energies

MDPI

Review

Systematic Review on Deep Reinforcement Learning-Based Energy Management for Different Building Types

Ayas Shaqour and Aya Hagishima *

Interdisciplinary Graduate School of Engineering Sciences, Kyushu University, Kasuga City 816-8580, Japan
* Correspondence: ayahagishima@kyudai.jp; Tel.: +81-92-583-7646

Abstract: Owing to the high energy demand of buildings, which accounted for 36% of the global share in 2020, they are one of the core targets for energy-efficiency research and regulations. Hence, coupled with the increasing complexity of decentralized power grids and high renewable energy penetration, the inception of smart buildings is becoming increasingly urgent. Data-driven building energy management systems (BEMS) based on deep reinforcement learning (DRL) have attracted significant research interest, particularly in recent years, primarily owing to their ability to overcome many of the challenges faced by conventional control methods related to real-time building modelling, multi-objective optimization, and the generalization of BEMS for efficient wide deployment. A PRISMA-based systematic assessment of a large database of 470 papers was conducted to review recent advancements in DRL-based BEMS for different building types, their research directions, and knowledge gaps. Five building types were identified: residential, offices, educational, data centres, and other commercial buildings. Their comparative analysis was conducted based on the types of appliances and systems controlled by the BEMS, renewable energy integration, DR, and unique system objectives other than energy, such as cost, and comfort. Moreover, it is worth considering that only approximately 11% of the recent research considers real system implementations.

Keywords: building energy demand; deep reinforcement learning; data-driven control; energy demand prediction; energy efficiency; energy management; residential building; office building; commercial building; data centre

Citation: Shaqour, A.; Hagishima, A. Systematic Review on Deep Reinforcement Learning-Based Energy Management for Different Building Types. *Energies* **2022**, *15*, 8663. https://doi.org/10.3390/en15228663

Academic Editor: Dimitrios Katsaprakakis

Received: 18 October 2022
Accepted: 15 November 2022
Published: 18 November 2022

1. Introduction

As of 2020, buildings accounted for 36% of global energy demand shares, and 37% of the total global energy-related CO_2 emissions. Residential buildings had the highest share at 22%, while non-residential buildings had 8%, and the final 6% were related to the construction industry [1]. Hence, buildings are at the frontiers of energy research, and will be key to realizing future smart grids and greener, sustainable energy systems. This can be achieved by developing efficient, smart, and adaptive buildings that go beyond the conventional role of passive energy consumers. Moreover, future smart buildings must adapt to the rising complexities of modern power grids because of the induced stochasticity of renewable energy generation and the decentralization of power supply. To accomplish these goals, investment in energy efficiency in buildings has recently been rapidly increasing; as of 2020, the total investment has reached $180 billion, increasing by 39.5% since 2015 [1]. Realizing these targets is based on many interlapping paradigms, as shown in Figure 1. Specifically, these include increasing their renewable power generation [2], having more efficient electrical products [3], improving their thermal design [4], and activating their role in the energy market [5]. All these measures are employed in conjunction to ensure that future buildings will be better aligned with global sustainability goals, have reduced energy consumption, and be smart active players in the energy market.

Building energy management systems (BEMS) are integral for realizing smart buildings. BEMS, which are based on advanced energy management, tie up all the other

paradigms [6]. The BEMS must be both high performing and conforming to human comfort levels. BEMS need to determine the best schedule for certain appliances [7] or efficient operational set points [8]. Moreover, BEMS control the utilization of the thermal body of a building to store energy [9] and when to pre-cool the building or pre-heat the building/water when there is a surplus of renewable energy. Finally, BEMS will manage buying or selling energy to the grid to minimize costs and increase profits [10]. Achieving these targets not only requires a deep understanding of each of the paradigms illustrated in Figure 1, but also on how they can be optimized and integrated using state-of-the-art BEMS.

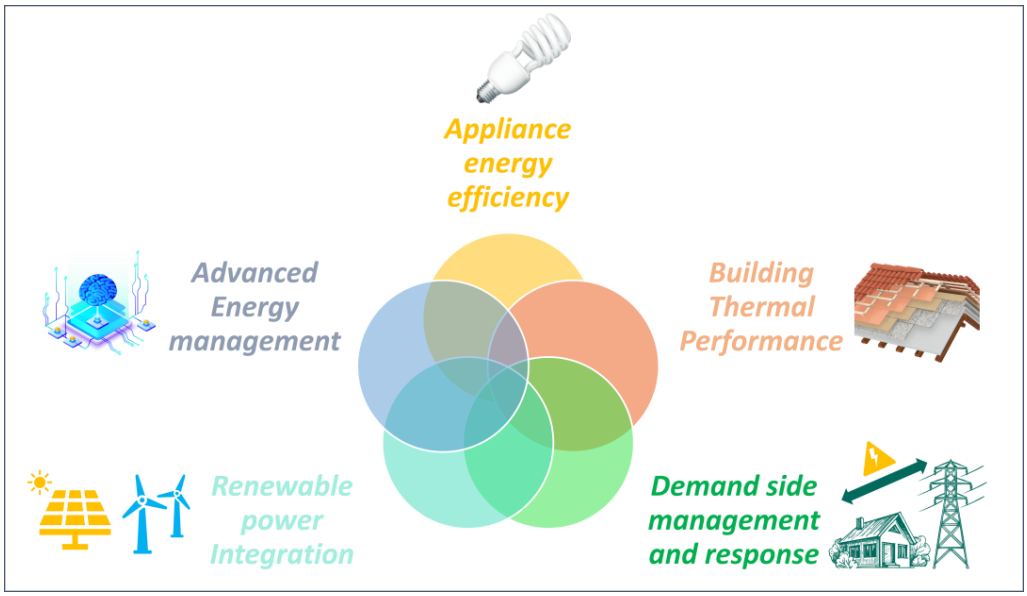

Figure 1. Interlapping Paradigms for building energy efficiency.

State-of-the-art BEMS is at the core of future smart buildings. Currently, with the breakthroughs of artificial intelligence (AI) and machine learning (ML) [11], the rapid development of IoT devices and sensing technology [12], and low-cost high-computational controllers, the inception and wide deployment of advanced BEMS is imminent. Previously, real-time BEMSs were operated using conventional control methods, such as rule-based methods and proportional-integral-derivative (PID) control, both of which are static and rely on heuristic rules. However, conventional BEMS face many limitations and challenges related to building modelling, satisfying and controlling multiple objectives, system generalization, and scaling.

As shown in Figure 2, constructing precise building models related to thermal characteristics is a complex task that relies on stochastic elements raised by the assumed schedules of appliance usage, human presence, and various elements in buildings [13]. The problem is amplified when real-time modelling is required for the BEMS to make decisions within a sub-second time window, especially for local control problems. This type of micro-real-time control will be more critical in future smart homes, where supervisory control with bigger time windows might fail to achieve optimal energy saving to the many fast-changing variables such as energy prices, renewable energy availability and human behaviors. In this case, the physics-based, complex white model that is generally more suited for design building standards and optimal building design cannot be used because of the high computational space and time complexity required [14]. Hence, simpler models, such as gray or black box models, are required [15]. Smart buildings with renewable energy, storage systems, electric vehicles, smart appliances, and heating, ventilating, and air-conditioning (HVAC) systems are required to coordinate energy management between these different entities, where operational constraints and objectives are multidimensional and intricate with high-dimensional solution spaces [16,17].

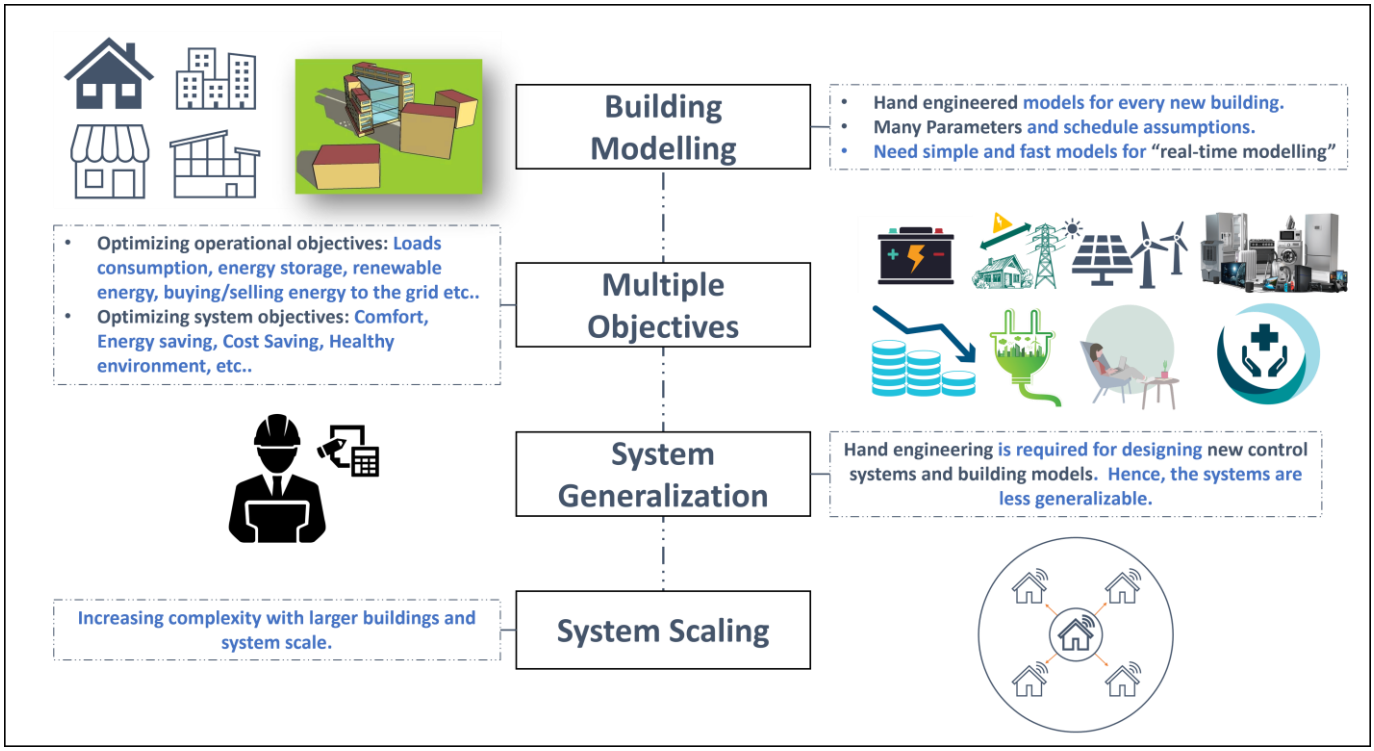

Figure 2. Challenges and limitations faced by conventional BEMS.

Furthermore, system objectives to satisfy comfort levels, energy savings, cost savings, and health and environmental goals introduce an additional layer of objectives and constraints that must be handled by the BEMS. In addition, because conventional modelling and control methods in BEMS are based on hand-engineered designs for each specific building and case, and such a system cannot be easily generalized to other buildings that generally have many different and unique designs [18]. Finally, scaling the target premise of a BEMS for multiple buildings increases the complexity of BEMS designs and further hinders the cost and time required for wide-scale deployment [19].

Owing to the previous challenges and the rise of ML methods in many fields since 2011, there has been increased interest in exploiting the benefits of such data-driven methods in BEMS. Applications range from improved forecasting of energy demand [20] and human behaviour in buildings [21] to anomaly detection [22], and classification of different building states [21]. In particular, reinforcement learning (RL) [23] and deep reinforcement learning (DRL), have attracted significant interest in the last five years. This is observed in both academia and industry because of its ability to solve the challenges mentioned in Figure 2. One notable application of RL-based BEMS was the data centres of Google by DeepMind, where a reduction in the cost of cooling reached 40% [24]. This highlights that advanced energy management, particularly relating to HVAC loads, which are the highest consumers of energy in buildings, is beyond small gains, and that there are large potential savings. As discussed by Yu et al., DRL can contribute to solving the challenges related to BEMS and can be summarized as follows [6]:

- Real-time Modelling: By utilizing neural networks, and complex environments, such models can be modeled with a lower computational cost following the training phase by forward propagation. Furthermore, DRL can also operate model-free or learn the representation of the environment without explicitly knowing its detailed model. This is similar to how human agents navigate the real world without knowing its detailed physical model but by learning how to interact with it.
- Handling multiple objectives: In DRL, through careful engineering of the objective function and multiple objectives can be maximized while satisfying the constraints.

- Generalization: Similar to human intelligence, being model-free and learning to maximize rewards in a stochastic environment increases the generalizability of BEMS.
- Scaling: Considering the previous points, scaling BEMS systems can be achieved in real-time using a less complex approach rather than relying on complex modelling and hand-engineered solutions for optimizing multiple variables.

It must be noted that utilizing neural networks and optimizing multiple objectives are not unique to DRL and are relatively shared with other control paradigms such as model predictive control (MPC) and proportional-integral-derivative (PID). Svetozarevic et al. provided a comprehensive discussion on their comparative analysis and characteristics [25]. Finally, the main objective of this work is to systematically investigate and summarize the recent advances of DRL applications in BEMS, while focusing on a building-type centric discussion.

1.1. Related Work

Owing to the promising benefits of data-driven methods, such as ML and DRL methods, there has been considerable attention in recent literature to review their various applications in BEMS-related areas. Each review focused on one or more aspects related to BEMS, and some considered broader applications in energy systems that are beyond buildings, as summarized in Table 1.

First, it was observed from a recent literature review that there is a high emphasis on method explanation and comparison owing to large variations in the RL/DRL models. This high emphasis on method-centric classification is clearly evident in the detailed work of Wang and Hong, where they classified the reviews based on the internal varying configurations of RL/DRL. For each part of the algorithm, they further investigated the chosen parameters and configurations used in recent research, such as algorithms, states, actions, rewards, and implementation environment [26].

Table 1. Recent reviews related to building energy modelling and BEMS.

