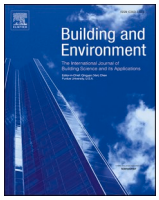
Building and Environment 244 (2023) 110766

Contents lists available at ScienceDirect

Building and Environment

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



# Dynamic indoor thermal environment using Reinforcement Learning-based controls: Opportunities and challenges

Arnab Chatterjee , Dolaana Khovalyg *

*Laboratory of Integrated Comfort Engineering (ICE), Ecole Polytechnique Fédérale de Lausanne (EPFL), CH-1700, Fribourg, Switzerland*

ARTICLE INFO

*Keywords:*
Reinforcement Learning
Dynamic indoor environment
HVAC controls
Energy efficiency
Thermal comfort
Temperature drifting

ABSTRACT

Currently, the indoor thermal environment in many buildings is controlled by conventional control techniques that maintain the indoor temperature within a prescribed deadband. The latest research provides evidence that more dynamic variations of the indoor thermal environment can promote health and trigger positive thermal alliesthesia , but such an environment requires a flexible and responsive control system that can adapt to the changes in real-time. As an emerging control technique, Reinforcement Learning (RL) has attracted growing research interest and demonstrated its potential to enhance building performance while addressing some limitations of other advanced control techniques. Thus, a comprehensive review explored the boundaries and limitations of a dynamic indoor environment and the possibilities to apply RL for building controls suitable for varying the indoor thermal environment. The first part discussed the studies on the permissible limits of temperature step changes and acceptable drifts to human occupants. It also debated the flexibility of the range of human thermal comfort and adaptation. In the next part, studies on RL for HVAC controls were explored, focusing on their application in creating a dynamic indoor thermal environment. The different algorithms, HVAC systems, co-simulation environment, action spaces, and energy-saving potentials were discussed. Overall, based on the review, this work outlined a potential pathway for the RL-based controller that can dynamically vary the indoor temperature. Suitable environmental parameters to be controlled, a choice of the RL-based algorithm, action space, and co-simulation environment are discussed.

## 1. Introduction

The building sector accounts for more than 40% of global energy use and 30% of greenhouse gas emissions, per the 2019 UNEP report [1]. The main reasons for this increase in energy use are population growth, rapid urbanization, and the increase in the ownership of personal appliances [2]. Cooling accounted for 3.5–7% of the total energy use in 2010, and an additional 15% of the space and water heating is produced by electricity [3]. At the heart of this energy use is making the indoor environment comfortable for the occupants. People in developed countries spend most of their lifetime indoors; for instance, the average American spends 93% of their time indoors, according to Environmental Protection Agency (EPA). It indicates that the design of the indoor environment is crucial for the well-being of the occupants. Usually, the indoor environment design of buildings is performed based on the appropriate national and international standards stipulating acceptable indoor environment conditions; the new and existing buildings are generally expected to adhere to them very closely [4]. Yet in reality, there are large numbers of buildings worldwide that do not comply with the standards. In this aspect, thermal comfort standards have a role to play, and, at present, they are in a transitional period with foreseeable further rapid modifications. Traditionally, at least in the context of the USA, the thermal comfort standards had in mind mechanically conditioned buildings, with temperatures held within narrow limits [5].

Additionally, the designs are based on long-term laboratory studies under steady-state indoor conditions. As per ASHRAE 55–2022 and ISO 7730, the operative temperature should not exceed 2.2 and 2 °C during any 1 h, respectively. However, it also specifies that this fluctuation requirement shall only apply to the situation where the indoor environment is not under the direct control of the occupant. Consequently, the target of HVAC control has been intended to maintain the indoor climate relatively constant. Given how energy intensive the buildings are, it stands to reason that in any move towards a greener future, the buildings sector will have a significant role to play [6,7]. The IPCC Working Group III also concluded that buildings hold the potential for maximum emissions reduction economically [8]. The attempted

* Corresponding author.
*E-mail address:* dolaana.khovalyg@epfl.ch (D. Khovalyg).

https://doi.org/10.1016/j.buildenv.2023.110766
Received 6 March 2023; Received in revised form 2 August 2023; Accepted 22 August 2023
Available online 26 August 2023

'greening' of buildings must consider the indoor comfort of the occupants involved, and comfort has a strong correlation with the health and productivity of the population [9]. Departing from the focus on near steady-state conditions of mechanically conditioned buildings opens avenues for reworking the standards towards altered realities of building energy and comfort [10,11].

The concept of a one-size-fits-all approach to providing thermal comfort for a given population using centralized mechanical systems is undesirable and fundamentally flawed. Diversity in the thermal preferences of building occupants resulting from variations in clothing level, physical activity, expectations, and physiology suggests that the criteria for evaluating occupants' thermal acceptability in office buildings may need to be recast. Although extant literature on transient or spatially non-uniform thermal comfort typically focuses on minimizing or eliminating discomfort, Schellen et al. demonstrated that positive pleasure or alliesthesia could be associated with temporal thermal transients [12]. Zhang et al. have shown that maximal thermal comfort in the built environment may increase our susceptibility to obesity and related disorders and, in parallel, requires high energy use in buildings [13]. Mild cold exposure increases body energy expenditure without shivering and compromising our comfort. Hence, rethinking our indoor climate by allowing ambient temperatures to drift may protect health and reduce energy use. Letting our body spend more energy to maintain thermal balance may positively affect health on a population scale.

Furthermore, temperature training by regular exposure to mild cold keeps the peripheral vascular system in motion and thereby helps to train the cardiovascular system. Occupants maintain an optimal body temperature within a relatively narrow range at about 37 °C under normal conditions [14]. Most people spend most of their time in an environment where the ambient temperature is below their body temperature. It is acceptable as the fundamental processes of metabolism and physiology are subject to entropic loss resulting in heat production. Therefore, a thermal heat sink of lower ambient temperature, where excess heat produced by metabolism can be dissipated, is beneficial for maintaining homeothermy. When the ambient temperature drops below the lower limit of the thermo-neutral zone (TNZ), thermogenic mechanisms are activated, resulting in extra heat production, preserving the core body temperature [14,15]. Conversely, when it rises above the upper TNZ limit, physiologic processes to amplify heat dissipation are activated, increasing the metabolic rate [16,17]. Without adequate energy intake compensation (reduction) to offset the lowered metabolic demand at thermoneutrality, an energy imbalance should contribute to weight gain. Thus, a dynamic indoor environment that increases the body's energy expenditure without physical activity can be the potential long-term solution to fight obesity in office buildings and provide thermally stimulating conditions.

It is crucial to highlight the difference between a dynamic indoor environment and standard HVAC control practices. A dynamic indoor environment refers to a constantly changing environment regarding its physical conditions (e.g., temperature, humidity, air quality) and occupancy patterns (e.g., number of occupants and their activities). This environment requires a flexible and responsive control system that can adapt to these changes in real-time to maintain comfortable and healthy conditions for the occupants. This dynamic environment influences the occupants' energy expenditure (EE) to promote healthy and comfortable indoor conditions. A dynamic indoor environment includes but is not limited to, changes in temperature, humidity, lighting, ventilation, and air quality. Creating a dynamic indoor environment aims to provide a comfortable, healthy, and energy-efficient indoor environment for the occupants. The dynamic indoor environment can be created in multiple ways.

- *Step changes* consist of a dynamic environment involving sudden temperature changes.
- *Drifts* refer to gradual and continuous changes in a system or environment over time.
- *Cyclic variations* consist of periodic fluctuations in a system or environment over regular intervals.
- *Transient changes* refer to sudden, short-lived changes in a system or environment that may occur due to external factors or events.
- *Non-uniform conditions* comprise situations where the environment or system is not uniform, i.e., differences across the system.

In contrast, a typical HVAC system control focuses mainly on maintaining the indoor temperature within a set range and does not consider the variability in occupancy patterns or other environmental factors. It also does not consider the variations in the energy expenditure of the occupants. The focus is on maintaining thermal comfort and indoor air quality, but the energy expenditure of the occupants is not considered. Thus, a standard HVAC system control is designed to maintain a static indoor environment, while a dynamic indoor environment aims to vary the indoor conditions in a controlled manner to influence the EE of the occupants. For example, a rule-based controller might set the temperature to a fixed set point and adjust the heating or cooling accordingly, regardless of how many people are in the room or what they are doing. A dynamic indoor environment requires a more sophisticated control system, such as a DRL (Deep Reinforcement Learning) based controller, which can learn from the environment and adapt the control strategy accordingly. By considering a range of factors, such as occupancy, temperature, and humidity, a DRL-based controller can optimize the control of the HVAC system and maintain comfortable and healthy conditions more energy-efficiently. Thus, the main difference between a dynamic indoor environment and an HVAC system control is the level of complexity and adaptability of the control system. A dynamic indoor environment requires a flexible and responsive control system that can learn from the environment and adapt to changing conditions, while a typical HVAC system control focuses on maintaining a fixed set point temperature, regardless of other factors.

It goes in hand with adaptive comfort, which refers to the concept that people can adapt to and feel comfortable in a wide range of indoor environmental conditions. It challenges the traditional approach to thermal comfort, which relies on fixed temperature and humidity thresholds, by recognizing that individual comfort preferences and responses to indoor conditions can vary based on factors such as clothing, activity levels, and personal habits. Hellwig et al. found that widening temperature bands in actively conditioned buildings can lead to energy conservation and increased satisfaction [18]. Ferrari et al. described the adaptive approach to determining temperature levels that follow the variability of the outdoor climate [19]. Candido suggested that the adaptive model of thermal comfort can help architects design more sustainable and stimulating indoor environments [20]. Schweiker et al. compared the adaptive comfort model with the predicted mean vote (PMV)-approach and the calculation of the human body exergy consumption (HBx-) rate, and found that the minimum HBx-rate corresponds well to the neutral temperature given by the adaptive comfort model [21]. Overall, the papers suggest that the adaptive comfort concept can be applied to actively conditioned buildings and can lead to improved energy performance and user wellbeing.

Although the study focuses on dynamic variations in air temperature, it is important to acknowledge the fact that indoor air quality (IAQ) is also a critical aspect of the indoor environment because it directly influences the health, comfort, and well-being of building occupants. Kamaruzzaman et al. [22] reported that poor IAQ can lead to dissatisfaction among occupants, which can negatively affect their work productivity and stress levels. Fang et al. [23] found that temperature and humidity can also impact occupants' perception of IAQ, with higher temperatures and humidity leading to less acceptable air quality. Moschandreas et al. [24] highlighted that occupant perception of IAQ is affected by various factors, including temperature, occupant density, and odor perception. The above research studies suggest that improving IAQ can lead to a more comfortable and productive indoor environment for occupants.

Returning to the discussion on dynamic indoor environment, smart thermostats are currently available to residential and commercial consumers to help them save energy by learning occupancy patterns and weather predictions. In developing countries, the demand for new air conditioning devices is expected to increase significantly in the coming years [25]. It is an opportunity to commercialize smart air conditioning devices or smart thermostats that can help to create a dynamic indoor environment for energy optimization, similar comfort levels, and long-term health benefits. Nedergaard et al. have illustrated that a dynamic thermal environment may be healthier for the human body [16]. Messaoud et al. also concluded that thermal environments that benefit the human ability of thermal adaptation should be considered healthier thermal environments. However, it is easier said than done. Many unanswered questions remain and need to be addressed; for example, what is the mechanism of an acceptable dynamic thermal environment? What is the most appropriate thermal environment for both comfort and health? How might such a thermal environment be constructed? How should dynamic thermal environments be evaluated? It is also challenging from the control perspective. Numerous factors influence the control policy in the built environment, i.e., indoor environmental conditions, occupancy, occupants' actual heat emission, and comfortable sensation. Implementing such a policy is challenging and usually hard to model as they may differ from case to case. A Deep RL-based framework for energy optimization and healthy thermal environment control in buildings is a promising solution to tackle this complexity.

As an emerging control technique, Reinforcement Learning (RL) has attracted growing research interest and demonstrated its potential to enhance building performance while addressing some limitations of other advanced control techniques [26]. Currently, many RL algorithms have been considered for the development of HVAC controls, for instance, Q-learning, Deep Q-networks (DQN), Actor-Critic, Deep Deterministic Policy Gradient (DDPG), and Asynchronous Advantage AcfDRLtor Critic (A3C). Q-learning is the most basic RL algorithm to learn the value of an action in a particular state. For any finite Markov decision process (FMDP), Q-learning finds an optimal policy to maximize the expected value of the total reward over successive steps, starting from the current state. Q-learning is only practical for minimal environments, and it quickly loses its feasibility when the number of states and actions in the environment increases. The DQN algorithm was developed by enhancing Q-Learning with deep neural networks and experience replay. DQN uses a deep neural network to approximate the values, which is fine as long as the relative importance is preserved. There are, however, certain limitations in using DQN that are fixed by using Actor-critic, DDPG, and A3C algorithms. Actor-Critic is the Temporal Difference (TD) method with a separate memory structure to explicitly represent the policy independent of the value function. The policy structure is known as the *actor* because it is used to select actions, and the estimated value function is known as the *critic* because it criticizes the actions made by the actor.

A significant advantage of using an *actor-critic* is that the action probabilities of the actor fully determine the exploration. One of the state-of-the-art model-free algorithms is DDPG, which concurrently learns a Q-function and a policy. Lillicrap et al. [27] combined ideas from DQN and DPG to create a very successful algorithm to solve continuous problems off-policy, the DDPG algorithm. It can learn efficient policies on most continuous problems, including the HVAC control, as all the control parameters are continuous (i.e., air temperature, air flow rate, and air relative humidity). The A3C algorithm is one of the newest algorithms developed in the field of DRL algorithms. Unlike DQN, which uses a single agent and a single environment, this algorithm uses multiple agents, each with its network parameters and a copy of the environment. These agents interact with their respective environments asynchronously, learning with each interaction. The algorithm predicts both the value function and the optimal policy function. Thus, it is another DRL algorithm that has the potential to provide good results in HVAC controls.

As the DRL algorithms have a high potential to be used in building applications, general reviews on RL to create controls for energy-efficient and comfortable buildings are available in the literature. For instance, there were quite a few studies [26,28,29] that conducted a comprehensive review of existing studies that applied ML and RL for building controls or occupant comfort in indoor built environments. They.

Provided a detailed breakdown of the existing RL studies that use a specific variation of each significant component of the RL. Moreover, Han, in particular, also briefly reviewed the empirical applications in occupant behavior in building controls using RL, how they have contributed to shaping the modeling paradigms, and how they might suggest a future research direction [30]. In a similar context, some studies compared several predictive control approaches that allow obtaining a high thermal comfort level by optimizing the use of an HVAC system through different cost functions [31]. Heidari et al. [32,33] developed occupant-centric controls using the RL algorithm to optimize residential energy systems and balance comfort, energy use, and hygiene in hot water systems. Some other studies reviewed the application of RL controls in the general context of buildings. For example, some papers reviewed control algorithms, including RL, to integrate thermal energy storage in buildings [34] and develop autonomous building energy management systems [35]. A comprehensive literature review exists comparing RL methods that have been investigated for Demand Response applications in smart grids [36,37]. We also found several literature reviews focusing on other aspects of RL control development or applications. Li and Wen analyzed various building energy models used in building control and operation studies [38]. Messaoud et al. reviewed the most fundamental concepts of ML categories and Algorithms applicable to IoT devices [39]. Dong et al. provided a systemic review of how indoor sensors influence managing optimal energy saving, thermal comfort, visual comfort, and indoor air quality in the built environment [40]. Wang and Ma reviewed several supervisory control strategies for building HVAC systems highlighting the advantages and easy implementation of model-free algorithms such as RL [41]. Table 1 summarizes the contents of the previous review papers, highlighting the novelty of the literature review conducted in this study.

It was seen that papers that talk about RL-based HVAC controls focused mainly on reducing energy use and also sometimes on maintaining or improving the thermal comfort of the occupants. There was a lack of studies using RL techniques to develop dynamic indoor environment conditions. The contribution of this chapter is (i) to review what the acceptable dynamic thermal conditions are, and (ii) to review what RL algorithm can be appropriate for the design of the controller enabling dynamic environment with triple objectives (energy, comfort, health). In the first part, studies related to the dynamic indoor environment have been presented, highlighting different ways of creating it (e.g., temperature step changes, drifts), its effect on the thermal comfort and adaptation of the occupants, health implications, and building energy use reduction. The objective was to gain knowledge regarding acceptable dynamic thermal conditions regarding the temperature ranges and limits of temperature step changes. These parameters would then be used as the inputs for the design of the RL-based controller. In the next part, the developments of RL concerning HVAC controls have been explored. The aim was to review the state-of-the-art RL algorithms and assess the advantages of using one over the other for creating a dynamic indoor environment. In the discussion, we summarize the literature review findings and use the knowledge gained about the characteristic of the dynamic indoor environment to define the properties of the RL-based control. The ultimate goal is to develop an RL-based control that can create a dynamic indoor environment. Finally, the significant obstacles in developing RL-based controls are highlighted. More studies should focus on bringing the concepts of a dynamic indoor environment and RL controls under the same hood.

**Table 1**
Comparison between this work and previous literature review on a similar topic.

| Ref No. | Main focus | System Type | Involved Methods/Algorithms | DRL methods for DIET control | Future directions in DRL-based DIET control |
|---|---|---|---|---|---|
| [26] | Building Energy Optimization | HVAC, batteries, home appliances, DHW, windows, lighting | RL | No | No |
| [28] | Occupant comfort control | HVAC, lighting | RL | No | No |
| [29] | Occupant comfort control | HVAC | PID, MPC, Rule-based, Q-learning | No | No |
| [30] | Occupant comfort control | HVAC, lighting, window, blind | RL, DRL | No | No |
| [31] | Building energy optimization & Occupant comfort | HVAC, lighting | MPC | No | No |
| [34] | Thermal energy storage in buildings | PCM-based HVAC | MPC, adaptive, NN, fuzzy logic, RL | No | No |
| [35] | Building energy management | HVAC, lighting, appliances, batteries, electrical grid | RL (Q-learning, SARSA, Actor-Critic) | No | No |
| [36] | Residential demand response | Smart grids | EA, RL, SPC, MCTS | No | No |
| [37] | Demand Response | HVAC, DHW, appliances, EV | RL (Q-learning, SARSA, Actor-Critic, Multi-player) | No | No |
| [39] | IoT in Smart Homes | RFID, Zigbee, WSN | ML | No | No |
| [40] | Indoor environment control | HVAC | Sensor-based control | No | No |
| [41] | Building Energy Optimization | HVAC | Supervisory control strategies, ML, RL | No | No |
| This work | Dynamic indoor environment control | HVAC | DRL (Q-learning, DQN, Actor-Critic, DDPG, A3C, SAC, TRPO, PPO) | Yes | Yes |

DIET – Dynamic Indoor Environment, PID – Proportional-Integral-Derivative, MPC – Predictive Model Control, EA – Evolutionary Algorithm, SPC – Set-point control, MCTS – Monte Carlo tree search, NN – Neural Network, RFID – Radio Frequency Identification, WSN – Wireless Sensor Network, PCM – Phase change material.

## 2. Methodology

A literature search was conducted on the academic search platform *Science Direct* using the topic structure and keywords shown in Equations (1) and (2), where the symbol "*" is used to search for terms in both singular and plural forms. The *Science Direct* platform could retrieve papers from the traditional built environment and computer science fields. In addition, many papers were selected from the references in the above search.

$$topic = (building)\ AND\ [(step\ change *\ \ OR\ drift * \ OR\ non-uniform\ condition * \ \ OR\ thermal\ adaptation)\ AND\ temperature] \quad \text{Eq. 1}$$

$$topic = (reinforcement\ learning)\ AND\ [(building *\ \ OR\ house *\ OR\ home *)\ AND\ control]\ OR\ (smart\ thermostat) \quad \text{Eq. 2}$$

With the search structure and keywords listed in Eq. (1), 129 articles were found. We manually filtered to remove review articles, some duplicates, and other studies unrelated to our field of interest for this review. We included a few additional articles unavailable in Science Direct, and the final selection contained 52 articles, summarized in Table 2. The table is organized concerning the type of variation (i.e., step changes, transient changes, drifts, conditioning type, non-uniform conditions), the variable that was the subject of change (i.e., $T_{air}$, $RH_{air}$, $V_{air}$), and the variation ranges. Also crucial in this context were the characteristics of the subjects (i.e., age, gender, weight, height, and ethnicity). Finally, the findings of the particular studies are reported in the last column, highlighting how the subjects felt or the physical reaction to the controlled chamber experiments. The papers on the dynamic environment came from various journals, with two-thirds of the publications from journals specializing in buildings, indoor environments, and energy. The remaining publications are from periodicals with a strong emphasis on thermal biology, physiology, and behavior, mainly focusing on the thermal adaptation and comfort evaluation of the human subjects involved. The papers were classified into the different aspects of the dynamic environment (i.e., step changes, drifts, non-uniform conditions, thermal comfort evaluation, and thermal adaptation). The literature review for the papers concerning the dynamic indoor environment was conducted in mid-2019. The primary reason for not extending the formal review to 2022 is repeated information found in the subsequent years and a sharp drop in human subject-based studies worldwide due to the start of the COVID-19 pandemic.

