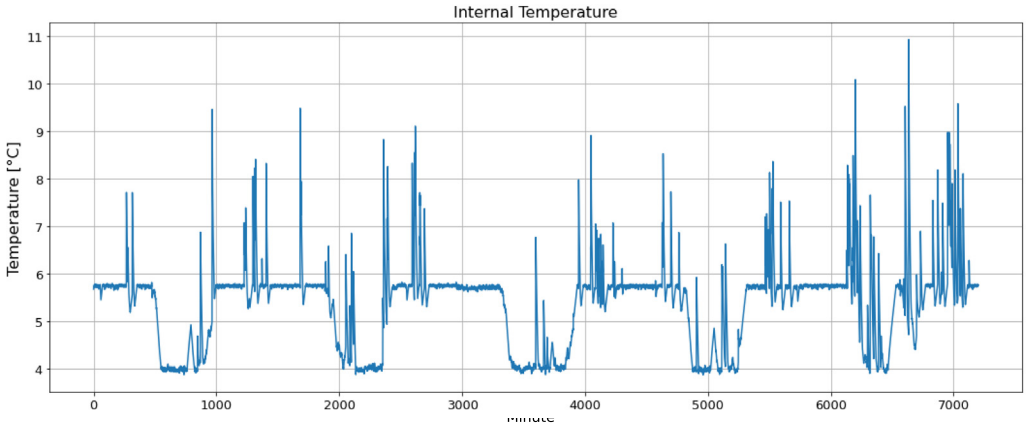
As already stated, the policy parameterization is made using the SoftMax Policy; the feature vector $x(s, a)$ of Eq. (3) is then constructed using the active tiles for the current state. All the inputs for algorithm 1 are now fulfilled. To train the agent, the algorithm was run for 120.000 iterations, which would be equivalent to 120.000 min (83 days).

4. Results

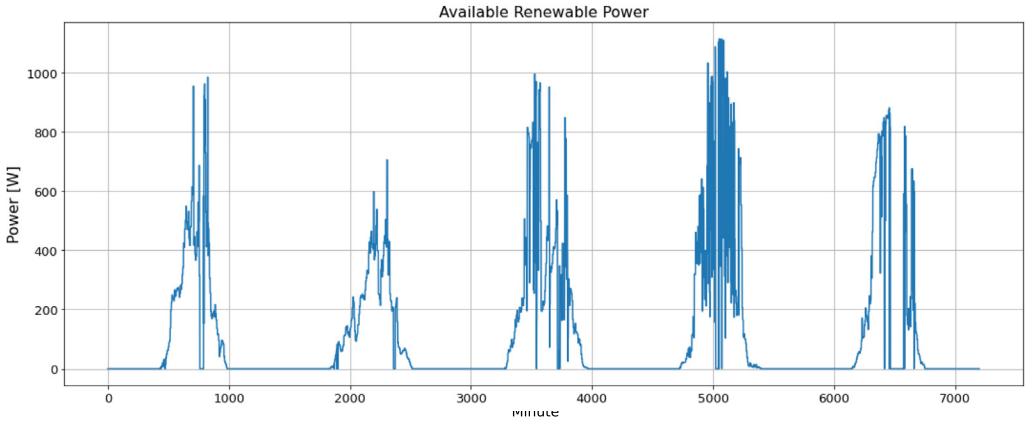
4.1. Training of the RL agent

The results are shown in Figs. 8 and 9. The first figure shows the temperature and the available RES power on the first five days of training. The agent is still learning a good policy, which is the reason for the uneven temperature variations. The second figure shows the last five days of simulation, the agent was able to successfully learn a policy that is able to keep the temperature between the intended range. The figure also demonstrates how the policy takes advantage of the available RES energy, lowering the fridge temperature to its minimum while it is available. With this behavior the agent is using the fridge as a thermal energy battery, using the available RES to keep the cold temperature and avoid using grid energy for the period of time that it takes the temperature to go back to 6 °C, which in average is 90 min. It is important to notice that the temperature spikes that overpass the 6 °C mark are due to the simulated opening of the fridge door. Before continuing with the analysis, the agent's obtained mean temperature was calculated in order to make sure it is comparable to the original control. The agent was capable of obtaining a mean temperature of 5.4 °C, 0.1 °C lower than the original control. This shows that even though the agent produces a different temperature profile, is capable of keeping the same average temperature.

