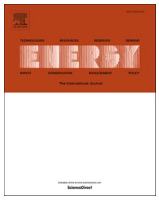
Energy 335 (2025) 137892

Contents lists available at ScienceDirect

Energy

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



Enhancing self-consumption ratio in a smart microgrid by applying a reinforcement learning-based energy management system

Parisa Hajialigol [a,*], Kingsley Nweye [b], Mohammadreza Aghaei [a], Behzad Najafi [c], Amin Moazami [a], Zoltan Nagy [b]

[a] Department of Ocean Operations and Civil Engineering, Norwegian University of Science and Technology (NTNU), Ålesund, Norway
[b] Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin, Austin, TX, USA
[c] Dipartimento di Energia, Politecnico di Milano, Milano, Italy

A R T I C L E I N F O

Keywords:
Reinforcement learning (RL)
Soft actor-critic (SAC)
Energy management system (EMS)
Smart microgrid (SMG)
Distributed energy resources (DER)
Electric vehicle (EV)

A B S T R A C T

This study presents an updated version of the CityLearn Gym environment by integrating a stochastic data-driven vehicle-to-building model. To this end, EVs are modeled as local mobile storage using stochastic behavior derived from a real-world charging dataset, considering uncertainties in EV arrival/departure times, battery capacity, and the arrival state of charge (SoC). Then, the model is integrated within CityLearn to use a reinforcement learning-based energy management system (EMS) to control and optimize a smart microgrid's energy consumption and storage systems. A real-world microgrid in Norway is used to evaluate system performance under three scenarios, including one where solar panel (PV) generation is shared across buildings. The main objective is to provide energy flexibility by enhancing the self-energy consumption of solar generation by finding the optimal control policy for storage systems, which are batteries and EVs. The proposed EMS is designed using the soft actor-critic (SAC) algorithm to coordinate among the different flexible sources by defining the priority resources and direct charging control signals. Three scenarios are investigated and the shared scenario, which in PV generation can be shared between buildings, has had the best performance. The performance of the EMS is evaluated by five key indicators. The results show that the self-consumption ratio of microgrid has been increased up to 23 % and daily peak power has been reduced by up to 20 % compared to RBC as a conventional method. This highlights the impact of storage systems, especially EVs, on the microgrid performance to increase the penetration of solar energy through the energy transition and the potential of RL in advancing intelligent EMS design for future energy systems.

Nomenclature

Abbreviation			
AI	Artificial Intelligence	SG	Smart Grid
BC	Blockchain	SMG	Smart Microgrid
DER	Distributed Energy Resource	SoC	State of Charge
DRL	Deep RL	V2B	Vehicle-to-Building
DSM	Demand-Side Management	V2G	Vehicle-to-Grid
EMS	Energy Management System	*Symbols*	
ESS	Energy Storage System	r	Reward
EV	Electric Vehicle	p^i	Penalty
GIB	Grid Interactive Building	E^i	Electrical demand
IEA	International Energy Agency	n^i_{EV}	Number of EVs in each building
IoT	Internet of Things	$n_{Building}$	Number of buildings in a MG
KPI	Key Performance Indicator	$SoC^i_{Battery}$	SoC of Batteries
MDP	Markov Decision Process	SoC^i_j	SoC of EVs
ML	Machine Learning	*SCR*	self-consumption ratio
MPC	Model Predictive Control	*ZNE*	zero net energy
PV	Photovoltaic	*ADP*	average daily peak
RBC	Rule-Based Control	*R*	ramping
RES	Renewable Energy Source	1-LF	1 - Load Factor
RL	Reinforcement Learning	τ	decay rate
SAC	Soft Actor-Critic	λ	discount factor
SARSA	State-Action-Reward-State-Action	α	learning rate
SEM	Smart Energy Meter	T	temperature

(continued on next column)

(*continued*)

* Corresponding author.
E-mail address: Parisa.hajialigol@ntnu.no (P. Hajialigol).

https://doi.org/10.1016/j.energy.2025.137892
Received 3 May 2024; Received in revised form 12 July 2025; Accepted 2 August 2025
Available online 5 August 2025