For the literature review on dynamic indoor environments, the focus was on selecting research papers that investigate how changes in indoor conditions, specifically air temperature, relative humidity, and air flow rate, influence various aspects of human well-being. The primary objective was to understand the impact of these dynamic environmental factors on the thermal comfort, health, and physical performance of individuals. By examining such studies, the aim was to gain insights into how indoor environments can be optimized to enhance occupant comfort and overall well-being. To ensure the relevance and applicability of the selected papers, specific information related to participant characteristics was also required. It was essential that each paper provided information on the number of participants involved in the study, their age, metabolic rate, clothing levels, and body mass index (BMI) or at least some of the above information. These parameters play a crucial role in understanding individual responses to changes in indoor environmental conditions, as different age groups, activity levels, and body compositions can have varying thermal comfort requirements.

Regarding the review on RL-based HVAC controls Conducted with the search structure and keywords listed in Eq. (2), 108 articles were found. After manually filtering and adding a few additional articles unavailable in Science Direct, the final selection contained 63 articles, summarized in Table 3. The table presents essential information from the existing studies on dynamic indoor environments, including details regarding the RL/DRL algorithm type used in the studies (i.e., q-learning, DQN, actor-critic, DDPG, A3C). At the same time, it was crucial to look at the type of HVAC systems (i.e., heat pump, air conditioner, VAV, VRF, electric heater, chiller) used in the studies. Other findings in terms of the building scale (i.e., single, multi), building type (i.e., residence, office, laboratory, data center, commercial), simulation software (Energy Plus, Trnsys, Matlab, BuildSim) for training, and the various controlled parameters are also reported in Table 3.

Quite a few papers developed RL control models without testing their performance, and those papers were not considered in this review. As the use of RL for building controls is a multidisciplinary field, half of the publications come from journals specializing in computer science, artificial intelligence, and controls. The remaining publications are from periodicals with a strong emphasis on the environment, energy, and buildings. The review's primary focus was on the different types of ML, RL, and DRL algorithms used in the studies. The literature review for the papers concerning RL controls was conducted in 2022. In addition to the

**Table 2**
Summary of existing studies on the dynamic indoor environment.

| Ref No. | Year | Journal Name | Subject Characteristics | | | | Met | Clo | Variation Type | Parameter | Parameter Range | Findings |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Number, gender | Age (y.o.) | Weight (kg) | Height (m) | | | | | | |
| [42] | 2007 | Building & Environment | 6 M, 6 F | 30–60 | n/a | n/a | 1.2 | 0.54 | Step Changes | $RH_{air}$ | 30–70% RH | No effect on subjective performance |
| [43] | 2016 | Applied Ergonomics | 14 M, 22 F | 21.8 ± 3.9 | 58.7 ± 11.9 | 1.64 ± 0.07 | 1 | 0.6 | Step changes | $T_{air}$ | 24–35 °C | Thermal step-changes affect thermal comfort |
| [44] | 2018 | Building & Environment | 467 | 21 ± 3 | 61 ± 12 | 1.67 ± 0.08 | 1 | 0.39 | Step Changes | $T_{air}$, $RH_{air}$, $V_{air}$ | 28–31 °C, 60–90%, 0–3 m/s | $V_{air}$ and particularly $RH_{air}$ have a significant impact on comfort |
| [45] | 2011 | Building & Environment | 12 M, 11 F | 24 ± 2.5 | 59.1 ± 8.6 | 1.69 ± 0.07 | 1.2 | 0.4–0.93 | Step changes | $T_{air}$ | 19.5–30 °C | Overall thermal sensation correlated with a head sensation |
| [46] | 2013 | Journal of Thermal Biology | 24 M, 24 F | 25 ± 1.3 | 63.8 ± 12.7 | 1.70 ± 0.18 | 1 | 0.5 | Step changes | $T_{air}$ | 26–38 °C | The effect of $T_{rad}$ was different from those of the Tair |
| [47] | 2014 | Building & Environment | 15 M, 15 F | 20 ± 1 | 56 ± 6.6 | 1.66 ± 0.06 | 1 | 0.57 | Step Changes | $T_{air}$, $RH_{air}$ | 20–29 °C, 50–70% | The acceptable conditions are 29.2 °C, 50% RH and 28.0 °C, 70% RH |
| [48] | 2015 | Building & Environment | 8 M, 8 F | 20.3 ± 2.1 | 59.6 ± 8.9 | 1.68 ± 0.07 | 1.2 | 1 | Step Changes | $T_{air}$ | 12.8–23.0 °C | Occupants expected to keep $T_{air}$ between 15.3 and 19.4 °C |
| [49] | 2016 | Building & Environment | 9 M, 9 F | 30 ± 3 | 66.9 ± 9.6 | 1.68 ± 0.06 | 1.2–1.7 | 1.2 | Step Changes | $T_{air}$ | [-20, 20] °C | The effect of down-steps on humans is more intensive than up-steps |
| [50] | 2017 | Building & Environment | 15 M, 15 F | 19.2 ± 0.8 | 53.0 ± 7.1 | 1.65 ± 0.08 | 1 | 0.57 | Step Changes | $T_{air}$ | 20–32 °C | Permissible temperature steps were no larger than 3 °C |
| [51] | 2016 | Energy & Buildings | 12 M, 12 F | 22 ± 1 | 61 ± 5 | 1.71 ± 0.05 | 1 | 0.5 | Step Changes | $T_{air}$ | 22–37 °C | The most sensitive body parts are the back, arm, and leg |
| [52] | 2005 | Int J of Biometeorology | 30 M | 21.7 ± 0.8 | 61 ± 7.2 | 1.71 ± 0.06 | 1 | 0.42 | Step Changes | $T_{air}$ | 22–37 °C | Subject suffers higher physiological strain during down-step vs. up-step |
| [53] | 2015 | Physiology & Behavior | 12 M, 12 F | 22 ± 1 | 61 ± 5 | 1.71 ± 0.05 | 1 | 0.5 | Step Changes | $T_{air}$ | 22–37 °C | Females are more prone to show thermal dissatisfaction with cold |
| [54] | 1993 | Journal of thermal biology | 6 M | n/a | n/a | n/a | n/a | n/a | Step Changes | $T_{air}$ | 20–30 °C | Down-steps and up-steps have different skin temperature effect |
| [55] | 2017 | Energy & Buildings | 9 M, 9 F | 30 ± 3 | 66.9 ± 9.6 | 1.68 ± 0.06 | 1.2–1.7 | 1.2 | Step Changes | $T_{air}$ | [-20, 20] °C | Change of skin temperature due to down-step vs. up-step |
| [56] | 2011 | Building & Environment | 8 M, 8 F | 20.8 ± 1.1 | 55.8 ± 12.8 | 1.65 ± 0.1 | 1.2 | 0.5 | Step Changes | $T_{air}$ | 20–32 °C | Temperature step of 4 °C or less are the upper permissible limits |
| [57] | 2013 | Building & Environment | 15 M | 22–30 | 56.4–93.6 | 1.62–1.78 | 1 | 0.03 | Step Changes | $T_{air}$ | 20–35 °C | The upper body skin temperature decreases by convective heat loss |
| [58] | 2014 | Building & Environment | 20 M | 23.8 ± 1.0 | 62.9 ± 5.6 | 1.75 ± 0.03 | 1 | 0.5 | Step Changes | $T_{air}$ | 25–32 °C | Skin heat fluxes can be used to predict the thermal sensation |
| [59] | 2017 | Building & Environment | 8 | college students | n/a | n/a | 1.1 | 0.57 | Step Changes | $T_{air}$ | 20–30 °C | People prefer lower rather than neutral temperature |
| [60] | 2016 | Energy & Buildings | 12 M, 12 F | 22 ± 1 | 66.5 ± 6.3 | 1.77 ± 0.05 | 1 | 0.5 | Step Changes | $T_{air}$ | 22–37 °C | Oral, skin temperature, HR, and HRV are sensitive to temperature steps |
| [61] | 2017 | Energy & Buildings | 10 M, 10 F | 20.3 ± 1 | 62.9 ± 5.73 | 1.70 ± 0.05 | 1.1 | 0.57 | Step Changes | $T_{air}$ | 20–32 °C | Thermal comfort improves during the transfer of environment |
| [62] | 2015 | Procedia Engineering | 12 M, 12 F | 21 ± 1 | n/a | n/a | 1 | 0.3 | Step Changes | $T_{air}$ | 22–37 °C | Thermal sensation before the temperature step differs from immediately after a step change. |
| [63] | 2017 | Procedia Engineering | 6 M, 6 F | 58 ± 5 | 66.0 ± 8.9 | 1.63 ± 0.06 | varied | varied | Step Changes | $T_{air}$ | n/a | Temperature alteration led to an increase in blood pressures |
| [64] | 1993 | Indoor Air | 12 M | 24 | 73.6 | 1.84 | 1 | 0.6–1.0 | Step Changes | $T_{air}$ | 12–29.7 °C | Thermal sensations are more sensitive to down-steps vs. up-steps |
| [65] | 2015 | Building & Environment | 12 M, 12 F | 22 ± 1 | 60.8 ± 7.8 | 1.71 ± 0.08 | 1 | 0.3 | Step Changes | $T_{air}$ | 22–37 °C | More time is needed for skin temperature to stabilize after up-steps vs. down-steps. |
| [66] | 2014 | Plos One | 12 M | 20–30 | n/a | n/a | 1.2 | n/a | Step Changes | $T_{air}$ | 12–22 °C | Occupants feel uncomfortable when the temperature difference is greater than 5 °C. |

(*continued on next page*)

**Table 2** (*continued*)

| Ref No. | Year | Journal Name | Subject Characteristics | | | | Met | Clo | Variation Type | Parameter | Parameter Range | Findings |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Number, gender | Age (y.o.) | Weight (kg) | Height (m) | | | | | | |
| [67] | 2016 | Indoor Air | 20 M | 22–30 | n/a | n/a | 1.1 | 1.1 | Conditioning type | $T_{air}$ | 26, 36 °C | Dynamic thermal environments are suitable for the human |
| [68] | 2007 | Building & Environment | 15 M, 15 F | 18–23 | n/a | n/a | 1.2 | 0.5–0.7 | Conditioning type | $T_{air}$ | 30, 35 °C | Pleasure during alternating heating and cooling |
| [69] | 2019 | Building & Environment | 501 M, 542 F | mostly 21-30 | n/a | n/a | 1.18–1.22 | 0.47–0.53 | Conditioning type | $T_{air}$ | 24.9–29.0 °C | Thermal adaptation in Naturally Ventilated and Hybrid spaces |
| [70] | 2017 | Building & Environment | 30 | 20–23 | 47.4–67.4 | 1.57–1.72 | 1.2 | 0.45–0.62 | Conditioning type | $T_{air}$ | 24–25.2 °C | Mean comfort temperatures are between 25.1 and 26.8 °C |
| [71] | 2016 | Building & Environment | 13 M, 15 F | 23–66 | n/a | n/a | 1.2 | n/a | Conditioning type | $T_{air}$ | 26.3–28.2 °C | There was no thermal discomfort in both FR or CL mode offices due to thermal adaptations |
| [72] | 2016 | Building & Environment | 325 | n/a | n/a | n/a | 1.2 | 0.51–0.62 | Conditioning type | $T_{air}$ | 23.1–26.6 °C | Comfort temperatures between 25.6 and 27.5 °C |
| [73] | 2017 | Physiology & Behavior | 19 F | 22.3 ± 1.9 | 62.7 ± 5.5 | 1.70 ± 0.07 | 1 | 0.04 | Transient Changes | $T_{air}$, Other | 26–32 °C | Room temperature independently affects body temperatures |
| [74] | 2015 | Building & Environment | 20 M | 22.4 ± 1.7 | 77.6 ± 8.5 | 1.81 ± 0.04 | 1.2 | 0.6–1.07 | Transient Changes | $T_{air}$, $V_{air}$ | 18–28 °C | Max & min temperature in summer and winter was 28 & 18 °C |
| [75] | 2017 | Journal of Thermal Biology | 12 M, 12 F | 22.5 ± 1 | 61.1 ± 5.1 | 1.70 ± 0.55 | 1 | 0.5 | Transient Changes | $T_{air}$ | 22–37 °C | Great predication of thermal sensation in transient conditions |
| [76] | 2001 | Physiology & Behavior | 8 F | 22.6 ± 1.5 | 63.9 ± 6.9 | 1.69 ± 0.04 | 1–3.3 | n/a | Transient Changes | $T_{air}$ | 22, 27 °C | The body regulates body temperature at different set points |
| [77] | 2018 | Energy & Buildings | 10 M, 10 F | 23.9 ± 1.1 | 56.9 ± 8.8 | 1.68 ± 0.09 | 1.2 | 0.32 | Transient Changes | $T_{air}$, $RH_{air}$, $V_{air}$ | 25.0–28.1 °C, 41.5–80.6% | The head, chest, back, and hand thermal sensation was higher than the overall thermal sensation. |
| [78] | 2015 | J of Thermal Biology | 36 M | 22.7 ± 0.5 | 64.7 ± 1.9 | 1.73 ± 0.03 | 1 | 0.3 | Transient Changes | $T_{air}$, $V_{air}$ | 16–32 °C, 0–5 m/s | Human thermal loads can represent skin temperature change |
| [79] | 2014 | Building & Environment | 9 M, 6 F | 23.9 ± 0.6 | 59.9 ± 2.6 | 1.69 ± 0.02 | 1.2 | n/a | Transient Changes | $T_{air}$ | 20–28 °C | Psychological adaption can speed up the process of thermal adaption |
| [80] | 2014 | Building & Environment | 11 M, 7 F | 29.9 ± 1.8 | 66.3 ± 5.8 | 1.69 ± 0.04 | 1 | 0.5 | Transient Changes | $T_{air}$ | 21–26 °C | Thermal adaptation must be taken into account for the design of buildings |
| [81] | 1967 | Environmental Research | 3 M | 22–24 | n/a | n/a | 1 | 0 | Transient Changes | $T_{air}$ | 12–48 °C | Thermal comfort lies in the temperature range for thermal neutrality (28–30 °C) |
| [82] | 2019 | Energy | 90 M, 20 F | 22 ± 1.8 | 68.1 ± 10.8 | 1.79 ± 0.04 | 1.0–1.3 | 0.5 | Transient Changes | $T_{air}$ | 19.3–32.7 °C | Short-term acclimation can improve thermal acceptability |
| [83] | 2016 | Energy & Buildings | 367 M, 373 F | 17–37 | n/a | n/a | varied | 0.96–1.1 | Transient Changes | $T_{air}$ | 14.9–26.4 °C | The neutral temperature in cities lies between 22.0 and 22.7 °C. |
| [84] | 2011 | HVAC&R Research | n/a | 6–16 | n/a | n/a | 1 | 0.5 | Transient Changes | $T_{air}$, $V_{air}$ | 25-20 °C | Providing some means of avoiding elevated temperatures would improve performance |
| [85] | 2018 | Indoor Air | 37 | 11 | n/a | n/a | 1.2 | n/a | Transient Changes | $T_{air}$ | 32-26 °C | Acclimatization can increase the optimal temperature for learning |
| [86] | 2016 | Indoor Air | 30 F | 23 | n/a | n/a | 1.2 | n/a | Transient Changes | $T_{air}$, $RH_{air}$ | 20–26 °C, 40–60% | Tair and RHair have a significant impact on the perception of indoor air quality |
| [87] | 2020 | Indoor Air | 6 M, 6 F | 18–30 | n/a | n/a | 1.2–1.4 | 0.22–0.64 | Transient Changes | $T_{air}$ | 23–27 °C | Moderately elevated temperatures may lead to reduced performance |
| [88] | 2007 | Building & Environment | 17 | n/a | n/a | n/a | 1.3 | 0.32 | Non-uniform Conditions | $T_{air}$ | 15.6–31.5 °C | Positive finger–forearm skin temperature gradient in warm/hot |
| [89] | 2013 | Energy & Buildings | 10 M | 24.7 ± 2.0 | 77.3 ± 8.5 | 1.82 ± 0.08 | 1.2 | 0.6 | Non-uniform conditions | $T_{air}$, Other | 23.7–25.5 °C | For predicting thermal comfort under non-uniform conditions, Tair only is not sufficient. |
| [90] | 2018 | Building & Environment | 20 M, 90 F | mostly 21-30 | n/a | n/a | 1.2 | 0.5 | Non-uniform Conditions | $T_{air}$, $RH_{air}$, $V_{air}$ | 19.3–26.6 °C | Occupant thermal perception was unaffected when the Tair differences were within ±2 °C. |
| [91] | 2017 | Procedia Engineering | 12 M, 12 F | 23 ± 2 | n/a | n/a | 1 | 0.4 | Non-uniform conditions | $T_{air}$, $V_{air}$ | 22–28 °C | Non-uniform conditions decreased thermal sensation to cool |
| [92] | 2017 | S of the Total Environment | 17 M, 19 F | 12–59 | 37–66 | 1.44–1.66 | 1 | 0.8–1 | Non-uniform Conditions | $T_{air}$, $RH_{air}$, $V_{air}$ | 14–18 °C | The local heating seat can improve users' thermal sensation |

(*continued on next page*)

**Table 2** (*continued*)

| Ref No. | Year | Journal Name | Subject Characteristics | | | | Met | Clo | Variation Type | Parameter | Parameter Range | Findings |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | Number, gender | Age (y.o.) | Weight (kg) | Height (m) | | | | | | |
| [93] | 2012 | Physiology & Behavior | 10 M, 10 F | 24.4 ± 1.8 | 71.0 ± 8.9 | 1.76 ± 0.08 | 1.2 | 0.6 | Non-uniform Condition | $T_{air}$ | 24, 25 °C | Females are more uncomfortable in the same environment as compared to male |
| [94] | 2014 | Physiology & Behavior | 16 F | 23 ± 4 | 65.5 ± 7.9 | 1.69 ± 0.06 | 1.2 | 0.58 | Drifts | $T_{air}$ | 24–32 °C | A more considerable discomfort in the upward ramp vs. downward ramp |
| [95] | 2016 | Energy & Buildings | 16 M, 20 F | 25.7 ± 4.4 | 68.5 ± 15.4 | 1.74 ± 0.08 | 1.2 | n/a | Drifts | $T_{air}$ | 25–29 °C | Adaptative actions adjustments are affected by the occupancy density |
| [96] | 2010 | Indoor Air | 16 M | 22-25, 67–73 | 80.3 ± 7.9 | 1.80 ± 0.08 | 1.2 | 1 | Drifts | $T_{air}$ | 17–25 °C | Temperature drifts up to ± 2 K/h in the 17–25 °C is acceptable. |
| [97] | 2016 | S&T for the Built Env | n/a | 18–25 | n/a | n/a | 1.2 | 0.5 | Cyclic Variations | $T_{air}$ | 26–29 °C | Participants' performance was not significantly affected by changes in Tair. |

DRL algorithms, application to the different types of HVAC systems used was also an important subject. In most of the studies, it was seen that the RL control was trained in a simulation environment.

For the literature review on DRL algorithms, a set of research papers that focused on using Machine Learning (ML) or DRL techniques to achieve energy use reduction or thermal comfort improvement in both single and multi-zone buildings were curated. The objective was to explore how these advanced algorithms can optimize building performance and enhance occupant comfort, which are critical aspects of sustainable building design and operation. To ensure the relevance and quality of the selected papers, particular emphasis on specific criteria was placed. Each paper had to demonstrate the application of ML or DRL algorithms for energy efficiency or comfort optimization in buildings. Moreover, the papers needed to provided comprehensive information regarding the type of building studied (e.g., residential, commercial, institutional), details about the HVAC system employed, and the specific DRL algorithm used. This information was used to assess the applicability of the findings to different building types and understand the versatility of DRL algorithms in different HVAC contexts.