In order to evaluate the performance of the agent in comparison with the original control of the fridge, the total energy used was calculated for both cases. The original control had an energy consumption of 60.5 kWh on the same time frame of the simulation, and of those only 15.2 kWh would have been generated by RES, meaning a 25% of the total energy consumed. In order to calculate this, the following was assumed:



(a) Temperature [°C].



(b) Available RES Power [W].

Fig. 9. Results of the last five days of training.

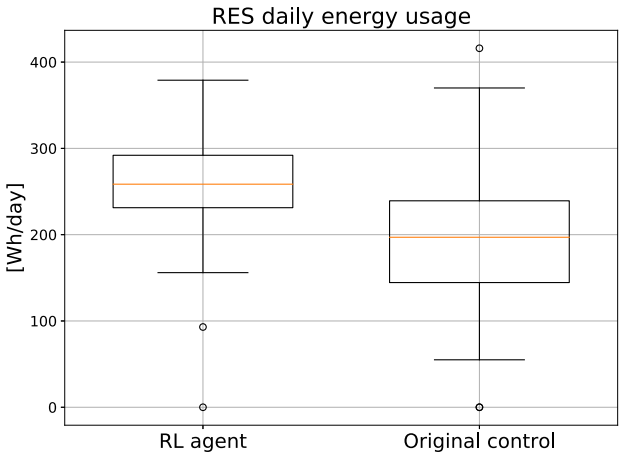


Fig. 10. Box plots for the daily RES usage [Wh/day].

- The mean power consumption of the fridge according to the monitoring system is 60 W, meaning that every minute the fridge is on 1 Wh of energy is used.
- If the power generated by the RES system is higher than 100 W, then it is considered that the fridge is fully running from this source.
- If the power generated by the RES system is lower than 100 W, then it is considered that the fridge is fully running from the conventional power grid.

Table 2
Right tailed t-tests comparing the daily RES usage of the original control and the RL agent control [Wh/day].

Value	RL agent	Original control	*t-value*	*p*
M	289.60	226.47	6.09	<0.001
SD	55.92	86.07		

The trained agent had a total energy consumption of 46.6 kWh, meaning a reduction of 23% compared with the original control. In addition, the agent also increased the amount of RES energy used to 21.2 kWh, 45.5% of the total energy consumed. The improvement on the total energy consumption is possible thanks to the capacity of the learned policy of keeping the temperature on its upper limit for a longer time. This decreases the temperature gradient between the interior and exterior of the fridge, reducing the necessary energy to keep this temperature.

Fig. 10 presents the box plots of the daily RES usage for both the original control and the RL agent. In order to evaluate the obtained results a right tailed t-test was performed ($H_o : \mu_{RL} = \mu_o$, $H_a : \mu_{RL} > \mu_o$, $\alpha = .05$). As shown by the mean, standard deviation and t-test results in Table 2 there is statistical evidence that the RL agent was able to increase the usage of RES energy in comparison with the original control.

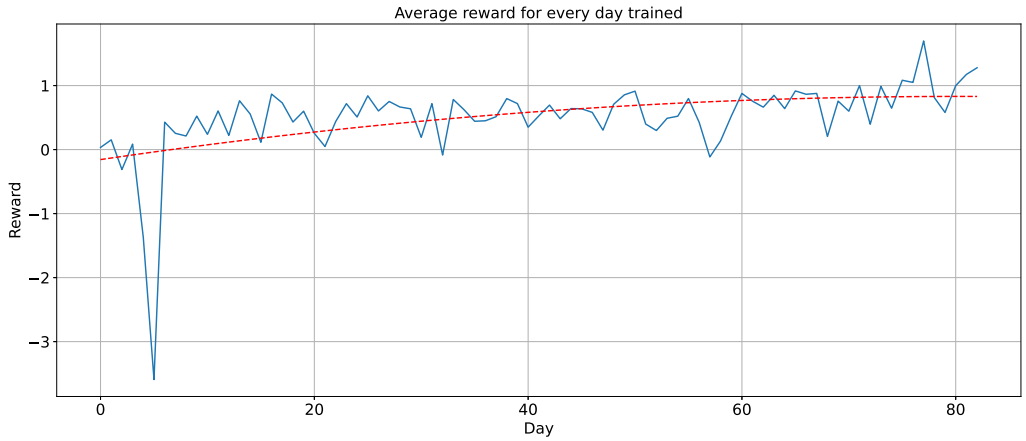


Fig. 11. Obtained daily average reward.