1. Introduction

To deal with some global challenges such as climate change, environmental hazards, and resource limitations, it is necessary to pursue sustainable development through the global energy transition. The Energy Transition Outlook 2022 by DNV introduces several scenarios for this energy transition [1], projecting that renewable energy sources (RES) such as solar and wind are expected to contribute 64 % of global electricity generation by 2050. Among these, solar energy, especially solar Photovoltaic (PV) systems, is expected to increase more than 20-fold from 2019 to 2050, making it the dominant source, representing approximately 70 % of the RES mix by 2050. In addition, the International Energy Agency (IEA) World Energy Outlook 2022 also emphasize the need for policies and regulations to facilitate the integration of solar energy into power systems, along with the development of supporting technologies such as energy storage and demand-side management systems that employ advanced technologies to optimize energy consumption and sustainability and reduce costs and emissions [2].

1.1. Motivation

The one that can greatly change energy usage by integrating sustainable resources, enhancing efficiency, and improving supply security is the smart microgrid (SMG). The concept of an SMG refers to an electrical distribution grid that incorporates smart energy meters (SEMs), RESs, energy storage systems (ESSs), and communication networks to supply local consumers [3]. An SMG can operate in parallel with the main grid to fully exploit distributed energy resources (DERs) or islanded to provide a reliability guarantee for local service when there is a failure in the main utility grid [4]. One of the recent popular energy systems is the electric vehicle (EV), which is rapidly growing due to the advantages of high efficiency, low emissions, and energy saving. EVs are expected to account for 80 % of passenger cars by 2050 [1]. This can offer new opportunities for energy transition. However, this increased adoption of EVs also leads to higher electricity consumption within the grid. Therefore, EVs have become the focus of much research and are favored by a lot of countries around the world, especially Norway. Norway is leading the adoption of EVs with an 88 % sales share in 2023 [5].

Since EVs typically remain parked in a charging station for a longer time than the actual charging time, it is possible to shift the EV charging load in time. The use of Vehicle-to-Grid (V2G) or Vehicle-to-Building (V2B) technology enables the discharge of energy from EV batteries to the grid or local buildings. All these smart and integrated systems add complexity and peak power for the grid, although they have a lot of environmental and economic benefits. This highlights the critical importance of a well-designed and effective energy management system (EMS) in achieving goals like energy consumption reduction, balancing supply and demand, increasing RES utilization or flexibility, and minimizing costs [6]. Therefore, further research efforts are needed to focus on improving EMS in real-world Smart Grids (SGs) and SMGs, which can be considered different new energy systems and their integration. This can be done by an Artificial Intelligence (AI) EMS. One of these AI methods is Reinforcement Learning (RL), an area of Machine Learning (ML) where an agent learns an optimal control policy through experience from interaction in an environment [7]. RL control has some advantages that make it a preferable control algorithm for use in complex energy systems. RL outperforms conventional methods by providing model-free, adaptive control that learns optimal energy management strategies directly from data and real-time interaction [8]. Unlike rule-based or optimization-based methods, which rely on predefined rules or require precise system models, RL-EMS dynamically adapts to uncertainties in PV generation, load demand, and EV behavior. It is scalable in multi-agent environments and optimizes long-term performance goals, such as maximizing PV self-consumption and minimizing peak loads. Some studies show how RL techniques can be applied to EMS in various domains, including smart MGs, with the increasing regional expansion of RESs.

1.2. Literature review

Several similar works are reviewed and categorized in three aspects: RL EMS in an SMG, EV model, and control systems, and a short review of different frameworks for implementing RL EMS. A comprehensive review of the advantages of RL algorithms in adapting to dynamic environments and optimizing complex systems was presented [9]. In addition, RL concepts, algorithms, and their practical implementation in power and energy systems were reviewed in another paper [10] highlighting the challenges of dealing with uncertain future system information and the strategies employed by RL algorithms to handle them, such as value function approximation and regular entropy learning. Moreover, it discussed the lack of an in-depth analysis or comparison of the performance of different RL algorithms in specific scenarios, in addition to using RL in real-world energy systems [10]. An RL-based EMS for smart energy buildings in a smart grid environment was discussed [9]. The system integrates various distributed energy sources, including RESs, ESSs, and V2G stations, to optimize energy consumption and reduce operating costs. The key contribution lay in modeling the EMS using Markov Decision Process (MDP) and proposing an RL-based algorithm to minimize energy costs under uncertain future information. Additionally, the challenge of coordinating energy management in a complex building was addressed [11]. It highlighted the need for advanced control strategies to maximize energy flexibility and minimize costs and greenhouse gas emissions. While the paper provided significant contributions, it only focused on a centralized Deep RL (DRL) approach, and future research could investigate decentralized DRL approaches and compare their performance. Moreover, a real-time dynamic optimal EMS solution for MGs based on DRL was developed [12]. This research showed the effectiveness and computational efficiency of the proposed method through a case study. Although it has a lack of exploring coordinated operation mechanisms for multiple MGs.

