# Recent Advances in Reinforcement Learning Applications for Building Energy Management: A Mini Review.

Shaqour, Ayas
Interdisciplinary Graduate School of Engineering Sciences, Kyushu University

Hagishima, Aya
Interdisciplinary Graduate School of Engineering Sciences, Kyushu University

https://doi.org/10.5109/5909098

出版情報：Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES). 8, pp.239-245, 2022-10-20. 九州大学大学院総合理工学府
バージョン：
権利関係：

# Recent Advances in Reinforcement Learning Applications for Building Energy Management: A Mini Review.

Ayas Shaqour[1*], Aya Hagishima[1]
[1]Interdisciplinary Graduate School of Engineering Sciences, Kyushu University, Japan.
*Corresponding author email: ayasshaqour@kyudai.jp

**Abstract:** *In 2019, buildings accounted for 55% of the global electricity demand, making them a key contributor to global emissions and a core target for energy efficiency, energy reduction, and policies and measures promoting renewable energy usage. Reinforcement learning (RL) is an agent-based modelling technique that has proven successful in many applications, particularly in artificial intelligence. RL has attracted research attention owing to its utilization in building energy management (BEM) applications. In this work, the latest research advances that utilize this method are investigated and discussed, primarily its usage in modelling complex building energy problems, building energy consumption control, optimization for comfort and cost savings, and the enhancement of demand forecasting algorithms. Furthermore, the combination of RL with other deep learning methods is discussed. As a state-of-the-art technology in smart grid building applications, RL is applied for control purposes and forecasting enhancement.*

**Keywords:** Building Energy Demand; Deep reinforcement learning; Energy consumption prediction; Energy efficiency; Energy Management.

## 1. INTRODUCTION

In 2019, the building sector accounted for 55% of the global electricity demand [1], accounting for almost 38% of global greenhouse gas emissions. Thus, the reduction of such high demands and an increase in renewable energy through various policies and technologies has become an urgent issue[2]. Building energy demand is a multidimensional problem that can be approached considering different intricate paradigms listed below:

1- Improvement of building thermal performance determined by materials and design of a building [3].
2- Energy efficiency improvement of appliances and facilities [4]
3- Changes in energy-related occupant behaviors through various interventions, including market-based pricing policies and the demand response approach[5][6].
4- Renewable energy integration [7].
5- Smart buildings incorporated with AI-based energy management and optimization [8].

To overcome these challenges, researchers have deployed the latest advancements in artificial intelligence (AI) and machine learning fields, particularly to tackle the building energy challenge from the macro-country levels to micro-building levels[9]. The ML approach can be classified into three main categories: The first is supervised learning, which maps the relationship between dependent variables and the target variable. It can be used for numerous building energy applications, such as predicting human behavior, risks, energy demand, and renewable energy generation, as well as identifying equipment types and activities[10][11][12][13]. The second is unsupervised learning, which is primarily used for autonomous pattern recognition in buildings and can be useful for identifying energy lifestyle patterns, consumer types, and anomaly detection[14][15]. The area of smart building energy management (BEM) systems, because of their innovative features compared to conventional control and forecasting methods. Hence, this article presents a brief overview of the methodology of RL and its recent applications in BEM. Section 2 presents a discussion on the basic methodology for RL and deep RL (DRL) and their varieties. Section 3 introduces the most recent research on RL/DRL applications for building energy control and forecasting, and Section 4 presents the conclusions of the paper. Third is reinforcement learning (RL) algorithms that have recently attracted research attention, particularly in the