Ref	Year Published	Coverage Span (Up to)	Review Objectives	Main Methods	BEMS Centric	Multiple Building Type-Centric
[27]	2022	~2021	Extensive RL Centric review related to BEMS. Primary emphasis is on the classification, types, and applications of RL Algorithms.	RL, DRL	Yes	No
[21]	2022	~2021	A system-level oriented review of the integration of learning methods for realizing intelligent buildings management.	ML, RL, DRL	Yes	No
[28]	2021	~2020	A broad review of RL applications in energy systems.	RL, DRL	No	No
[6]	2021	~2020	A system-scale-centric review of the application of DRL in BEMS.	DRL	Yes	No
[29]	2020	~2019	Energy management of AC in buildings via Computational Intelligence (CI) algorithms.	White Box, Black Box, Gray Box.	No	No
[30]	2020	~2019	General CI algorithms for BEMS of residential homes.	Mathematical optimization, GA, ML, RL, MPC.	Yes	No
[31]	2020	~2019	Modelling occupant behaviours.	Rule-based, Stochastic, and Data-Driven.	No	No
[32]	2020	~2019	General DRL applications in the power system.	DRL	No	No

Table 1. *Cont.*

Ref	Year Published	Coverage Span (Up to)	Review Objectives	Main Methods	BEMS Centric	Multiple Building Type-Centric
[26]	2020	~2019	RL-based BEMS application, focusing on a detailed intrinsic review of the different RL methods, variations, configurations, and simulation vs. real environment analysis.	RL	Yes	No
[33]	2019	2019	RL-based occupant comfort control in buildings.	RL	Yes	No
[34]	2019	~2018	RL-based BEMS focuses on method variations and an energy appliance-centric review.	RL	Yes	No
This Study	-	~2022	DRL-based BEMS review centric to applications for different types of buildings with a focus on the details of promising recent advances and limitations.	DRL	Yes	Yes

1.2. Motivation and Objectives

While recent reviews are detailed and informative of RL applications for BEMS, they do not consider building-type-centric discussions. Different building types, such as residential, commercial, and others, should have different specific characteristics, challenges, limitations, and potential, particularly from a data-driven approach perspective. Thus, it will be extremely useful for researchers in this field to realize building type-centric landscapes and discussions related to this area of research, particularly in terms of challenges and opportunities. Furthermore, with the rapid growth of this field, as discussed in Section 2, realizing the most recent creative and innovative research direction in this promising area of research can contribute to the existing literature. Therefore, the present study aims to contribute to the following:

- To systematically review recent advances and innovations in data-driven DRL-based BEMS.
- To conduct a building type-centric review and analysis.
- To discuss the limitations and challenges related to each building type.
- To realize the promising directions of DRL-based BEMS research, especially from a building-type-centric perspective.

The remainder of this paper is organized as follows: Section 2 discusses the basic classification of RL and DRL methods, as well as the PRISMA approach. Section 3 discusses the recent research for each building type. Section 4 discusses the main conclusions observed and future research recommendations. Finally, Section 5 presents the conclusions of the study.

2. Methodologies: Deep Reinforcement Learning and PRISMA

This research primarily focuses on conducting a systematic review of the DRL applications in BEMS, focusing on the perspective of each building type. The present systematic review is based on the PRISMA framework, which is a method for conducting systematic reviews, followed by a checklist and general flow chart. While the PRISMA framework is extremely detailed with a long general checklist, this study will follow these guidelines as much as possible from the specific perspective of this field. Furthermore, a brief overview of DRL methods is discussed before the review steps and methods.

2.1. A Brief Overview of RL and DRL

RL first emerged in the 1950s and was related to the optimal control theory used for the formulation of control systems for specific target variables [23]. At its core, RL is based on the Markov decision processes (MDPs) proposed by Bellman. They were utilized to formulate control problems and depict the representation of the environment in which

the RL agent learns its optimal behaviour policy [35]. MDP problems entail the Markov property, where transitions in their environments are a function of the present actions taken by the agent and states, and do not depend on past actions and states. They can be formulated as a tuple of five elements: states, actions, rewards, transition probability function, and the initial distribution state S, A, R, P, ρ_0. It can be observed from the top-left of Figure 3 [6,36,37], starting from an initial state ρ_0, for every time step, the agent observes the state (S) from its environment and then takes action (A) with an optimal strategy (policy) to maximize its future rewards (R) [38]. The goal of the agent is to learn the optimal policy (π) to navigate the environment through a certain episode of states and actions (τ) that can maximize the expected return of future rewards. Detailed formulations of various types of RL algorithms were discussed by Sutton and Barto [23]. One crucial difference between RL and DRL is that conventional RL algorithms generally use tabular methods to build the value table of each state–action pair in the environment, which the agent learns and updates. However, tabular methods are not suitable for problems with large state and action spaces, such as videos and images, or problems in which the environment can have an extremely large variance of new unseen states. Hence, function approximators were introduced to solve this problem using different supervised methods, particularly deep learning methods, as shown in Figure 3; hence, the name Deep RL [23].

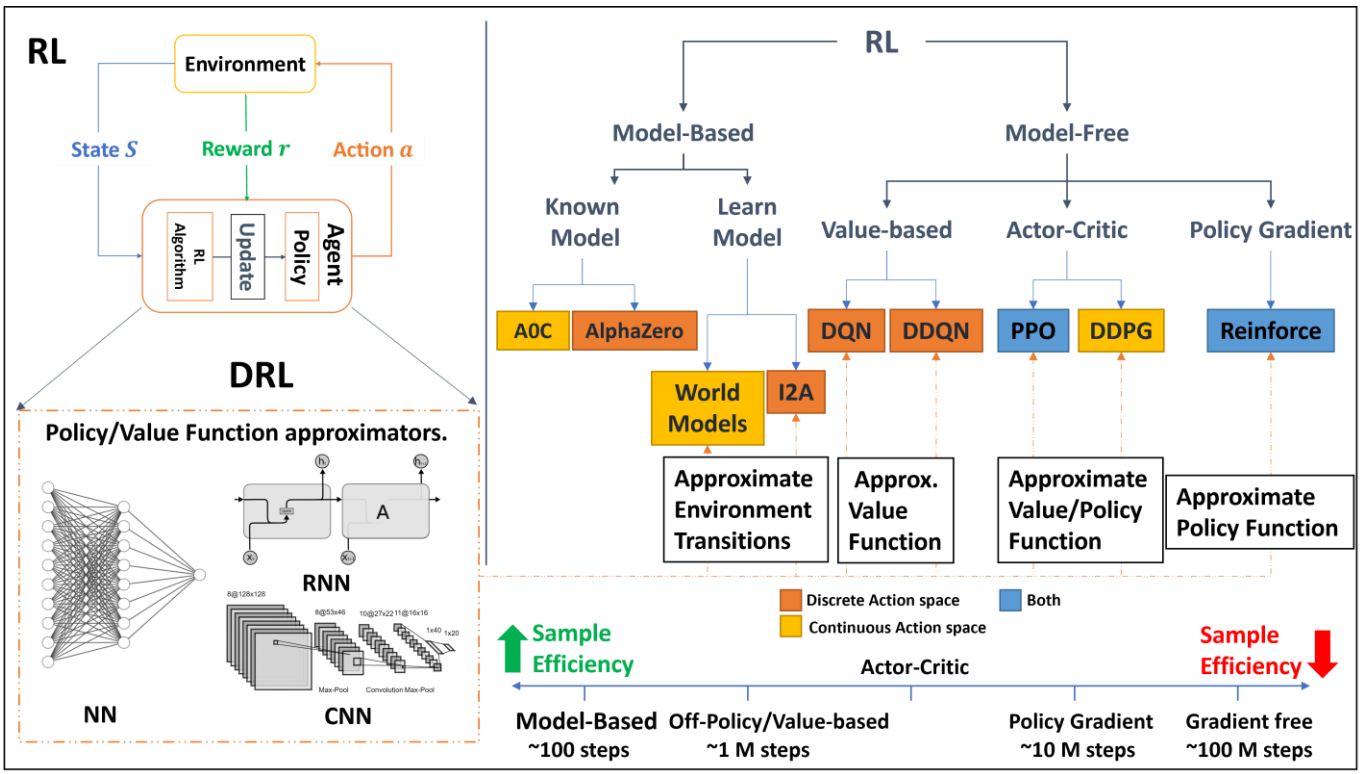

Figure 3. Brief overview of RL, DRL, and their variants.

Owing to the rapid development of the ML field, there have been many recently developed variants of DRL, some of which are depicted in Figure 3, with is a brief highlight of their principal characteristics. In general, RL can be classified as either model-based or model-free RL. Model-based RL is based on the fact that the environment and system dynamics are either known or learned via an ML algorithm outside the RL agent. Model-based RL occurs when the environment or system dynamics are known or learned and encoded in a neural network; for example, where the agent can utilize a transition probability function to predict or estimate the future state of the environment and reward prior to taking each action. These models tend to be more sample-efficient, requiring less experience to reach good performance, such as the popular Alphazero [39] or Alphazero continuous (A0C) [40],

which work with an already known model, such as the game of Go or chess, or ones that learn environment models, such as world models [41] and imagination-augmented agents (I2A) [42]. On the other hand, when environments are too complex or have exceedingly large state-action spaces, model-free RL is generally preferred, although these methods tend to require more samples (more experience and data) to achieve good performance, as they need to learn the best policy by purely learning from their actions in the environment without previously knowing its effect, as it does not use a given model. However, it can build its own representation of the environment. Some of the popular model-free algorithms can be observed in Figure 3 as being value-based/off policies, such as deep Q-networks (DQN) [43] and double deep Q-learning (DDQN) [44]; or actor-critic-based, such as proximal policy optimization (PPO) [45] and deep deterministic policy gradient (DDPG) [46]; and on-policy or policy gradient, such as the REINFORCE method [47].

As indicated in Table 1, the recent literature has done a great job of providing a method-centric review and the different types of algorithms used for each specific application in the context of BEMS. In general, in the context of a BEMS, selecting the best algorithm is initially based on the BEMS environment; if it is too complex to model and can face unforeseen changes, then model-free approaches should be used. Furthermore, as shown in Figure 3, the algorithms have a specific character in their action spaces being either continuous, discrete, or both, which is a crucial factor in selecting the correct algorithm. Finally, it can be noticed that in general, some algorithms are extensions of another, where the improved version can be selected to test for improved performance.

2.2. PRISMA Review Framework

The PRISMA framework proposed by Denyer and Tranfield [48] was originally designed for management and organizational studies. It has recently been depicted as a standardized roadmap for implementing systematic reviews. Its primary objective is to identify published research in a specific area, systematically select and screen related studies, and finally reach a conclusion with clarity on the target inquiry of the review. The principles of PRISMA are as follows:

1. Formulate a question for the review.
2. Find the related studies.
3. Select and evaluate the studies.
4. Analyse the findings.
5. Report the results.

This subsection discusses the first three steps, and the remainder of the paper describes the last two steps.

2.2.1. Question Formulation Using CIMO Logic

First, to identify the review question, the CIMO-logic framework is implemented, which stands for (context, intervention, mechanism, and outcome) as proposed by Denner et al., to capture the four core pillars of a well-designed systemic review [49]. Table 2 presents the CIMO-logic-based proposition for this review.

2.2.2. Locating and Screening Relative Studies

Based on Table 2, the search keywords were populated according to the proposed CIMO logic using logical operators in the Scopus database. Table 3 lists the different types of buildings that are currently investigated in the literature related to RL-based BEMS. Initially, for each building type, multiple words were used in the search query, and it was found that the most researched types include residential buildings followed by offices. Finally, a few different types of buildings, such as educational facilities, data centres, and an excluded study had investigated a hotel building. These are only papers that mentioned the building type in the abstract, title, or keywords, based on the different words that were initially used for each type. Figure 4 presents the number of publications per year, where the recent surge of interest in RL-based BEMS is evident. It must also be noted that the keyword

"RL" also captures studies that are based on DRL algorithms that explicitly mention DRL in their title, abstract or keywords. Furthermore, almost half of the publications were not captured with the initial keyword search based on building type. A detailed screening of the literature was conducted using the PRISMA framework based on Moher et al. [50] as shown in Figure 5.

Table 2. CIMO-logic for DRL application in BEMS for different building types.

Context (C): *Where? In Which Specific Area Is the Intervention Applied?*	**Intervention (I):** *What Is the Intervention of Interest?*	**Mechanisms (M):** *What Are the Target Methods of Such Interventions*	**Outcome (O):** *What Are the Main Expected Outcomes?*
Different Building Types: • Residential • Office • University Campus/School • Data centres • Other Commercial	Energy management and optimization of: • Appliances • Electric vehicles • HVAC systems • Renewable energy • Demand response	Modelling and controlling using deep reinforcement learning.	• Reducing energy consumption • Reducing energy costs • Satisfying human comfort levels • Satisfying health constraints

Table 3. Initial search query results for RL and DRL-based BEMS per building type.