In addition, each paper had to clearly describe the simulation framework used to model the building and HVAC system. This ensured that the research was based on robust simulation methodologies, enabling a controlled assessment of the algorithm's effectiveness. Considering these factors, the aim was to present a comprehensive and reliable review of the current state-of-the-art in applying DRL algorithms for energy efficiency and thermal comfort improvement in the building sector. By examining the selected papers, this review provides an overview of the latest advancements in ML and DRL-based approaches for optimizing building performance and occupant comfort. The findings and methodologies presented can serve as a basis for further research and innovation in sustainable building design and operation. Moreover, the review highlights the potential of DRL algorithms to revolutionize the building industry, opening up possibilities for intelligent, energy-efficient, and occupant-centric building systems that align with the growing demand for sustainable and environmentally friendly construction practices. In addition to the studies found in Science Direct, we included a few additional articles not available in this database but found through other sources, such as Google Scholar and other academic journals.

## 3. Results - Dynamic Indoor Environment

This section presents the findings from the literature review conducted on the papers focusing on a dynamic environment. As shown in Fig. 1, the publication on this topic jumped between 2016 and 2017 and then stayed stable. Regarding the variation types, almost half of the studies investigated the effect of step changes in the indoor environmental variable. The other notable variation types were transient changes, drifts, conditioning types, and non-uniform conditions. In most (75%) of the studies, the subject of the change was air temperature. Almost 80% of the reviewed studies were controlled climate chamber experiments. In most papers, the sample size was between 10 and 50 subjects, with a relatively equal distribution of males and females. A couple of field studies also focused on investigating the thermal adaptation of subjects that studied more than 500 individuals. Asians and Caucasians were the most common ethnic groups investigated in the controlled experiments. In Table 1, we classified the articles based on the variable type and analyzed them concerning the changed indoor environmental variable, the variation range, and the occupants' reaction. Thus, most of the papers reviewed after filtering consisted of research that conducted studies with human subjects by exposing them to different indoor environmental variations.

### 3.1. Temperature step changes and drifts

As the literature shows, temperature step changes are the most

common and easiest way to create a dynamic environment [42–66]. In addition, the other variations in the literature review were temperature drifts, transient changes, and non-uniform conditions. The non-uniform conditions [88–93] refer to the indoor environment where radiant heating/cooling or personal conditioning systems were used. In addition, quite a few studies investigate the effect of different conditioning types [67–72], which consist of the indoor environment with air conditioning, natural ventilation, and mixed mode.

In studies involving step changes, acceptable temperature steps were proposed as step magnitudes no larger than 3 °C as the performance of the 3 °C steps was better than that of the 6 °C steps in terms of thermal sensation, comfort, and acceptability [50]. In Horikoshi et al. [57], similar conclusions were reached, i.e., a temperature step of 4 °C or less is the permissible upper limit. Du et al. [59], the researchers concluded that occupants feel uncomfortable when temperature step changes are more significant than 5 °C. In large step-change conditions, more than 45 min was needed for mean skin temperature to achieve steady after down-step and instep skin temperature contributed most to this result [51]. It has been seen that the physiological strain during the down-step from a higher operative temperature is much different as compared to up-steps, even when the magnitude of the step changes is the same [50,52,55,56]. Also, the instant change in skin temperature caused by down-step is remarkably more prominent than that caused by up-steps.

Moreover, overshoot in thermal sensation occurs in both down-step and up-step changing stages. The maximum changes in overall human thermal sensation due to step-change environment show asymmetry in response. The maximum changes in overall thermal sensation at the down-step stage are much more significant than those of the up-step change, indicating that humans are more sensitive to cold stimuli [58]. In the studies focusing on transient conditions, it was repeatedly concluded that people could thermally adapt to the changes in the environment, and it must be taken into account in the design of the buildings [75,77,80,81,83]. Under controlled and non-uniform conditions, variation in thermal sensation due to the spatial variation of the environment is an essential factor in determining thermal comfort. Whereas under dynamic conditions, thermal sensation change with time significantly affects thermal comfort [98]. A transient thermal environment can be established by changing air temperature or the magnitude of air movement. Studies of human responses showed that airflow similar to natural air movement has the highest preference. The physical process through which air movement affects human thermal states is convective heat transfer, and, thus, the difference between human and ambient temperatures and air speeds are key variables [78]. In temperature drifting conditions, subjects feel less or equally comfortable as in constant temperature; the conditions do not lead to unacceptable situations.

A transient environment can result in substantial energy savings while an acceptable indoor climate could be maintained [68]. As already recommended by the standards, temperature drifts up to ±2 °C/h in the range of 17–25 °C were assessed as applicable and will not lead to unacceptable conditions [96]. In addition, as per studies, cyclic temperature variations are not likely to significantly negatively influence the occupants' learning performance [98,99]. The critical finding from the papers belonging to this section was that the permissible upper limit of step changes is 4 °C, which should be done in about 45 min. That essential input would be used later in developing the RL controls.

### 3.2. Thermal comfort and adaptation

The outdoor temperature dramatically influences the indoor thermal environment, even in conditioned buildings. In studies, it has been found that the lower the outdoor temperature and the longer the cold season was, the higher the usage rate of the district heating system in winter and the longer the heating season was. The outdoor temperature affects adaptive behavior, even in heated buildings [100–102]. Also, absolute humidity could be a good indicator of indoor conditions [103]. Considering the above inference, having a dynamic indoor environment that can vary as a function of the outdoor temperature conditions makes sense. The low outdoor temperature helps people adapt to the cold climate [104]. This result indicates that maintaining a high indoor temperature during the winter is not only a waste of energy but also nullifies people's adaptation to the environment [91,104,105].

In this context, the thermal transformations of the free-living Tuvan nomadic pastoralists of South-Central Siberia were studied in January 2020. In total, 8 yurts and 12 adults were studied, as presented in [106]. Their total energy expenditure, activity, skin, and exposure temperature were monitored. The case study aimed to illustrate the extreme thermal exposure conditions experienced by the Tuvan nomadic pastoralists and the extent of thermal adaptation the human body can exhibit. The results of this case study demonstrate the ability of human subjects to thermally adapt to a wide range of variations in their indoor environment, especially in the air temperature. The data collected during the study revealed that the subjects could maintain thermal comfort despite the wide range of air temperature, relative humidity, and radiant asymmetry variations. The extensive range of skin temperature variations suggests that it might be an adaptative action due to exposure to extreme cold conditions at specific points during the day.

In tropical countries, the comfort range of both seasons was 15–34 °C. The subjects demonstrated considerable adaptability to indoor temperature variations due to adaptive measures [107]. Operative temperatures between 24.5 and 30 °C in Malaysia and 26–28 °C in Indonesia are comfortable in the summer. These results are higher than the recommended temperatures in each country [72]. People resort to adaptive behaviors which can widen the range of comfortable temperatures. In Malaysia and Japan, there is a tendency to overcooling in buildings when operating within the temperatures specified in existing local guidelines. The current guidelines for mechanically conditioned buildings might also underestimate occupants' thermal preference in hot-humid climates, where a greater degree of heat tolerance could be found in the general population [70]. In a study conducted in NUS in Singapore, the neutral operative temperature in an air-conditioned space of 26.7 °C suggested that the typical setpoint of 24 °C could be increased by two degrees without affecting the average comfort level of occupants [69].

In China, people have increasingly begun to install air conditioning systems. Residents' expectations to improve their living conditions have already led to increased use of standalone heating devices—with a dramatic 4.4-fold growth in heating energy use from 2001 to 2011 [108]. In colder parts of China, the neutral temperature is 15.8 °C in winter and 28 °C in summer. It shows that the occupants can adapt to various temperatures [108]. The acceptable conditions for transitional spaces in the hot-humid area of China have upper limits of an air temperature of 29.2 °C [47]. Cao et al. found that the neutral indoor operative temperature during the summer in Beijing was 26.8 °C, while in the winter, it was 20.7 °C [104]. The acceptable temperature ranges for the classrooms in Shaanxi, Gansu, and Qinghai Provinces were 12.7–16.9 °C, 11.9–17.1 °C, and 15.8–18.7 °C, respectively temperature [59]. In a field study by Buonocore et al. [46], occupants expected to keep air temperature between 15.3 and 19.4 °C. People even prefer a slightly lower temperature than a neutral temperature [59].

North America has the highest compliance with the ASHRAE comfort zone but holds a narrower operative temperature range than Asia. Research by Zhang et al. has concluded that the average indoor temperature value in other parts of the world is much higher than in China [109]. Most of the outside-comfort-zone points of Europe are due to overheating, while the outside-comfort-zone points of China are mainly caused by overcooling [109]. The indoor design temperature should consider the local climate, clothing thermal resistance, physiological characteristics, and psychological adaptability of occupants. Buildings often have discomfort due to high temperatures in the wintertime, which is indicative of overheating [102,110,111]. Although it is cold outdoors in the winter, indoor spaces in North China tend to be warm

due to space heating [104]. Many asymmetries exist from the point of view of human response to temperature variations. Radiant temperature affects the human skin temperature more than the air temperature. The faster the radiant temperature changes, the larger the increment or decrement of skin temperature [46]. The human body also has physiological adaptation to the environment, and the physiological adaptation is different between the increase and decrease in environmental temperature [46]. The local sensation is a function of both local and overall (whole-body) skin temperature and the rate of change over time of local skin temperature and body core temperature [112]. The most sensitive body parts are the back, arm, and leg. Typically, more than 45 min is needed for mean skin temperature to achieve steady after the down-step.

Also, the human body has psychological and physiological adaptations to the environment [51,55]. Under uniform and steady-state conditions, the thermal index can predict the PMV sensation well. Contrarily, non-uniform environments can achieve significantly different thermal sensation votes compared to predicted PMV values. The differences are probably caused by combined local discomfort factors [89]. The relationship between Thermal Sensation votes and human mean skin temperature in a transient environment is not as linear as in a controlled environment. Heat loss from the human skin surface can be used to predict dynamic thermal sensation instead of the heat transfer of the whole human body [66]. Both comfort and discomfort come from dynamic contrast. People feel significantly better if the poor thermal environment improves even a little [59]. The temperature changes provide cold or hot stimulation to the human body, which can increase the pleasure of feelings. Comfort results from eliminating discomfort, suggesting that a stable neutral environment may not be the best [59].

### 3.3. Human body energy expenditure and health impact

In the context of a building's indoor environment, the constant argument is that the occupants always need to feel comfortable. Contrarily, research studies have shown that maximal thermal comfort in the built environment may increase our susceptibility to obesity and related disorders. In parallel, it requires high energy use in buildings [13]. Mild cold exposure increases body energy expenditure without shivering and compromising our comfort. The body must work harder to regulate its temperature, leading to an increased metabolic rate [113]. Hence, rethinking our indoor climate by allowing ambient temperatures to drift may improve health and reduce energy use. Letting our body spend more energy to maintain thermal balance may positively affect health on a population scale.

Furthermore, temperature training by regular exposure to mild cold keeps the peripheral vascular system in motion and thereby helps to train the cardiovascular system. Allowing indoor temperatures to drift more than permitted under current standards can substantially reduce energy use by the built environment. More frequent cold exposure alone will not save the world, but it is a severe factor to consider in creating a sustainable environment with a healthy lifestyle [13]. Most studies focusing on the effect of mild cold exposure on EE have used a temperature of around 14–16 °C, which subjects often do not appreciate. Prolonged exposure to a less cold environment may be a potential solution. It was shown that thermal comfort changed from uncomfortable to comfortable during the experiments involving 10 days of cold acclimation. This change in comfort was significant and may even increase with a longer acclimation duration. Those experiments used low (15 °C) temperatures, but others [114] have used 17–19 °C with significant effects on energy metabolism approaching realistic indoor temperature conditions.

Additionally, circumstantial evidence from a Dutch newspaper search from 1872 showed that around 1870, a temperature of 13–15 °C was experienced as comfortable [115]. It links to the conclusion of [116] that in the UK, an increase in winter mean indoor dwelling temperatures of 1.3 °C per decade occurred between 1978 and 1996. Thus, it promotes that the thermal comfort perception depends significantly on the personal history of exposure to extreme environments and can be improved by long-term exposure to dynamic indoor environments. Dynamic and locally varying temperatures can quite easily be implemented in practice. Modern buildings already use dynamic temperature drifts and different local indoor climate zones. In the future, there is a need to undertake monitoring studies in living laboratories and actual daily living conditions, preferably comparing various thermal strategy interventions. It can be accomplished by using living laboratory environments and studying neighborhood effects. The latter can ideally be used for research involving long-term effects on health and well-being in combination with other lifestyle interventions [117].

The body can maintain its core temperature, called homeothermy, through metabolic processes and physiological adjustments. In ambient temperatures lower than the body temperature, the excess heat produced by metabolism can be dissipated, thereby preserving homeothermy. When the ambient temperature falls below the lower limit of the TNZ, the body activates thermogenic mechanisms to produce extra heat and maintain its core temperature. As such, creating a dynamic indoor environment that increases the body's energy expenditure without physical activity could be a potential long-term solution to combat obesity in office buildings while also providing thermally stimulating conditions. The other benefit of a dynamic indoor environment that provides sensory stimulation, such as changing temperature, extends to improving cognitive function by engaging the brain and reducing boredom [118]. Additionally, it can help to alleviate mood and reduce stress levels in the human body [119]. Thus, a dynamic environment benefits energy use reduction and thermal stimulation and can ensure long-term healthy indoor conditions by increasing the human body's energy expenditure.

The effects of cold exposure on fat burning and overall health have been extensively studied in recent years. Firstly, it is essential to understand that our bodies have two types of fat: white adipose tissue (WAT) and brown adipose tissue (BAT). WAT is the more well-known type of fat that stores energy, while BAT generates heat by burning calories. Exposure to cold temperatures has been shown to activate BAT, leading to increased calorie burning and potential fat loss. Lee et al. [120] found that 10 days of exposure to cold temperatures (14–15 °C) increased BAT activity in healthy individuals, leading to a 42% increase in calorie burning. However, the study also noted that the increase in calorie burning was only seen in participants with a healthy BMI and no underlying health conditions. Yoneshiro et al. [121] found that long-term exposure to frigid temperatures (5–8 °C) can increase BAT activity and decrease WAT. The study also noted increased BAT activity and insulin sensitivity, suggesting potential health benefits beyond fat loss.

Exposure to cold temperatures can stimulate the body to burn more calories to maintain its core temperature, potentially leading to weight loss. However, the extent to which cold exposure contributes to weight loss and whether it is a healthy and sustainable strategy is still under debate. It is supported by several studies, including a study by Lee et al. [122], which found that cold exposure increased BAT activation, energy expenditure, and fat oxidation in healthy men. Van Marken Lichtenbelt et al. [16] found that exposing healthy men to cold temperatures (14–15 °C) for 6 h a day over 10 days increased their energy expenditure by about 13%. Yoneshiro et al. [121] showed that exposing overweight and obese individuals to cold temperatures (15–16 °C) for 2 h a day for six weeks led to a small but significant decrease in body fat. However, the study did not assess the long-term effects of cold exposure on weight loss or the participants' overall health. Lee et al. [123] published that exposure to a temperature of 17 °C for 2 h per day for six weeks led to a slight increase in BAT activity in healthy individuals. Yoneshiro et al. [124] found that prolonged exposure to temperatures around 17° slightly increased energy expenditure and fat oxidation in healthy young men. In a separate publication, Yoneshiro et al. [125] reported that exposure to mild cold (16–17 °C) for 2 h per day for six weeks increased energy expenditure and decreased body fat mass in healthy young men.

While some studies suggest that mild cold exposure can improve insulin sensitivity and may have other health benefits [126], it is essential to consider the potential risks associated with long-term exposure to colder temperatures. There is limited research on the long-term health effects of mild cold exposure, and further investigation is needed to make a solid conclusion on this issue.

While these studies suggest that exposure to cold temperatures can lead to increased fat burning and potential health benefits, it is essential to note that prolonged exposure to cold temperatures can also have adverse health effects, such as decreased immune function and increased risk of cardiovascular disease. Lindemann et al. [127] found that older women's physical performance decreased in a moderately cold indoor environment. Okamoto-Mizuno et al. [128] reported that cold exposure during sleep could affect cardiac autonomic response without affecting sleep stages and subjective sensations. Song et al. [129] concluded that prolonged indoor cold exposure could increase the acceptability of the human body to cold exposure but also elevate the vigilance to warm exposure. Héroux [130] found that intermittent cold exposure could lead to decreased body and muscle growth, and overall insulation. These findings suggest that exposure to cold indoor environments can have negative effects on human health and performance, and that further research is needed to fully understand the extent of these effects. Regarding occupant acceptability, it is unlikely that long-term exposure to cold temperatures, i.e., below 16 °C, would be acceptable or practical for most people. However, exposure to a less cold environment with temperatures between 16 and 21 °C may be a more feasible and potentially beneficial option. Exposure to a temperature between 16 and 21 °C may increase BAT activity and thus increase calorie burning. However, the evidence is not as strong as for colder temperatures, and the effects may not be as significant. Overall from the little literature available, it is evident that the advantages of mild cold exposure in terms of the long-term benefits outweigh its potential drawbacks. This review generally sets a good tone for developing the DIET Controller that focuses primarily on heating applications and exposes occupants to mild cold environments during winter.

## 4. Results - Reinforcement Learning-based Controls

This section analyzes the various RL control methods based on their selection of control actions and algorithms used to extract valuable knowledge from the environment and interact with the occupants. RL deals with learning via interaction and feedback. Essentially, an agent (or several) is built to perceive and interpret the environment in which it is placed; furthermore, it can take actions and interact with it. The publications on this topic peaked in 2019 and stayed stable in the past years. The application of RL-based controls in various HVAC systems has been explored in the research studies considering supply systems such as heat pumps, air conditioners, chillers, electric heaters, DOAS, and emission systems, for instance, radiant heating and cooling systems. Still, the most common is the simple HVAC systems with heating and cooling setpoints modeled in Energy Plus.

Regarding the DRL algorithms, Q-learning remains the most explored, with 38% of the papers using it in their HVAC controls (see Fig. 2). Nevertheless, Deep Deterministic Policy Gradient (DDPG) and Asynchronous Advantage Actor Critic (A3C) have shown much promise in recent years. As seen in Figs. 2–5, multi-zone office buildings were most commonly used for the building model with air temperature setpoint as the frequently used controlling the action. Moreover, about 85% and 52% of the studies reported reduced HVAC energy use and thermal comfort maintenance or improvement. In Table 3, we classified the articles based on the DRL algorithm. We analyzed them concerning the HVAC system considered in the model, the conditioning type, the simulation software, the building type and scale, and the outcomes regarding energy use reduction and comfort improvement or maintenance. Moreover, the table also reports the controlling action that was used in the different DRL algorithms.