RL algorithm, specifically the Iterative Q (FQI), was used to optimize the charging of EVs in an office building with rooftop PV [13]. The goal was to maximize the self-consumption of locally generated solar energy by shifting EV charging from morning to afternoon when PV production is higher. However, it did not investigate the impact of different EV charging strategies on grid stability, investigated dynamic pricing mechanisms or demand response strategies, and did not evaluate the integration of battery storage systems to enhance self-consumption and grid interaction capabilities further. An overview of SG and the role of DER in modern power infrastructure was provided [14]. It highlighted the importance of technologies such as AI, Internet of Things (IoT), and Blockchain (BC) in increasing the performance and efficiency of SGs. These technologies facilitate the transition from conventional fossil-fuel-rich grids to DER-based SGs, enabling a two-way flow of power and data between peers in power system grids. However, the paper lacks specific examples or case studies to demonstrate the practical implementation and benefits of these technologies in real-world smart grid scenarios.

Afterward, a new approach using AI and ML algorithms to improve energy management in EVs was introduced in operation [15]. It used real-time datasets and various ML techniques, including classification, regression, RL, and clustering, to identify suitable profile patterns to optimize energy efficiency in electric vehicles. The proposed method included the development of EV models based on specification analysis, description of vehicle characteristics, and formulation of energy efficiency considerations. This study showed the effectiveness of ML algorithms in improving energy efficiency in different stages of EVs. Nine real-world challenges for RL control in grid-interactive buildings (GIBs) were discussed [16]. It emphasized the need for standardization

of environments in building control research and advocated the expression of research within this framework to facilitate fair comparisons between state-of-the-art controllers such as model predictive control (MPC) and RL control.

In addition, a variation on an experience-based approach was introduced to train DRL agents for energy consumption management in buildings with ESSs [17]. Addressing aspects such as computational complexity, scalability, or sensitivity to hyperparameters would increase the completeness of the paper. Further studies can focus on evaluating the robustness of the proposed approach in real-world environments and evaluating its applicability in different building energy management scenarios. Then, a new approach was proposed that coordinates distributed generation and storage systems with the demand side to jointly control energy systems, responding to environmental changes. The performance was investigated under extreme weather conditions, showing significant reductions in energy demand and peak power, increasing self-consumption rate and grid independence, and maintaining suitable indoor thermal comfort conditions [18].

All in all, most existing research in smart energy buildings focused on calculating energy schedules using predicted day-ahead information rather than real-time data and real-world scenarios. In addition, real-world feasibility and scalability evaluation of the proposed approaches in diverse buildings and EV scenarios can be valuable for more investigation. Further research is needed to validate its performance in real-world scenarios and explore potential strategies for different penetration of solar energy generation and ESS capacity. Furthermore, it would be useful to address potential scalability issues or computational requirements when deploying DRL-based EMS in larger or more complex microgrids. Some papers are reviewed in detail and presented in Table 1.