Search Query	Building Type	Hits
TITLE-ABS-KEY (reinforcement-learning AND energy AND building)	All	470
TITLE-ABS-KEY (deep AND reinforcement-learning AND energy AND building)	All	225
TITLE-ABS-KEY (reinforcement-learning AND energy AND building AND (residential OR home OR appartement OR district))	Residential	105
TITLE-ABS-KEY (reinforcement-learning AND energy AND building AND office)	Office	70
TITLE-ABS-KEY (reinforcement-learning AND energy AND building AND (campus OR laboratory OR educational OR school))	Educational	14
TITLE-ABS-KEY (reinforcement-learning AND energy AND building AND (data AND (center or centre))	Datacenter	9
TITLE-ABS-KEY (reinforcement-learning AND energy AND building AND commercial) AND NOT TITLE-ABS-KEY (office OR residential OR home OR appartement OR campus OR educational OR school OR (data AND (center OR centre)) OR laboratory OR district)	Other Commercial	9

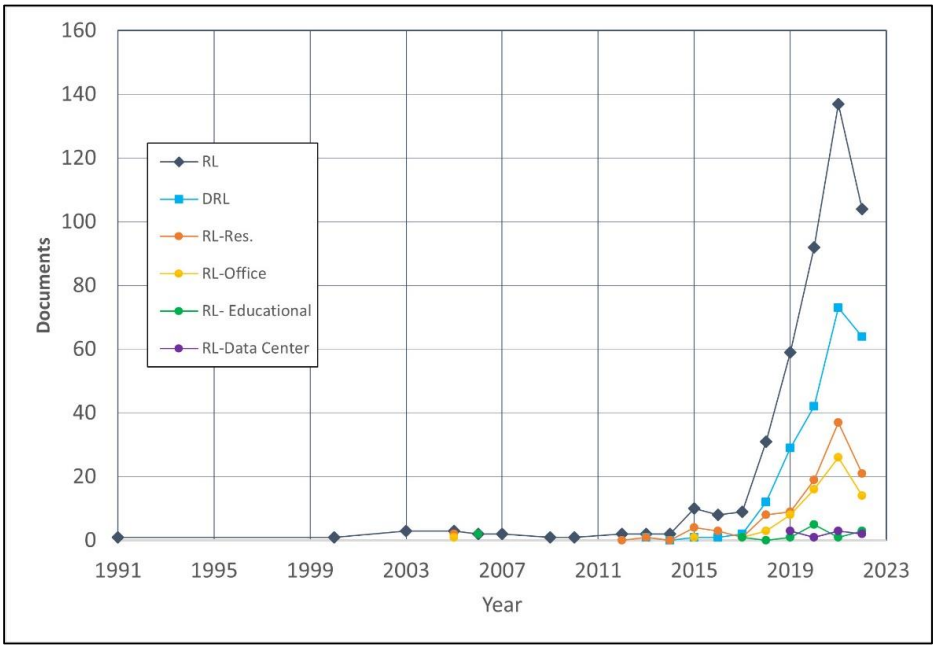

Figure 4. RL and DRL-based BEMS publications per year for different types of buildings.

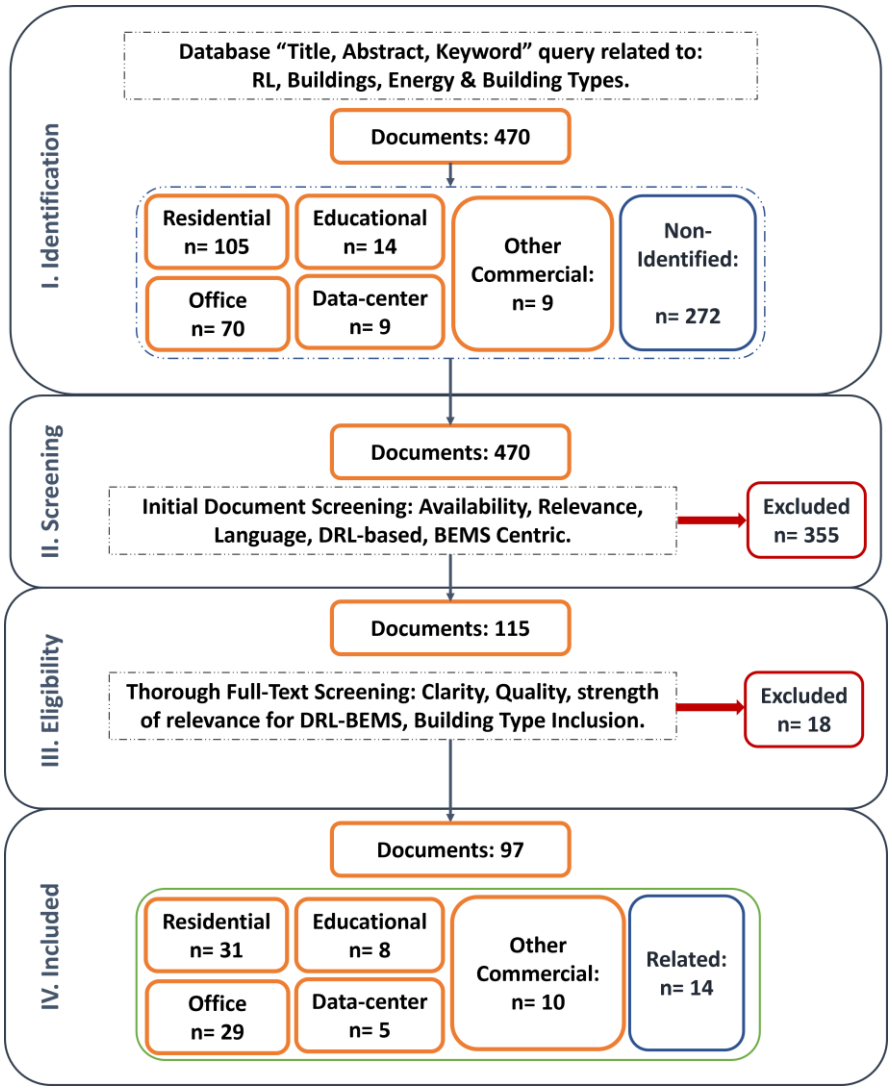

Figure 5. PRISMA-based literature screening methodology.

3. Recent Advances in DRL-Based BEMS per Building Type

The DRL-based BEMS field has grown rapidly in the last five years, with numerous creative ideas and innovations for integrating advanced data-driven control methods in the development of fully enabled smart buildings. Although residential buildings are by far the largest energy consumers, other building types, such as offices and educational buildings, are also being investigated. It would be useful to realize the different directions of research, types of applications, and innovative ideas being implemented for each building type. In particular, it is crucial from a data-centric perspective, as being able to train and use data-driven methods requires large amounts of data, particularly when deploying such systems in the real world. Therefore, it would be interesting to understand how these challenges are satisfied in different types of buildings.

3.1. Residential Buildings

As previously mentioned, residential buildings account for almost 22% of global energy demand, making them one of the most energy-consuming building types. Additionally, while some types of commercial buildings are primarily used during the day by employees, especially in the post-COVID-19 era, they can be more grid-friendly from the perspective of being more aligned with solar energy availability depending on the work culture of the country. It is primarily because there is more solar energy during

the day, whereas residential energy demand increases after work hours, peaking in the evening, which can be a sensitive period for grid operators to compensate for the supply demand change. This could be one underlying factor why this type of building is receiving significant attention in this field, particularly from a DR perspective. Table 4 presents recent research conducted on DRL-based BEMS in residential buildings.

Table 4. Recent application of DRL-based BEMS on residential buildings.

Ref	Year	Building Study Scale	BEMS	ESS	PV	DR	DRL	Estimator	Unique Objective	Real System	Energy */Cost Saving
[51]	2022	Single	HVAC	x	x	o	DQN	DNN	-	-	19.40%
[52]	2022	Single	HVAC	x	x	o	DQN, DDPG	DNN	-	√	25.9–32%
[53]	2022	Single	HVAC, EV, Appliances	o	o	o	ACKTR	Kronecker-Factored	-	-	25.37% ▲
[54]	2022	Single	HVAC	o	o	o	Clustering-DDPG	DNN	-	-	41%
[55]	2022	Single	Appliances	x	o	o	DQN	DNN	Peak demand	-	30%
[56]	2022	Single	HVAC, WHP	x	o	x	DDQN	DNN	Health	-	7–60% *,▲
[57]	2022	Single	HVAC, EV, Appliances	o	x	o	A2C	DNN	-	-	23%
[58]	2022	Single	HVAC, Appliances	o	o	o	MDRL	DNN	-	-	25.80%
[59]	2022	Single	HVAC	x	x	x	DDQN	DNN	Health	-	23.80% ▲
[60]	2022	Single	HVAC, EV, Appliances	o	x	o	DQN	DNN	-	-	21.30%
[61]	2021	Single	HVAC	x	x	o	DQN	DNN	-	-	19.48%
[62]	2021	Single	WHP	x	x	o	DQN	DNN	-	-	19–35%
[63]	2021	Single	HVAC	x	x	o	DDQN-PER	DNN	Health	-	3.51–8.56%
[64]	2021	Single	HVAC	x	x	o	DDPG	DNN	-	-	12.7–50% ▲
[65]	2021	Single	HVAC, EV, Appliances	o	o	o	TD3, DQN, DPG	DNN	-	-	5.93–12.45%
[66]	2020	Single	Appliances	x	x	o	DQN	CNN	Peak demand	-	11.66% ▲
[67]	2020	Single	HVAC	x	x	o	DQN	DNN	-	-	43.89%
[13]	2020	Single	HVAC, Battery	o	o	o	DDPG	DNN	-	-	8.10–15.21%
[68]	2022	Single	HVAC	x	x	x	RLMPC vs. (DDQN, MPC)	DNN	-	-	-
[25]	2022	Single	WHP, EV	o	o	o	DDPG	DNN	-	√	30% *
[69]	2021	Single	TES	o	o	o	REINFORCE	DNN	-	-	50%
[70]	2021	Single	HVAC	x	x	x	REINFORCE	Monte-Carlo PG	-	-	13–64%
[71]	2020	Single	HVAC	x	x	o	DQN	DNN	-	√	21% ▲, 30%
[72]	2020	Multi.	HVAC	x	x	x	DQN	BCNN	-	-	53% *
[73]	2022	Multi	CHP, Boiler	x	x	o	DRLEM	-	-	-	3.30%
[74]	2022	Multi	HVAC, Appliances, Battery	o	o	o	A2C	DNN	Peak	-	5–35%
[75]	2022	Multi	HVAC, WHP, Appliances	o	o	o	SAC	DNN	-	-	3–7%

Table 4. *Cont.*

Ref	Year	Building Study Scale	BEMS	ESS	PV	DR	DRL	Estimator	Unique Objective	Real System	Energy */Cost Saving
[76]	2021	Multi	TES, Battery	o	o	x	MARLISA-_DACC	DNN	Emissions	-	-
[77]	2021	Multi	HVAC	x	x	x	DQN	DNN	-	-	5–12%
[78]	2020	Multi	TES	o	o	o	SAC	DNN	Peak	-	-
[17]	2018	Multi	HVAC, EV, Appliances	x	o	o	DQN, DDPG	DNN	Peak	-	27.40%

o: included, x: not included, * Energy saving, ▲ Over all energy/cost saving (Others are an improvement over a baseline Controller). Table Abbreviations: (ACKTR) Actor-critic kronecker-factored trust region; (A2C) advantage actor-critic; (BCNN) Bayesian-Convolutional-Neural-Networks; (CNN) Convolutional neural network; (DDQN-PER) Double deep Q-learning prioritized experience replay; (DNN) Deep neural network; (EV) Electric vehicle; (RLMPC) Reinforcement Learning Model Predictive Control; (SAC) Soft actor-critic; (TES) Thermal energy storage; (TD3) Twin Delayed DDPG; (WHP) Water heating pump.

As indicated in Table 4, DRL-based BEMS research can consider one or multiple buildings to measure the performance of DRL algorithms under different scenarios or to test a multiple-agent DRL approach for managing energy flow, considering multiple buildings or zones simultaneously [75]. Glatt et al. introduced a decentralized actor-critic reinforcement learning algorithm MARLISA; however, they focused on integrating a centralized critic (*MARLISA_DACC*) to coordinate energy storage systems (ESS) control, such as batteries and thermal energy storage (TES), between various buildings in a manner that enhances DR performance and reduces carbon footprints [76]. With the increase in the scale of residential buildings, multiple-agent approaches can learn to share information and act in a positively correlated manner to maximize the BEMS performance over single-agent approaches. Ahrarinouri et al. utilized a distributed reinforcement learning energy management (DRLEM) to control the energy flow of combined heat and power (CHP) and boilers between multiple buildings, where the connection between the multiple agents reduced the heat losses and costs by 18.3% and 3.3%, respectively, and increased energy sharing in peak time by 23% [73]. Hence, distributed, and multi-agent approaches will be key methods in further research on residential neighbourhoods and buildings, where renewable energy and EV can be coordinated between different houses to reduce renewable energy curtailment and maximize profits in peer-to-peer local energy trading hubs.

The large variety of appliances and BEMS targets are major opportunities in deploying DRL-based BEMS in residential buildings, and there is a high potential for DR because of their contribution to both morning and evening peak demands [79], and detached houses having space for renewable energy integration. In the recently reviewed literature in Table 4, 77% of the studies considered demand response systems, where the varying electricity price was integrated into the objectives of the control logic, while 42% and 45% had also considered the integration of ESS and PV renewable energy, respectively. Furthermore, while 74% of the systems were deployed to manage HVAC systems related to BEMS targets, 32% of the studies included different types of shiftable/fixable appliances, and 19% investigated the inclusion of electric vehicles (EVs). Table 4 classifies the general BEMS target systems in residential buildings, while Table 5 includes a detailed list of appliances that were directly controlled, apart from HVAC systems and TE; noticeable appliances include dishwashers, washing machines, and EVs. The diversity of BEMS targets in residential buildings is noticeable and considerably high, giving it a unique potential and research perspective. This is probably related to the fact that homeowners might have higher relative demand flexibility than office buildings; for example, owing to direct cost benefits. The operating environment tends to have higher levels of stress and no direct benefits to individuals to compromise their comfort, where the benefit is for business owners.

Table 5. Residential appliances controlled using DRL-based BEMS.