**Table 3**
Summary of existing studies on RL-based controls.

| Ref | Year | Journal | HVAC System | Conditioning type | Control level (Scale) | Building type | Control type | Control Algorithm | Training in Simulation/ Measurement | Controlled parameters | Deployed in real? | Energy savings | Comfort Improvement |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| [140] | 2017 | DAC '17 | Ideal Load System | C, H | M | Office | RL | Q-Learning | Simulation | Supply air temperature | No | Yes, 5–50% | n/a |
| [141] | 2016 | Sustain. Energy, Grids & Networks | Air Conditioner | C, H | S | Laboratory | RL | Q-learning | Thermal model | HVAC on-off | Yes | n/a | n/a |
| [142] | 2022 | Energy and Buildings | Chiller cooling system | C | M | Office | RL | Q-learning | Simulation | Chilled water temperature set point | No | Yes | n/a |
| [143] | 2021 | Energy and Buildings | VAV System | C, H | M | University | RL | Q-learning | Simulation, Energy Plus | Thermostat set point | No | Yes | Yes |
| [144] | 2010 | Control Eng. Practice | Simple HVAC system | C, H | S | n/a | RL | Q-learning | Simulation | Room air temperature | No | Yes | Yes |
| [145] | 2014 | J of Ambient Intel. & Smart Environ. | Simple HVAC system | H | S | n/a | RL | Q-learning | Simulation, MATLAB | HVAC on-off | No | Yes | Yes |
| [146] | 2019 | Energy Procedia | VAV System | C, H | S | n/a | RL | Q-learning | Simulation, Energy Plus | Supply air temperature | No | n/a | Yes |
| [147] | 2021 | Energy and AI | Heat Pump | H | M | Residence | RL | Q-learning | Simulation, Building Simulator | Heat pump operation modes | No | Yes, 8–16% | Yes |
| [148] | 2020 | Applied Energy | Air Conditioner | C, H | S | Residence | RL | Q-learning | Simulation | AC Set Point temperature | No | Yes | Yes |

(continued on next page)

**Table 3** (*continued*)

| Ref | Year | Journal | HVAC System | Conditioning type | Control level (Scale) | Building type | Control type | Control Algorithm | Training in Simulation/ Measurement | Controlled parameters | Deployed in real? | Energy savings | Comfort Improvement |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| [149] | 2021 | IEEE Trans. on sustain. computing | Simple HVAC System | C, H | M | Office | RL | Q-learning | Simulation, Energy Plus | Supply air flow rate | Yes | Yes, 26.9% | n/a |
| [150] | 2016 | arXiv | Air Conditioner | C, H | S | Laboratory | RL | Q-learning | Thermal model | HVAC on-off | Yes | n/a | n/a |
| [151] | 2015 | Applied Energy | Heat Pump | C, H | M | Residence | RL | Q-learning | Simulation, MATLAB/Simulink | Heat supply level | No | Yes | n/a |
| [152] | 2015 | IEEE Conference Proceedings | Simple HVAC system | C | M | n/a | RL | Q-learning | Simulation, Energy Plus | Zone set point temperature | No | Yes | Yes |
| [153] | 2013 | REHVA World Congress (CLIMA) | Simple HVAC system | C | S | n/a | RL | Q-learning | Simulation | Zone set point temperature | No | Yes, 50% | Yes |
| [154] | 2016 | Sustain. Energy, Grids & Networks | Air conditioner | H | S | Laboratory | RL | Fitted Q-iteration | Simulation and Measurement | Indoor air temperature | No | n/a | n/a |
| [155] | 2017 | Energy Procedia | Heat Pump | C, H | M | Residence | RL | Fitted Q-iteration | Simulation | Room target temperature | No | n/a | n/a |
| [156] | 2017 | IEEE Transactions on smart grid | Heat Pump | H | M | Residence | RL | Fitted Q-iteration | Simulation | Heat pump power level | No | Yes, 19% | n/a |
| [157] | 2018 | IEEE Transactions on smart grid | Simple HVAC system | C, H | S | n/a | RL | CNN-based Q-network | Simulation | Zone air temperature | No | Yes | n/a |
| [158] | 2019 | Building & Environment | Air Conditioner | C, H | S | Laboratory | DRL | Double Q-learning | Simulation | Supply air temperature, fan air flowrate | Yes | Yes, 4–5% | Yes |
| [159] | 2022 | Applied Energy | DOAS | C | S | Office | DRL | Dueling Q-network | Simulation, Modelica | Room temperature set point, Ceiling fan speed | No | Yes, 13.9% | Yes, 11% |
| [160] | 2020 | Building & Environment | AHU | C, H | S | Office | DRL | LSTM based Q-learning | Simulation | Damper position, fan speed, heating valve status | No | Yes, 27–30% | Yes |
| [161] | 2019 | IEEE Transactions on smart grid | Air Conditioner | C | M | Residence | RL | ANN based Q-learning | Simulation, MATLAB | Air conditioner power ratings | No | Yes | No |
| [162] | 2019 | S & T for Built Environment | AHU | C | M | Office | DRL | DQN | Simulation, Energy Plus | Damper position, cooling water temperature | No | Yes, 15.7% | No |
| [163] | 2019 | Energy & Buildings | VRF System | C, H | M | Office | DRL | DQN | Simulation | Set temperature, air flowrate, and humidifier on/off | No | Yes, 12.4–32.2% | n/a |
| [164] | 2021 | Energy and Buildings | Simple HVAC system | C, H | S | n/a | DRL | DQN | Simulation | Zone temperature set point | No | Yes, 6% | n/a |
| [165] | 2022 | Applied Thermal Engineering | VAV System | C, H, V | M | Office | DRL | DQN | Simulation, Energy Plus | Supply air temperature, Chilled water temperature | No | Yes | Yes |
| [166] | 2022 | Energy and Buildings | Cooling water system | C | M | Office | DRL | DQN | Simulation, Energy Plus | Cooling load distribution, water pumps | Yes | Yes, 11% | n/a |
| [167] | 2022 | Building and Environment | VAV System | C, H, V | M | Commercial | DRL | DQN | Simulation, Energy Plus | Heating and cooling set point | No | Yes, 13% | Yes, 9% |
| [168] | 2019 | BuildSys '19 | Simple HVAC system | C, H | M | n/a | DRL | DDQN | Simulation, Energy Plus | Temperature set-point of HVAC system | No | Yes, 8.1–14.26% | No |
| [169] | 2022 | ACM/IEEE Conference | Multi-zone HVAC | C, H, V | M | Office | BRL | Batch DQN | Simulation | Zone air temperature, supply airflow set point | No | Yes, 7.2% | Yes, 16.7% |
| [170] | 2019 | IEEE Transactions on smart grid | HVAC load dataset | C, H | M | Data Center | DRL | DQN, DDPG | Simulation | air conditioner on/off, electric vehicle on/off | No | Yes | n/a |

(*continued on next page*)

**Table 3** (*continued*)

| Ref | Year | Journal | HVAC System | Conditioning type | Control level (Scale) | Building type | Control type | Control Algorithm | Training in Simulation/ Measurement | Controlled parameters | Deployed in real? | Energy savings | Comfort Improvement |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| [171] | 2019 | arXiv | Ideal Load System | C, H | S | n/a | DRL | DDPG | Simulation, TRNSYS | Set point air temperature and humidity | No | Yes | Yes |
| [172] | 2021 | Applied Energy | Simple HVAC System | C, H | M | Residence | DRL | DDPG | Simulation | Zone set point temperature | No | Yes, 15% | Yes, 79% |
| [173] | 2020 | IEEE Internet of Things journal | Ideal Load System | C, H | S | n/a | DRL | DDPG | Simulation, TRNSYS | Set point air temperature and humidity | No | Yes, 4.31–9.15% | Yes |
| [174] | 2020 | IEEE Transactions on Cybernetics | Chiller cooling system | C | M | Data Center | DRL | DDPG | Simulation, Energy Plus | Cooling air temperature, chilled water temperature | No | Yes, 11% | Yes |
| [175] | 2019 | arXiv | Simple HVAC system | C | M | Residence | DRL | DDPG | Simulation | HVAC input power | No | Yes | Yes |
| [176] | 2022 | Energy | Simple HVAC system | C, H | M | n/a | DRL | DDPG | Simulation | HVAC Power | No | Yes, 12.8% | No |
| [177] | 2021 | Electric Power Systems Research | Load dataset | C, H | M | Residence | DRL | DDPG | Simulation | Indoor air temperature | No | Yes | n/a |
| [178] | 2022 | Connection Science | Multi-zone HVAC | C | M | Residence | DRL | DDPG | Simulation | Set point air temperature | No | Yes, 3.5–5% | Yes, 65–68% |
| [179] | 2019 | Energy Procedia | VAV system | C, H | M | Office | DRL | Policy gradient | Simulation | Supply air temperature | No | n/a | n/a |
| [180] | 2020 | IEEE Transactions on smart grid | AHU | C, H, V | M | Office | DRL | Actor-Critic | Simulation | AHU damper position, air supply rate | No | Yes, 56–75% | Yes |
| [181] | 2020 | Applied Energy | Electric heater | H | S | n/a | DRL | Actor-Critic | Simulation, Energy Plus | Zone air temperature set point | Yes | Yes | Yes |
| [182] | 2017 | MDPI - Processes | VAV System | C, H | S | Office | DRL | Actor-Critic | Simulation | Set point air temperature | No | Yes, 5.03% | Yes, 15.5% |
| [183] | 2018 | ASHRAE Conference | Radiant heating system | H | M | Office | DRL | A3C | Simulation, Energy Plus | Heating system supply water temperature | No | Yes, 15% | No |
| [184] | 2018 | BuildSys '18 | Radiant heating system | H | M | Office | DRL | A3C | Simulation, BuildSim | Supply water temperature set point | Yes | Yes, 16.6–18.2% | n/a |
| [185] | 2019 | Energy & Buildings | Radiant heating system | H | M | Office | DRL | A3C | Simulation, Energy Plus | Indoor air temperature, Supply water temperature | Yes | Yes, 16.7% | Yes |
| [186] | 2018 | arXiv | AHU | C, H | M | Data Center | DRL | TRPO | Simulation | air temperature, supply fan air mass flow rate | No | Yes, 22% | No |
| [187] | 2020 | Energy and AI | Simple HVAC system | C, H | M | Office | DRL | PPO | Simulation, Energy Plus | Zone temperature set points | No | Yes, 22% | Yes |
| [188] | 2022 | Applied Energy | Chiller cooling system | C, V | M | Data Center | DRL | SAC | Simulation, Energy Plus | Supply air temperature, mass flow rate | No | Yes, 5% | n/a |
| [189] | 2022 | Building and Environment | Simple HVAC system | C, V | M | Office | DRL | MAAC | Simulation, Energy Plus | HVAC temperature set point, Fan speed | No | Yes, 1–4% | Yes, 64–72% |
| [190] | 2021 | Applied Energy | Chiller cooling system | C, V | M | Data Center | DRL | SAC, TD3, TRPO, PPO | Simulation, Energy Plus | Supply air temperature set point, air mass flow rate | No | Yes, 13% | Yes |
| [191] | 2019 | ACM Conference Proceedings | Radiant heating system | H | M | Commercial | RL | Gnu-RL | Simulation, Energy Plus | Supply water temperature | Yes | Yes, 16.7% | Yes |

(*continued on next page*)

**Table 3** (*continued*)

| Ref | Year | Journal | HVAC System | Conditioning type | Control level (Scale) | Building type | Control type | Control Algorithm | Training in Simulation/ Measurement | Controlled parameters | Deployed in real? | Energy savings | Comfort Improvement |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| [192] | 2022 | Applied Energy | VAV System | C, H, V | M | Residence | RL | HDCMARL | Simulation | Duct dampers, Cooling coil valves | No | Yes, 32% | n/a |
| [193] | 2007 | Building and Environment | Heat Pump | C, H, V | S | n/a | RL | State-Action function | Simulation, MATLAB/Simulink | Heat Pump on-off, air ventilation on-off | No | No | Yes |
| [194] | 2020 | Energy | Simple HVAC System | C, H | M | Office | DL | ANN | Simulation, MATLAB | Supply air temperature | No | Yes, 30.72% | n/a |
| [195] | 2019 | Applied Energy | Air Conditioner | C | S | Office | DL | SLFF-ANN | Measurement | Current air temperature | Yes | Yes, 36.5% | Yes |
| [196] | 2016 | Energy & Buildings | n/a | C | M | Residence | DL | ANN, MLP | Simulation, MATLAB | Indoor air temperature, relative humidity | No | n/a | n/a |
| [197] | 2020 | Journal of Cleaner Production | Simple HVAC system | C, H | S | n/a | DL | MLP | Simulation, Energy Plus | Indoor air temperature, relative humidity | No | Yes | Yes |
| [198] | 2015 | Energy & Buildings | AHU | C, H | M | Office | DL | ANN-based on-off | Mathematical model | Zone air temperatures | No | No | n/a |
| [199] | 2019 | Building & Environment | Heat Pump | C, H | M | Office | DL | LSTM, BPNN, SVM, DT | Measurement | Indoor air temperature | Yes | n/a | n/a |
| [200] | 2012 | Energy & Buildings | VRF System | C, H | M | University | DL | RBF ANN | Measurement | Zone air temperature, relative humidity | Yes | Yes, 50% | Yes |
| [201] | 2005 | Energy & Buildings | Air Conditioner | C, H | S | Office | IEEMS | Fuzzy logic | Simulation | HVAC on-off | Yes | Yes, 38% | Yes |
| [202] | 2016 | Applied Energy | Simple HVAC system | C, H | S | Laboratory | MPC | MPC-DTS | Measurement | Supply air temperature | Yes | Yes, 25% | Yes |
| [203] | 2019 | Building and Environment | Simple HVAC system | C, V | M | Laboratory | ML | GNB, DT, SVM, MLP | Measurement | Room air temperature | Yes | Yes, 4–25% | n/a |
| [204] | 2014 | IFAC | n/a | C, H | M | Office | ML | Occupant feedback | Mathematical model | Room set-point temperature | No | n/a | n/a |

C – Cooling, H – Heating, V – Ventilation, S – Single-zone, M – Multi-zone.

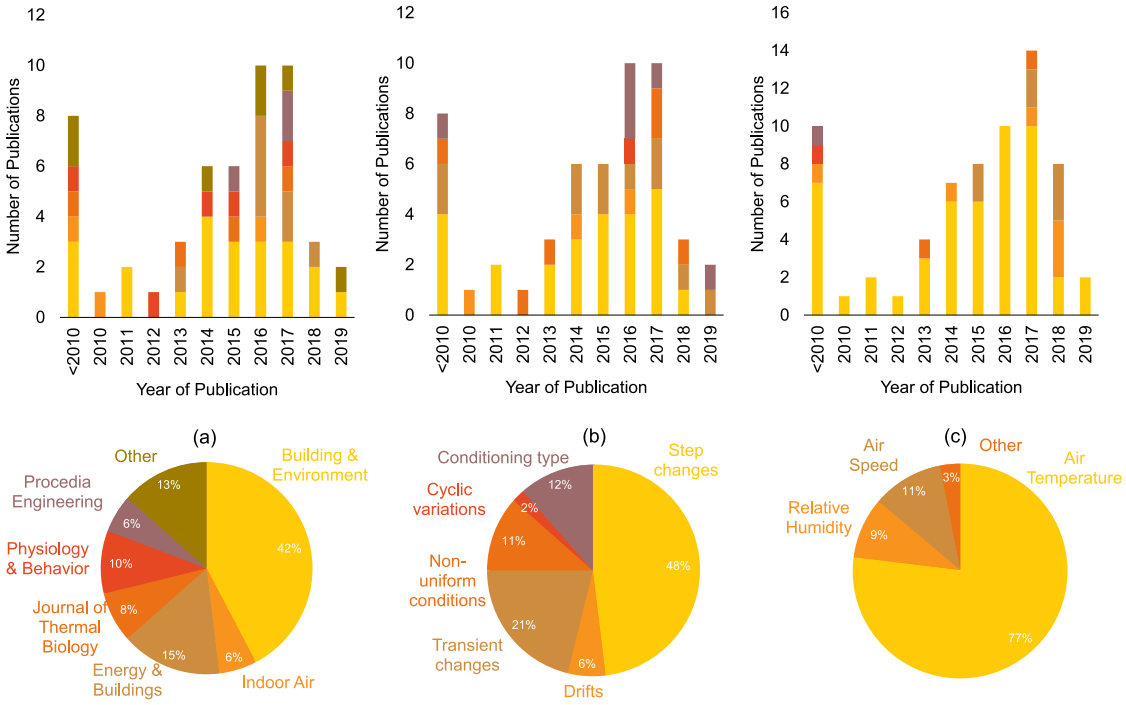



**Fig. 1.** Yearly distribution of the articles on the dynamic environment concerning (a) publication journals, (b) Variation type, and (c) variable.

Concerning HVAC system control, there is also pioneering research in the literature focusing on improving energy efficiency and economy using powerful Deep RL approaches. In Wei et al. [131], a deep Q network (DQN) for coordinated control of shared data center and HVAC loads is built, using a neural network to estimate Q values of paired state actions. In Claessens et al. [132], a Convolutional Neural Network (CNN) approximates the state action value function to capture better the spatial and temporal correlations in the input state data in the convolution operation. In Mocuna et al. [133], deep policy gradient (DPG) methods are studied to control many responsive needs, such as air conditioners, electric vehicles, and dishwashers. Wang et al. [134] adopted an essential strategy for optimizing thermal comfort and HVAC energy use. Zhang et al. [135] establish a practical HVAC control framework based on the benefit of stakeholder assessments for building-wide energy models. Ahn et al. [136] use DQN to achieve optimal control coordination between different HVAC systems.

The above studies demonstrate the effectiveness of the applied Deep RL method in optimizing HVAC thermal control strategies compared to designed benchmarks. In Gao et al. [137], the author uses his DDPG method to achieve continuous thermal control of his HVAC without

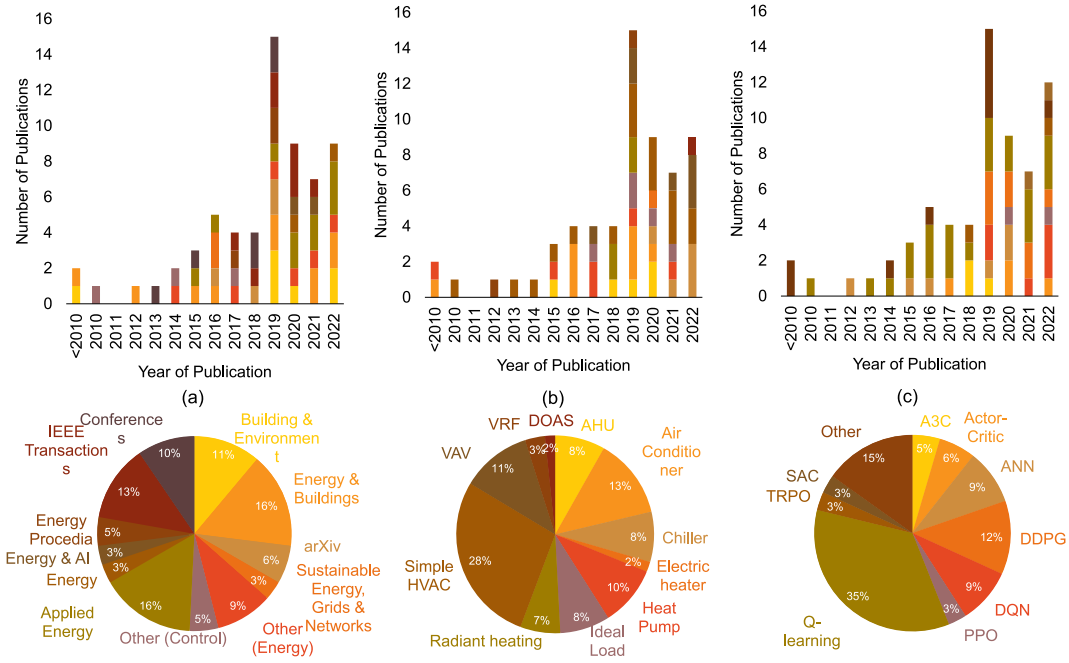



**Fig. 2.** Yearly distribution of the articles on RL controls concerning (a) publication journals, (b) HVAC system type, and (c) DRL algorithm.

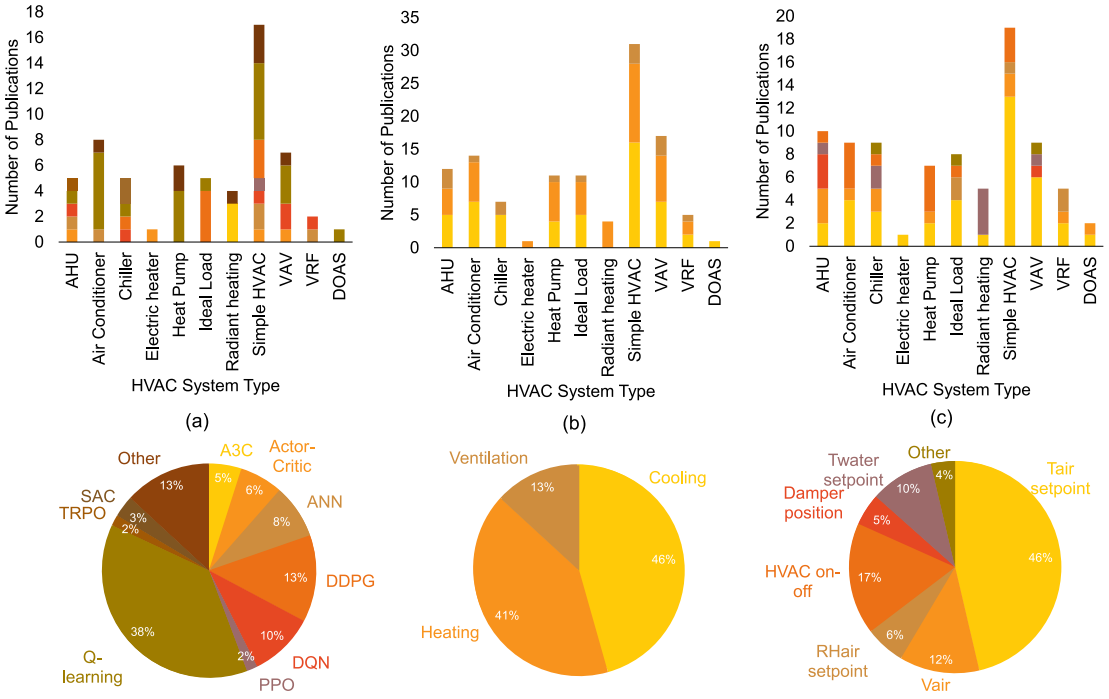



**Fig. 3.** (a) DRL algorithm, (b) conditioning type, and (c) controller parameters used in different HVAC systems.

arbitrariness. However, this study still focuses on its HVAC control for a single area previously treated by the discretionary method. Furthermore, the applied approach was only compared with other RL methods, and no reference cases were developed to verify the optimality and generality of the control strategy obtained. In Yu et al. [138], the multi-agent deep RL method with an attention mechanism is applied to minimize the energy use of HVAC systems in multi-zone commercial buildings. They are updated in parallel during training. Although this study provides interesting insights, it is worrisome that for the proposed algorithm, the number of neural networks that need to be trained increases with the number of regions, which may lead to excessive computational load. In Zou et al. [139], a long short-term memory (LSTM) network was combined with DDPG to simulate better the actual operation of multiple air handling units (AHUs).