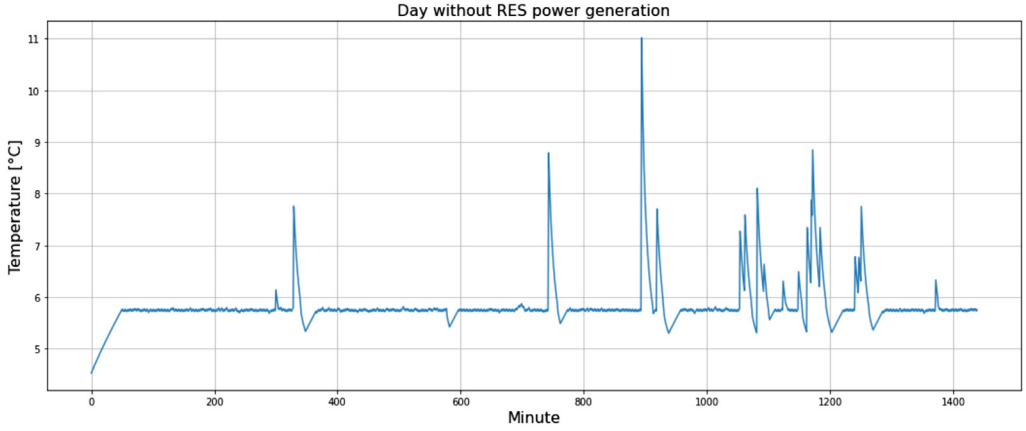


Fig. 12. Internal temperature on a day without any RES power.

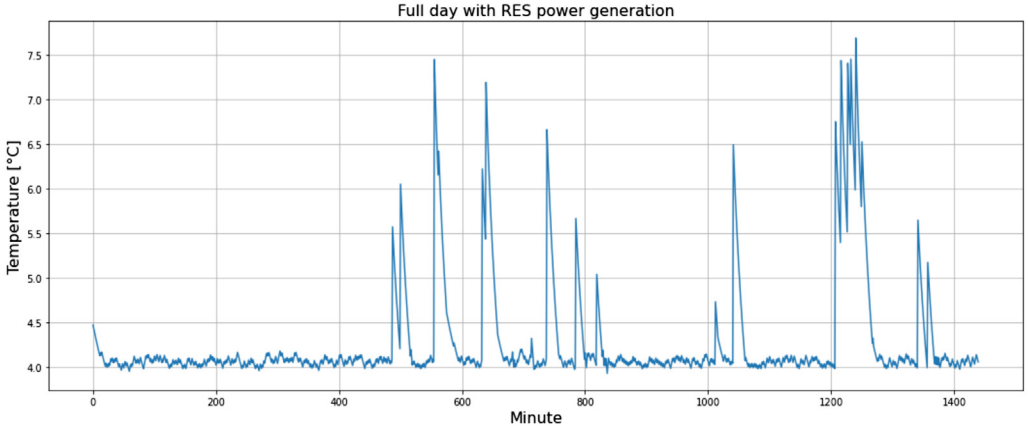


Fig. 13. Internal temperature on a day with continuous RES power.

To evaluate the training process of the agent, the average daily reward is shown in Fig. 11. As expected and shown by the red trend line, the average reward increases as the training process advances.