Optimizing EV charging scheduling in the SG considering various uncertain variables remains a challenging problem. As the penetration of EVs increases, it becomes more critical to address these uncertainties. Some research investigated the integration of various energy systems into the grid and here it has reviewed papers with a focus on the integration of V2G and V2B as a new opportunity for MG. For example, Mathankumar et al. optimized energy usage in different stages of EV operation using AI–ML algorithms for EMS [15]. Also, the EV charging control problem was addressed in the presence of PV generation using DRL methods. This study proposed mathematical formulations of environments with discrete, continuous and parametric action spaces along with corresponding DRL algorithms to solve them [24]. A metadata-based EV routing optimization framework aimed at minimizing road energy consumption was also developed [25]. The proposed strategy used the SARSA (State-Action-Reward-State-Action) RL algorithm to learn the optimal EV travel policy. The simulation results showed that the proposed framework can reduce the energy demand by 11.04 % and 5.72 % for some EV trips. In this regard, future research directions could focus on optimizing the hardware implementation of AI algorithms in vehicles, addressing computational challenges, and exploring strategies to reduce data interruptions from external APIs [25]. Another optimal operation method for ESSs in PV storage charging stations was proposed based on intelligent RL [26]. It examined the limitations of existing model-based stochastic optimization methods by considering the complex operational characteristics of ESSs and the uncertainty of PV power generation and EV charging load characteristics. The purpose of the proposed method was to maximize the income of PV charging stations by optimizing charging and discharging strategies of energy storage in real time.

In another study, the performance of demand-side resilient management was assessed in response to extreme weather conditions and future climate scenarios. However, the paper lacks significant effects on self-consumption and load matching index, because the flexible algorithm was not adjusted when the load was supplied by the PV system. Further research was suggested to evaluate the effectiveness of integrating a battery system with the proposed flexible algorithm, especially in scenarios where PV generation is insufficient [27]. Moreover, a novel stochastic optimization approach was proposed that combines scenario-based uncertainty description with piecewise hybrid correction rules for resort decision-making. The proposed EMS was applied to a real-world rural MG case study and a laboratory-scale MG, demonstrating reliability and fuel efficiency compared to state-of-the-art optimization approaches. However, it lacks detailed insight into the specific mathematical formulation and implementation of the stochastic optimization approach. In addition, this paper could provide more information about the computational complexity of the proposed model and its scalability for larger MG systems with a higher number of scenarios and generators [28]. The profitability of PV and battery systems installed on municipal buildings was also investigated [29], considering the influence of various factors such as load profile, system size, and electricity tariffs. It addressed the increasing deployment of hybrid PV and battery systems globally, driven by declining battery costs and changing electricity market dynamics. Results showed that small batteries may be profitable when combined with a large PV plant, especially when no feed-in tariff is granted [29].

Table 1
Literature review of RL EMS in microgrids.

Ref.	Objective Function	RL algorithm	Microgrid components	KPIs	Key contribution
[9]	Minimize the operational costs	Q-learning	Grid, PVs, EV charging, a storage system, building loads	Average daily cost, Daily energy bought from the grid	Sensitivity Analysis of Learning Parameters, using real-world data, using time-of-use (tou) and real-time energy pricing policies.
[12]	Maximize the profit	Q-learning	Grid, PVs, wind and diesel generators, a storage system, demand	Average Energy Generation Profit of a DER (AEP) Fairness Factor (FF) Electricity Purchased from Maingrid (EPM)	Sensitivity Analysis of Learning Parameters, four configurations to assess the performance, KPI comparison
[19]	Minimize the operational costs	Q-learning + Decision Tree	PVs and diesel generators, Batteries, loads	Total error (Err_{total})	A combination of RL and decision trees, a comparison of a decision tree with different methods
[20]	Minimize the costs	Multi-agent Q-learning	Grid, gas network, PVs, micro-CHP, EV battery, loads	Energy Consumption	A scenario-based method with the real data, a comparison an optimization-based study
[21]	Maximizing the overall system efficiency and optimizing the dispatch of local resources	DQN, Double DQN, SARSA, REINFORCE, Actor–critic, A3C, PPO	Grid, PVs, Batteries, thermostatically controlled loads, price-responsive loads	Average daily rewards Training time (min) Total and daily profit	A comparison of seven deep RL algorithms
[22]	Promoting learning efficiency and to ensure benefits equilibrium	Multi-agent deep Q-network (MADQN)	Grid, PVs, wind turbine (WT), micro-turbine (MT), Batteries, loads	?	Two scenarios, Comparison of Agents' Benefits Equilibrium with MARL and SARL, Comparison of Long-term Performance with Different Approaches
[23]	Minimizing the real-time operation cost	Proximal policy optimization (PPO)	Grid, PVs, wind generators, a storage system, load	Operation Time (s) error	A comparison with other algorithms: DDPG, DQN, and SP