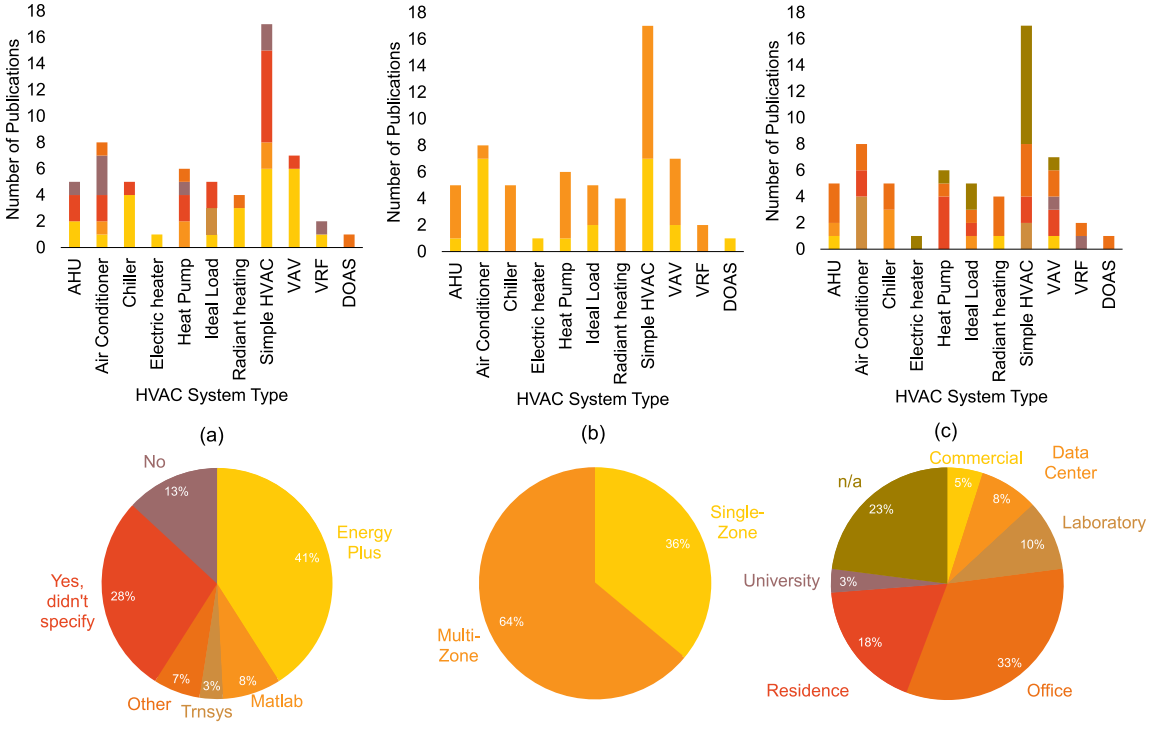



**Fig. 4.** (a) Simulation software, (b) building model scale, and (c) building type used in different HVAC systems.

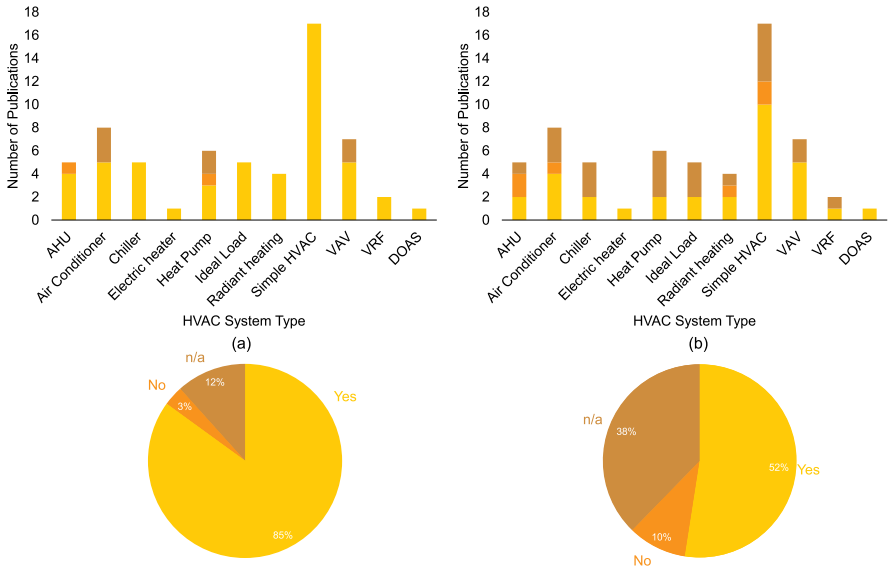



**Fig. 5.** (a) Energy use reduction and (b) thermal comfort improvement for different HVAC systems.

One of the most critical branching points in an RL algorithm is whether the agent has access to a model of the environment. The main upside of model-based methods is that they allow the agent to plan by thinking ahead, seeing what would happen for a range of possible choices, and explicitly deciding between options. Agents can then distill the results from planning into a learned policy. When this works, it can substantially improve sample efficiency over methods that do not have a model. The main downside is that a ground-truth model of the environment is usually not available to the agent. If an agent wants to use a model in this case, it has to learn it purely from experience, creating several challenges. The biggest one is that bias in the model can be exploited by the agent, resulting in an agent that performs well concerning the learned model but behaves sub-optimally in the natural environment. Model learning is fundamentally complex, so even intense effort can fail to pay off. On the other hand, algorithms that do not use models are called model-free. While model-free methods forego the potential gains in sample efficiency from using a model, they tend to be easier to implement and tune. Model-free methods are more popular and have been more extensively developed and tested. It is also true in the applications of RL in HVAC control development. Thus, the papers reviewed in the article mainly consist of model-free algorithms such as Q-learning, DQN, DDPG, and A3C.

### 4.1. Algorithms

In this subsection, the findings in terms of the RL algorithms have been summarized in Table 4 and with some detailed discussion. The purpose is to examine how current research has methodologically explored the algorithms and to propose potential future work.

Several studies [140–161] leveraged the Q-learning algorithm to develop the model-free controller. The traditional Q-learning algorithm has been used relatively equally for heating and cooling applications. The building models comprised single and multi-zone offices, residences, and laboratories. The training of the models was primarily

**Table 4**
Summary of the differences between popular ML and RL algorithms for dynamic indoor environment control application.

| Algorithm | Learning Type | Architecture | Advantage | Disadvantage |
|---|---|---|---|---|
| ANN | Supervised | Feedforward Neural Network | Simple architecture can handle simple problems | Limited to static environments, requires labeled training data |
| CNN | Supervised/ Unsupervised | Convolutional Neural Network | Can extract features from raw data, suitable for image processing | Limited to grid-like input data, may require pre-training |
| RNN | Supervised/ Unsupervised | Recurrent Neural Network | Can handle sequential data, suitable for time-series analysis | May suffer from vanishing gradients, requires careful architecture design |
| Q-learning | Model-free | Neural Network | Simple, easy to implement | Limited to small state-action spaces, requires a good reward function |
| DQN | Model-free | Deep Neural Network | Can handle complex and large state-action spaces, more efficient | Requires a large amount of training data, sensitive to hyper parameters |
| Actor-critic | Model-free | Two networks: actor and critic | Can learn continuous action spaces, more stable than DQN | More complex architecture, slower training |
| DDPG | Model-free | Two networks: actor and critic | Can learn continuous action spaces, less sensitive to hyper parameters than DQN | More complex architecture, requires more training data than actor-critic |
| SAC | Model-free | Two networks: actor and critic | Can handle continuous action spaces, more sample efficient than DDPG | More complex architecture, slower training, requires careful hyper parameter tuning |
| A3C | Model-free | Asynchronous actor-critic | More sample efficient, can scale to multiple CPUs | More complex architecture, slower training, sensitive to hyper parameters |
| TRPO | Model-free | Trust Region Policy Optimization | More stable training, can handle non-linear policies | Slower than other algorithms, requires more hyper parameter tuning |
| PPO | Model-free | Proximal Policy Optimization | More stable than other policy gradient methods | Requires more hyper parameter tuning than DDPG and SAC |
| TD3 | Model-free | Twin Delayed DDPG | More stable training, can handle continuous action spaces | More complex architecture, requires more hyper parameter tuning than DDPG and SAC |

carried out in Energy Plus and Matlab, with the most common controlling actions being air temperature set points and HVAC system on-off. Regarding the above research outcomes, the studies reported reduced energy use in the 5–50% range. While most studies [140–153] used the traditional Q-learning algorithm, some used a modified version. Studies [154–156] used fitted Q-iteration to model heat pump controls and air conditioners for heating and cooling applications using laboratories or residences as the building model. In Claessens et al. [157] and Lu et al. [161], the researchers adopted a neural network-based Q-learning, more precisely using Convolutional Neural Networks (CNN) and Artificial Neural Networks (ANN), respectively. There were also some other variations, for instance, the double Q-learning algorithm [158], the dueling Q-learning algorithm [160], and Q-learning with experience replay [146]. The advantages of using Q-learning for dynamic indoor environment control are as follows:

- Q-learning is a model-free RL algorithm, which means it can learn to make decisions without requiring a complete environment model, making it suitable for complex, dynamic indoor environments.
- It can learn to optimize control policies through trial-and-error interactions with the environment, leading to better energy efficiency, comfort, and cost savings.
- It can adapt to changing conditions in real-time, making it suitable for unpredictable or rapidly changing environments.

However, it also comes with its share of disadvantages when used in the context of dynamic indoor environment control, which is as follows:

- Q-learning requires significant computation resources, particularly for large and complex environments.
- This algorithm can be sensitive to the quality of the reward function, which must be designed carefully to encourage the desired behavior.
- It is not always guaranteed to converge to an optimal policy and may require significant tuning of hyper parameters to achieve good performance.

The second most commonly used DRL algorithm in the literature was the deep Q-networks (DQN). Quite a few studies [162–170] employed the DQN algorithm for modeling the HVAC control in multi-zone office buildings or data centers. The traditional DQN algorithm has been used primarily for heating and cooling, but there were also studies where it was used for ventilation. The building models used consisted of both single and multi-zone models with Variable Air Volume (VAV), Variable Refrigerant Flow (VRF) systems, and Air Handling Units (AHU). The training of the models was primarily carried out in Energy Plus, with the most common controlling actions being air temperature set points, heating or cooling water temperature set points, damper position, and HVAC system on-off.

Regarding the above research outcomes, the studies reported reduced energy use in the 8–32% range. Ding et al. [168], in particular, adopted an interesting approach by using Dueling Double Deep Q-network (DDQN) to model a simple HVAC system in Energy Plus for heating and cooling. As per their results, DDQN reduced energy use by 8–14% compared to state-of-the-art rule-based controllers. The main advantages of using DQN for dynamic indoor environment control compared to Q-learning are as follows:

- DQN incorporates a deep neural network that can learn complex representations of the environment and control policies, enabling it to handle more extensive and complex indoor environments.
- It is more efficient regarding computational resources, using function approximation to generalize across states and actions.
- DQN is more stable than Q-learning, as it uses experience replay to learn from a pool of past experiences, reducing the impact of sequential correlations in the data.

Disadvantages of using DQN for dynamic indoor environment control compared to Q-learning:

- DQN can suffer from the problem of overestimation of action values, which can lead to suboptimal control policies.
- It can be sensitive to the hyper parameters of the deep neural network, such as the learning rate and the architecture, which must be tuned carefully to achieve good performance.
- Learning a good policy requires much training data, which may not be feasible in some indoor environments with DQN.

The following DRL algorithm that we focus on is the actor-critic. We reviewed three studies [180–182] that used the actor-critic algorithm for single and multi-zone office buildings. They considered the VAV system, electric heater, and AHU for heating, cooling, and ventilation applications. The model training was done in Energy Plus, with air temperature set points, airflow rate modulation, and damper position being the controlling actions. The studies reported reduced energy use in the 5–75% range and improved thermal comfort by 15%. Actor-critic and DQN are model-free Reinforcement Learning algorithms for dynamic indoor environment control. Here are some of the advantages and disadvantages of using actor-critic over DQN,

Advantages of actor-critic:

- Actor-critic is well suited for environments with continuous action spaces, whereas DQN requires discretization of the action space, leading to reduced performance.
- It typically has more stable training than DQN, which can be prone to instability caused by overestimating Q-values.
- Actor-critic learns policies directly, whereas DQN learns Q-values, which must then be converted into policies through an additional step.

Disadvantages of actor-critic:

- Actor-critic has a more complex architecture than DQN, which can make it harder to implement and more computationally expensive.
- It has slower training than DQN, which can be a drawback if time is critical.
- It typically requires more data to perform better than DQN, making it sometimes less efficient.

Overall, actor-critic may be better than DQN for dynamic indoor environment control applications where continuous action spaces are involved, and stability is critical. However, the increased complexity and slower training of actor-critic may make it less attractive in some instances where these factors are essential. DDPG is a state-of-the-art algorithm that improves some of the shortcomings of DQN and actor-critic. Gao et al. [171,173] leveraged the DDPG algorithm to model the HVAC control in single-zone buildings for heating and cooling purposes. Mocanu et al. [170], Du et al. [172], Li et al. [174], and several other studies [175–178] did the same for multi-zone residential buildings and data centers and showed that DDPG could effectively provide 4–15% energy savings in comparison to on-off, Q-learning, DQN and SARSA based controllers while maintaining the occupant thermal comfort. Trnsys and Energy Plus were the most used software for modeling the building with air temperature set points, heating or cooling water temperature set points, and HVAC power level as the controlled parameters. DDPG (Deep Deterministic Policy Gradient) is a model-free RL algorithm well suited for dynamic indoor environment control applications with continuous action spaces. Some of the advantages of using DDPG include the following:

- DDPG is a model-free algorithm that does not require a system dynamics model. It makes it well-suited for controlling systems with complex or unknown dynamics.

- It can handle continuous action spaces, which is vital for many control tasks without discrete actions.
- DDPG uses actor-critic architecture, which helps to stabilize the learning process. The actor network generates actions, and the critic network evaluates the actions; this separation of responsibilities often results in more stable learning.
- It is an off-policy algorithm, which means it can learn from past experiences stored in a replay buffer. It allows the agent to learn from diverse experiences, which can lead to more robust and generalizable policies.
- DDPG can handle high-dimensional state spaces using a neural network as a function approximator. It allows the agent to learn complex, non-linear policies that effectively address many inputs.
- It can be combined with other methods, such as curiosity-driven exploration, hyper parameter tuning, and parallelization, improving the controller's performance.

Overall, DDPG may be a better choice than DQN and actor-critic for dynamic indoor environment control applications where continuous action spaces are involved, and sample efficiency is a critical concern. However, the increased complexity and data requirements of DDPG and the potential for overfitting may make it less attractive in some instances where these factors are essential. Ultimately, the choice between these algorithms depends on the indoor environment control application's specific requirements.

Zhang et al. [183–185] have used the A3C algorithm to develop Reinforcement Learning-based controls. In these studies, the HVAC system was a radiant heating system for multi-zone office buildings. The papers showed that by employing A3C, there is a possibility of 15–18% energy savings compared to baseline rule-based controllers while maintaining or improving the thermal comfort perception of the occupants. In the studies [186–190], the authors investigated the performance of new and emerging DRL algorithms like soft actor-critic (SAC), Proximal Policy Gradient (PPO), Trust Region Policy Optimization (TRPO) and Twin-delayed DDPG (TD3). These algorithms were mainly tested for multi-zone office buildings and data centers with a chiller cooling system and AHU being the modeled HVAC system. In all these studies, the researchers used Energy Plus for the simulations and controlled the air temperature set point and flow rate. The papers reported reduced energy use by 5–22% and improved thermal comfort by 64–72%. Overall, A3C, TD3, and SAC may be better choices than DDPG for dynamic indoor environment control applications where sample efficiency, stability, and multi-modal policies are essential. However, these algorithms may require more training time and be more computationally expensive than DDPG, which can sometimes be a drawback.

Apart from the DRL algorithms mentioned above, other control-oriented studies used different ML-based algorithms for energy use reduction and thermal comfort improvement. Between 2012 and 2020, seven studies [194–200] leveraged the traditional or modified version of Artificial Neural Networks (ANN) to develop controllers for heating and cooling applications in residences and offices. In almost all of these papers, either Energy Plus or Matlab was used for the simulation. The controlled parameters were the air temperature or humidity set point, reducing energy use by 30–50%. There are a few disadvantages of using traditional ML algorithms for dynamic indoor environment control compared to using DRL algorithms:

- ML algorithms are typically limited to supervised learning. They rely on labeled training data to learn from, which can be challenging to obtain for dynamic indoor environment control applications.
- They are typically not designed to interact sequentially. It means that they cannot consider the long-term effects of their actions on the environment, which is vital for dynamic indoor environment control.
- ML algorithms typically learn a static policy that maps inputs to outputs, i.e., they cannot adapt to environmental changes, which is essential for dynamic indoor environment control.
- Many ML algorithms are designed to work with discrete action spaces, which can be limiting for dynamic indoor environment control applications with continuous action spaces.

In contrast, DRL algorithms can address these limitations by enabling agents to learn from experience and sequentially interact with the environment. DRL algorithms can also learn dynamic policies that adapt to changes in the environment and handle continuous action spaces. However, DRL algorithms can be more complex and computationally expensive than traditional ML algorithms, which can sometimes be a disadvantage. Ultimately, the choice between ML and DRL algorithms depends on the specific requirements of the dynamic indoor environment control application.

### 4.2. HVAC system controls

In RL, action is the controller's decision regarding controlling the environment. In the case of HVAC control, the action could be adjusting the indoor temperature set point, supply air temperature, and fan speed. HVAC control is complex for two reasons. First, there are different components: a terminal, an air-handling unit, a heating/cooling source, and a condenser; for each component, there are different device types; for instance, the terminal could be a variable air volume (VAV) box or baseboard radiator. Second, for each device, there are different levels of controls. The controller could directly control the actuator level or the set point (supervisory control). Conventional controllers are needed to control the actuator to track the set point if the supervisory control is selected. As seen in Table 3, various HVAC systems have been the subject of the development of RL-based controls. It comprises single-zone air conditioners, multi-zone supplying air handling units, Heat Pumps, Radiant heating systems, Variable Air volume (VAV) systems, and Variable Refrigerant Flow (VRF) systems. Many studies modeled simple HVAC or ideal air load systems in simulation software such as Energy Plus.

In terms of the controlling action, there are endless possibilities in the context of HVAC controls with RL, as seen in Table 3. The most commonly controlled parameter was the zone set point or air supply temperature, seen chiefly with Air conditioners or simple HVAC systems modeled in Energy Plus. Again, in the case of air conditioners and heat pump systems, turning on off or modifying the system's power level (if available) was frequently observed in the studies. In studies with radiant water heating systems, the most common variable to be controlled was the supply hot water temperature set point. Last, a few research works, mostly involving AHU, leveraged the damper position or the fan speed as the modulating action.

Interestingly, out of the projects reviewed, only about 28% of the studies were deployed or tested in real life. Otherwise, all the other studies focused on training the controls via co-simulation. For this purpose, a combination of building modeling software and programming languages was used. The single or multi-zone building models were developed in building energy modeling software such as Energy Plus, Trnsys, and BuildSim. The controller logic was coded in Python or MATLAB using the RL packages like PyTorch and TensorFlow. Energy Plus is a whole building energy simulation program that engineers, architects, and researchers use to model energy use (heating, cooling, ventilation, lighting, and plug and process loads) and building water use. In many studies, Energy Plus was used along with BCVTB. The Building Controls Virtual Test Bed (BCVTB) is a software environment that allows users to couple different simulation programs for co-simulation and to couple simulation programs with actual hardware. For example, the BCVTB enables the simulation of a building in Energy Plus and the HVAC and control system in MATLAB or Python.