Appliance	#No.	Reference
HVAC	19	[13,17,52–54,56–58,60,61,63–65,67,71,72,74,75,77]
Washing machine	8	[53,55,57,58,60,65,66,74]
Dish washer	8	[17,53,55,57,58,60,65,66]
Electric vehicle (EV)	6	[17,25,53,57,60,65]
Water heating pump (WHP)	5	[25,56,59,62,75]
Underfloor heating	2	[49,68]
Clothes dryer	2	[53,66]
Vacuum cleaner	1	[66]
Passive heating and cooling	1	[70]
Boiler	1	[73]
Light	1	[60]
Ventilation	1	[63]
Grinder	1	[66]

For the DRL methods, the most-utilized algorithm was DQN, while DDQN and DDPG were notable. Many studies include a comparison between the different types of DRL to determine the best method based on realizing system objectives. Meanwhile, others investigated hybrid methods, such as the mixed deep reinforcement learning (MDRL) introduced by Huang et al. [58], which combines both DQN and DDPG for enhanced performance, and the RLMPC implemented by Arroyo et al. [68] which combines both the MPC and DDQN methods in a manner that leverages the benefits of both methods. Two recent unique variations of DRL were also observed. First, the actor-critic approach using the Kronecker-factored trust region (ACKTR) introduced by Chu et al. [53] increased the sampling efficiency and integrated discrete and continuous action spaces that exhibited high potential. The second algorithm is a combination of clustering and DDPG developed by Zenginis et al., which homogeneously partitions the training data using a clustering method and then trains different agents of each subset of the training data, achieving higher energy efficiency over a single agent [54]. While these methods are not directly related to the type of building, exhibiting such methods can aid researchers in choosing recently advanced implementations of DRL on the basis of their application and building type. Finally, DNNs have been the most used value/policy function estimators, whereas very few used other methods, such as CNN. In general, owing to the mixed type of state variables, DNNs can effectively map state–action spaces and can be considered the default estimator; however, this indicates that there can be potential for testing other methods.

The primary objectives of most BEMS systems are typically the same in terms of comfort and reducing energy/cost. In terms of energy and cost, they are highly correlated, where a reduction in one depicts a reduction in the other, although different studies report their primary objective improvements in terms of energy or cost based on whether DR is considered; hence, the price of energy analysis is included. Other secondary objectives, highlighted by some studies, include health factors such as indoor CO_2 levels, and the reduction of peak demand, which usually refers to the improvement over a rule-based baseline controller or a comparison between single and multiple-agent methods. Hence, the high energy-saving percentages do not necessarily depict the overall energy reduction, making it harder to cross-compare studies based on these numbers. Nevertheless, they highlight the advantages of energy savings in residential buildings utilizing DRL. Finally, real implementations are significantly lacking, with only three studies (<10%) out of 31 having validated their models outside of a simulation environment, which highlights a clear research gap.

3.2. Office Buildings

Office buildings face the challenge of a limited variety of appliances apart from HVAC systems, mainly because they are located in cities and high-rise buildings with limited

space for installing renewable energy. While keeping these facts in perspective, the recent application of DRL-based BEMS in offices can be observed in Table 6.

Table 6. Recent applications of DRL-based BEMS in office buildings.

Ref	Year	BEMS	ESS	PV	DR	DRL	Estimator	Unique Objective	Real System	Energy */Cost Savings
[80]	2022	CHP, Battery, PV	o	o	x	DDPG	DNN	-	-	-
[81]	2022	HVAC	x	x	x	DQN	DNN	-	-	-
[82]	2022	HVAC, TES	o	o	o	SAC	DNN	Self-consumption/ Sufficiency	-	39.5–84.3%
[83]	2022	HVAC	x	x	x	PPO	DNN	-	-	48.97% *
[84]	2022	HVAC, PCSs	x	x	x	MAAC		-	-	0.7–4.18% *,▲
[85]	2022	Chiller, TES	o	x	o	SAC	DNN	Discomfort	-	-
[86]	2022	Battery, fan coil units	o	o	o	Dueling DQN	DNN	Discomfort	-	8%
[87]	2022	HVAC	x	x	x	A3C	DNN	-	-	16.10% *
[88]	2022	HVAC	x	x	x	A3C	DNN	-	-	12.80% *
[89]	2022	HVAC	x	x	x	BDQ	DNN	-	-	14% *,▲
[90]	2022	HVAC	x	x	x	PPO, A2C	DNN	Discomfort	-	4–22%*
[91]	2022	HVAC	x	x	x	DQN	DNN	Emissions	-	
[92]	2021	HVAC	x	x	x	DQN	DNN	-	-	**6% ***
[93]	2021	HVAC	x	x	x	DQN	DNN	Health	-	
[94]	2021	HVAC	x	x	o	SAC	DNN	-	-	9.70%
[95]	2021	HVAC, Blind	x	x	x	BDQN, SAC, PPO	DNN	-	-	11.0–31.8%
[96]	2021	EV	x	o	o	PPO	DNN	-	-	62.5% ▲
[97]	2021	HVAC	x	x	x	PPO	DNN	-	-	4.5–13.2%
[98]	2021	HVAC	x	x	x	SAC	DNN	Temperature violation	-	-
[99]	2021	HVAC, Battery	o	o	o	DDPG	DNN	-	-	39.60%
[100]	2020	HVAC	x	x	x	DQN	DNN	Health	-	15.70% *
[101]	2020	Water Heating	x	x	x	DDQN	DNN	-	-	5–12% ▲
[102]	2020	HVAC, Battery, EV, EWH	o	o	o	DQN	DNN	-	-	-
[103]	2020	HVAC	x	x	x	DDPG	DNN	-	-	27–30% ▲
[104]	2019	HVAC, Light, Blind	x	x	x	BDQ	DNN	-	-	8.1–14.26%
[105]	2019	HVAC	x	x	x	DQN	DNN	-	-	12.4–32.2% *
[106]	2019	HVAC	x	x	x	A3C	DNN	-	√	16.70% *
[107]	2018	HVAC	x	x	x	A3C	DNN	-	√	16.6–18.2% *
[108]	2018	HVAC	x	x	x	A3C	DNN	-	√	15% ▲

o: included, x: not included, * Energy saving, ▲ Over all energy/cost saving (Others are an improvement over a baseline Controller). Table Abbreviations: (A3C) Asynchronous advantage actor-critic; (BDQ) Branching-Dueling Q-network; (CHP) Combined heat and power; (EWH) Electric water heater; (MAAC) Multi-agent actor-critic; (PCS) Personal comfort systems.

The number of recent office building-related studies is comparable to that of residential buildings. The first difference can be noticed when observing the appliance category type, which is primarily related to HVAC systems. Only two studies investigated EVs, while few other control targets were investigated, such as TES, blind control, light control, and personal comfort systems (PCSs). HVAC systems are the main energy consumers in offices and have the flexibility and potential to save energy. In addition to HVAC control, recent

innovations can be found for BEMS integrated with EVs. Liang et al. included EVs in their BEMS that utilized a safe reinforcement learning (SRL) strategy to mitigate the effect of extreme weather events and increase building resilience and proactivity [102]. Meanwhile, Mbuwir et al. used EVs as their core and only a BEMS target in an office building, which revealed that by utilizing a multi-agent DRL; specifically, a promising saving potential of up to 62.5% can be achieved [96]. Furthermore, it can be noticed that only 24% of research considered DR systems, and only 21% included PV or energy storage systems.

The methods of DRL utilized in office buildings are more diversified than those observed in residential buildings, including the asynchronous advantage actor-critic (A3C) and the soft-actor critic (SAC), where their comparison has indicated improved performance over baseline, rule-based controllers, although one downside is that their comparison to other DRL has not always been considered. Zhang et al. introduced a branching–dueling Q-network (BDQN) and compared it to both PPO and SAC, where they reported that BDQN converged to the highest reward, followed by SAC, revealing higher sample complexity than their counterpart, although they performed slower than PPO, and consumed less memory. Hence, this revealed a trade between time, RAM usage, and reward. Another comparison between the advantage actor-critic (A2C) and PPO was conducted by Lee et al., where A2C exhibited better performance [90]. Such a comparison is useful in guiding researchers to choose the best subset of algorithms from the current large pool of DRL algorithms.

A critical observation related to office buildings is the significance of indoor thermal comfort in realizing the high productivity of workers. This can be observed in four studies that highlighted the reduction in discomfort or temperature violations as a system objective. Because there is less DR inclusion in the BEMS, a higher number of studies have reported energy savings rather than cost savings in comparison to residential buildings. Finally, only three studies conducted by Zhang et al. implemented and validated their models in real systems [106].

3.3. Educational Buildings

As depicted in Table 7, which shows recent research on educational buildings, they are mainly either schools or university facilities and laboratories. The target of the BEMS primarily focused on HVAC systems, and one study investigated TES control and other ventilation systems by controlling windows and air cleaners. Only two recent works included demand response systems with integrated energy storage, mainly TES. As for the objectives, health was considered by An et al., who deployed DQN to control ventilation in two laboratory rooms to achieve reduced economic loss and $PM_{2.5}$-related health risks [109]. This is an interesting co-benefit perspective to quantify not only energy and cost reduction, but also to quantify the impact on human health and integrate the findings into the BEMS objective. Furthermore, Chemingui et al. included the reduction of indoor contamination as a core target of their BEMS. This was realized by optimizing the HVAC system managing 21 zones in a school model, achieving 44% increased thermal comfort, 21% reduction in energy consumption, and low indoor CO_2 concentration [110]. Considering real implementations, three studies conducted real model validation: one in a laboratory setting, one in a university building, and another in a school setting. Laboratories are suitable for real-system validation, although acquiring data to train the agent can be challenging if the data does not already exist. In An et al., the approach was first to conduct an offline training phase based on an apartment model coupled with particle dynamics for $PM_{2.5}$ modelling, after which the trained agent was tested in a laboratory room with different $PM_{2.5}$ [109]. Schmidt et al. conducted a 43-day experiment in a Spanish school by deploying a BEMS utilizing a fitted Q-iteration and Bayesian regularized neural network coupled with genetic optimization. They confirmed that by maintaining comfort levels similar to the reference period, energy consumption decreased by almost 33%, and while prioritizing higher comfort, only a 5% energy increase was observed [111].

Table 7. Recent applications of DRL-based BEMS in educational buildings.

Ref	Year	Type	Scale	BEMS	ESS	PV	DR	DRL	Estimator	Unique Objective	Real System	Energy */Cost Savings
[112]	2022	University	Single	HVAC	o	x	o	PPO-Clip	DNN	-	-	9.17%
[113]	2022	University	Multi	TES	o	x	o	SAC	DNN	Load-Factor	-	6.72%
[109]	2022	University	Lab.	Ventilation	x	x	x	DQN	DNN	Health	√	2.4–43.7%
[114]	2022	University	Single	HVAC	x	x	x	SAC	DNN	-	√	-
[115]	2021	University	Multi	HVAC	x	x	x	DDPG	DNN	-	-	15.40% *,▲
[110]	2020	School	Single	HVAC	x	x	x	DDPG	DNN	Health	-	21% *,▲
[116]	2020	University	Single	HVAC	x	x	x	PPO	DNN	-	-	10.80% *
[111]	2017	School	Single	HVAC	x	x	x	fitted Q-iteration	-	-	√	33% *,▲

o: included, x: not included, * Energy saving, ▲ Overall energy/cost saving (others are an improvement over a baseline controller).

Finally, a recent innovative idea introduced by Zhou et al. combines DRL with deep learning for building energy prediction. It was not included in Table 7 because it is indirectly related to the BEMS. They utilized DDPG to add an additional learning layer to an LSTM forecaster by having the agent learn to tune the hyperparameters of the LSTM as new training data arrive. They demonstrated that when there is a high variation in the new training data, the prediction accuracy can be increased by up to 23.5% [117].

3.4. Datacenters

As listed in Table 8, few studies have investigated data centres. It was observed that the BEMS does not consider DR, renewable energy, or storage systems and is primarily focused on HVAC systems. In general, the main objective of the BEMS is to lower energy demand while meeting operational constraints, while comfort can be slightly compromised in other building types. As a system target, the operational efficiency of data centers is more sensitive as it can compromise the data center's main operation.

Table 8. Recent applications of DRL-based BEMS in datacenters.

Ref	Year	BEMS	ESS	PV	DR	DRL	Estimator	Unique Objective	Real System	Overall Energy Saving
[118]	2022	HVAC	x	x	x	SAC	DNN	Operation	-	3–5.5%
[119]	2022	HPC/AI Cluster	x	x	x	DQN	DNN	Operation	√	40%
[120]	2021	HVAC	x	x	x	SAC, PPO, TD3, TRPO	DNN	Operation	-	10%
[121]	2019	HVAC	x	x	x	DQN	DNN	Operation	-	-
[122]	2019	HVAC	x	x	x	Model-Based DRL, PPO	DNN	Operation	-	17.1–21.8%

o: included, x: not included. Table Abbreviations: (TRPO) Trust Region Policy Optimization.

One unique study implemented by Narantuya et al. utilized a multi-agent DRL (mDRL) based on a DQN to optimize computational resource allocation in high-performance computing (HPC)/AI systems. Their system was further deployed in real-time, reducing the task completion time by 20% and the energy consumption by 40% [119]. Finally, Beimann et al. conducted a comparative analysis of four different DRL methods for the control of a simulated HVAC system of a data centre. Their computational experimental results revealed that SAC has exceptionally high sample efficiency, reaching stable performance with 10 times less data required in comparison to PPO, TRP, and TD3; hence,

it is recommended for future utilization, particularly in noisy environments. Moreover, it was reported that all models can achieve an energy reduction of approximately 10% in comparison to a baseline controller [120].

3.5. Other Commercial Buildings

Finally, Table 9 includes commercial buildings that are not classified as educational, offices or data centres. Such types of buildings are introduced as either commercial buildings, storehouses, industrial parks, or a mix of (retail and restaurant buildings, offices, and residential) [123,124].

Table 9. Recent applications of DRL-based BEMS in other commercial buildings.

Ref	Year	Scale	BEMS	ESS	PV	DR	DRL	Estimator	Unique Objective	Real System	Energy */Cost Savings
[125]	2022	Single	HVAC	x	x	x	MA-CWSC, DQN	DNN	-	-	11.10% *
[126]	2022	Storehouse	HVAC	x	x	x	DDQN	DNN	-	-	34.20% *
[127]	2022	Industrial Park	HVAC	x	x	o	Dueling SAC	DNN	-	-	2.80% ▲
[123]	2022	Multi	HVAC, WHP, Inverter, Battery	o	o	o	PPO	DNN	Over/Under voltage	-	-
[128]	2022	Single	HVAC	x	x	x	DDQN	DNN	-	-	50% *
[16]	2021	Single	HVAC	x	x	o	MAAC	DNN	Health	-	56.50–75.25%
[124]	2021	Multi	HVAC, TES	o	o	o	SAC	DNN	-	-	7% *, 4%
[129]	2021	Multi	HVAC, TES	o	o	o	SAC	DNN	Peak	-	23% ▲
[130]	2020	Single	HVAC	x	x	o	A3C, Apex-DQN	DNN	-	-	-
[131]	2019	Single	HVAC	x	x	o	PPO	DNN	-	-	22% *

o: included, x: not included, * Energy saving, ▲ Over all energy/cost saving (others are an improvement over a baseline controller). Table Abbreviations: (MA-CWSC) Multi-Agent deep reinforcement learning method for the building Cooling Water System Control.