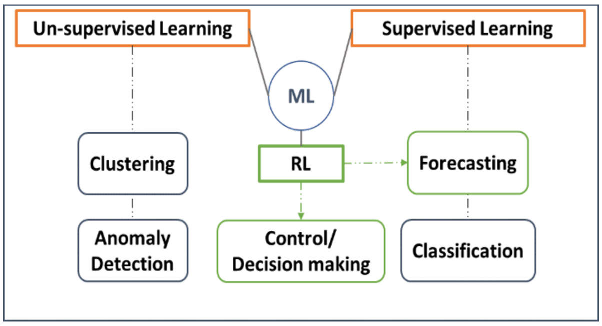

Fig. 1. Major applications of ML methods in Building energy Demand Management

## 2. REINFORCEMENT LEARNING OVERVIEW AND FORMULATION

The RL methodology is based on optimal control theory and first emerged in the 1950s. It was utilized to formulate system control problems in which the target system behavior variable was reduced with time[16]. The Markov decision processes (MDPs) introduced by Bellman constitute a core part of the RL theory [17]and MDPs were introduced to formulate problems related to optimal control. One of the most important characteristics of RL is that it can operate with or without a system model (model-based vs. model-free), which is a significant advantage over conventional control methodologies[18]. Recently, model-free RL has been gaining increasing attention. It can be applied to complex systems whose dynamics are too difficult to capture and model[19]. One of the recent breakthrough applications of RL and DL is the deep mind model applied to Google’s data center. Following its application, the cooling costs were reduced by 40%, and their proposed model-free solution had the following advantages over conventional methods[20]:

1. The nonlinear, intricate, and complex nature of the data center environment and its equipment makes engineering formulations of such systems unfeasible.
2. A large number of scenarios that could occur in such environments, both internally and externally, hinder the speed of adapting to such static-engineering solutions.
3. The fact that different datacenters with different configurations would require specific fine-tuned models, where RL and DL solutions are more generalizable.

This is one example of how such methods can be leveraged to reduce costs and increase energy efficiency in different types of building settings.

## 3. RL FORMULATION

The core aspects of RL include the environment, agents, and their interactions, as shown in **Error! Reference source not found.**. This figure depicts how agents at each time step in the real world observe a state ($s_t$) from their surroundings, take action ($a_t$) according to a certain goal (policy), and finally receive feedback (reward $r_t$) based on how the state changes. With sufficient experience, the RL agent learns the best strategy (optimal policy) to navigate the environment and maximize its rewards[19]. At the core of any RL lies the formalization of MDPs, which entails the Markov property, which states that the transitions are only based on recent states and actions and are irrelevant to any prior history. The MDP is a tuple of five elements $\langle S, A, R, P, \rho_0 \rangle$; with $S$ and $A$ being the set of all valid states and actions, respectively; $R$ represents the reward function, where at any moment in time, $r_t$ is defined as a function of state and action $R(s_t, a_t, s_{t+1})$, $P$ represents the transition probability function that defines the probability of transitioning into state ($s'$) starting at state $s$ and performing action $a$ $P(s'|s, a)$, and finally $\rho_0$ indicates the initial state distribution that is used to define the first state of the problem. A transition of states $s_{t+1}$ is based on the model of the environment that might be unknown to the agent, and the last action $a_t$ is based on the agent's policy, which can either be deterministic $f(s_t, a_t)$ or stochastic $P(.|s_t, a_t)$[19]. First, based on a certain episode, $\tau$ is defined as a sequence of actions and states in the world $(s_0, a_0, s_1, a_1, \ldots)$, and the cumulative reward over that episode can be defined in several ways, such as a finite-horizon (fixed T) undiscounted return:

$$R(\tau) = \sum_{t=0}^{T} r_t \qquad (1)$$

or an infinite-horizon discounted return:

$$R(\tau) = \sum_{t=0}^{\infty} \gamma^t r_t \qquad (2)$$

where $\gamma \in (0,1)$ indicates a discount factor for convergence purposes over infinite sums, bias towards early rewards, and other mathematical reasons. For policy $\pi$ and return $R(\tau)$ over T steps in time, the expected return $J(\pi)$ of the RL problem is formulated as

$$\boldsymbol{J(\pi) = \int_{\tau} P(\tau \mid \pi) R(\tau) = \underset{\tau\sim\pi}{\mathrm{E}}[R(\tau)]} \qquad (3)$$

where $\boldsymbol{P(\tau \mid \pi)}$ indicates the transition probability of a T-step sequence of state/action $\boldsymbol{\tau}$ and the principal optimization challenge in a RL paradigm is to find the optimal policy $\boldsymbol{\pi}^*$ that maximizes the expected return. The right side of Equation 3 contains a common abbreviation for the expected return depicted by $\mathbf{E}[\boldsymbol{R(\tau)}]$, where $(\boldsymbol{\tau \sim \pi})$, meaning that $\boldsymbol{\tau}$ follows a policy $\boldsymbol{\pi}$.