In addition, the BCVTB allows expert simulation users to expand individual programs' capabilities by linking them to other programs. Another commonly used building simulation tool is Trnsys, an extremely flexible graphically based software environment used to simulate the

behavior of transient systems. It consists of an extensive library of components, each of which models the performance of one part of the system. The standard library includes approximately 150 models ranging from pumps to multizone buildings, weather data processors to economics routines, and essential HVAC equipment to cutting-edge emerging technologies. Co-simulation enables combining complementary features available in the coupled tools, i.e., defining complex building models in Energy Plus or Trnsys and writing detailed codes for different control algorithms in Python. Co-simulation facilitates a fully integrated design analysis, which would not have been possible if any of the BPS tools had been used individually.

Simulation based methodologies come with their own set of limitations and challenges. Simulation-based approaches rely on mathematical models and assumptions to represent real-world systems. The accuracy of the results depends on the quality of the models and the data used for calibration. Some simulations can be computationally intensive, requiring significant computing resources and time, especially for large and complex systems. To make simulations computationally feasible, certain simplifications are often made, which may not fully capture all real-world complexities and interactions. Validating and calibrating simulation models can be challenging, and discrepancies between simulated and observed data may occur.

On the other hand, implementing RL algorithms in real world comes with its own set of problems. RL algorithms often require large amounts of data for effective learning, and real-world data collection can be expensive and time-consuming. In critical applications, RL algorithms must be designed with safety in mind to avoid unintended consequences or hazardous actions. Balancing exploration (trying new actions) and exploitation (choosing known actions) is challenging, especially in safety-critical systems or environments with limited resources. Training RL agents can be sample-inefficient, requiring many interactions with the environment, which may not be practical in real-world scenarios. Additionally it is quite complicated to integrate RL algorithms with the existing HVAC system, which can require extensive background knowledge and expertise in both.

Research should focus on developing more accurate and realistic simulation models to better represent complex real-world systems. Future work should aim to develop RL algorithms that are computationally efficient and scalable to handle real-world applications. Developing data-efficient RL algorithms that require fewer samples for learning will be essential for real-world implementation. Addressing safety concerns and ensuring robustness of RL algorithms will be critical for their deployment in safety-critical applications. Combining simulation-based and real-world data-driven approaches can leverage the strengths of both paradigms for more practical and effective solutions. Future research should explore human-in-the-loop RL, where human expertise and guidance can be combined with RL algorithms to enhance decision-making in complex real-world environments.

### 4.3. Energy savings potential

A primary benefit of a dynamic environment is increased HVAC energy savings. For small, medium, and large office buildings, selecting daily optimal set points would lead to 10–37% savings, depending on the climate. Daily optimal dead band selection of 3–6 °C could result in an average energy use reduction of 10–21%, respectively, compared to baseline 3 °C. Daily optimal set point selection in ranges of 22.5 ± 1 °C, 22.5 ± 2 °C, and 22.5 ± 3 °C can potentially result in an average saving of 8%, 13%, and 16%, respectively [205]. A widened thermostat set point range results in significant energy savings if implemented correctly. Hot climates benefit more from increased cooling set points, whereas cold temperatures benefit more from decreased heating set points. As a result of the simulation study conducted by Chatterjee et al. it was found that the dynamic heating profiles can reduce the heating energy consumption by 33–73% and the total energy demand by 16–36% as compared to the baseline setup using SIA 2024–2015 in office buildings in Switzerland [206]. A wide thermostat set point range such as 18.3–27.8 °C can save 32%–73% of HVAC energy use [207]. Ghahramani et al. [208] demonstrated that the optimal building level daily energy control policy results in average savings of 28–51%, depending on the climate. In addition, if thermal comfort requirements were uniformly distributed, the daily optimal set point selection, subject to thermal comfort constraints, led to 18–38% energy savings, depending on the climate. These savings are conservative as thermal comfort preferences are often skewed toward energy-efficient set points.

In most of the reviewed papers focusing on developing RL-based HVAC controls, the main emphasis was on energy savings or reduction of electricity costs. About 80% of the documents reviewed in this article reported increased energy savings by the developed RL-based controller compared to the traditional rule-based controller. Wei et al. [133], Nikovski et al. [146], Yu et al. [173], and Xu et al. [193] have shown the highest reduction in energy costs, about 50–75%, as compared to rule-based or heuristic controllers. Yu et al. [165] used Multi actor attention-critic algorithms to control the damper position and air supply rate to optimize the performance of Air Handling Units in Office buildings. Almost half of the papers stated energy savings in the 5–30% range. In addition, 62% of the research projects also considered maintenance or improving thermal comfort conditions as an additional objective of the HVAC controller. Furthermore, half of the studies developed a controller that achieved both goals. Du et al. [156], in particular, used the DPPG algorithm to reduce comfort violations by 79% compared to rule-based strategies.

The findings on the effectiveness of RL-based controls in multi-zone office buildings and data centers present exciting opportunities for fostering novel and sustainable building and infrastructure development ventures. RL algorithms have demonstrated their capability to efficiently tackle complex tasks, making them valuable tools for optimizing energy use and enhancing occupant comfort in large-scale building systems.

From a sustainability aspect, these findings have several implications. First, the significant energy reductions achieved by RL-based controls have a direct impact on environmental conservation. By curbing energy consumption in office buildings and data centers, which are substantial energy consumers, these algorithms contribute to mitigating greenhouse gas emissions and alleviating the strain on energy resources. These energy savings are aligned with global sustainability goals, such as reducing carbon footprints and combatting climate change.

Moreover, the adoption of RL-based controls in building and infrastructure development ventures can yield substantial economic benefits. The observed energy use reduction translate to reduced operating costs for building owners and operators, enhancing their financial viability and long-term profitability. The potential for cost savings encourages investments in sustainable building projects, driving market demand for innovative and energy-efficient solutions. In terms of human-centric design, the improvements in thermal comfort achieved by the actor-critic algorithm are crucial for occupant well-being and productivity. A comfortable indoor environment fosters occupant satisfaction, leading to enhanced performance and reduced absenteeism. This human-centric approach aligns with sustainable design principles, where buildings are not only energy-efficient but also optimize the well-being and comfort of their occupants.

## 5. Discussion

The previous sections presented works that focused on the different aspects of dynamic indoor environments and RL-based controls. Section 3 introduced the permissible limits of temperature step changes and drifts acceptable to human occupants. It also discussed the flexibility of the range of human thermal comfort and adaptation in different parts of the world. Moreover, an important point was made regarding how creating a dynamic indoor environment can benefit the human body in

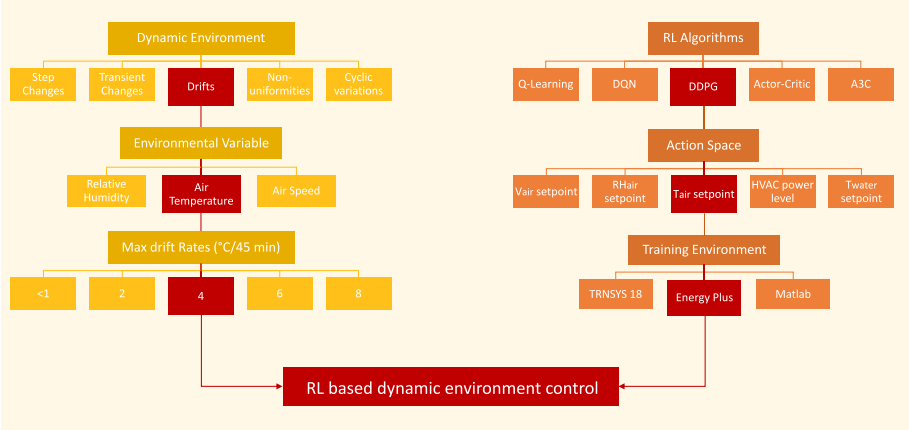



**Fig. 6.** Linking of the ideas for the creation of dynamic indoor environment controls using RL.

the long term. Section 4 explored the studies that worked with RL and focused on HVAC controls. The different algorithms, HVAC systems, co-simulation environment, action spaces, and energy-saving potentials were discussed. There was lack of research that intended to bring the two concepts together. Research studies have proved that creating a dynamic indoor environment that leads to thermal exploration on the mild cold side can have health benefits in the long term. RL algorithms, particularly DQN, DDPG, and A3C, have shown their potential to contribute to developing HVAC controls. Thus, it creates the opportunity to build HVAC control strategies based on RL to create a dynamic indoor environment. It would be a novel approach to designing the control of the building micro-climate.

As mentioned before, creating such control is not an easy task. Based on the literature review conducted in the previous section, RL-based authorities have shown promise in tackling such complicated tasks. At the beginning of the chapter, an important question was raised, i.e., how one can create a dynamic indoor environment optimizing occupants' energy, comfort, and health. This section summarizes the literature review's findings to answer the above question and obtain the essential elements of RL control. A graphical representation of this summary is provided in Fig. 6. The air temperature was the simplest and most frequently changed variable to enable a dynamic environment. Considering the thermal adaptation of the people, the acceptable temperature range was found to be 15–30 °C. Drifts were the most favorable mode of change in air temperature, with 4 °C being the maximum permissible limit for the difference in 45 min. Finally, the outcomes of the creation of a dynamic indoor environment are three-fold, (i) reduction of energy use, (ii) long-term health benefits for the occupants, and (iii) creation of thermally acceptable conditions.

The information summarized in the previous paragraph was used as input for the RL control. The first step was to decide on the DRL algorithm to work with, and DDPG emerged as an obvious choice. In building controls, temperature, humidity, and airspeed, which are the predominant control variables, are all continuous. Compared with other commonly used methods, such as Q-Learning and Deep Q-Learning, DDPG can avoid discretizing the control variables, which can improve control precision. Previous research studies have shown that a DDPG thermal control policy can achieve higher thermal comfort and energy efficiency than baseline methods. DDPG offers a number of advantages for developing RL-based controllers, including the ability to handle continuous action spaces, stable learning, off-policy learning, and high-dimensional state spaces. It can also be combined with other techniques to further improve performance. The inclusion of Fig. 6 in our study serves a significant purpose in providing readers with a comprehensive overview of the literature review findings. This flowchart effectively summarizes the diverse configurations and possibilities explored in the dynamic indoor environment and RL domain. By depicting different types of dynamic environments, environmental variables, RL algorithms, and action spaces, the figure highlights the versatility and adaptability of the methodologies explored in the literature. The flowchart's design facilitates a clear understanding of the research landscape, enabling readers to grasp the range of options available in the field. It presents a visual representation of the various combinations that can be employed to achieve dynamic indoor environments and RL control strategies, thus empowering researchers and practitioners to tailor their approach based on specific project requirements and objectives.

Moreover, the flowchart acknowledges the particular combination we chose for our study, namely the utilization of air temperature drifts as the dynamic environment and employing the DDPG algorithm with air temperature set point changes for RL implementation. This specific configuration served as a valuable reference point for our investigation

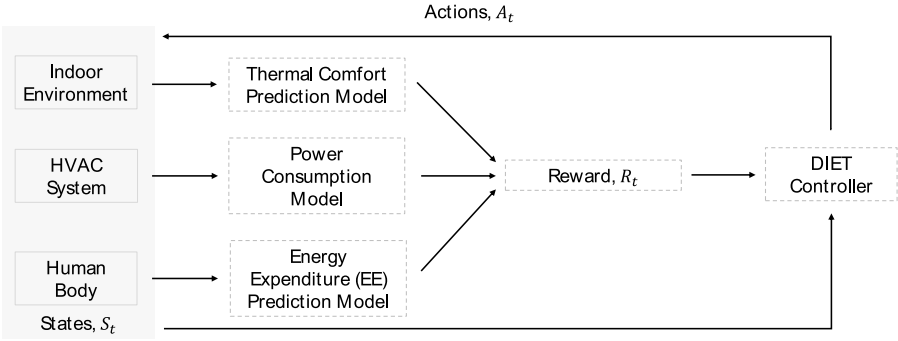



**Fig. 7.** Flowchart showing the modeling approach for the DIET Controller.

and contributed to the meaningful insights generated in our research. In one of the most evident versions, the air temperature set-point can be chosen for the parameter controlled by the DRL controller, given that it was the most common variable to create a dynamic environment. In the case of the reward functions, there could be three parts (Fig. 7), which are (i) the HVAC energy use prediction model, (ii) the human body energy expenditure prediction model, and (iii) PMV model-based thermal acceptability evaluation. It constitutes the central element of the DRL-based Controller named DIET (**D**ynamic **I**ndoor **E**nvironment**T**) Controller. Crucially, the figure emphasizes that other studies can opt to explore alternative combinations of dynamic indoor environments and RL control strategies, allowing for a diverse range of methodologies to be assessed and compared. This acknowledgment high-lights the importance of context-specific considerations and the need to adapt solutions to different building types, climates, and occupant needs.

The DIET Controller model, as defined above, would describe the states, actions, and rewards for the RL framework. The building thermal control would be developed as a cost-minimization problem, and DDPG would be primarily used for training the thermal control policy. One crucial point is that these RL controls are entirely HVAC system-dependent. The controlled parameters change with the HVAC System; for example, it is usually the power levels for an electric heating system or air handling units, and the room/supply air temperature set points are controlled. Another property that affects the control depending on the HVAC system is the response time of different emission systems. For a radiant heating/cooling system, a lower update frequency of the controls is required owing to the longer response time, while the control updates can be more frequent for an electric heater.

Using DRL-based control algorithms to regulate air temperature setpoints in a building can have several implications on ventilation rate and IAQ. These implications arise from the dynamic nature of DRL algorithms, which continuously adjust control actions based on feedback from the environment, and the strong coupling between temperature control and ventilation. DRL algorithms may optimize the air temperature setpoints to achieve specific energy efficiency or comfort goals. However, they might not explicitly consider the impact on ventilation rates, leading to potential deviations from recommended ventilation standards. In some cases, DRL algorithms may prioritize temperature control over ventilation, resulting in reduced ventilation rates if not properly constrained. Conversely, DRL algorithms may increase ventilation rates in response to high indoor temperatures to improve thermal comfort, which could lead to higher energy consumption if not optimized.

The trade-offs between temperature control and IAQ can be challenging to balance. For instance, increasing ventilation rates to improve IAQ can lead to higher energy usage for heating or cooling. DRL algorithms may need to adapt the control strategy to balance temperature control objectives and IAQ requirements based on real-time conditions and occupancy patterns. These algorithms can dynamically adjust temperature set points based on external factors (e.g., outdoor temperature, occupancy, solar radiation), which may influence ventilation rates. Occupant behavior can significantly influence IAQ and ventilation rates. DRL algorithms may incorporate occupancy patterns and preferences to optimize temperature set points and ventilation rates accordingly. Integrating occupant feedback and comfort preferences can lead to a more personalized and efficient control strategy. DRL algorithms excel in adapting to dynamic environments, which is crucial for maintaining IAQ under changing conditions (e.g., occupancy variations and outdoor air quality). The ability of DRL algorithms to learn and adjust in real-time can help optimize IAQ while ensuring energy efficiency.

In general, RL offers several opportunities in dynamic indoor environment control, including:

- ***Improved Energy Efficiency***: RL algorithms can be designed to optimize energy consumption in indoor environments, leading to improved energy efficiency and reduced energy costs.
- ***Enhanced Comfort***: As seen in the previous sections, RL algorithms can also be designed to improve thermal comfort, air quality, and other environmental factors, leading to enhanced comfort for occupants.
- ***Personalization***: RL algorithms can be designed to learn and adapt to the specific preferences and behavior of individual occupants, leading to a more personalized indoor environment. In this study this feature is particularly important given that the DRL-based controller should be able to create an environment that is not only energy saving but also occupant-centric.
- ***Real-time Control***: Real-time control of indoor environments can be obtained from RL-based controls, allowing the system to respond quickly to changes in the environment and maintain optimal conditions.
- ***Improved Environmental Performance***: RL algorithms can be designed to optimize environmental performance, such as reducing greenhouse gas emissions, improving air quality, and reducing energy consumption.
- ***Automated Control***: RL algorithms can provide automated control of indoor environments, reducing the need for manual intervention and freeing up time for other tasks.
- ***Improved Scalability***: RL algorithms can be designed to be scalable, allowing them to be used in large and complex indoor environments, such as commercial buildings and high-rise buildings.

Overall, RL has the potential to provide significant improvements in the control and management of dynamic indoor environments, leading to enhanced comfort, energy efficiency, and environmental performance. By leveraging the strengths of RL, researchers and practitioners can continue to advance the state of the art in dynamic indoor environment control and provide innovative solutions to meet the challenges of this field. However, at the same time RL-based dynamic indoor environment control faces several challenges, some of which are as follows:

- ***Model Uncertainty***: One of the main challenges in RL-based dynamic indoor environment control is the uncertainty in modeling the indoor environment. This includes uncertainty in predicting future indoor temperatures, air quality, and other environmental parameters, which can lead to suboptimal control policies.
- ***Non-linear Systems***: Indoor environments are highly non-linear systems, which can make it difficult to model and predict their behavior. This can lead to significant challenges in designing RL algorithms that can effectively control these systems.
- ***State and Action Space***: The state and action space in RL-based dynamic indoor environment control can be large and complex, making it difficult to design efficient and effective control policies.
- ***Computational Cost***: RL algorithms can be computationally intensive, requiring a lot of computational resources and time to converge to an optimal solution. This can be a significant challenge in real-time applications where the system needs to respond quickly to changes in the environment.
- ***Interaction with Occupants***: The behavior of occupants can have a significant impact on the indoor environment, and RL algorithms need to account for this in their control policies. This can be a complex challenge, as occupant behavior can be highly variable and difficult to predict.
- ***Exploration-Exploitation Trade-off***: RL algorithms need to balance the trade-off between exploration (trying new actions to find better solutions) and exploitation (using the current best solution) to converge to an optimal solution. This can be challenging in real-world applications, where the system needs to respond quickly to changes in the environment.

## 6. Conclusions

This paper reviewed a broad set of studies that focused on (i) various aspects of a dynamic indoor environment and (ii) the use of Reinforcement Learning for HVAC controls. Based on the papers reviewed in the context of the dynamic indoor environment, the permissible limits of temperature step changes and drifts acceptable to human occupants are explored. It also discussed the flexibility of the range of human thermal comfort and adaptation in different parts of the world. Furthermore, an important point was made regarding how creating a dynamic indoor environment can benefit the human body in the long term. The literature review suggested that thermal comfort perception is a function of personal experience in thermal exposure and can be improved by long-term thermal adaptation. As per the current building thermal comfort regulating standards, the temperature in the buildings is held within narrow limits. It leads to the occupants being exposed to conditions within the thermo-neutral zone. Without adequate energy intake compensation (reduction) to offset the lowered metabolic demand at thermoneutrality, an energy imbalance should result in contributing to weight gain. In these conditions, a dynamic indoor environment that increases the body's energy expenditure without physical activity can potentially contribute to solving long-term obesity in office buildings. Although it is not straightforward to create such an environment, the RL-based controls, particularly leveraging DDPG or A3C algorithms, have shown promise in similar tasks and can potentially be the answer to the task at hand.

Moving on to the current situation concerning RL-based building controls, they have attracted much research interest. However, RL controllers are still in the Research & Development stage and have limited acceptance in actual buildings. Of the 64 reviewed studies, only 28% of the RL controllers were implemented and tested in real buildings. The major obstacles limiting the application of RL controllers to real-world building controls include (1) the training process being time-consuming and data-intensive and (2) the security of the controls needing to be addressed. Also, it is not very well known how to implement transfer learning so that pilots trained from a small number of buildings can be generalized. A data-rich, open-source, and interoperable virtual testbed is needed to enable cross-study validation and benchmarking of RL controller performance. Finally, more studies should be focused on bringing the concepts of a dynamic indoor environment and RL controls under the same hood.

## CRediT authorship contribution statement

**Arnab Chatterjee:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis. **Dolaana Khovalyg:** Writing – review & editing, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

## Declaration of competing interest

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

## Data availability

Data will be made available on request.