The impact of random modeling of occupants' behavior on the energy consumption of office buildings was examined [30]. This study evaluated the uncertainty in the annual electrical energy consumption of an office building by using different stochastic models to model occupant occupancy and actions. The study method can be extended to different building types, although the improvement in modeling complexity requires longer simulation times [30]. Furthermore, a real-time EMS for hybrid EVs was developed by using RL to increase fuel consumption and minimize battery degradation with statistical features of driving conditions, enabling the creation of an adaptive environment model. However, the proposed strategy has limitations, including the need for a more comprehensive battery model to manage thermal safety and the potential challenge of managing large operational spaces. Therefore, future research directions include integrating RL algorithms to address scalability issues and conducting real vehicle experiments to validate the control strategy in practical driving scenarios [31].

A stochastic bottom-up model was introduced to generate residential EV charging load profiles, taking into account the outdoor temperature [32]. This eliminates the need for EV data to assess grid effects and optimal charging strategies. The model used real-world data from residential charging in Norway to provide hourly load profiles for different numbers and types of EVs, assuming immediate charging after connection. However, it lacks specific validation of model accuracy against experimental data. A detailed case study from Norway was also presented [33] focusing on the power flexibility potential of EVs in apartment buildings, considering their loads and PV generation. Overall, this research highlighted the significant potential of residential EV charging in increasing electricity flexibility and provided practical insights for real-world applications. However, it could be further enhanced by examining additional factors such as the integration of ESSs, dynamic pricing mechanisms, and demand response programs. In addition, conducting field experiments or pilot projects to verify the findings in real environments will increase the practical application of the proposed solutions.

There are also some similar papers in order to use RL techniques to optimize EV charging schedules, with considerations for grid conditions and RESs in Table 2. However, there is a recurring research gap in investigating the scalability and real-time applicability of RL-EMS in SGs and EV charging systems, especially concerning multi-agent or deep RL models. There is further research needed to enhance the application of DRL in EV charging control; for example, incorporating additional sources of flexibility such as considering V2G capabilities, extending the method to multiple EV and PV installations, and exploring other RL methods for multi-objective optimization. The literature review shows that investigating the scalability and robustness of the RL algorithm in different EV routing scenarios and real-world deployment scenarios will be valuable for practical implementation.

It is necessary to take a look at the available simulation environment. Different simulation environments, such as CityLearn, MPCPy [38], BOPTEST [39], and Energym [40] are utilized to benchmark building control algorithms. These environments have various energy systems controlled to achieve load satisfaction and energy flexibility. While they rely on robust simulation engines like EnergyPlus and Modelica, which may pose challenges for users due to their dependencies. Moreover, these are designed for specific system-level or building-level environments and do not facilitate district-level control or objectives [8]. CityLearn, conversely, supports multiple buildings, district-level control, and multi-agent control without requiring a co-simulation engine, making it configurable for the benchmarking purposes outlined in this study [41].

1.3. Contribution

As the literature review establishes the knowledge gap, there has been a growing interest in data-driven approaches in the modeling and control of EV charging infrastructure. Consequently, researchers seek to deploy predictive analysis methods to solve EV charging management problems with uncertainty. Therefore, the question arises of how to optimize the relationship between generation capacities, storage options, and variable demands in an MG. Because, in addition to the RESs, EVs, in particular, could play a significant role in grid balancing. Thus, in this work, a stochastic data-driven EV model is developed and integrated with an AI-based EMS within the CityLearn framework. The EMS deploys the Soft Actor-Critic (SAC) algorithm to learn optimal scheduling and charging policies that maximize PV self-consumption in the microgrid and the reward is defined based on the optimal use of storage systems to achieve this objective. Incorporating real-time data about the SMG controlled by an RL-EMS can provide a more practical and realistic approach to tackling these challenges. Therefore, the model is validated by employing a real-world case study which is an SMG consisting of various types of buildings and energy systems. Then, different scenarios are designed and employed in the model to analyze the effect of PV capacity and storage systems in enhancing the self-consumption ratio. Results show that the proposed RL-based EMS significantly improves self-consumption and other KPIs compared to both a baseline (no control) and a conventional rule-based control (RBC) strategy.