### 3.1 RL Variants and Deep RL

There are many variants of RL that are optimized for different problems, the two main ones include [21]:

1. Model-based: either the model of the environment is given or learn the model (approximate environment transitions). Model-based approaches are more sample efficient (time steps of ~ 100 to converge).
2. Model-Free: either value-based (approximate Q-function), policy-based (approximate policy), or a combination of both (approximate policy-value function). Value-based RL, such as Q-learning, is more sample efficient (time steps ~ 1M) than policy-based approaches (time steps ~ +10M).

Deep RL (DRL) is a branch of RL algorithms that utilizes deep learning algorithms, such as feedforward neural networks, convolutional neural networks, or recurrent neural networks, to approximate either the environment transitions, Q-function, or policy-value function [22]. An example of a popular DRL is deep Q-learning (DQN) [23], which is based on a Q-learning algorithm, as shown in Fig. 2. In Q-learning, the quality (estimated return) of a pair of action states is stored in a lookup table and is based on the estimation of the Bellman optimality equation [21,24] with a Q-function:

$$\boldsymbol{Q_{t+1}(s_t, a_t) = Q_t(s_t, a_t) + \alpha\left[r_t + \gamma \underset{a}{max} Q_t(s_{t+1,} a) - Q_t(s_t, a_t)\right]} \qquad (4)$$

where $\mathbf{Q_{t+1}}$ indicates the estimated new Q-value based on an action $\mathbf{a_t}$ taken in state $\mathbf{s_t}$, $\boldsymbol{\alpha}$ indicates the learning rate, $\boldsymbol{\gamma}$ indicates the discount factor, and $\boldsymbol{\underset{a}{max} Q_t(s_{t+1,} a)}$ denotes the maximum reward that can be gained in the new state $\mathbf{s_{t+1}}$ by considering the best action $\boldsymbol{a}$. In a DQN, the Q-function is approximated using any deep learning algorithm, as shown in Fig. 2. DQN is one of the many variants of DRL that utilizes deep learning algorithms to approximate policy/value functions, which is extremely useful in cases where the state-action space is too large or continuous, thus increasing the difficulty of obtaining its Q-table in the case of Q-learning as an example.

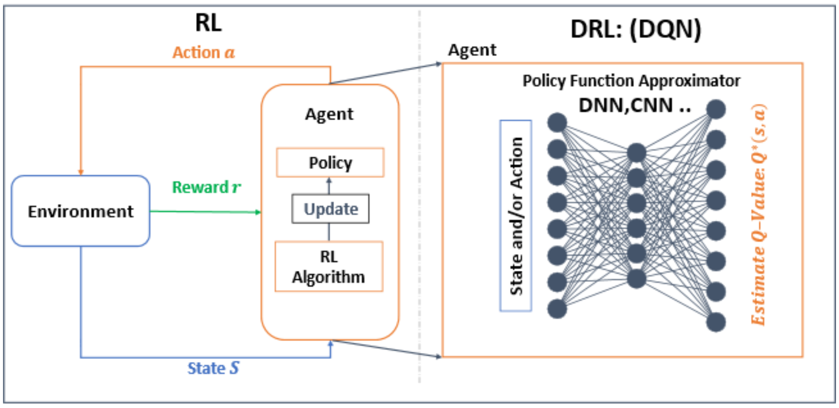

Fig. 2. RL and DRL overview

RL and DRL algorithms can be further classified based on their action spaces as discrete (DQN), continuous (deep deterministic policy gradient (DDPG) [25], reinforcement [26]), or both (proximal policy optimization (PPO)) [27]. The type of action space is a crucial characteristic that can dictate the selection of the RL algorithm used, which is determined by the target application, particularly in BEM systems.