## References


[1] UNEP, Buildings and Climate Change, Summary for decision-makers, 2019.
[2] P. Nejat, F. Jomehzadeh, M. Taheri, M. Gohari, M. Muhd, A global review of energy consumption, CO2 emissions and policy in the residential sector (with an overview of the top ten CO2 emitting countries), Renew. Sustain. Energy Rev. 43 (2015) 43–62.
[3] IEA, Transition to Sustainable Buildings, 2013.
[4] D. Khovalyg, O. Kazanci, H. Halvorsen, I. Gundlach, W. Bahnfleth, et al., Critical review of standards for indoor thermal environment and air quality, Energy Build. 213 (1098) (2020) 19.
[5] L. Pérez-Lombard, J. Ortiz, J. Coronel, I. Maestre, A review of HVAC systems requirements in building-energy regulations, Energy Build. 43 (2011) 255–268.
[6] A. Heller, M. Uhd, P. Fischer-Nilesn, J. Frederiksen, H. Juhler-Verdoner, E. Hansen, et al., Smart buildings: combining energy efficiency, Flexibility and Comfort (2015).
[7] V. Fabi, M. Sugliano, R. Andersen, S. Corgnati, Validation of occupants' behaviour models for indoor quality parameter and energy consumption prediction, Procedia Eng. 121 (2015) 1805–1811.
[8] M. Levine, D. Ürge Vorsatz, K. Blok, L. Geng, D. Harvey, S. Lang, et al., Residential and commercial buildings, in: Climate Change 2007: Mitigation, Cambridge University Press, Cambridge, 2007, pp. 387–446.
[9] A. Leaman, B. Bordass, Assessing building performance in use 4: the Probe occupant surveys and their implications, Build. Res. Inf. 29 (2) (2001) 129–143.
[10] A. Mishra, M. Loomans, J. Hansen, Thermal comfort of heterogeneous and dynamic indoor conditions - an overview, Build. Environ. 109 (2016) 82–100.
[11] J. Nedergaard, T. Bengtsson, B. Cannon, Three years with adult human brown adipose tissue, Ann. N. Y. Acad. Sci. 1212 (2011) E20–E36.
[12] T. Parkinson, R. de Dear, Thermal pleasure in built environments: physiology of alliesthesia, Build. Res. Inf. 43 (2015) 288–301.
[13] W. van marken Lichtenbelt, B. Kingma, A. van der Lans, L. Schellen, Cold exposure – an approach to increasing energy expenditure in humans, Trends Endocrinol. Metabol. 25 (2014).
[14] L. Jansky, Shivering," *Physiology and Pathophysiology of Temperature Regulation*, 1998.
[15] B. Cannon, J. Nedergaard, Brown adipose tissue: function and physiological significance, Physiol. Rev. 84 (2004) 277–359.
[16] W. van Marken Lichtenbelt, J. Vanhommerig, N. Smulders, J. Drossaerts, G. Kemerink, N. Bouvy, P. Schrauwen, G. Teule, Cold-activated Brown adipose tissue in healthy men, N. Engl. J. Med. 360 (2009) 1500–1508.
[17] K. Virtanen, et al., Functional brown adipose tissue in healthy adults, N. Engl. J. Med. 360 (2009) 1518–1525.
[18] R. Hellwig, D. Teli, A. Boerstra, The potential of the adaptive thermal comfort concept in long-term actively conditioned buildings for improved energy performance and user wellbeing, IOP Conf. Ser. Earth Environ. Sci. 588 (3) (2020).
[19] S. Ferrari, V. Zanotto, Adaptive comfort: analysis and application of the main indices, Build. Environ. 49 (2012) 25–32.
[20] C. Candido, Adaptive comfort: passive design for active occupants, Revista de Engenharia Civil IMED 2 (1) (2015).
[21] M. Schweiker, M. Shukuya, Adaptive comfort from the viewpoint of human body exergy consumption, Build. Environ. 51 (2012) 351–360.
[22] S. Kamaruzzaman, N. Sabrani, The effect of indoor air quality (IAQ) towards occupants ' psychological performance in office buildings, IAQ in Office Building 4 (2011).
[23] L. Fang, G. Clausen, P. Fanger, Impact of temperature and humidity on the perception of indoor air quality, Indoor Air 8 (2) (1998) 80–90.
[24] D. Moschandreas, P. Chu, Occupant perception of indoor air and comfort in four hospitality environments, AIHA J. 63 (1) (2002) 47–54.
[25] M. McNeil, V. Letschert, Future Air Conditioning Energy Consumption in Developing Countries and what Can Be Done about it: the Potential of Efficiency in the Residential Sector, 2008.
[26] Z. Wang, T. Hong, Reinforcement learning for building controls: the opportunities and challenges, Appl. Energy 269 (2020).
[27] T. Lillicrap, J. Hunt, A. Pritzel, N. Heess, et al., Continuous Control with Deep Reinforcement Learning, arXiv, 2015.
[28] M. Hana, R. Maya, X. Zhang, X. Wanga, S. Pan, D. Yan, Y. Jin, et al., A review of reinforcement learning methodologies for controlling occupant comfort in buildings, Sustain. Cities Soc. 51 (2019).
[29] S. Lee, P. Karava, Towards Smart Buildings with Self-Tuned Indoor Thermal Environments- A Critical Review, Energy and Buildings, 2020.
[30] M. Han, J. Zhao, X. Zhang, J. Shen, Y. Li, The Reinforcement Learning Method for Occupant Behavior in Building Control: A Review, Energy and Built Environment, 2020.
[31] M. Castilla, J. Álvarez, M. Berenguel, F. Rodríguez, J. Guzmán, M. Pérez, A comparison of thermal comfort predictive control strategies, Energy Build. 43 (2011) 2737–2746.
[32] A. Heidari, F. Marechal, D. Khovalyg, An occupant-centric control framework for balancing comfort, energy use and hygiene in hot water systems: a model-free reinforcement learning approach, Appl. Energy 312 (2022), 118833.
[33] A. Heidari, F. Marechal, D. Khovalyg, Reinforcement Learning for proactive operation of residential energy systems by learning stochastic occupant behavior and fluctuating solar energy: balancing comfort, hygiene and energy use, Appl. Energy 318 (2022), 119206.
[34] Z. Yu, G. Huang, F. Haghighat, H. Li, G. Zhang, Control strategies for integration of thermal energy storage into buildings: state-of-the-art review, Energy Build. 106 (2015) 203–215.
[35] K. Mason, S. Grijalva, A review of reinforcement learning for autonomous building energy management, axXiv (2019).
[36] I. Dusparic, A. Taylor, A. Marinescu, F. Golpayegani, S. Clarke, Residential demand response: experimental evaluation and comparison of self-organizing techniques, Renew. Sustain. Energy Rev. 80 (2017) 1528–1536.

[37] J. Vázquez-Canteli, Z. Nagy, Reinforcement learning for demand response: a review of algorithms and modeling techniques, Appl. Energy 235 (2019) 1072–1089.
[38] X. Li, J. Wen, Review of building energy modeling for control and operation, Renew. Sustain. Energy Rev. 37 (2014) 517–537.
[39] S. Messaoud, A. Bradai, S. Bukhari, P. Quang, et al., A survey on machine learning in Internet of Things: algorithms, strategies, and applications, Internet of Things 12 (2020).
[40] B. Dong, V. Prakash, F. Feng, Z. O'Neill, A review of smart building sensing system for better indoor environment control, Energy Build. 199 (2019) 29–46.
[41] S. Wang, Z. Ma, Supervisory and optimal control of building HVAC systems: a review, HVAC R Res. 14 (2007) 3–32.
[42] H. Tsutsumi, S. Tanabe, J. Harigaya, Y. Iguchi, G. Nakamura, Effect of humidity on human comfort and productivity after step changes from warm and humid environment, Build. Environ. 42 (2007) 4034–4042.
[43] N. Dahlan, Y. Gital, Thermal sensations and comfort investigations in transient conditions in tropical office, Appl. Ergon. 54 (2016) 169–176.
[44] C. Buonocore, R. Vecchi, V. Scalco, R. Lamberts, Influence of relative air humidity and movement on human thermal perception in classrooms in a hot and humid climate, Build. Environ. 146 (2018) 98–106.
[45] Q. Jin, L. Duanmu, H. Zhang, X. Li, H. Xu, Thermal sensations of the whole body and head under local cooling and heating conditions during step-changes between workstation and ambient environment, Build. Environ. 46 (2011) 2342–2350.
[46] Y. Liu, L. Wang, J. Liu, Y. Di, A study of human skin and surface temperatures in stable and unstable thermal environments, J. Therm. Biol. 38 (7) (2013) 440–448.
[47] Y. Zhang, J. Zhang, H. Chen, X. Du, Q. Meng, Effects of step changes of temperature and humidity on human responses of people in hot-humid area of China, Build. Environ. 80 (2014) 174–183.
[48] Z. Yu, B. Yang, N. Zhu, Effect of thermal transient on human thermal comfort in temporarily occupied space in winter - a case study in Tianjin, Build. Environ. 93 (2015) 27–33.
[49] J. Xiong, Z. Lian, H. Zhang, Effects of exposure to winter temperature step-changes on human subjective perceptions, Build. Environ. 107 (2016) 226–234.
[50] Z. Zhang, Y. Zhang, E. Ding, Acceptable temperature steps for transitional spaces in the hot-humid area of China, Build. Environ. 121 (2017) 190–199.
[51] J. Xiong, X. Zhou, Z. Lian, J. You, Y. Lin, Thermal perception and skin temperature in different transient thermal environments in summer, Energy Build. 128 (2016) 155–163.
[52] K. Nagano, K. Takaki, M. Hirakawa, Y. Tochihara, Effects of ambient temperature steps on thermal comfort requirements, Int. J. Biometeorol. 50 (2005) 33–39.
[53] J. Xiong, Z. Lian, X. Zhou, J. You, Y. Lin, Investigation of gender difference in human response to temperature step changes, Physiol. Behav. 151 (2015) 426–440.
[54] T. Horikoshi, Y. Fukaya, Responses of human skin temperature and thermal sensation to step change of air temperature, J. Therm. Biol. 18 (5–6) (1993) 377–380.
[55] J. Xiong, Z. Lian, H. Zhang, Physiological response to typical temperature step-changes in winter of China, Energy Build. 138 (2017) 687–694.
[56] C. Chen, R.C.S. Hwang, Y. Lu, Effects of temperature steps on human skin physiology and thermal sensation response, Build. Environ. 46 (2011) 2387–2397.
[57] S. Takada, S. Matsumoto, T. Matsushita, Prediction of whole-body thermal sensation in the non-steady state based on skin temperature, Build. Environ. 68 (2013) 123–133.
[58] H. Liu, J. Liao, D. Yang, X. Du, P. Hua, Y. Yang, B. Li, The response of human thermal perception and skin temperature to step-change transient thermal environments, Build. Environ. 73 (2014) 232–238.
[59] W. Ji, B. Cao, M. Luo, Y. Zhu, Influence of short-term thermal experience on thermal comfort evaluations: a climate chamber experiment, Build. Environ. 114 (2017) 246–256.
[60] J. Xiong, Z. Lian, X. Zhou, J. You, Y. Lin, Potential indicators for the effect of temperature steps on human health and thermal comfort, Energy Build. 113 (2016) 87–98.
[61] W. Ji, B. Cao, Y. Geng, Y. Zhu, B. Lin, Study on human skin temperature and thermal evaluation in stepchange conditions: from non-neutrality to neutrality, Energy Build. 156 (2017) 29–39.
[62] J. Xiong, Z. Lian, X. Zhou, Investigation of subjectively assessed health symptoms and human thermal perceptions in transient thermal environments, Procedia Eng. 121 (2015) 212–216.
[63] J. Xiong, Z. Lian, H. Zhang, Investigation of the elderly's response to winter temperature steps in severe cold area of China, Procedia Eng. 205 (2017) 309–313.
[64] R. de Dear, J. Ring, P. Fanger, Thermal sensations resulting from sudden ambient temperature changes, Indoor Air 3 (1993) 181–192.
[65] J. Xiong, Z. Lian, X. Zhou, J. You, H. Lin, Effects of temperature steps on human health and thermal comfort, Build. Environ. 94 (P1) (2015) 144–154.
[66] X. Du, B. Li, H. Liu, D. Yang, W. Yu, J. Liao, et al., The response of human thermal sensation and its prediction to temperature step-change (cool-neutral-cool), PLoS One 9 (8) (2014) 1–10.
[67] Y. Zhu, Q. Ouyang, B. Cao, X. Zhou, J. Yu, Dynamic thermal environment and thermal comfort, Indoor Air 26 (2016) 125–137.
[68] R. Zhao, Investigation of transient thermal environments, Build. Environ. 42 (2007) 3926–3932.
[69] S. Lau, J. Zhang, Y. Tao, A comparative study of thermal comfort in learning spaces using three different ventilation strategies on a tropical university campus, Build. Environ. 148 (2019) 579–599.
[70] S. Zaki, S. Damiati, H. Rijal, A. Hagishima, A. Abd Razak, Adaptive thermal comfort in university classrooms in Malaysia and Japan, Build. Environ. 122 (2017) 294–306.
[71] M. Mustapa, S. Zaki, H. Rijal, A. Hagishima, M. Ali, Thermal comfort and occupant adaptive behaviour in Japanese university buildings with free running and cooling mode offices during summer, Build. Environ. 105 (2016) 332–342.
[72] S. Damiati, S. Zaki, H. Rijal, S. Wonorahardjo, Field study on adaptive thermal comfort in office buildings in Malaysia, Indonesia, Singapore, and Japan during hot and humid season, Build. Environ. 109 (2016) 208–223.
[73] M. te Kulve, L. Schlangen, L. Schellen, A. Frijns, W. van Marken Lichtenbelt, The impact of morning light intensity and environmental temperature on body temperatures and alertness, Physiol. Behav. 175 (2017) 72–81.
[74] M. Chludzinska, A. Bogdan, The effect of temperature and direction of airflow from the personalised ventilation on occupants' thermal sensations in office areas, Build. Environ. 85 (2015) 277–286.
[75] X. Zhou, J. Xiong, Z. Lian, Predication of skin temperature and thermal comfort under two-way transient environments, J. Therm. Biol. 70 (2017) 15–20.
[76] W. van Marken Lichtenbelt, M. Westerterp-Plantenga, P. van Hoydonck, Individual variation in the realtion between body temperature and energy expenditure in response to elevated ambient temperature, Physiol. Behav. 73 (2001) 235–242.
[77] Z. Fang, H. Liu, B. Li, M. Tan, O. Olaide, Experimental investigation on thermal comfort model between localthermal sensation and overall thermal sensation, Energy Build. 158 (2018) 1286–1295.
[78] Y. Shimazaki, A. Yoshida, T. Yamamoto, Thermal responses and perceptions under distinct ambient temperature and wind conditions, J. Therm. Biol. 49–50 (2015) 1–8.
[79] W. Liu, H. Huangfu, J. Xiong, Q. Deng, Feedback effect of human physical and psychological adaption on time period of thermal adaption in naturally ventilated building, Build. Environ. 76 (2014) 1–9.
[80] M. Fadeyi, Initial study on the impact of thermal history on building occupants' thermal assessments in actual air-conditioned office buildings, Build. Environ. 80 (2014) 36–47.
[81] A. Gagge, J. Stolwijk, J. Hardy, Comfort and thermal sensations and associated physiological responses at VArious ambient temperatures, Environ. Res. 1 (1967) 1–20.
[82] D. Chong, N. Zhu, W. Luo, Z. Zhang, Broadening human thermal comfort range based on short-term heat acclimation, Energy 176 (2019) 418–428.
[83] B. Cao, M. Luo, M. Li, Y. Zhu, Too cold or too warm? A winter thermal comfort study in differentclimate zones in China, Energy Build. 133 (2016) 469–477.
[84] P. Wargocki, D. Wyon, The effects of moderately raised classroom temperatures and classroom ventilation rate on the performance of schoolwork by children (RP-1257), HVAC R Res. 13 (2) (2011) 193–220.
[85] J. Porras-Salazar, D. Wyon, B. Piderit-Moreno, S. Contreras-Espinoza, P. Wargocki, Reducing classroom temperature in a tropical climate improved the thermal comfort and the performance of elementary school pupils, Indoor Air 28 (2018) 892–904.
[86] L. Fang, P. Wargocki, T. Witterseh, G. Clausen, P. Fanger, Filed study on the impact of temperature, humidity and ventilation on perceived air quality, Proceedings of Indoor Air 99 (2) (1999) 107–112.
[87] L. Lan, L. Xia, R. Hejjo, D. Wyon, P. Wargocki, Perceived air quality and cognitive performance decrease at moderately raised indoor temperatures even when clothed for comfort, Indoor Air 30 (2020) 841–859.
[88] D. Wang, H. Zhang, E. Arens, C. Huizenga, Observations of upper-extremity skin temperature and corresponding overall-body thermal sensations and comfort, Build. Environ. 42 (2007) 3933–3943.
[89] L. Schellen, M. Loomans, M. De Wit, W. Olesen, et al., Effects of different cooling principles on thermal sensation and physiological responses, Energy Build. 62 (2013) 116–125.
[90] M. Loomans, A. Mishra, M. Derks, J. Kraakman, H. Kort, Occupant response to transitions across indoor thermal environments in two different workspaces, Build. Environ. 144 (2018) 402–411.
[91] Y. Wang, Z. Lian, A study on the thermal comfort under non-uniform thermal environment, Procedia Eng. 205 (2017) 2531–2536.
[92] Q. Deng, R. Wang, Y. Li, Y. Miao, J. Zhao, Human thermal sensation and comfort in a non-uniform environment with personalized heating, Sci. Total Environ. 578 (2017) 242–248.
[93] L. Schellen, M. Loomans, M. de Wit, B. Olesen, W. van Marken Lichtenbelt, The influence of local effects on thermal sensation under non-uniform environmental conditions — gender differences in thermophysiology, thermal comfort and productivity during convective and radiant cooling, Physiol. Behav. 107 (2012) 252–261.
[94] C. Jacquot, L. Schellen, B. Kingma, M. van Baak, W. van Marken Lichtenbelt, Influence of thermophysiology on thermal behavior: the essentials of categorization, Physiol. Behav. 128 (2014) 180–187.
[95] M. Schweiker, A. Wagner, The effect of occupancy on perceived control, neutral temperature, and behavioral patterns, Energy Build. 117 (2016) 246–259.
[96] L. Schellen, W. Van Marken Lichtenbelt, M. Loomans, J. Toftum, M. De Wit, Differences between young adults and elderly in thermal comfort, productivity, and thermal physiology in response to a moderate temperature drift and a steady-state condition, Indoor Air 20 (4) (2010) 273–283.