4.2. Agent behavior in different scenarios

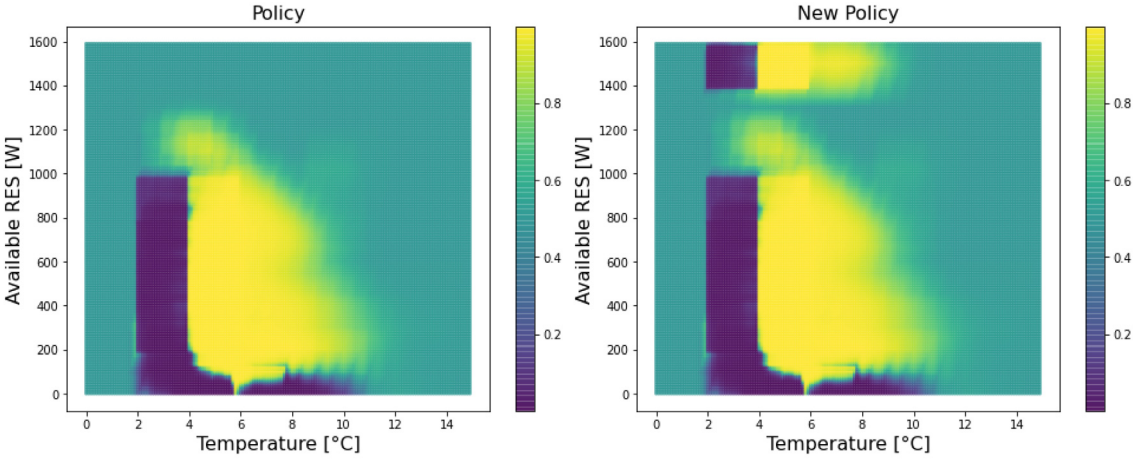
To make sure that the trained agent is able to perform on non-ideal conditions, or conditions that were not present during training, the following three scenarios were evaluated.

4.2.1. Day without RES

A theoretical day without any available RES energy was tested. Fig. 12 shows the results of this test, as expected, the agent kept the temperature close to the upper limit, waiting for available RES to lower the temperature.

4.2.2. Day with RES available for 24 h

For this, the agent is faced with a theoretical day where RES energy is available for 24 h. This could happen, for example, if additional to the solar charging station, a wind turbine was added



(a) Policy after initial training. (b) Policy after new data is presented to the agent.

Fig. 14. Policy change with new data.

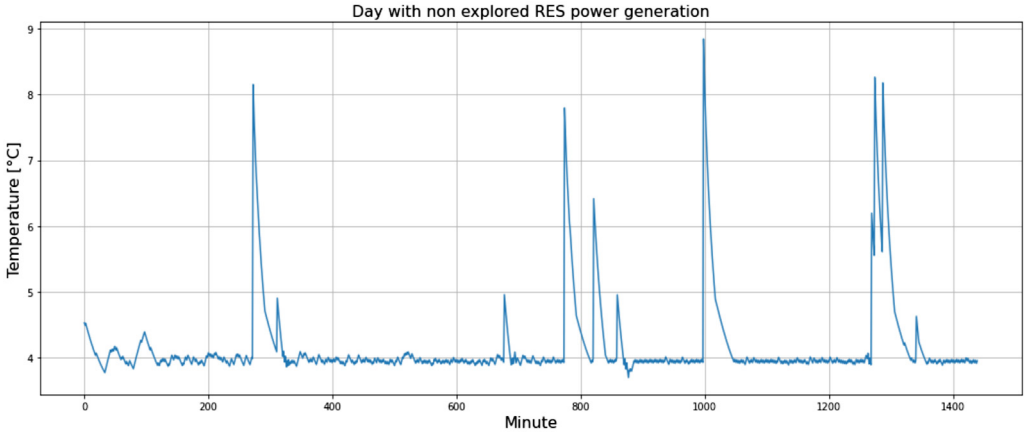


Fig. 15. Internal temperature on a day with non-visited states.

to the system to supply energy at night. As shown in Fig. 13, the agent keeps the temperature at the lower limit for the entire day. Doing this, ensures that the fridge uses the maximum amount of RES energy.

4.2.3. Day with non-visited states

Finally, the agent was tested on a theoretical day where the RES available power was higher than the highest on the original training data. The highest power available while training was 1310 W. For this test, a constant power of 1500 W was presented to the agent during the theoretical day. Fig. 14(a) shows the policy learned by the agent on the original training. The colors represent the probability of turning the fridge on for any given state. This figure is also a good representation of the close to deterministic nature of the SoftMax policy. The states that were visited are either completely yellow or purple, but notice that 0 and 1 are not the limits of the legends. As expected, when the available RES power is 1500 W then the policy is not defined (the probability is 0.5, meaning a random choice). Fig. 14(b) presents the new policy after the agent was presented with the theoretical day. The policy was updated and now is able to properly select actions when the RES power is close to 1500 W. Finally, Fig. 15 shows the temperature behavior during the test. During the first 200 min, as expected, the temperature was less stable as the agent was still learning a policy for the non explored states. However, in the last 200 min, the agent had already learned the policy, and thus, the temperature is kept stable at lower values (see Fig. 14).