## 4. RECENT APPLICATION OF RL AND DRL IN BUILDING ENERGY MANAGEMENT

Recently, there have been applications of RL and DRL algorithms for BEM. In general, the main goal of these algorithms is decision making, where the decision (action) can be a control signal, control threshold, algorithm selection, human behavior, rescheduling choices, etc., for the many elements that exist in the building premise. First, the challenges of conventional BEM are briefly discussed in comparison to RL BEM. Second, the latest applications and research are introduced.

### 4.1 Conventional vs RL based BEM: challenges and opportunities.

Learning and adaptation is a key characteristic of RL; hence, such functionalities can be adapted and evolved with time and usage, which can provide an advantage over static conventional methods [11]. One such case is conventional ML-based forecasting, which can be enhanced with an added layer of learning beyond its weights and biases, as discussed in subsection 3.1. Moreover, they can overcome the limitations of conventional proportional-integral-derivative (PID) feedback-based control systems that rely on standard, static, and heuristic rules, such as the ASHRAE Guideline 36. PID-based control systems also suffer from the limitation of integrating future predictions in the control loop, which can be overcome using model predictive control (MPC) systems [28]. MPCs have a predictive aspect “P” and are based on identifying the model “M” that characterizes different types of dynamics within a building environment. Models for building applications are not easily generalizable relative to other applications of MPC, such as car and airplane applications, primarily because buildings can have several unique designs. Building energy modelling has been commonly performed based on detailed physics-based models using tools, such as EnergyPlus, which can consume a relatively large simulation time (t > ~ sec/min), scaling up with larger and more complex building types [29]. Although such modelling is typically performed for constructing building design standards and pre-optimizing building systems, it might not be optimal for real-time BEM with a very high response time (t < ~ s) owing to the complex physics-based modelling approach. In buildings, the ubiquitous challenges in both modelling and control decisions are usually intricate and feed into one another, as discussed by Yu et al.[8], which can be summarized as follows:

- **Real Time Modelling:** Constructing accurate and efficient real-time thermal dynamics models for buildings is challenging because of their complex and stochastic elements[30]. Complexity is caused by the many elements and parameters that must be fed into the modelling approaches. Stochasticity exists due to uncertainty concerning extrinsic elements in modern buildings, such as renewable energy generation, dynamic energy pricing, temperature, and intrinsic elements, such as occupancy, HVAC schedule, and indoor temperatures [31].
- **Multi-objective optimization:** The subsystems existing in smart buildings, such as cooling, heating, and energy storage, have temporal and spatial operational constraints, where the control decision of one subsystem can affect the future decisions of all others, thereby increasing the complexity of managing and coordinating decision making [32]. Furthermore, based on efficient computational complexity in terms of both space and time, it is challenging to implement real-time BEM in conventional control methods when dealing with large solution spaces, particularly for large-scale BEM [33].
- **Generalization:** Difficulty in engineering standard and generalizable solutions applicable to BEM problems under varying environments and building designs/configurations using conventional modelling and control approaches[34].

Through online learning and other methods, such as transfer learning, model-free RL can overcome the need for complex building models, while also having the potential for being generalizable. Furthermore, multi-objective optimization can be efficiently solved by designing appropriate reward functions [8]. Finally, most of the existing research has focused on simulation environments, which is discussed in Tables 1 and 2, and only a few studies have focused on real implementation and validation, as discussed by Wang and Hong[28], in which only 11% of the studies were conducted in real buildings. Real implementations face challenges related to the requirement of large amounts of data and training time, security issues, and the need for transfer learning to achieve better generalization[28].

### 4.2 Building energy demand forecasting

High-performing forecasts of future energy demands are crucial to achieving optimal operation as well as supply demand balance, particularly when renewable energy is integrated, or demand response programs are active. Table 1 lists the recent applications of a variety of RL/DRL algorithms to increase the performance of prediction. While recent research in this area is very limited, as RL is primarily used for decision-making control-based problems, which is evident from the table, researchers are combining RL with the DL algorithms to increase the prediction performance by:

- Training the RL agent to select the best model at each prediction horizon from a set of available models based on a certain context.
- Training the RL agent to change the hyperparameters in real time to obtain an adaptive DL model, specifically when the training data exhibits high variation.
- An autoencoder is used to improve the state-space representation and then a DRL is used to directly predict the energy demand.