[97] M. Miura, T. Ikaga, Human response to the indoor environment under fluctuating temperature, Science and Technology for the Built Environment 22 (6) (2016) 820–830.
[98] Y. Zhang, R. Zhao, Relationship between thermal sensation and comfort in non-uniform and dynamic environments, Build. Environ. 44 (7) (2009) 1386–1391.
[99] F. Zhang, R. de Dear, University students' cognitive performance under temperature cycles induced by direct load control events, Indoor Air 27 (1) (2016) 78–93.
[100] H. Yan, L. Yang, W. Zheng, D. Li, Influence of outdoor temperature on the indoor environment and thermal adaptation in Chinese residential buildings during the heating season, Energy Build. 116 (2016) 133–140.
[101] B. Yang, T. Olofsson, F. Wang, W. Lu, Thermal comfort in primary school classrooms: a case study under subarctic climate area of Sweden, Build. Environ. 135 (2018) 237–245.
[102] Z. Wang, A. Li, J. Ren, Y. He, Thermal adaptation and thermal environment in university classrooms and offices in Harbin, Energy Build. 77 (2014) 192–196.
[103] K. Lee, D. Lee, The relationship between indoor and outdoor temperature in two types of residence, Energy Proc. 78 (2015) 2851–2856.
[104] B. Cao, Y. Zhu, Q. Ouyang, X. Zhou, L. Huang, Field study of human thermal comfort and thermal adaptability during the summer and winter in Beijing, Energy Build. 43 (5) (2011) 1051–1056.
[105] M. Derks, A. Mishra, M. Loomans, et al., Understanding thermal comfort perception of nurses in a hospital ward work environment, Build. Environ. 140 (2018) 119–127.
[106] A. Sellers, D. Khovalyg, G. Plasqui, W. van Marken Lichtenbelt, High daily energy expenditure of Tuvan nomadic pastoralists living in an extreme cold environment, Sci. Rep. 12 (2022), 20127.
[107] A. Jindal, Thermal comfort study in naturally ventilated school classrooms in composite climate of India, Build. Environ. 142 (2018) 34–46.
[108] C. Xu, S. Li, X. Zhang, S. Shao, Thermal comfort and thermal adaptive behaviours in traditional dwellings: a case study in Nanjing, China, Build. Environ. 142 (2018) 153–170.
[109] N. Zhang, B. Cao, Z. Wang, Y. Zhu, B. Lin, A comparison of winter indoor thermal environment and thermal comfort between regions in Europe, North America, and Asia, Build. Environ. 117 (2017) 208–217.
[110] E. Diaz Lozano Patiño, M. Vakalis, M. Touchie, E. Tzekova, et al., Thermal comfort in multi-unit social housing buildings, Build. Environ. 144 (2018) 230–237.
[111] B. Yang, T. Olofsson, F. Wang, L. Weizhuo, Thermal comfort in primary school classrooms: a case study under subarctic climate area of Sweden, Build. Environ. 135 (2018) 237–245.
[112] H. Zhang, E. Arens, C. Huizenga, T. Han, Thermal sensation and comfort models for non-uniform and transient environments: Part I: local sensation of individual body parts, Build. Environ. 45 (2) (2010) 380–388.
[113] J. Gallis, A. O'Neil, T. Hayes, M. Hession, C. Gray, The potential of the indoor environment to increase physical activity and reduce sedentary behavior in office workers, Int. J. Behav. Nutr. Phys. Activ. 13 (23) (2016).
[114] T. Yoneshiro, S. Aita, M. Matsushita, T. Kayahara, Recruited brown adipose tissue as an antiobesity agent in humans, J. Clin. Invest. 123 (8) (2013) 3404–3408.
[115] K. Knip, Waaron de septemberhitte, NRC (2016).
[116] A. Mavrogianni, M. Ucci, A. Marmot, J. Wardle, Historic variations in winter indoor domestic temperatures and potential implications for body weight gain, Indoor Built Environ. 22 (2011) 360–375.
[117] W. van marken Lichtenbelt, M. Hanssen, H. Pallubinsky, B. Kingma, L. Schellen, Healthy excursions outside the thermal comfort zone, Build. Res. Inf. 45 (2017) 819–827.
[118] The Impact of the Built Environment on Health, The Harvard T.H. Chan School of Public Health, 2021.
[119] Indoor Air Quality and Health, World Health Organization, 2018.
[120] P. Lee, J. Greenfield, K. Ho, M. Fulham, P. Ainslie, Cold-activated brown adipose tissue is an independent predictor of higher bone mineral density in women, J. Clin. Endocrinol. Metab. 101 (9) (2016) 3520–3526.
[121] T. Yoneshiro, et al., Brown adipose tissue activation by cold stimulation in humans: a study in a Japanese cohort, Obesity 21 (3) (2013) 287–294.
[122] P. Lee, J. Linderman, S. Smith, R. Brychta, J. Wang, C. Idelson, Irisin and FGF21 are cold-induced endocrine activators of brown fat function in humans, Cell Metabol. 19 (2) (2014) 302–309.
[123] P. Lee, S. Smith, J. Linderman, A. Courville, R. Brychta, W. Dieckmann, et al., Temperature-acclimated brown adipose tissue modulates insulin sensitivity in humans, J. Clin. Endocrinol. Metab. 99 (1) (2014) 2013–2388.
[124] T. Yoneshiro, S. Aita, M. Matsushita, Y. Okamatsu-Ogura, T. Kameya, et al., Age-related decrease in cold-activated brown adipose tissue and accumulation of body fat in healthy humans, Obesity 21 (9) (2013) 1779–1785.
[125] T. Yoneshiro, S. Aita, M. Matsushita, Y. Okamatsu-Ogura, T. Kameya, Y. Kawai, M. Miyagawa, et al., Brown adipose tissue, whole-body energy expenditure, and thermogenesis in healthy adult men, J. Clin. Investig. 123 (7) (2013) 1–7.
[126] M. Hanssen, A. van der Lans, B. Brans, et al., Short-term cold acclimation improves insulin sensitivity in patients with type 2 diabetes mellitus, Nat. Med. 21 (8) (2015) 863–865.
[127] U. Lindemann, J. Oksa, D. Skelton, N. Beyer, J. Klenk, J. Zscheile, C. Becker, Effect of cold indoor environment on physical performance of older women living in the community, Age Ageing 43 (2014) 571–575.
[128] K. Okamoto-Mizuno, K. Mizuno, Effects of thermal environment on sleep and circadian rhythm, J. Physiol. Anthropol. 31 (14) (2012).
[129] C. Song, L. Huang, Y. Liu, Y. Dong, X. Zhou, J. Liu, Effects of indoor thermal exposure on human dynamic thermal adaptation process, Build. Environ. 179 (2020), 106990.
[130] O. Héroux, The effect of intermittent indoor cold exposure on white rats, Can. J. Biochem. Physiol. 38 (6) (1960).
[131] T. Wei, S. Ren, Q. Zhu, Deep Reinforcement Learning for Joint Datacenter and HVAC Load Control in Distributed Mixed-Use Buildings, IEEE Trans Sustainable Comput, 2019.
[132] B. Claessens, P. Vrancx, F. Ruelens, Convolutional neural networks for automatic state-time feature extraction in reinforcement learning applied to residential load control, IEEE Trans. Smart Grid 9 (2016) 3259–3269.
[133] E. Mocanu, D. Mocanu, P. Nguyen, A. Liotta, M. Webber, M. Gibescu, et al., On-line building energy optimization using deep reinforcement learning, IEEE Trans. Smart Grid 10 (2018) 3698–3708.
[134] Y. Wang, K. Velswamy, B. Huang, A long-short term memory recurrent neural network based reinforcement learning controller for office heating ventilation and air conditioning systems, Processes 5 (2017) 46–63.
[135] Z. Zhang, A. Chong, Y. Pan, C. Zhang, K. Lam, Whole building energy model for HVAC optimal control: a practical framework based on deep reinforcement learning, Energy Build. 199 (2019) 472–490.
[136] K. Ahn, C. Park, Application of deep Q-networks for model-free optimal control balancing between different HVAC systems, Sci Technol Built Environ 26 (2019) 61–74.
[137] G. Gao, J. Li, Y. Wen, DeepComfort: Energy-Efficient Thermal Comfort Control in Buildings via Reinforcement Learning, IEEE Internet of Things J, 2020.
[138] L. Yu, Y. Sun, Z. Xu, C. Shen, D. Yue, T. Jiang, et al., Multi-agent Deep Reinforcement Learning for HVAC Control in Commercial Buildings, IEEE Trans Smart Grid, 2020.
[139] Z. Zou, X. Yu, S. Ergan, Towards optimal control of air handling units using deep reinforcement learning and recurrent neural network, Build. Environ. 168 (2020) 1–15.
[140] T. Wei, Y. Wang, Q. Zhu, Deep reinforcement learning for building HVAC control, in: DAC, 2017, p. 17. Austin, Texas.
[141] G. Costanzo, S. Iacovella, F. Ruelens, T. Leurs, B. Claessens, Experimental analysis of data-driven control for a building heating system, Sustainable Energy, Grids and Networks 6 (2016) 81–90.
[142] S. Qiu, Z. Li, D. Fan, R. He, et al., Chilled water temperature resetting using model-free reinforcement leanring: engineering application, Energy Build. 255 (2022), 111694.
[143] Z. Deng, Q. Chen, Reinforcement learning of occupant behavior model for cross-building transfer learning to various HVAC control systems, Energy Build. 238 (2021), 110860.
[144] Z. Yu, A. Dexter, Online tuning of a supervisory fuzzy controller for low-energy building system using reinforcement learning, Control Eng. Pract. 18 (2010) 532–539.
[145] P. Fazenda, K. Veeramachaneni, P. Lima, U. O'Reilly, Using reinforcement learning to optimize occupant comfort and energy usage in HVAC systems, J. Ambient Intell. Smart Environ. 6 (2014) 675–690.
[146] R. Jia, M. Jin, K. Sun, T. Hong, C. Spanos, Advanced building control via deep reinforcement learning, Energy Proc. 158 (2019) 6158–6163.
[147] P. Lissa, C. Deane, M. Schukat, F. Seri, M. Keane, E. Barrett, Deep reinforcement learning for home energy management system control, Energy and AI 3 (2021).
[148] C. Lork, W. Li, Y. Qin, Y. Zhou, et al., An uncertainty-aware deep reinforcement learning framework for residential air conditioning energy management, Appl. Energy 276 (2020) 115–126.
[149] T. Wei, S. Ren, Q. Zhu, Deep reinforcement learning for joint datacenter and HVAC load control in distributed mixed-use buildings, IEEE Transactions on Sustainable Computing 6 (3) (2021) 370–385.
[150] G. Costanzo, S. Iacovella, F. Ruelens, T. Leurs, B. Claessens, Experimental Analysis of Data-Driven Control for a Building Heating System, arXiv, 2016.
[151] L. Yang, Z. Nagy, P. Goffin, A. Schlueter, Reinforcement learning for optimal control of low exergy buildings, Appl. Energy 156 (2015) 577–586.
[152] B. Li, L. Xia, A multi-grid reinforcement learning method for energy conservation and comfort of HVAC in buildings, in: IEEE International Conference on Automation Science and Engineering, CASE), Gothenburg, Sweden, 2015.
[153] D. Nikovski, J. Xu, M. Nonaka, A method for computing optimal set-point schedule for HVAC systems, in: REHVA World Congress, CLIMA), 2013.
[154] G. Costanzo, S. Iacovella, F. Ruelens, T. Leurs, B. Claessens, Experimental analysis of data-driven control for a building heating system, Sustainable Energy, Grids and Networks 6 (2016) 81–90.
[155] J. Vázquez-Canteli, J. Kämpf, Z. Nagy, Balancing comfort and energy consumption of a heat pump using batch reinforcement learning with fitted Q-iteration, Energy Proc. 122 (2017) 415–420.
[156] F. Ruelens, B. Claessens, S. Vandael, B. Schutter, R. Babuška, R. Belmans, Residential demand response of thermostatically controlled loads using batch reinforcement learning, IEEE Trans. Smart Grid 8 (5) (2017) 2149–2159.
[157] B. Claessens, P. Vrancx, F. Ruelens, Convolutional neural networks for automatic state-time feature extraction in reinforcement learning applied to residential load control, IEEE Trans. Smart Grid 9 (4) (2018) 3259–3269.
[158] W. Valladares, M. Galindo, J. Gutiérrez, W. Wu, et al., Energy optimization associated with thermal comfort and indoor air control via a deep reinforcement learning algorithm, Build. Environ. 155 (2019) 105–117.
[159] Y. Lei, S. Zhan, E. Ono, Y. Peng, et al., A practical deep reinforcement learning framework for multivariate occupant-centric control in buildings, Appl. Energy 324 (2022), 119742.

[160] Z. Zou, X. Yu, S. Ergan, Towards optimal control of air handling units using deep reinforcement learning and recurrent neural network, Build. Environ. 168 (2020).
[161] R. Lu, S. Hong, M. Yu, Demand response for home energy management using reinforcement learning and artificial neural network, IEEE Trans. Smart Grid 10 (6) (2019) 6629–6639.
[162] K. Ahn, C. Park, Application of deep Q-networks for model-free optimal control balancing between different HVAC systems, Science and Technology for the Built Environment 26 (2020) 61–74.
[163] Y. Yoon, H. Moon, Performance Based Thermal Comfort Control (PTCC) Using Deep Reinforcement Learning for Space Cooling, Energy & Buildings, 2019.
[164] Z. Jinag, M. Risbeck, V. Ramamurti, S. Murugesan, et al., Building HVAC control with reinforcement learning for reduction of energy cost and demand charge, Energy Build. 239 (2021), 110833.
[165] X. Fang, G. Gong, G. Li, L. Chun, et al., Deep reinforcement learning optimal control strategy for temperature setpoint real-time reset in multi-zone building HVAC system, Appl. Therm. Eng. 212 (2022), 118552.
[166] Q. Fu, X. Chen, S. Ma, N. Fang, B. Xing, J. Chen, Optimal control method of HVAC based on multi-agent deep reinforcement learning, Energy Build. 270 (2022), 112284.
[167] X. Deng, Y. Zhang, Y. Zhang, H. Qi, Towards optimal HVAC control in non-stationary building environments combining active change detection and deep reinforcement learning, Build. Environ. 211 (2022), 108680.
[168] X. Ding, W. Du, A. Cerpa, OCTOPUS: deep reinforcement learning for holistic smart building control, in: BuildSys '19, 2019. New York, USA.
[169] H. Liu, B. Balaji, S. Gao, R. Gupta, D. Hong, Safe HVAC control via batch reinforcement learning, in: 2022 ACM/IEEE 13th International Conference on Cyber-Physical Systems (ICCPS), 2022. Milano, Italy.
[170] E. Mocanu, D. Mocanu, P. Nguyen, A. Liotta, M. Webber, et al., On-Line building energy optimization using deep reinforcement learning, IEEE Trans. Smart Grid 10 (4) (2019) 3698–3708.
[171] G. Gao, J. Li, Y. Wen, Energy-Efficient Thermal Comfort Control in Smart Buildings via Deep Reinforcement Learning, arXiv, 2019.
[172] Y. Du, H. Zandi, O. Kotevska, K. Kurte, J. Munk, et al., Intelligent multi-zone residential HVAC control strategy based on deep reinforcement learning, Appl. Energy 281 (2021) 116–131.
[173] G. Gao, J. Li, Y. Wen, DeepComfort: energy-efficient thermal comfort control in buildings via reinforcement learning, IEEE Internet Things J. 7 (9) (2020) 8472–8484.
[174] Y. Li, Y. Wen, D. Tao, K. Guan, Transforming cooling optimization for green data center via deep reinforcement learning, IEEE Trans. Cybern. 50 (5) (2020) 2002–2013.
[175] L. Yu, W. Xie, D. Xie, Y. Zou, D. Zhang, et al., Deep Reinforcement Learning for Smart Home Energy Management, IEEE Internet of Things Journal, 2019.
[176] X. Liu, M. Ren, Z. Yang, G. Yan, et al., A multi-step predictive deep reinforcement learning algorithm for HVAC control systems in smart buildings, Energy 259 (2022), 124857.
[177] Y. Du, F. Li, J. Munk, K. Kurte, O. Kotevska, et al., Multi-task deep reinforcement learning for intelligent multi-zone residential HVAC control, Elec. Power Syst. Res. 2021 (2021), 106959.
[178] Z. Ding, Q. Fu, J. Chen, H. Wu, Y. Lu, F. Hu, Energy-efficient control of thermal comfort in multi-zoneresidential HVAC via reinforcement learning, Connect. Sci. 34 (1) (2022) 2364–2394.
[179] R. Jia, M. Jin, K. Sun, T. Hing, C. Spanos, Advanced building control via deep reinforcement learning, Energy Proc. 158 (2019) 6158–6163.
[180] L. Yu, Y. Sun, Z. Xu, C. Shen, D. Yue, et al., Multi-agent deep reinforcement learning for HVAC control in commercial buildings, IEEE Trans. Smart Grid (2020) 1–14.
[181] A. Hosseinlooa, A. Ryzhov, A. Bischi, H. Ouerdane, Data-driven control of micro-climate in buildings: an event-triggered reinforcement learning approach, Appl. Energy 277 (2020), 115451.
[182] Y. Wang, K. Velswamy, B. Huang, A long-short term memory recurrent neural network based reinforcement learning controller for office heating ventilation and air conditioning systems, Processes 46 (5) (2017).
[183] Z. Zhang, A. Chong, Y. Pan, C. Zhang, S. Lu, et al., A deep reinforcement learning approach to using whole building energy model for HVAC optimal control, in: Building Performance Modeling Conference and SimBuild, 2018. Chicago, USA.
[184] Z. Zhang, K. Lam, Practical implementation and evaluation of deep reinforcement learning control for a radiant heating system, in: BuildSys '18, Shenzen, China, 2018.
[185] Z. Zhang, A. Chong, Y. Pan, C. Zhang, K. Lam, Whole building energy model for HVAC optimal control: a practical framework based on deep reinforcement learning, Energy Build. 199 (2019) 472–490.
[186] T. Moriyama, G. De Magistris, M. Tatsubori, T. Pham, A. Munawar, R. Tachibana, Reinforcement Learning Testbed for Power-Consumption Optimization, arXiv, 2018.
[187] D. Azuatalam, W. Lee, F. de Nijs, A. Liebman, Reinforcement learning for whole-building HVAC control and demand response, Energy and AI 2 (2020) 1000–1020.
[188] M. Mahbod, C. Chng, P. Lee, C. Chui, Energy saving evaluation of an energy efficient data center using a model-free reinforcement learning approach, Appl. Energy 322 (2022), 119392.
[189] L. Yu, Z. Xu, T. Zhang, X. Guan, D. Yue, Energy-efficient personalized thermal comfort control in office buildings based on multi-agent deep reinforcement learning, Build. Environ. 223 (2022), 109458.
[190] M. Biemann, F. Scheller, X. Liu, L. Huang, Experimental evaluation of model-free reinforcement learning algorithms for continuous HVAC control, Appl. Energy 298 (2021), 117164.
[191] B. Chen, Z. Cai, M. Berges, Gnu-RL: a precocial reinforcement learning solution for building HVAC control using a differentiable MPC policy, in: BuildSys '19, 2019. New York, USA.
[192] R. Homod, H. Togun, A. Hussein, F. Al-Mousawi, et al., Dynamics analysis of a novel hybrid deep clustering for unsupervised learning by reinforcement of multi-agent to energy saving in intelligent buildings, Appl. Energy 313 (2022), 118863.
[193] K. Dalamagkidis, D. Kolokotsa, K. Kalaitzakis, G. Stavrakakis, Reinforcement learning for energy conservation and comfort in buildings, Build. Environ. 42 (2007) 2686–2698.
[194] Y. Gao, S. Li, X. Fu, W. Dong, B. Lu, Z. Li, Energy management and demand response with intelligent learning for multi-thermal-zone buildings, Energy 210 (2020).
[195] T. Chaudhuri, Y. Soh, H. Li, L. Xie, A feedforward neural network based indoor-climate control framework for thermal comfort and energy savings in buildings, Appl. Energy 248 (2019) 44–53.
[196] L. Mba, P. Meukam, A. Kemajou, Application of artificial neural network for predicting hourly indoor air temperature and relative humidity in modern building in humid region, Energy Build. 121 (2016) 32–42.
[197] Y. Chen, Z. Tong, Y. Zheng, H. Samuelson, L. Norford, Transfer learning with deep neural networks for model predictive control of HVAC and natural ventilation in smart buildings, J. Clean. Prod. 254 (2020), 119866.
[198] H. Huang, L. Chen, E. Hu, A neural network-based multi-zone modelling approach forpredictive control system design in commercial buildings, Energy Build. 97 (2015) 86–97.
[199] C. Xu, H. Chen, J. Wang, Y. Guo, Y. Yuan, Improving prediction performance for indoor temperature in public buildings based on a novel deep learning method, Build. Environ. 148 (2019) 128–135.
[200] P. Ferreira, A. Ruano, S. Silva, E. Conceicao, Neural networks based predictive control for thermal comfort and energy savings in public buildings, Energy Build. 55 (2012) 238–251.
[201] D. Kolokotsa, K. Niachou, V. Geros, K. Kalaitzakis, et al., Implementation of an integrated indoor environment and energy management system, Energy Build. 37 (2005) 93–99.
[202] X. Chen, Q. Wang, J. Srebric, Occupant feedback based model predictive control for thermal comfort and energy optimization: a chamber experimental evaluation, Appl. Energy 164 (2016) 341–351.
[203] Y. Peng, Z. Nagy, A. Schlüter, Temperature-preference learning with neural networks for occupant-centric building indoor climate controls, Build. Environ. 154 (2019) 296–308.
[204] S. Gupta, K. Kar, S. Mishra, J. Wen, Building temperature control with active occupant feedback, in: Proceedings of the 19th World Congress, 2014. Cape Town, South Africa.
[205] A. Ghahramani, K. Zhang, K. Dutta, et al., Energy savings from temperature setpoints and deadband: quantifying the influence of building and system properties on savings, Appl. Energy 165 (2016) 930–942.
[206] A. Chatterjee, D. Khovalyg, Energy Savings with dynamic heating profiles in office buildings, in: Windsor Conference, 2020.
[207] T. Hoyt, E. Arens, H. Zhang, Extending air temperature setpoints: simulated energy savings and design considerations for new and retrofit buildings, Build. Environ. 88 (2015) 89–96.
[208] A. Ghahramani, K. Dutta, B. Becerik-Gerber, Energy trade off analysis of optimized daily temperature setpoints, J. Build. Eng. 19 (2018) 584–591.