All of the studies listed in Table 9 investigated HVAC systems as the main BEMS target, while two studies included TES and one considered WHP and renewable energy inverters. DR systems were also included in seven studies, particularly in those with larger scales, such as industrial parks or multiple buildings. One notable method introduced was the dueling SAC-based memory-augmented DRL by Zhao et al. to overcome the limitation of time lag in district heating systems in an industrial park. Their novel methodology reduced the energy costs by 2.8% [127]. Furthermore, two multi-agent approaches were observed. First, Fu et al. utilized a multi-agent DRL method for developing a cooling water system control (MA-CWSC) to control the frequency of the cooling tower and cooling water pump in many chillers. Compared with the single-agent DQN, the proposed model had faster training and simpler action space, resulting in an 11.1% energy saving over the rule-based baseline [125]. Second, Yu et al. introduced a multi-agent actor-critic (MAAC) algorithm for a multi-zone HVAC system. Their objective was not only to minimize energy costs but also reduce the indoor CO_2 concentration in the building [16].

In terms of secondary objectives, Pigott et al. considered voltage regulations for a simulated IEEE-33 bus connected to nine buildings. The building types are diverse and include 37 fast-food restaurants, four medium offices, five retail stores, a mall, and 145 residential houses. These models were based on the recent CityLearn framework, which is a platform dedicated to multi-agent models in smart grids, and hence contains both building and power-flow models. Utilizing multiple DRL agents, their model nominally reduced the under-voltage instances and overvoltage occurrences by 34% [123]. Moreover,

Pinto et al. considered both peak demand and peak-to-average ratio, which were reduced by 23% and 20%, respectively, by using a centralized SAC agent controlling four different building types (small/medium offices, retail, and restaurant). Finally, in terms of real system validation, none was observed [129].

4. Discussion and Future Research Recommendations

As reported in the most recent literature, DRL-based BEMS have attracted significant research interest for investigation and proof of concept across a variety of building types. It is crucial to understand how the unique characteristics of different types of buildings correlate with both the opportunities and challenges of a new area of research, which is both promising and rapidly advancing.

4.1. Scaling up in Residential Buildings with Transfer Learning and Multi-Agent DRL

Regarding residential buildings, the main characteristics that were observed in the related literature include the high amount of research, large variety of appliances to explore, and large focus on DR integration as well as renewable energy. These characteristics are highly correlated with the unique characteristics of residential buildings:

- They are the highest energy consumer across all building types.
- They have high potential for renewable energy installations for detached houses owing to the availability of space.
- Their DR systems directly impact the user in terms of cost and benefit.
- They have a wide range of flexible and unique appliances.

This notion has been related to residential buildings in recent years and has been observed to be more integrated into their research in this area. Hence, many of these aspects have been well explored, and recent ideas are related to testing newer algorithms or increasing the complexity of the objectives to include health aspects or peak reductions. The next mile that has been less explored is the BEMS that are related to multiple buildings or residential districts, and on a higher scale, how it could be integrated with other building types, and the power grid. Future residential areas with a good share of renewable energy, EVs, and other energy storage systems will serve as vital energy trading hubs to maximize the benefits of renewable energy, primarily as they can sell surplus energy to other homes and buy it when it is cheaper than grid power. These systems, known as peer-to-peer energy trading systems [132] or community-level virtual power plants [133], can be further investigated using multi-agent and distributed DRL methods. Another promising direction for district residential areas is the utilization of transfer and federated learning. These approaches aim to transfer the learned control knowledge and model of a home's energy demand to either another home or global model. As investigated by Lee et al., having agents upload their local models to a server, aggregating them into one global mode, forwarding the model back to each home, and finally, performing local training can enhance the performance of the agents [74]. There are many opportunities to investigate these methods, particularly to resolve the requirement for large amounts of data for training the DRL agents.

4.2. Emphasizing Thermal Discomfort in the Office and Improving Baseline DRL Experience

As for office buildings, while having a considerable interest in the research community, very few studies have considered the DR aspect and EVs. Furthermore, the targets of the BEMS in the majority of research include HVAC systems or water thermal systems, while some have investigated lights, window blinds, and personal comfort systems. Personal comfort was observed as a unique characteristic that has emerged and has been highlighted in the literature related to office buildings. Only a few studies related to office buildings have reported the objective of reducing thermal discomfort. This is a remarkable observation in the office space, particularly from the perspective of DR systems, where the direct beneficiaries are the business owners; hence, ensuring that employees' comfort is not violated is crucial in these building types while trying to optimize energy efficiency.

Similar to residential buildings, the concept of transfer learning has been explored recently and can be further expanded. Zhang et al. reported that utilizing transfer learning for multi-agent DRL for multiple zone control of an HVAC system can reduce the energy consumption by 40.4% [83]. This improvement was observed even if the system policies originated in other buildings and were not locally retrained. Further investigations in this direction can allow databases to be obtained as baseline experiences for DRL controllers. If the baseline rises sufficiently high, this could result in requiring far less data, and in the best-case scenario, fully removing the need for offline training, thereby improving both generalization and scalability.

4.3. Achieving More with Less in the Laboratory and Exploring the Potential in Schools

There were three types of educational buildings: university offices and facilities, laboratory experiment setups, and schools. Regarding real-system validations, experimental setups serve as a good starting point to observe the potential of DRL-based BEMS outside simulation environments. This may require the research designer to decrease the complexity of their experiment in terms of BEMS targets according to the reproducible environment. As observed in An et al., their work targeted apartment ventilation using only windows and air cleaners, which was replicated in a laboratory setup for validation. Moreover, they demonstrated that even though the RL agent was trained on a residential building's data, it exhibited robust performance in the laboratory setup [109]. This validates both the generalization ability of DRL methods and the potential direction for additional studies to include such experimental validations in their investigations, which is still lacking in the literature. The long, 43-day real experiments conducted by Schmidt et al. in the Sierra Elvira School in Granada, Spain, demonstrated the effectiveness of their proposed RL methodology for HVAC control [111]. On the one hand, many schools might not have an HVAC system for every classroom, and on the other hand, there might not be clear standards or regulations on how such technologies can be utilized or experimented with in schools. Therefore, further investigations should be conducted in this area.

4.4. Integrating DR and Renewable Energy in Data Centre Applications

DeepMind's success in applying RL for cooling Google's datacenters in 2016 contributed significantly to the attraction for its applications in the context of BEMS. Existing research experiments have adopted the latest state-of-the-art methods of DRL in the cooling context of data centers, while one recent study applied DRL to the operation of the HPC/AI cluster in a building to reduce its energy demand. However, none of the studies included renewable energy, energy storage, or DR potential in their DRL. Whether such elements are effective in the context of a DRL-based BEMS could be further explored.

4.5. Beyond Building Types: Grid Level Integration of Various Buildings

The next step in scaling up DRL-based BEMS entails exploring their integration to power grids, as outlined in the recent work of Pigott et al. [123]. Hence, the objectives of such systems are explored beyond comfort and energy savings to achieve wide-scale operational power grid goals. Pigott et al. included 192 buildings of various types integrated within a distribution network and reported improvements in both under-and overvoltage instances [123]. Moreover, they discussed many other metrics to capture grid objectives, such as flattening the ramping of load demands, reducing the peak, and minimizing the overall net consumption. With the lack of such power-grid scale studies, further exploration of large-scale optimization with DRL-based BEMS, particularly utilizing multi-agent models, can be further investigated.

4.6. Challenges: A Hunger for Data and a Lack of Real Validation

The significant challenge of any advanced data-driven framework, particularly related to deep learning, is the requirement for large amounts of data to provide its full potential. This challenge was observed in many studies that required large amounts of training data

for offline training of their DRL agents to achieve a good performance level for deployment. An untrained DRL agent is similar to a newborn, without any knowledge of the real world. Hence, pre-training or offline training of the DRL agent before deployment in the real world is a critical step. Accordingly, the agent can operate sufficiently well to ensure comfort, health, and energy targets. Following the initial training phase, the agent can be trained online and adapted to a changing environment. Heidari et al. reported that with sufficient offline training, online or continuous learning is not necessary [56]. This further depicts the importance of having a sufficiently large dataset in the target environment before conducting research or being able to deploy such agents in the real world, which may not always be the case. The challenge of data-hungry paradigms can further feed into the problem of the lack of real-world implementations and validations. Approximately 11% of the reviewed literature published in 2022 has included real implementations or validations, which highlights a critical gap and should be emphasized. To implement such real systems, the related experiments can only be performed if a large amount of data related to the environment is available. As observed in the literature, this challenge can be approached from three main perspectives, as shown in Figure 6.

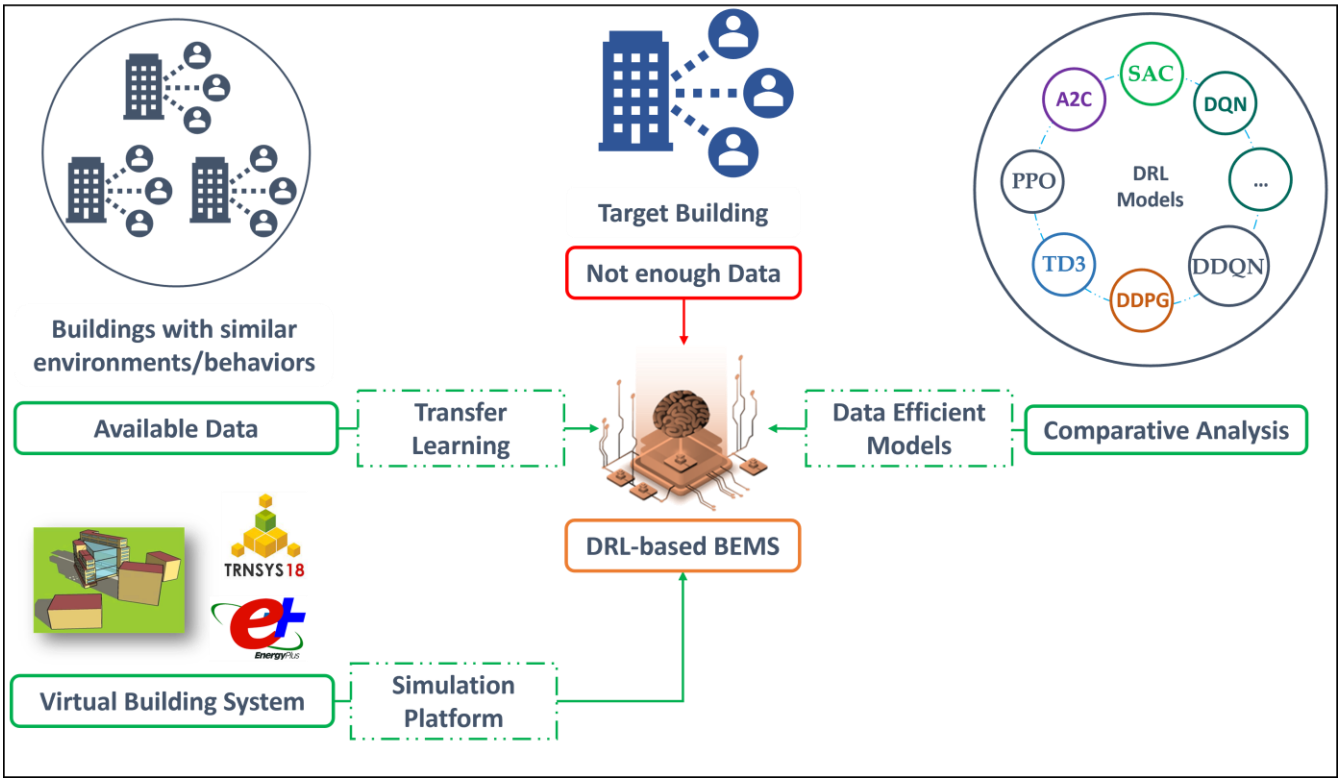

Figure 6. Approaches for the lack of data challenge.

While recent research in this area is starting to investigate the potential of transfer learning approaches, there is a lack of comparative analysis of the data efficiency potential of the different available DRL approaches in the context of BEMS. Very few studies, such as Beimann et al., have highlighted the relative amount of training data required for each model to achieve stable performance [120]. Hence, future research should investigate the data efficiency of DRL models in different contexts and building types to realize the best models that can mitigate the requirement for large amounts of data. Finally, virtual building systems can be used to simulate the behavior of buildings with unknown or little data for offline pretraining of the agent until it reaches a good base performance [134].

4.7. An Extra Adaptive Layer for Machine Learning Models with DRL for BEMS

Time-series forecasting has been extensively applied in the context of energy and buildings to forecast future energy demands, renewable power, surrounding weather conditions, human behaviours, and occupancies. ML and DL methods are used for this purpose, where they learn to map the input variables to the target variable by optimizing the internal model parameters. More recently, researchers have utilized DRL to add an extra layer of learnability to these methods. This was done by allowing them to further optimize their hyperparameters, as the model was retrained with new data, such as the previously discussed work by Zhou et al. [117]. Furthermore, Ramos et al. utilized RL to choose a suitable forecaster (DNN vs. KNN) at each time step, which is most likely to provide the best prediction based on different contexts. This is another form of adaptivity that can further reduce the prediction errors by utilizing RL [135]. Lie et al. used DRL to directly forecast future energy consumption and reported improved performance over conventional supervised methods [136,137]. Optimizing time-series forecasts will be key for advancing BEMS applications, and the combination of DRL with ML/DL models can be further investigated, even beyond forecasting applications to classification and clustering tasks.