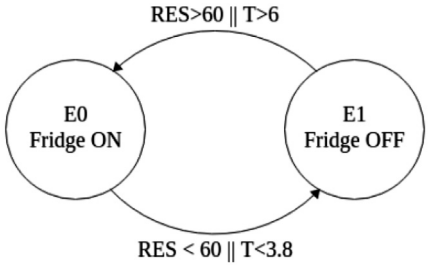


Fig. 16. Finite state machine control.

4.3. Agent behavior against standard control models

A finite state machine (FSM) controller was formulated in order to compare this traditional approach to the trained agent. As shown in Fig. 16 the machine consists of two states (on and off) and two transitions. The FSM, just like the agent, is designed to keep the temperature between 3.8 °C and 6 °C. It also takes into consideration the available RES in order to maximize its usage.

Fig. 17 shows the temperature profile accomplished by both control methods when evaluated for five days under the same environment (same RES and same fridge openings). Even though the temperature profiles are similar and both had the same mean

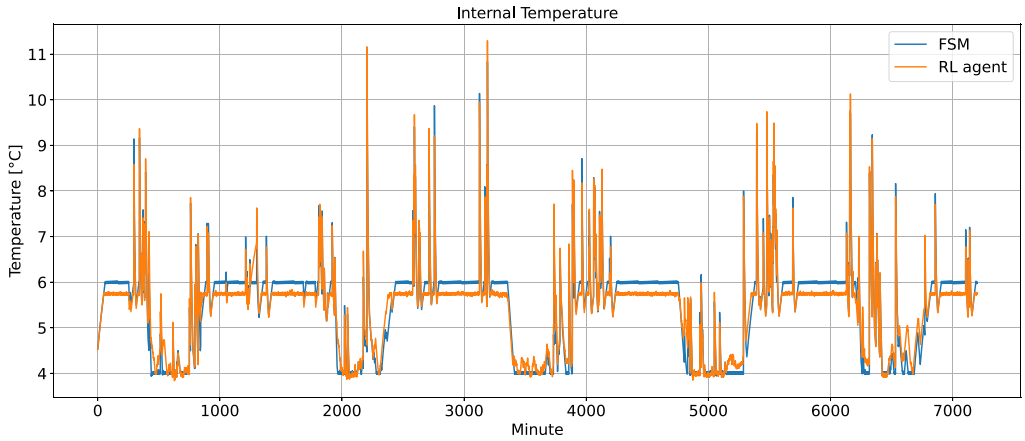


Fig. 17. Temperature comparison between finite state machine and RL agent control.

temperature (5.4 °C), the RL agent used less energy than the FSM to accomplish these temperatures with 2.85 kWh and 3.27 kWh respectively, equivalent to 12.84% less energy.

5. Conclusions

A fridge energy management system consisting of three different components was proposed with the objective of maximizing the consumption of energy generated by RES and reducing the total energy usage. The first component is a monitoring system, which supervises a fridge in order to obtain data on its temperature, power consumption and door state. For the second component, the collected data is used to create a model of the fridge. This model uses a predictive model to determine the future temperature of the fridge. Lastly, the model is used to train a RL agent that is in charge of controlling the fridge.

The trained agent was able to learn a policy that kept the internal temperature in the correct range while making sure to use as much RES energy as possible. The original control of the fridge took only 25% of its energy from RES, in contrast, our RL agent managed to take 45.5% from these sources, an increase of 82%. In addition to this, the agent successfully reduced the total amount of energy used by the simulated fridge by 23%. These results show good potential for the use of RL in this kind of control application. Additionally the agent was also tested in three different scenarios that had not been considered while training: a day without RES, a day with 24 h of available RES, and a day with states not presented during training. The agent excelled in these tests since it was able to keep the fridge temperature on the accepted margins, and was capable of learning and further improving its policy from the new experience. Finally the agent was tested against a finite state machine control, the RL method was able to reduce energy consumption by 8.3% while also reducing the mean temperature of the fridge by 0.26 °C.