A new and interesting perspective is presented on the integrations of these methods, which particularly relates to the first two points, that is, RL can make the DL algorithms more dynamic and evolve in real-time on a higher dimension than learnable parameters. This shows the potential for exploring the integration of RL with other DL algorithms, not only for prediction purposes, as discussed, but also for classification or unsupervised approaches. For many prediction applications, DDPG is the choice of DRL algorithm because of its continuous action space, which is suitable for predicting continuous variables, such as energy demand, and for having high accuracy, as it requires higher computational resources. Finally, the direct usage of DRL is being tested for prediction applications and DRL might be able to outperform other ML methods, as observed in Table 1.

### 4.3 Building energy demand control

Smart and direct management of key energy-consuming appliances, especially thermal loads, such as water heating and heating, ventilation, and air conditioning (HVAC) systems, are key to achieving energy savings in buildings. Hence, most recent RL-based BEM research, as shown in Table 2, focuses on identifying the optimal strategies of such smart load control, not only to achieve energy and cost savings but also to satisfy comfort levels. The key points observed during the study are as follows:

- Thermal loads either HVAC or water heating are the primary control targets through their setpoints or ON/OFF operations.
- RL has been used to solve the control challenge while satisfying operational constraints, such as the thermal comfort level, energy demand, and electricity price.
- The DQN is the most extensively utilized algorithm for these applications because of its discrete action space, while other algorithms are being tested, such as the DDQN, PPO, and SAC.
- The effectiveness of such systems is measured by the cost savings and energy demand reduction.
- A model can first be trained offline in a simulated environment for better initial performance, and then deployed for real-time online learning.

Table 1. RL and DRL for building energy demand forecasting

| Ref. | Year | Target Building | RL Methods | RL Target Decision | Description & Results |
|---|---|---|---|---|---|
| [35] | 2022 | University Campus Building | DDPG | LSTM-model hyperparameter | When the new training data has high variance, day ahead energy demand and peak demand prediction can gain up to 23% increased accuracy. |
| [36] | 2022 | Office Building | Deep-forest-based DQN (DF–DQN) | An hour ahead energy demand | The proposed model decreased the mean absolute percentage error MAPE, the mean absolute error MAE, and root mean squared error RMSE by 5.5%, 7.3%, and 8.9%, respectively than DRL models. |
| [37] | 2022 | Building and Industry | Multiarmed Bandit | Choice of K-nearest Neighbors (KNN) or Artificial Neural Networks (ANN) | Depending on different contexts, the RL algorithm can learn to switch between KNN and ANN methods at each time horizon to improve performance. |
| [38] | 2020 | Office Building | Asynchronous Advantage Actor-Critic (A3C,) DDPG, and Recurrent Deterministic Policy Gradient (RDPG) | The ground source heat pump (GSHP)-5 minutes ahead/1 hour demand energy forecasting | Compared the prediction performance of three of the most common DRL algorithms with three conventional machine learning algorithms, for both 5 min ahead and 1-hour head horizons. DDPG and RDPG exhibited enhanced performance while A3C did not show an advantage over the other algorithms. |
| [17] | 2019 | Office Building | DDPG | (GSHP)-5 minute ahead demand energy forecasting | The combination of the DRL with Autoencoder (AE) feature extraction to predict the HVAC system energy improves performance by 22.46% and 25.96% for the MAE and RMSE respectively, thereby outperforming support vector machines (SVM) and neural networks (NN). |