4.8. CityLearn: A Multi-Agent RL Environment for Large-Scale BEMS Research

With the rapid growth in this field, the research community is contributing to accelerating research and implementation efforts for large-scale BEMS integration with RL. CityLearn is a recently developed environment that has been notably leveraged by many of the reviewed studies [75,76,78,123,129]. CityLearn is a python-based, open-source environment based on the OpenAI gym for conducting multi-agent RL-based BEMS simulations in cities. Its main objective is to facilitate the comparison of different RL algorithms in a standardized manner in the context of a multi-agent BEMS. Furthermore, it includes prebuilt models of different heating and cooling loads, as well as solar arrays and batteries [138]. Finally, since its inception in 2020, the yearly CityLearn challenge has attracted different teams and researchers to investigate the wide-scale deployments of RL-based BEMS in smart cities.

4.9. Reproducibility: Open Sourcing the Code and Data for Data-Driven BEMS

As with any research field, especially computational and data-driven research, being able to reproduce the results and conduct a comparative analysis to measure method improvements is a core part of the scientific endeavor [139,140]. It was observed that much of the research is conducted on different data sets and system configurations; hence, performance metrics can be less meaningful when cross-comparing them between studies, especially to validate new methods' performance. Open data is critical for validating improved DRL methods in the context of BEMS, and open codes can aid in reproducing such results as well as accelerating the progress of this research area similar to the previously mentioned CityLearn environment. Out of the large data base analyzed, very few researchers have notably open sourced and shared their codes such as Zhang et al. [95,106], Touzani et al. [99], and Svetozarevic et al. [25], while Marzullo et al. introduced a full open-source simulation environment for advanced building control performance testing [91]. Meanwhile a few other researchers noted the availability of their data on request [69,84,88,89]. As can be noticed, there are very few open-source contributions in DRL-based BEMS, as well as standardized open-data sets such as the Pecan Street database utilized by Yu et al. [13]. Hence, it is recommended for future researchers to consider the contribution to open source, in terms of their codes for reproducibility as well as for creating and utilizing standardized data sets.

5. Conclusions

In this study, a systematic review based on the PRISMA methodology was conducted for research on DRL-based BEMS in the context of different building types. Five major building types were identified from a pool of 470 papers: residential buildings, offices,

educational buildings, data centres, and other commercial buildings. The main goal was to investigate the relationship between the unique characteristics of each building type and their recent research landscape in the context of DRL-based BEMS. In doing so, the unique research directions can be more clearly identified, and the innovations applied in one building-type context may prove useful to another. First, it was observed that residential and office buildings were the most explored types of buildings, with residential buildings being the main energy consumers among other types. Second, detailed characteristics of each building in this context were identified, such as the emphasis on reducing discomfort in offices, lack of DR and renewable energy in data centre building research, and the direction toward realizing district-level BEMS in residential buildings. Third, the main challenge related to the need for large amounts of data was discussed, where recent studies approached this challenge using transfer learning and data-efficient DRL models. Finally, there is still a clear gap in real implementations and system validations, where only 11% of the recent works have been reported so far.

Author Contributions: Conceptualization, A.S. and A.H.; methodology, A.S. and A.H.; investigation, A.S.; data curation, A.S.; writing—original draft preparation, A.S.; writing—review and editing, A.H.; supervision, A.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Nomenclature

AI	Artificial Intelligence
A0C	Alphazero continuous
A2C	Advantage actor-critic
A3C	Asynchronous advantage actor-critic
ACKTR	Actor-critic kronecker-factored trust region
BCNN	Bayesian-Convolutional-Neural-Networks
BDQ	Branching-Dueling Q-network
BEMS	Building energy management systems
CI	Computational intelligence
CHP	Combined heat and power
DNN	Deep neural network
DRL	Deep reinforcement learning
DRLEM	Distributed reinforcement learning energy management
DPG	Deep policy gradient
DDPG	Deep deterministic policy gradient
DDQN	Double deep Q-learning
DDQN-PER	Double deep Q-learning prioritized experience replay
DQN	Deep Q-learning
EWH	Electric water heater
ESS	Energy storage system
EV	Electric vehicle
TES	Thermal energy storage
GA	Genetic Algorithm
HPC	High-performance computing
HVAC	Heating, ventilating, and air-conditioning
I2A	Imagination-augmented agents
KNN	K-nearest neighbors
MAAC	Multi-agent actor-critic
MDP	Markov decision process
MDRL	Mixed deep reinforcement learning
ML	Machine learning

MPC	Model predictive control
PCS	Personal comfort systems
PID	Proportional-integral-derivative
PPO	Proximal policy optimization
RL	Reinforcement learning
RLMPC	Reinforcement learning model predictive control
SAC	Soft actor-critic
SRL	Safe reinforcement learning
TD3	Twin delayed DDPG
TRPO	Trust Region Policy Optimization.
WHP	Water heat pump

References

1. United Nations Environment Programme. *2021 Global Status Report for Buildings and Construction: Towards a Zero-Emission, Efficient and Resilient Buildings and Construction Sector*; United Nations Environment Programme: Nairobi, Kenya, 2021.
2. Shaqour, A.; Farzaneh, H.; Yoshida, Y.; Hinokuma, T. Power Control and Simulation of a Building Integrated Stand-Alone Hybrid PV-Wind-Battery System in Kasuga City, Japan. *Energy Rep.* **2020**, *6*, 1528–1544. [CrossRef]
3. Cao, X.; Dai, X.; Liu, J. Building Energy-Consumption Status Worldwide and the State-of-the-Art Technologies for Zero-Energy Buildings during the Past Decade. *Energy Build.* **2016**, *128*, 198–213. [CrossRef]
4. Kamal, M.A. Material Characteristics and Building Physics for Energy Efficiency. *Key Eng. Mater.* **2015**, *666*, 77–87. [CrossRef]
5. Shaqour, A.; Farzaneh, H.; Almogdady, H. Day-Ahead Residential Electricity Demand Response Model Based on Deep Neural Networks for Peak Demand Reduction in the Jordanian Power Sector. *Appl. Sci.* **2021**, *11*, 6626. [CrossRef]
6. Yu, L.; Qin, S.; Zhang, M.; Shen, C.; Jiang, T.; Guan, X. A Review of Deep Reinforcement Learning for Smart Building Energy Management. *IEEE Internet Things J.* **2021**, *8*, 12046–12063. [CrossRef]
7. Chen, C.; Wang, J.; Heo, Y.; Kishore, S. MPC-Based Appliance Scheduling for Residential Building Energy Management Controller. *IEEE Trans. Smart Grid* **2013**, *4*, 1401–1410. [CrossRef]
8. Afram, A.; Janabi-Sharifi, F.; Fung, A.S.; Raahemifar, K. Artificial Neural Network (ANN) Based Model Predictive Control (MPC) and Optimization of HVAC Systems: A State of the Art Review and Case Study of a Residential HVAC System. *Energy Build.* **2017**, *141*, 96–113. [CrossRef]
9. Ben Romdhane, S.; Amamou, A.; ben Khalifa, R.; Saïd, N.M.; Younsi, Z.; Jemni, A. A Review on Thermal Energy Storage Using Phase Change Materials in Passive Building Applications. *J. Build. Eng.* **2020**, *32*, 101563. [CrossRef]
10. Li, T.; Dong, M. Residential Energy Storage Management with Bidirectional Energy Control. *IEEE Trans. Smart Grid* **2019**, *10*, 3596–3611. [CrossRef]
11. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
12. Lawal, K.; Rafsanjani, H.N. Trends, Benefits, Risks, and Challenges of IoT Implementation in Residential and Commercial Buildings. *Energy Built Environ.* **2022**, *3*, 251–266. [CrossRef]
13. Yu, L.; Xie, W.; Xie, D.; Zou, Y.; Zhang, D.; Sun, Z.; Zhang, L.; Zhang, Y.; Jiang, T. Deep Reinforcement Learning for Smart Home Energy Management. *IEEE Internet Things J.* **2020**, *7*, 2751–2762. [CrossRef]
14. Wei, T.; Wang, Y.; Zhu, Q. Deep reinforcement learning for building HVAC control. In Proceedings of the 54th Annual Design Automation Conference, Austin, TX, USA, 18–22 June 2017; Part 128280. [CrossRef]
15. Serda, M.; Becker, F.G.; Cleary, M.; Team, R.M.; Holtermann, H.; The, D.; Agenda, N.; Science, P.; Sk, S.K.; Hinnebusch, R.; et al. Comparative Analysis of White-, Gray- and Black-Box Models for Thermal Simulation of Indoor Environment: Teaching Building Case Study. *Uniw. Śląski* **2018**, *7*, 173–180. Available online: https://publications.ibpsa.org/conference/paper/?id=simbuild2018_C025 (accessed on 18 October 2022).
16. Yu, L.; Sun, Y.; Xu, Z.; Shen, C.; Yue, D.; Jiang, T.; Guan, X. Multi-Agent Deep Reinforcement Learning for HVAC Control in Commercial Buildings. *IEEE Trans. Smart Grid* **2021**, *12*, 407–419. [CrossRef]
17. Mocanu, E.; Mocanu, D.C.; Nguyen, P.H.; Liotta, A.; Webber, M.E.; Gibescu, M.; Slootweg, J.G. On-Line Building Energy Optimization Using Deep Reinforcement Learning. *IEEE Trans. Smart Grid* **2019**, *10*, 3698–3708. [CrossRef]
18. Gao, G.; Li, J.; Wen, Y. DeepComfort: Energy-Efficient Thermal Comfort Control in Buildings Via Reinforcement Learning. *IEEE Internet Things J.* **2020**, *7*, 8472–8484. [CrossRef]
19. Xu, Z.; Jia, Q.S.; Guan, X.; Xie, X. A new method to solve large-scale building energy management for energy saving. In Proceedings of the IEEE International Conference on Automation Science and Engineering (CASE), New Taipei, Taiwan, 18–22 August 2014; pp. 940–945. [CrossRef]
20. Shaqour, A.; Ono, T.; Hagishima, A.; Farzaneh, H. Electrical Demand Aggregation Effects on the Performance of Deep Learning-Based Short-Term Load Forecasting of a Residential Building. *Energy AI* **2022**, *8*, 100141. [CrossRef]
21. Alanne, K.; Sierla, S. An Overview of Machine Learning Applications for Smart Buildings. *Sustain. Cities Soc.* **2022**, *76*, 103445. [CrossRef]
22. Araya, D.B.; Grolinger, K.; ElYamany, H.F.; Capretz, M.A.M.; Bitsuamlak, G. An Ensemble Learning Framework for Anomaly Detection in Building Energy Consumption. *Energy Build.* **2017**, *144*, 191–206. [CrossRef]