There are also limitations to the presented work. The agent was trained to work on fixed speed compressors meaning that newer fridge models that work with variable speed compressors would not be able to be properly controlled. Additionally, the ANN was trained using data that would not be representative for countries located further from the equator line, where weather does change throughout the year. Finally, atypical situations where not taken into consideration while training the agent, for example if the fridge was unintentionally kept opened for long periods of time the agent would try to lower the temperature by keeping the compressor on even though the temperature would not be able to drop, making the fridge use energy unnecessarily.

As for future work, the authors consider that there is still improvement that can be made to the approach in order to get superior results. For example, a continuous action RL algorithm, where the agent would be able to not only turn the fridge on and off but also select for how long the fridge should stay in that state, could be more suited for a real-world implementation of the proposed control system. Secondly, the control system could be more complete and obtain better performance if an energy storage system would be integrated, this way, the surplus energy could be stored and used when RES power is not available. Finally, the energy management system must be implemented on a real fridge to test its behavior and real-life viability.

CRediT authorship contribution statement

Juan Pablo Giraldo-Pérez: Conceptualization, Methodology, Investigation, Software, Visualization, Writing – original draft. **Ricardo Mejía-Gutiérrez:** Conceptualization, Validation, Supervision. **Jose Aguilar:** Conceptualization, Validation, Writing – review & editing.

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

Data will be made available on request.

Acknowledgments

Authors would like to thank Universidad EAFIT and the alliance “ENERGETICA 2030", which is a Research Program, with code 58667 from the “*Colombia Científica*" initiative, funded by The World Bank through the call “778-2017 Scientific Ecosystems”. The research program is managed by the Colombian Ministry of Science, Technology and Innovation (Minciencias) with contract No. FP44842-210-2018.

References

[1] P.A. Owusu, S. Asumadu-Sarkodie, A review of renewable energy sources, sustainability issues and climate change mitigation, Cogent Eng. 3 (1) (2016) 1–14.
[2] G. DNV, Energy transition outlook 2017, 2017, DNV GL, Høvik, Norway.