Table 2. RL and DRL for building energy demand control

| Ref. | Year | Target Building | RL Methods | RL Target Decision | Description & Results |
|---|---|---|---|---|---|
| [39] | 2022 | Building HVAC model | DQN | HVAC based on variable air volume (VAV) temperature set point reset sequence. | Their model exhibited improved performance over fixed set point temperature control in a building simulated using energy plus. |
| [40] | 2022 | Residential houses | Double DQN (DDQN) | Heat pump ON/OFF, space heating setpoint | The proposed DRL model employs solar energy generation, weather variables, and hot water usage to control the space heating while balancing energy demand, comfort levels, and water hygiene. Accordingly, a reduction of 60% energy was estimated by coupling the proposed control system with solar energy. |
| [41] | 2022 | University Building | PPO | HVAC setpoint | Formulated a DR system that considers the dynamic time-of-use pricing, considering energy demand, thermal comfort, and environmental features to learn the optimal control of the thermostat set point. The results revealed cost savings of 9.17% compared to the constant setpoint. |
| [24] | 2021 | Simulated Smart Residential Buildings | DQN | Heating Control | The proposed control scheme resulted in a 15—30% comfort increase and reduced costs by 5—12%. It was also concluded that in the case of multiple smart buildings, decentralized control outperforms central control. |
| [42] | 2021 | Residential heating system | Soft Actor-Critic (SAC) | Heating Control | The DRL model coupled with a probabilistic window opening behavior method that was employed to capture the occupancy and building interaction achieved an estimated energy saving of 2—6%. |
| [43] | 2020 | Office Building | DQN | Water supply heating set point | 5—12% energy saving was estimated in comparison with conventional control schemes. |
| [44] | 2019 | University Campus Building | DQN | Variable refrigerant flow (VRF) system and a humidifier control | Evaluated the efficiency of a model that combines the Gaussian process regression (GPR) for predicting thermal comfort performance (PMV) in real-time and a DRL model to determine the optimal control policy to minimize energy consumption while sustaining thermal comfort under dynamic environmental conditions. |

## 5. CONCLUSION

RL has been introduced in the context of building energy demand prediction and energy management to overcome the ubiquitous challenges of conventional building modelling. With this background, this paper provides a brief overview of the different types of RL, particularly DRL methods. DRL utilizes DL methods as policy approximators and is classified as model-based or model-free, with discrete or continuous action spaces. These characteristics of DRL are critical for selecting the appropriate algorithm based on the application type and target variable of control or forecast. As pointed out in recent literature, DQN was primarily used for control applications because it has a discrete action space, whereas DDPG with a continuous action space was utilized for prediction purposes. Although recent research has reported innovative attempts to combine DL and DRL for demand forecasting, little research has been conducted to exploit such combinations in the context of BEM. Considering the high potential of such a combination of methodologies, it should be further explored even outside of the prediction category. Thermal loads were identified as the main target variable for optimal control operations in buildings, where RL was employed to control their thermal set-points and operation. Control is performed by solving multi-objective optimizations, such as energy saving, comfort, and cost reduction. Finally, most recent research has been conducted under simulation environments for different building types and available data; hence, testing and validating RL-based BEM for real buildings is very limited and has its own implementation challenges, which makes it an interesting direction for future research.

## 7. REFERENCES

[1] UN Environment Programme, Global Alliance for Buildings and Construction, 2020 Global Status Report For Buildings And Construction, Https://Wedocs.Unep.Org/Bitstream/Handle/20.500.11822/34572/GSR_ES.Pdf. (2022).

[2] United States Department Of Energy, An Assessment Of Energy Technologies And Research Opportunities, website: https://Www.Energy.Gov/Sites/Prod/Files/2017/03/F34/Qtr-2015-Chapter5.Pdf. (2015).

[3] M.A. Kamal, Material Characteristics and Building Physics for Energy Efficiency, Key Engineering Materials. 666 (2015) 77–87.

[4] X. Cao, X. Dai, J. Liu, Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade, Energy and Buildings. 128 (2016) 198–213.

[5] A. Shaqour, H. Farzaneh, Analyzing the Impacts of a Deep-Learning Based Day-Ahead Residential Demand Response Model on The Jordanian Power Sector in Winter Season, Proceedings of International Exchange and Innovation Conference on Engineering & Sciences (IEICES). 7 (2021) 247–254.

[6] L. Malehmirchegini, H. Farzaneh, Modeling and Prioritizing Price–Based Demand Response Programs in The Wholesale Market in Japan, 7 (2021) 7.