23. Sutton, R.S.; Barto, A.G. *Reinforcement Learning: An Introduction*; MIT Press: Cambridge, MA, USA, 2018; ISBN 0262352702.
24. Deep Mind. DeepMind AI Reduces Google Data Centre Cooling Bill by 40%. Available online: https://www.deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-by-40 (accessed on 11 October 2022).
25. Svetozarevic, B.; Baumann, C.; Muntwiler, S.; di Natale, L.; Zeilinger, M.N.; Heer, P. Data-Driven Control of Room Temperature and Bidirectional EV Charging Using Deep Reinforcement Learning: Simulations and Experiments. *Appl. Energy* **2022**, *307*, 118127. [CrossRef]
26. Wang, Z.; Hong, T. Reinforcement Learning for Building Controls: The Opportunities and Challenges. *Appl. Energy* **2020**, *269*, 115036. [CrossRef]
27. Fu, Q.; Han, Z.; Chen, J.; Lu, Y.; Wu, H.; Wang, Y. Applications of Reinforcement Learning for Building Energy Efficiency Control: A Review. *J. Build. Eng.* **2022**, *50*, 104165. [CrossRef]
28. Perera, A.T.D.; Kamalaruban, P. Applications of Reinforcement Learning in Energy Systems. *Renew. Sustain. Energy Rev.* **2021**, *137*, 110618. [CrossRef]
29. Rajasekhar, B.; Tushar, W.; Lork, C.; Zhou, Y.; Yuen, C.; Pindoriya, N.M.; Wood, K.L. A Survey of Computational Intelligence Techniques for Air-Conditioners Energy Management. *IEEE Trans. Emerg. Top. Comput. Intell.* **2020**, *4*, 555–570. [CrossRef]
30. Leitao, J.; Gil, P.; Ribeiro, B.; Cardoso, A. A Survey on Home Energy Management. *IEEE Access* **2020**, *8*, 5699–5722. [CrossRef]
31. Carlucci, S.; de Simone, M.; Firth, S.K.; Kjærgaard, M.B.; Markovic, R.; Rahaman, M.S.; Annaqeeb, M.K.; Biandrate, S.; Das, A.; Dziedzic, J.W.; et al. Modeling Occupant Behavior in Buildings. *Build. Environ.* **2020**, *174*, 106768. [CrossRef]
32. Zhang, Z.; Zhang, D.; Qiu, R.C. Deep Reinforcement Learning for Power System: An Overview. *CSEE J. Power Energy Syst.* **2019**, *6*, 213–225. [CrossRef]
33. Han, M.; May, R.; Zhang, X.; Wang, X.; Pan, S.; Yan, D.; Jin, Y.; Xu, L. A Review of Reinforcement Learning Methodologies for Controlling Occupant Comfort in Buildings. *Sustain. Cities Soc.* **2019**, *51*, 101748. [CrossRef]
34. Mason, K.; Grijalva, S. A Review of Reinforcement Learning for Autonomous Building Energy Management. *Comput. Electr. Eng.* **2019**, *78*, 300–312. [CrossRef]
35. Han, M.; Zhao, J.; Zhang, X.; Shen, J.; Li, Y. The Reinforcement Learning Method for Occupant Behavior in Building Control: A Review. *Energy Built Environ.* **2021**, *2*, 137–148. [CrossRef]
36. Alexandropoulos, G.C.; Stylianopoulos, K.; Huang, C.; Yuen, C.; Bennis, M.; Debbah, M. Pervasive Machine Learning for Smart Radio Environments Enabled by Reconfigurable Intelligent Surfaces. *Proc. IEEE* **2022**, *110*, 1494–1525. [CrossRef]
37. Fridman, L. Introduction to Deep RL. Available online: https://deeplearning.mit.edu/ (accessed on 16 October 2022).
38. Achiam, J. Spinning up in Deep Reinforcement Learning. Available online: https://github.com/openai/spinningup (accessed on 15 July 2022).
39. Silver, D.; Hubert, T.; Schrittwieser, J.; Antonoglou, I.; Lai, M.; Guez, A.; Lanctot, M.; Sifre, L.; Kumaran, D.; Graepel, T.; et al. Mastering Chess and Shogi by Self-Play with a General Reinforcement Learning Algorithm. *arXiv* **2017**, arXiv:1712.01815.
40. Moerland, T.M.; Broekens, J.; Plaat, A.; Jonker, C.M. A0C: Alpha Zero in Continuous Action Space. *arXiv* **2018**, arXiv:1805.09613.
41. Ha, D.; Schmidhuber, J. World Models. *Forecast. Bus. Econ.* **2018**, 201–209. [CrossRef]
42. Racanière, S.; Weber, T.; Reichert, D.P.; Buesing, L.; Guez, A.; Rezende, D.; Badia, A.P.; Vinyals, O.; Heess, N.; Li, Y.; et al. Imagination-augmented agents for deep reinforcement learning. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5691–5702. [CrossRef]
43. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; Riedmiller, M. Playing Atari with Deep Reinforcement Learning. *arXiv* **2013**, arXiv:1312.5602.
44. Van Hasselt, H.; Guez, A.; Silver, D. Deep Reinforcement Learning with Double Q-Learning. In Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI 2016), Phoenix, AZ, USA, 12–17 February 2015; pp. 2094–2100. [CrossRef]
45. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Openai, O.K. Proximal Policy Optimization Algorithms. *arXiv* **2017**, arXiv:1707.06347.
46. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous Control with Deep Reinforcement Learning. *arXiv* **2015**, arXiv:1509.02971.
47. Williams, R.J. Simple Statistical Gradient-Following Algorithms for Connectionist Reinforcement Learning. *Mach. Learn.* **1992**, *8*, 229–256. [CrossRef]
48. Denyer, D.; Tranfield, D. Producing a systematic review. In *The Sage Handbook of Organizational Research Methods*; Hardcover; Sage Publications Ltd.: Thousand Oaks, CA, USA, 2009; pp. 671–689, ISBN 978-1-4129-3118-2.
49. Denyer, D.; Tranfield, D.; van Aken, J.E. Developing Design Propositions through Research Synthesis. *Organ. Stud.* **2008**, *29*, 393–413. [CrossRef]
50. Liberati, A.; Altman, D.G.; Tetzlaff, J.; Mulrow, C.; Gøtzsche, P.C.; Ioannidis, J.P.A.; Clarke, M.; Devereaux, P.J.; Kleijnen, J.; Moher, D. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLoS Med.* **2009**, *6*, e1000100. [CrossRef]
51. Blad, C.; Bøgh, S.; Kallesøe, C.S. Data-Driven Offline Reinforcement Learning for HVAC-Systems. *Energy* **2022**, *261*, 125290. [CrossRef]
52. Du, Y.; Li, F.; Kurte, K.; Munk, J.; Zandi, H. Demonstration of Intelligent HVAC Load Management with Deep Reinforcement Learning: Real-World Experience of Machine Learning in Demand Control. *IEEE Power Energy Mag.* **2022**, *20*, 42–53. [CrossRef]

53. Chu, Y.; Wei, Z.; Sun, G.; Zang, H.; Chen, S.; Zhou, Y. Optimal Home Energy Management Strategy: A Reinforcement Learning Method with Actor-Critic Using Kronecker-Factored Trust Region. *Electr. Power Syst. Res.* **2022**, *212*, 108617. [CrossRef]
54. Zenginis, I.; Vardakas, J.; Koltsaklis, N.E.; Verikoukis, C. Smart Home's Energy Management Through a Clustering-Based Reinforcement Learning Approach. *IEEE Internet Things J.* **2022**, *9*, 16363–16371. [CrossRef]
55. Lu, J.; Mannion, P.; Mason, K. A Multi-Objective Multi-Agent Deep Reinforcement Learning Approach to Residential Appliance Scheduling. *IET Smart Grid* **2022**, *5*, 260–280. [CrossRef]
56. Heidari, A.; Maréchal, F.; Khovalyg, D. Reinforcement Learning for Proactive Operation of Residential Energy Systems by Learning Stochastic Occupant Behavior and Fluctuating Solar Energy: Balancing Comfort, Hygiene and Energy Use. *Appl. Energy* **2022**, *318*, 119206. [CrossRef]
57. Shuvo, S.S.; Yilmaz, Y. Home Energy Recommendation System (HERS): A Deep Reinforcement Learning Method Based on Residents' Feedback and Activity. *IEEE Trans. Smart Grid* **2022**, *13*, 2812–2821. [CrossRef]
58. Huang, C.; Zhang, H.; Wang, L.; Luo, X.; Song, Y. Mixed Deep Reinforcement Learning Considering Discrete-Continuous Hybrid Action Space for Smart Home Energy Management. *J. Mod. Power Syst. Clean Energy* **2022**, *10*, 743–754. [CrossRef]
59. Heidari, A.; Maréchal, F.; Khovalyg, D. An Occupant-Centric Control Framework for Balancing Comfort, Energy Use and Hygiene in Hot Water Systems: A Model-Free Reinforcement Learning Approach. *Appl. Energy* **2022**, *312*, 118833. [CrossRef]
60. Forootani, A.; Rastegar, M.; Jooshaki, M. An Advanced Satisfaction-Based Home Energy Management System Using Deep Reinforcement Learning. *IEEE Access* **2022**, *10*, 47896–47905. [CrossRef]
61. Kurte, K.; Amasyali, K.; Munk, J.; Zandi, H. Comparative analysis of model-free and model-based HVAC control for residential demand response. In Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Coimbra, Portugal, 17–18 November 2021; ACM: New York, NY, USA, 2021; pp. 309–313.
62. Amasyali, K.; Munk, J.; Kurte, K.; Kuruganti, T.; Zandi, H. Deep Reinforcement Learning for Autonomous Water Heater Control. *Buildings* **2021**, *11*, 548. [CrossRef]
63. Yang, T.; Zhao, L.; Li, W.; Wu, J.; Zomaya, A.Y. Towards Healthy and Cost-Effective Indoor Environment Management in Smart Homes: A Deep Reinforcement Learning Approach. *Appl. Energy* **2021**, *300*, 117335. [CrossRef]
64. Liu, B.; Akcakaya, M.; McDermott, T.E. Automated Control of Transactive HVACs in Energy Distribution Systems. *IEEE Trans. Smart Grid* **2021**, *12*, 2462–2471. [CrossRef]
65. Ye, Y.; Qiu, D.; Wang, H.; Tang, Y.; Strbac, G. Real-Time Autonomous Residential Demand Response Management Based on Twin Delayed Deep Deterministic Policy Gradient Learning. *Energy* **2021**, *14*, 531. [CrossRef]
66. Mathew, A.; Roy, A.; Mathew, J. Intelligent Residential Energy Management System Using Deep Reinforcement Learning. *IEEE Syst. J.* **2020**, *14*, 5362–5372. [CrossRef]
67. McKee, E.; Du, Y.; Li, F.; Munk, J.; Johnston, T.; Kurte, K.; Kotevska, O.; Amasyali, K.; Zandi, H. Deep reinforcement learning for residential HVAC control with consideration of human occupancy. In Proceedings of the 2020 IEEE Power & Energy Society General Meeting (PESGM), Montreal, QC, Canada, 2–6 August 2020; pp. 1–5.
68. Arroyo, J.; Manna, C.; Spiessens, F.; Helsen, L. Reinforced Model Predictive Control (RL-MPC) for Building Energy Management. *Appl. Energy* **2022**, *309*, 118346. [CrossRef]
69. Zsembinszki, G.; Fernández, C.; Vérez, D.; Cabeza, L.F.; Cannavale, A.; Martellotta, F.; Fiorito, F. Deep Learning Optimal Control for a Complex Hybrid Energy Storage System. *Buildings* **2021**, *11*, 194. [CrossRef]
70. Park, B.; Rempel, A.R.; Lai, A.K.L.; Chiaramonte, J.; Mishra, S. Reinforcement Learning for Control of Passive Heating and Cooling in Buildings. *IFAC-Papers* **2021**, *54*, 907–912. [CrossRef]
71. Kurte, K.; Munk, J.; Kotevska, O.; Amasyali, K.; Smith, R.; McKee, E.; Du, Y.; Cui, B.; Kuruganti, T.; Zandi, H. Evaluating the Adaptability of Reinforcement Learning Based HVAC Control for Residential Houses. *Sustainability* **2020**, *12*, 7727. [CrossRef]
72. Lork, C.; Li, W.T.; Qin, Y.; Zhou, Y.; Yuen, C.; Tushar, W.; Saha, T.K. An Uncertainty-Aware Deep Reinforcement Learning Framework for Residential Air Conditioning Energy Management. *Appl. Energy* **2020**, *276*, 115426. [CrossRef]
73. Ahrarinouri, M.; Rastegar, M.; Karami, K.; Seifi, A.R. Distributed Reinforcement Learning Energy Management Approach in Multiple Residential Energy Hubs. *Sustain. Energy Grids Netw.* **2022**, *32*, 100795. [CrossRef]
74. Lee, S.; Choi, D.H. Federated Reinforcement Learning for Energy Management of Multiple Smart Homes with Distributed Energy Resources. *IEEE Trans. Ind. Inf.* **2022**, *18*, 488–497. [CrossRef]
75. Pinto, G.; Kathirgamanathan, A.; Mangina, E.; Finn, D.P.; Capozzoli, A. Enhancing Energy Management in Grid-Interactive Buildings: A Comparison among Cooperative and Coordinated Architectures. *Appl. Energy* **2022**, *310*, 118497. [CrossRef]
76. Glatt, R.; da Silva, F.L.; Soper, B.; Dawson, W.A.; Rusu, E.; Goldhahn, R.A. Collaborative energy demand response with decentralized actor and centralized critic. In Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Coimbra, Portugal, 17–18 November 2021; ACM: New York, NY, USA, 2021; pp. 333–337.
77. Gupta, A.; Badr, Y.; Negahban, A.; Qiu, R.G. Energy-Efficient Heating Control for Smart Buildings with Deep Reinforcement Learning. *J. Build. Eng.* **2021**, *34*, 101739. [CrossRef]
78. Kathirgamanathan, A.; Twardowski, K.; Mangina, E.; Finn, D.P. A Centralised soft actor critic deep reinforcement learning approach to district demand side management through CityLearn. In Proceedings of the Proceedings of the 1st International Workshop on Reinforcement Learning for Energy Management in Buildings & Cities, Online, 17 November 2020; ACM: New York, NY, USA, 2020; pp. 11–14.