[3] UPME, Plan energético nacional 2020- 2050, 2019, https://www1.upme.gov.co/DemandaEnergetica/PEN_documento_para_consulta.pdf. (Online; Accessed 07 April 2022).
[4] C. Voyant, G. Notton, S. Kalogirou, M.L. Nivet, C. Paoli, F. Motte, A. Fouilloy, Machine learning methods for solar radiation forecasting: A review, Renew. Energy 105 (2017) 569–582, http://dx.doi.org/10.1016/j.renene.2016.12.095.
[5] K. Mukilan, K. Thaiyalnayaki, Y.D. Dwivedi, J. Samson Isaac, A. Poonia, A. Sharma, E.A. Al-Ammar, S.M. Wabaidur, B.B. Subramanian, A. Kassa, Prediction of rooftop photovoltaic solar potential using machine learning, Int. J. Photoenergy 2022 (2022) 1–8, http://dx.doi.org/10.1155/2022/1541938.
[6] S. Kamal, P.S. Ramapraba, A. Kumar, B.C. Saha, M. Lakshminarayana, S. Sanal Kumar, A. Gopalan, K.G. Erko, Optimization of solar panel deployment using machine learning, Int. J. Photoenergy 2022 (2022) 1–7, http://dx.doi.org/10.1155/2022/7249109.
[7] L. Zhang, M. He, Prediction of solar cell materials via unsupervised literature learning, J. Phys. Condens. Matter 34 (9) (2022) http://dx.doi.org/10.1088/1361-648X/ac3e1e.
[8] M. Papinutto, R. Boghetti, M. Colombo, C. Basurto, K. Reutter, D. Lalanne, J.H. Kämpf, J. Nembrini, Saving energy by maximising daylight and minimising the impact on occupants: An automatic lighting system approach, Energy Build. 268 (2022) 112176, http://dx.doi.org/10.1016/j.enbuild.2022.112176.
[9] A. Vij, S. Vijendra, A. Jain, S. Bajaj, A. Bassi, A. Sharma, IoT and machine learning approaches for automation of farm irrigation system, Procedia Comput. Sci. 167 (2019) (2020) 1250–1257, http://dx.doi.org/10.1016/j.procs.2020.03.440.
[10] I. Yazici, O.F. Beyca, D. Delen, Deep-learning-based short-term electricity load forecasting: A real case application, Eng. Appl. Artif. Intell. 109 (December 2020) (2022) 104645, http://dx.doi.org/10.1016/j.engappai.2021.104645.
[11] M.L. Abdulrahman, K.M. Ibrahim, A.Y. Gital, F.U. Zambuk, B. Ja'Afaru, Z.I. Yakubu, A. Ibrahim, A review on deep learning with focus on deep recurrent neural network for electricity forecasting in residential building, Procedia Comput. Sci. 193 (2021) 141–154, http://dx.doi.org/10.1016/j.procs.2021.10.014.
[12] E.U. Haq, X. Lyu, Y. Jia, M. Hua, F. Ahmad, Forecasting household electric appliances consumption and peak demand based on hybrid machine learning approach, Energy Rep. 6 (2020) 1099–1105, http://dx.doi.org/10.1016/j.egyr.2020.11.071.
[13] A. Althnian, D. AlSaeed, H. Al-Baity, A. Samha, A.B. Dris, N. Alzakari, A. Abou Elwafa, H. Kurdi, Impact of dataset size on classification performance: An empirical evaluation in the medical domain, Appl. Sci. (Switzerland) 11 (2) (2021) 1–18, http://dx.doi.org/10.3390/app11020796.
[14] J. Prusa, T.M. Khoshgoftaar, N. Seliya, The effect of dataset size on training tweet sentiment classifiers, in: Proceedings - 2015 IEEE 14th International Conference on Machine Learning and Applications, ICMLA 2015, IEEE, 2016, pp. 96–102, http://dx.doi.org/10.1109/ICMLA.2015.22.
[15] O. Kotevska, K. Kurte, J. Munk, T. Johnston, E. McKee, K. Perumalla, H. Zandi, RL-HEMS: Reinforcement learning based home energy management system for HVAC energy optimization, ASHRAE Trans. 126 (2020) 421–429.
[16] X. Deng, Y. Zhang, H. Qi, Towards optimal HVAC control in non-stationary building environments combining active change detection and deep reinforcement learning, Build. Environ. 211 (August 2021) (2022) http://dx.doi.org/10.1016/j.buildenv.2021.108680.
[17] Z. Jiang, M.J. Risbeck, V. Ramamurti, S. Murugesan, J. Amores, C. Zhang, Y.M. Lee, K.H. Drees, Building HVAC control with reinforcement learning for reduction of energy cost and demand charge, Energy Build. 239 (2021) 110833, http://dx.doi.org/10.1016/j.enbuild.2021.110833.
[18] E.U. Haq, C. Lyu, P. Xie, S. Yan, F. Ahmad, Y. Jia, Implementation of home energy management system based on reinforcement learning, Energy Rep. 8 (2022) 560–566, http://dx.doi.org/10.1016/j.egyr.2021.11.170.
[19] D. Azuatalam, W.L. Lee, F. de Nijs, A. Liebman, Reinforcement learning for whole-building HVAC control and demand response, Energy AI 2 (2020) 100020, http://dx.doi.org/10.1016/j.egyai.2020.100020.
[20] Y. Du, H. Zandi, O. Kotevska, K. Kurte, J. Munk, K. Amasyali, E. Mckee, F. Li, Intelligent multi-zone residential HVAC control strategy based on deep reinforcement learning, Appl. Energy 281 (October 2020) (2021) 116117, http://dx.doi.org/10.1016/j.apenergy.2020.116117.