[7] A. Shaqour, H. Farzaneh, Y. Yoshida, T. Hinokuma, Power control and simulation of a building integrated stand-alone hybrid PV-wind-battery system in Kasuga City, Japan, Energy Reports. 6 (2020) 1528–1544.

[8] L. Yu, S. Qin, M. Zhang, C. Shen, T. Jiang, X. Guan, A Review of Deep Reinforcement Learning for Smart Building Energy Management, IEEE Internet of Things Journal. 8 (2021) 12046–12063.

[9] H. Farzaneh, L. Malehmirchegini, A. Bejan, T. Afolabi, A. Mulumba, P.P. Daka, Artificial intelligence evolution in smart buildings for energy efficiency, Applied Sciences (Switzerland). 11 (2021) 1–26.

[10] D.B. Araya, K. Grolinger, H.F. ElYamany, M.A.M. Capretz, G. Bitsuamlak, An ensemble learning framework for anomaly detection in building energy consumption, Energy and Buildings. 144 (2017) 191–206.

[11] K. Alanne, S. Sierla, An overview of machine learning applications for smart buildings, Sustainable Cities and Society. 76 (2022) 103445.

[12] T.F. Megahed, S.M. Abdelkader, A. Zakaria, Energy management in zero-energy building using neural network predictive control, IEEE Internet of Things Journal. 6 (2019) 5336–5344.

[13] A. Shaqour, T. Ono, A. Hagishima, H. Farzaneh, Electrical demand aggregation effects on the performance of deep learning-based short-term load forecasting of a residential building, Energy and AI. 8 (2022) 100141.

[14] X. Chen, C. Zanocco, J. Flora, R. Rajagopal, Constructing dynamic residential energy lifestyles using Latent Dirichlet Allocation, Applied Energy. 318 (2022) 119109.

[15] D. Wang, T. Enlund, J. Trygg, M. Tysklind, L. Jiang, Toward Delicate Anomaly Detection of Energy Consumption for Buildings: Enhance the Performance From Two Levels, IEEE Access. 10 (2022) 31649–31659.

[16] R.S. Sutton, A.G. Barto, Reinforcement learning: An introduction, MIT press, 2018.

[17] T. Liu, C. Xu, Y. Guo, H. Chen, A novel deep reinforcement learning based methodology for short-term HVAC system energy consumption prediction, International Journal of Refrigeration. 107 (2019) 39–51.

[18] M. Han, J. Zhao, X. Zhang, J. Shen, Y. Li, The reinforcement learning method for occupant behavior in building control: A review, Energy and Built Environment. 2 (2021) 137–148.

[19] Achiam, Joshua, Spinning Up in Deep Reinforcement Learning, (2018). website: https://github.com/openai/spinningup (accessed July 15, 2022).

[20] Deep Mind, DeepMind AI Reduces Google Data Centre Cooling Bill by 40%, website: Https://Www.Deepmind.Com/Blog/Deepmind-Ai-Reduces-Google-Data-Centre-Cooling-Bill-by-40. (2016).

[21] L. Fridman, Introduction to Deep RL, Deeplearning. Mit. Edu. (2019).

[22] G. C. Alexandropoulos, K. Stylianopoulos, C. Huang, C. Yuen, M. Bennis, M. Debbah, Pervasive Machine Learning for Smart Radio Environments Enabled by Reconfigurable Intelligent Surfaces, (2022). http://arxiv.org/abs/2205.03793 (accessed July 15, 2022).

[23] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller, Playing Atari with Deep Reinforcement Learning, (n.d.).

[24] A. Gupta, Y. Badr, A. Negahban, R.G. Qiu, Energy-efficient heating control for smart buildings with deep reinforcement learning, Journal of Building Engineering. 34 (2021) 101739.

[25] T.P. Lillicrap, J.J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, Continuous Control With Deep Reinforcement Learning, (n.d.). https://arxiv.org/pdf/1509.02971.pdf (accessed July 15, 2022).