79. Torriti, J.; Zhao, X.; Yuan, Y. The Risk of Residential Peak Electricity Demand: A Comparison of Five European Countries. *Energies* **2017**, *10*, 385. [CrossRef]
80. Gao, Y.; Matsunami, Y.; Miyata, S.; Akashi, Y. Operational Optimization for Off-Grid Renewable Building Energy System Using Deep Reinforcement Learning. *Appl. Energy* **2022**, *325*, 119783. [CrossRef]
81. Fang, X.; Gong, G.; Li, G.; Chun, L.; Peng, P.; Li, W.; Shi, X.; Chen, X. Deep Reinforcement Learning Optimal Control Strategy for Temperature Setpoint Real-Time Reset in Multi-Zone Building HVAC System. *Appl. Eng.* **2022**, *212*, 118552. [CrossRef]
82. Brandi, S.; Gallo, A.; Capozzoli, A. A Predictive and Adaptive Control Strategy to Optimize the Management of Integrated Energy Systems in Buildings. *Energy Rep.* **2022**, *8*, 1550–1567. [CrossRef]
83. Zhang, T.; Aakash Krishna, G.S.; Afshari, M.; Musilek, P.; Taylor, M.E.; Ardakanian, O. Diversity for transfer in learning-based control of buildings. In Proceedings of the Thirteenth ACM International Conference on Future Energy Systems, Online, 28 June–1 July 2022; ACM: New York, NY, USA, 2022; pp. 556–564.
84. Yu, L.; Xu, Z.; Zhang, T.; Guan, X.; Yue, D. Energy-Efficient Personalized Thermal Comfort Control in Office Buildings Based on Multi-Agent Deep Reinforcement Learning. *Build. Environ.* **2022**, *223*, 109458. [CrossRef]
85. Brandi, S.; Fiorentini, M.; Capozzoli, A. Comparison of Online and Offline Deep Reinforcement Learning with Model Predictive Control for Thermal Energy Management. *Autom. Constr.* **2022**, *135*, 104128. [CrossRef]
86. Shen, R.; Zhong, S.; Wen, X.; An, Q.; Zheng, R.; Li, Y.; Zhao, J. Multi-Agent Deep Reinforcement Learning Optimization Framework for Building Energy System with Renewable Energy. *Appl. Energy* **2022**, *312*, 118724. [CrossRef]
87. Zhang, W.; Zhang, Z. Energy Efficient Operation Optimization of Building Air-Conditioners via Simulator-Assisted Asynchronous Reinforcement Learning. *IOP Conf. Ser. Earth Environ. Sci* **2022**, *1048*, 012006. [CrossRef]
88. Zhong, X.; Zhang, Z.; Zhang, R.; Zhang, C. End-to-End Deep Reinforcement Learning Control for HVAC Systems in Office Buildings. *Designs* **2022**, *6*, 52. [CrossRef]
89. Lei, Y.; Zhan, S.; Ono, E.; Peng, Y.; Zhang, Z.; Hasama, T.; Chong, A. A Practical Deep Reinforcement Learning Framework for Multivariate Occupant-Centric Control in Buildings. *Appl. Energy* **2022**, *324*, 119742. [CrossRef]
90. Lee, J.Y.; Rahman, A.; Huang, S.; Smith, A.D.; Katipamula, S. On-Policy Learning-Based Deep Reinforcement Learning Assessment for Building Control Efficiency and Stability. *Sci. Technol. Built Environ.* **2022**, *28*, 1150–1165. [CrossRef]
91. Marzullo, T.; Dey, S.; Long, N.; Leiva Vilaplana, J.; Henze, G. A High-Fidelity Building Performance Simulation Test Bed for the Development and Evaluation of Advanced Controls. *J. Build. Perform. Simul.* **2022**, *15*, 379–397. [CrossRef]
92. Verma, S.; Agrawal, S.; Venkatesh, R.; Shrotri, U.; Nagarathinam, S.; Jayaprakash, R.; Dutta, A. EImprove—Optimizing energy and comfort in buildings based on formal semantics and reinforcement learning. In Proceedings of the 58th ACM/IEEE Design Automation Conference (DAC), Online, 5–9 December 2021; pp. 157–162.
93. Jneid, K.; Ploix, S.; Reignier, P.; Jallon, P. Deep Q-network boosted with external knowledge for HVAC control. In Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Coimbra, Portugal, 17–18 November 2021; ACM: New York, NY, USA, 2021; pp. 329–332.
94. Kathirgamanathan, A.; Mangina, E.; Finn, D.P. Development of a Soft Actor Critic Deep Reinforcement Learning Approach for Harnessing Energy Flexibility in a Large Office Building. *Energy AI* **2021**, *5*, 100101. [CrossRef]
95. Zhang, T.; Baasch, G.; Ardakanian, O.; Evins, R. On the joint control of multiple building systems with reinforcement learning. In Proceedings of the Twelfth ACM International Conference on Future Energy Systems, Online, 28 June–1 July 2021; ACM: New York, NY, USA, 2021; pp. 60–72.
96. Mbuwir, B.v.; Vanmunster, L.; Thoelen, K.; Deconinck, G. A Hybrid Policy Gradient and Rule-Based Control Framework for Electric Vehicle Charging. *Energy AI* **2021**, *4*, 100059. [CrossRef]
97. Zhang, X.; Chintala, R.; Bernstein, A.; Graf, P.; Jin, X. Grid-interactive multi-zone building control using reinforcement learning with global-local policy search. In Proceedings of the American Control Conference (ACC), Online, 25–28 May 2021; pp. 4155–4162.
98. Coraci, D.; Brandi, S.; Piscitelli, M.S.; Capozzoli, A. Online Implementation of a Soft Actor-Critic Agent to Enhance Indoor Temperature Control and Energy Efficiency in Buildings. *Energies* **2021**, *14*, 997. [CrossRef]
99. Touzani, S.; Prakash, A.K.; Wang, Z.; Agarwal, S.; Pritoni, M.; Kiran, M.; Brown, R.; Granderson, J. Controlling Distributed Energy Resources via Deep Reinforcement Learning for Load Flexibility and Energy Efficiency. *Appl. Energy* **2021**, *304*, 117733. [CrossRef]
100. Ahn, K.U.; Park, C.S. Application of Deep Q-Networks for Model-Free Optimal Control Balancing between Different HVAC Systems. *Sci. Technol. Built Environ.* **2020**, *26*, 61–74. [CrossRef]
101. Brandi, S.; Piscitelli, M.S.; Martellacci, M.; Capozzoli, A. Deep Reinforcement Learning to Optimise Indoor Temperature Control and Heating Energy Consumption in Buildings. *Energy Build.* **2020**, *224*, 110225. [CrossRef]
102. Liang, Z.; Huang, C.; Su, W.; Duan, N.; Donde, V.; Wang, B.; Zhao, X. Safe Reinforcement Learning-Based Resilient Proactive Scheduling for a Commercial Building Considering Correlated Demand Response. *IEEE Open Access J. Power Energy* **2021**, *8*, 85–96. [CrossRef]
103. Zou, Z.; Yu, X.; Ergan, S. Towards Optimal Control of Air Handling Units Using Deep Reinforcement Learning and Recurrent Neural Network. *Build. Environ.* **2020**, *168*, 106535. [CrossRef]
104. Ding, X.; Du, W.; Cerpa, A. OCTOPUS: Deep reinforcement learning for holistic smart building control. In Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, New York, NY, USA, 13–14 November 2019; ACM: New York, NY, USA, 2019; pp. 326–335.

105. Yoon, Y.R.; Moon, H.J. Performance Based Thermal Comfort Control (PTCC) Using Deep Reinforcement Learning for Space Cooling. *Energy Build.* **2019**, *203*, 109420. [CrossRef]
106. Zhang, Z.; Chong, A.; Pan, Y.; Zhang, C.; Lam, K.P. Whole Building Energy Model for HVAC Optimal Control: A Practical Framework Based on Deep Reinforcement Learning. *Energy Build.* **2019**, *199*, 472–490. [CrossRef]
107. Zhang, Z.; Lam, K.P. Practical implementation and evaluation of deep reinforcement learning control for a radiant heating system. In Proceedings of the 5th Conference on Systems for Built Environments, Shenzen, China, 7–8 November 2018; ACM: New York, NY, USA, 2018; pp. 148–157.
108. Zhang, Z.; Chong, A.; Pan, Y.; Zhang, C.; Lu, S.; Lam, K. A Deep reinforcement learning approach to using whole building energy model for HVAC optimal control. In Proceedings of the ASHRAE/IBPSA-USA Building Performance Analysis Conference and SimBuild, Chicago, IL, USA, 26–28 September 2018.
109. An, Y.; Niu, Z.; Chen, C. Smart Control of Window and Air Cleaner for Mitigating Indoor PM2.5 with Reduced Energy Consumption Based on Deep Reinforcement Learning. *Build. Environ.* **2022**, *224*, 109583. [CrossRef]
110. Chemingui, Y.; Gastli, A.; Ellabban, O. Reinforcement Learning-Based School Energy Management System. *Energies* **2020**, *13*, 6354. [CrossRef]
111. Schmidt, M.; Moreno, M.V.; Schülke, A.; Macek, K.; Mařík, K.; Pastor, A.G. Optimizing Legacy Building Operation: The Evolution into Data-Driven Predictive Cyber-Physical Systems. *Energy Build.* **2017**, *148*, 257–279. [CrossRef]
112. Li, Z.; Sun, Z.; Meng, Q.; Wang, Y.; Li, Y. Reinforcement Learning of Room Temperature Set-Point of Thermal Storage Air-Conditioning System with Demand Response. *Energy Build.* **2022**, *259*, 111903. [CrossRef]
113. Qin, Y.; Ke, J.; Wang, B.; Filaretov, G.F. Energy Optimization for Regional Buildings Based on Distributed Reinforcement Learning. *Sustain. Cities Soc.* **2022**, *78*, 103625. [CrossRef]
114. Jung, S.; Jeoung, J.; Hong, T. Occupant-Centered Real-Time Control of Indoor Temperature Using Deep Learning Algorithms. *Build. Environ.* **2022**, *208*, 108633. [CrossRef]
115. Li, J.; Zhang, W.; Gao, G.; Wen, Y.; Jin, G.; Christopoulos, G. Toward Intelligent Multizone Thermal Control with Multiagent Deep Reinforcement Learning. *IEEE Internet Things J.* **2021**, *8*, 11150–11162. [CrossRef]
116. Naug, A.; Quiñones-Grueiro, M.; Biswas, G. Continual adaptation in deep reinforcement learning-based control applied to non-stationary building environments. In *Proceedings the 1st International Workshop on Reinforcement Learning for Energy Management in Buildings & Cities, Online, 17 November 2020*; ACM: New York, NY, USA, 2020; pp. 24–28.
117. Zhou, X.; Lin, W.; Kumar, R.; Cui, P.; Ma, Z. A Data-Driven Strategy Using Long Short Term Memory Models and Reinforcement Learning to Predict Building Electricity Consumption. *Appl. Energy* **2022**, *306*, 118078. [CrossRef]
118. Bin Mahbod, M.H.; Chng, C.B.; Lee, P.S.; Chui, C.K. Energy Saving Evaluation of an Energy Efficient Data Center Using a Model-Free Reinforcement Learning Approach. *Appl. Energy* **2022**, *322*, 119392. [CrossRef]
119. Narantuya, J.; Shin, J.S.; Park, S.; Kim, J.W. Multi-Agent Deep Reinforcement Learning-Based Resource Allocation in HPC/AI Converged Cluster. *Comput. Mater. Contin.* **2022**, *72*, 4375–4395. [CrossRef]
120. Biemann, M.; Scheller, F.; Liu, X.; Huang, L. Experimental Evaluation of Model-Free Reinforcement Learning Algorithms for Continuous HVAC Control. *Appl. Energy* **2021**, *298*, 117164. [CrossRef]
121. Van Le, D.; Liu, Y.; Wang, R.; Tan, R.; Wong, Y.-W.; Wen, Y. Control of air free-cooled data centers in tropics via deep reinforcement learning. In Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, New York, NY, USA, 13–14 November 2019; ACM: New York, NY, USA, 2019; pp. 306–315.
122. Zhang, C.; Kuppannagari, S.R.; Kannan, R.; Prasanna, V.K. Building HVAC scheduling using reinforcement learning via neural network based model approximation. In Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, New York, NY, USA, 13–14 November 2019; ACM: New York, NY, USA, 2019; pp. 287–296.
123. Pigott, A.; Crozier, C.; Baker, K.; Nagy, Z. GridLearn: Multiagent Reinforcement Learning for Grid-Aware Building Energy Management. *Electr. Power Syst. Res.* **2021**, *213*, 108521. [CrossRef]
124. Deltetto, D.; Coraci, D.; Pinto, G.; Piscitelli, M.S.; Capozzoli, A. Exploring the Potentialities of Deep Reinforcement Learning for Incentive-Based Demand Response in a Cluster of Small Commercial Buildings. *Energies* **2021**, *14*, 2933. [CrossRef]
125. Fu, Q.; Chen, X.; Ma, S.; Fang, N.; Xing, B.; Chen, J. Optimal Control Method of HVAC Based on Multi-Agent Deep Reinforcement Learning. *Energy Build.* **2022**, *270*, 112284. [CrossRef]
126. Sun, Y.; Zhang, Y.; Guo, D.; Zhang, X.; Lai, Y.; Luo, D. Intelligent Distributed Temperature and Humidity Control Mechanism for Uniformity and Precision in the Indoor Environment. *IEEE Internet Things J.* **2022**, *9*, 19101–19115. [CrossRef]
127. Zhao, H.; Wang, B.; Liu, H.; Sun, H.; Pan, Z.; Guo, Q. Exploiting the Flexibility Inside Park-Level Commercial Buildings Considering Heat Transfer Time Delay: A Memory-Augmented Deep Reinforcement Learning Approach. *IEEE Trans. Sustain. Energy* **2022**, *13*, 207–219. [CrossRef]
128. Xu, D. Learning Efficient Dynamic Controller for HVAC System. *Mob. Inf. Syst.* **2022**, *2022*, 4157511. [CrossRef]
129. Pinto, G.; Deltetto, D.; Capozzoli, A. Data-Driven District Energy Management with Surrogate Models and Deep Reinforcement Learning. *Appl. Energy* **2021**, *304*, 117642. [CrossRef]
130. Zhang, X.; Biagioni, D.; Cai, M.; Graf, P.; Rahman, S. An Edge-Cloud Integrated Solution for Buildings Demand Response Using Reinforcement Learning. *IEEE Trans. Smart Grid* **2021**, *12*, 420–431. [CrossRef]
131. Azuatalam, D.; Lee, W.L.; de Nijs, F.; Liebman, A. Reinforcement Learning for Whole-Building HVAC Control and Demand Response. *Energy AI* **2020**, *2*, 100020. [CrossRef]

132. Kanakadhurga, D.; Prabaharan, N. Demand Response-Based Peer-to-Peer Energy Trading among the Prosumers and Consumers. *Energy Rep.* **2021**, *7*, 7825–7834. [CrossRef]
133. Monie, S.; Nilsson, A.M.; Widén, J.; Åberg, M. A Residential Community-Level Virtual Power Plant to Balance Variable Renewable Power Generation in Sweden. *Energy Convers. Manag.* **2021**, *228*, 113597. [CrossRef]
134. Hanumaiah, V.; Genc, S. Distributed Multi-Agent Deep Reinforcement Learning Framework for Whole-Building HVAC Control. *arXiv* **2021**, arXiv:2110.13450.
135. Ramos, D.; Faria, P.; Gomes, L.; Vale, Z. A Contextual Reinforcement Learning Approach for Electricity Consumption Forecasting in Buildings. *IEEE Access* **2022**, *10*, 61366–61374. [CrossRef]
136. Liu, T.; Xu, C.; Guo, Y.; Chen, H. A Novel Deep Reinforcement Learning Based Methodology for Short-Term HVAC System Energy Consumption Prediction. *Int. J. Refrig.* **2019**, *107*, 39–51. [CrossRef]
137. Liu, T.; Tan, Z.; Xu, C.; Chen, H.; Li, Z. Study on Deep Reinforcement Learning Techniques for Building Energy Consumption Forecasting. *Energy Build.* **2020**, *208*, 109675. [CrossRef]
138. Vázquez-Canteli, J.R.; Kämpf, J.; Henze, G.; Nagy, Z. CityLearn v1.0: An OpenAI Gym environment for demand response with deep reinforcement learning. In Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys 2019), New York, NY, USA, 13–14 November 2019; pp. 356–357. [CrossRef]
139. Baker, M. Is There a Reproducibility Crisis? *Nature* **2016**, *533*, 452–454. [CrossRef] [PubMed]
140. Peng, R.D. Reproducible Research in Computational Science. *Science* **2011**, *334*, 1226–1227. [CrossRef] [PubMed]