[21] K. Mason, S. Grijalva, A review of reinforcement learning for autonomous building energy management, Comput. Electr. Eng. 78 (2019) 300–312, http://dx.doi.org/10.1016/j.compeleceng.2019.07.019.
[22] S. Lee, D.H. Choi, Reinforcement learning-based energy management of smart home with rooftop solar photovoltaic system, energy storage system, and home appliances, Sensors (Switzerland) 19 (18) (2019) http://dx.doi.org/10.3390/s19183937.
[23] J. Sang, H. Sun, L. Kou, Deep reinforcement learning microgrid optimization strategy considering priority flexible demand side, Sensors 22 (6) (2022) http://dx.doi.org/10.3390/s22062256, URL https://www.mdpi.com/1424-8220/22/6/2256.
[24] M. Ren, X. Liu, Z. Yang, J. Zhang, Y. Guo, Y. Jia, A novel forecasting based scheduling method for household energy management system based on deep reinforcement learning, Sustainable Cities Soc. 76 (2022) 103207, http://dx.doi.org/10.1016/j.scs.2021.103207.
[25] J. Aguilar, A. Garces-Jimenez, M. R-Moreno, R. García, A systematic literature review on the use of artificial intelligence in energy self-management in smart buildings, Renew. Sustain. Energy Rev. 151 (2021) 111530, http://dx.doi.org/10.1016/j.rser.2021.111530.
[26] K. Stepanovic, J. Wu, R. Everhardt, M. de Weerdt, Unlocking the flexibility of district heating pipeline energy storage with reinforcement learning, Energies 15 (9) (2022) 1–26, http://dx.doi.org/10.3390/en15093290.
[27] P. Stanfel, K. Johnson, C.J. Bay, J. King, Proof-of-concept of a reinforcement learning framework for wind farm energy capture maximization in time-varying wind, J. Renew. Sustain. Energy 13 (4) (2021) 043305, http://dx.doi.org/10.1063/5.0043091.
[28] S. Jo, C. Jong, C. Pak, H. Ri, Multi-agent deep reinforcement learning-based energy efficient power allocation in downlink MIMO-noma systems, IET Commun. 15 (12) (2021) 1642–1654, http://dx.doi.org/10.1049/cmu2.12177, arXiv:https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/cmu2.12177.
[29] C. Corpoema-Cusa, Caracterización energética del sector residencial urbano y rural en Colombia, 2012, http://bdigital.upme.gov.co/handle/001/1111. (Online; Accessed 06 June 2022).
[30] S. Nesmachnow, G. Colacurcio, D.G. Rossit, J. Toutouh, F. Luna, Optimizing household energy planning in smart cities: A multiobjective approach, Rev. Fac. Ingeniería Univ. Antioquia (101) (2020) 8–19, http://dx.doi.org/10.17533/udea.redin.20200587.
[31] M. Shibl, L. Ismail, A. Massoud, Electric vehicles charging management using machine learning considering fast charging and vehicle-to-grid operation, Energies 14 (19) (2021) http://dx.doi.org/10.3390/en14196199.
[32] S. Lee, D.H. Choi, Energy management of smart home with home appliances, energy storage system and electric vehicle: A hierarchical deep reinforcement learning approach, Sensors (Switzerland) 20 (7) (2020) http://dx.doi.org/10.3390/s20072157.
[33] F. Alfaverh, M. Denai, Y. Sun, Demand response strategy based on reinforcement learning and fuzzy reasoning for home energy management, IEEE Access 8 (2020) 39310–39321, http://dx.doi.org/10.1109/ACCESS.2020.2974286.
[34] D.L. McCollum, Machine learning for energy projections, Nat. Energy 6 (2) (2021) 121–122, http://dx.doi.org/10.1038/s41560-021-00779-9.
[35] S. Bourhnane, M.R. Abid, R. Lghoul, K. Zine-Dine, N. Elkamoun, D. Benhaddou, Machine learning for energy consumption prediction and scheduling in smart buildings, SN Appl. Sci. 2 (2) (2020) 1–10, http://dx.doi.org/10.1007/s42452-020-2024-9.
[36] N.D. Andritsos, V. Stasinou, D. Tserolas, E. Giaouris, Temperature distribution and hygienic status of domestic refrigerators in Lemnos island, Greece, Food Control 127 (February) (2021) 108121, http://dx.doi.org/10.1016/j.foodcont.2021.108121.
[37] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, MIT Press, 2018.
[38] F. Tony Liu, K. Ming Ting, Z.-H. Zhou, Isolation forest ICDM08, 2008, Icdm, URL https://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/icdm08b.pdf{%}0Ahttps://cs.nju.edu.cn/zhouzh/zhouzh.files/publication/icdm08b.pdf?q=isolation-forest.
[39] A. Ovca, T. Škufca, M. Jevšnik, Temperatures and storage conditions in domestic refrigerators - Slovenian scenario, Food Control 123 (2021) http://dx.doi.org/10.1016/j.foodcont.2020.107715.