[26] J. Zhang, J. Kim, B. O'donoghue, S. Boyd, Sample Efficient Reinforcement Learning with REINFORCE, (2020). www.aaai.org (accessed July 15, 2022).

[27] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O.K. Openai, Proximal Policy Optimization Algorithms, (n.d.).

[28] Z. Wang, T. Hong, Reinforcement learning for building controls: The opportunities and challenges, Applied Energy. 269 (2020) 115036.

[29] T. Hong, F. Buhl, P. Haves, EnergyPlus Run Time Analysis, (2008).

[30] T. Wei, Y. Wang, Q. Zhu, Deep Reinforcement Learning for Building HVAC Control, Proceedings - Design Automation Conference. Part 128280 (2017).

[31] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, Z. Sun, L. Zhang, Y. Zhang, T. Jiang, Deep Reinforcement Learning for Smart Home Energy Management, IEEE Internet of Things Journal. 7 (2020) 2751–2762.

[32] L. Yu, Y. Sun, Z. Xu, C. Shen, D. Yue, T. Jiang, X. Guan, Multi-Agent Deep Reinforcement Learning for HVAC Control in Commercial Buildings, IEEE Transactions on Smart Grid. 12 (2021) 407–419.

[33] E. Mocanu, D.C. Mocanu, P.H. Nguyen, A. Liotta, M.E. Webber, M. Gibescu, J.G. Slootweg, On-Line Building Energy Optimization Using Deep Reinforcement Learning, IEEE Transactions on Smart Grid. 10 (2019) 3698–3708.

[34] G. Gao, J. Li, Y. Wen, DeepComfort: Energy-Efficient Thermal Comfort Control in Buildings Via Reinforcement Learning, IEEE Internet of Things Journal. 7 (2020) 8472–8484.

[35] X. Zhou, W. Lin, R. Kumar, P. Cui, Z. Ma, A data-driven strategy using long short term memory models and reinforcement learning to predict building electricity consumption, Applied Energy. 306 (2022) 118078.

[36] Q. Fu, K. Li, J. Chen, J. Wang, Y. Lu, Y. Wang, Building Energy Consumption Prediction Using a Deep-Forest-Based DQN Method, Buildings 2022, Vol. 12, Page 131. 12 (2022) 131.

[37] D. Ramos, P. Faria, L. Gomes, P. Campos, Z. Vale, Selection of features in reinforcement learning applied to energy consumption forecast in buildings according to different contexts, Energy Reports. 8 (2022) 423–429.

[38] T. Liu, Z. Tan, C. Xu, H. Chen, Z. Li, Study on deep reinforcement learning techniques for building energy consumption forecasting, Energy and Buildings. 208 (2020) 109675.

[39] X. Fang, G. Gong, G. Li, L. Chun, P. Peng, W. Li, X. Shi, X. Chen, Deep reinforcement learning optimal control strategy for temperature setpoint real-time reset in multi-zone building HVAC system, Applied Thermal Engineering. 212 (2022) 118552.

[40] A. Heidari, F. Maréchal, D. Khovalyg, Reinforcement Learning for proactive operation of residential energy systems by learning stochastic occupant behavior and fluctuating solar energy: Balancing comfort, hygiene and energy use, Applied Energy. 318 (2022) 119206.

[41] Z. Li, Z. Sun, Q. Meng, Y. Wang, Y. Li, Reinforcement learning of room temperature set-point of thermal storage air-conditioning system with demand response, Energy and Buildings. 259 (2022) 111903.

[42] S. Brandi, D. Coraci, D. Borello, A. Capozzoli, Energy Management of a Residential Heating System Through Deep Reinforcement Learning, Smart Innovation, Systems and Technologies. 263 (2022) 329–339.

[43] S. Brandi, M.S. Piscitelli, M. Martellacci, A. Capozzoli, Deep reinforcement learning to optimise indoor temperature control and heating energy consumption in buildings, Energy and Buildings. 224 (2020) 110225.

[44] Y.R. Yoon, H.J. Moon, Performance based thermal comfort control (PTCC) using deep reinforcement learning for space cooling, Energy and Buildings. 203 (2019) 109420.