Table 2
Literature review of RL EV management.

Ref.	Objective Function	RL algorithm	Microgrid components	KPIs	Key contribution
[15]	Minimizing the energy losses of EVs in each stage	Four different AI-ML combinations with IoT	EVs	EV profile data (Speed, Velocity, Battery Level and Torque)	Using IoT real-time dataset
[34]	Maximize the benefits of EV parking lots	Fuzzy logic controller + DRL methods	PVs, EV charging, and FCEVs	Power flow, mass flow, and daily reward	TD3-based DRL for controlling power flow, mass flow, and retail electricity prices.
[35]	Minimizing costs by optimizing the charging schedules	Multi-agent Q-learning	PVs, EV charging, and Grid	Average reward	Application of MADRL to EV charging scheduling, flexibility and adaptability in dynamic nature of EV charging by a sequential decision-making
[24]	Maximizing PV self-consumption and EV SoC at departure	Double deep Q-networks learning (DDQN), DDPG, and parametrized deep Q-networks learning (P-DQN), RBC, and MPC	PVs, EV charging, building loads, and Grid	Grid power, Training Time (min), SoC	A comparison of different control methods with RL
[36]	Addressing multiple issues related to EV charging scheduling	Deep Q-networks (DQN)	EVs	Reward	A simulator to generate training data and scenarios
[37]	Minimizing costs	Double deep Q-networks learning (DDQN), Partially Observable Markov Game (POMG)	PVs and diesel generators, EVs, loads, and Grid	Reward, Daily routing and scheduling decisions	Hybrid Continuous-Discrete Action Space, Adaptability to Dynamic Conditions, Integration of Transportation and Power Networks

Then, compared to prior studies, the major contributions of this paper are summarized as follows.

- Development of a stochastic data-driven model of two-way EV charging stations.
- Integration of the EV model with an RL-based EMS within the CityLearn environment
- Validation through a real-world Norwegian case study

To achieve these goals, the rest of this paper is organized as follows: First, we explain the SMG architecture, the proposed V2B model, the Citylearn environment, and the evaluation process in section 2. Then, the training hyperparameter settings and the case study are well-defined in section 3 and the results are reported and contextualized in section 4. Finally, the conclusion and recommendations for future work are presented in section 5.

2. Framework and implementation

The problem statement was well-defined in section 1 and it is time to define the SMG architecture and structural framework. Then, the methodology consisting of the proposed EV model, CityLearn environment incorporating RL techniques, and designing a reward function will be explained afterward. At the end of this section, some parameter training and Key Performance Indicators (KPIs) for evaluating the performance of the RL-based energy management algorithm within the CityLearn environment will be introduced. These steps collectively form the research approach to achieve the desired objectives.

2.1. Smart microgrid architecture

This study assesses the EMS of an SMG, focusing on the objective of increasing RES penetration by controlling the stationary batteries and EV batteries as local mobile storage. The architecture of the SMG, as shown in Fig. 1, involves connections to the grid and various DERs, such as a PV system, ESSs, and V2B stations as an efficient storage system in an EMS.

In this research, there are four kinds of datasets in terms of impact on the SMG based on the control actions by the EMS.

- "Electrical demand" represents the electrical energy demand required by any smart buildings for their internal energy consumption. In this research, “Electrical demand” refers to the measured delivered electrical energy for all purposes including electricity for HVAC, lighting, and equipment (plug load). The amount of electrical demand is in kWh at each time.
- "PV generation" signifies the energy generated by the PV system in kWh.
- "V2B demand" refers to the energy demand of the EV charging station in kWh. The V2B station has a bidirectional energy exchange capability with the building. Hence, there are two possible values for V2B, which can be positive (more EVs requiring charging) or negative (more EVs requiring discharging). The EV’s battery State of Charge (SoC) is unknown and subject to control.
- "Battery Energy Storage" acts as an energy buffer for load shifting by exchanging energy at the proper time. The value of SoC is also unknown and subject to control.

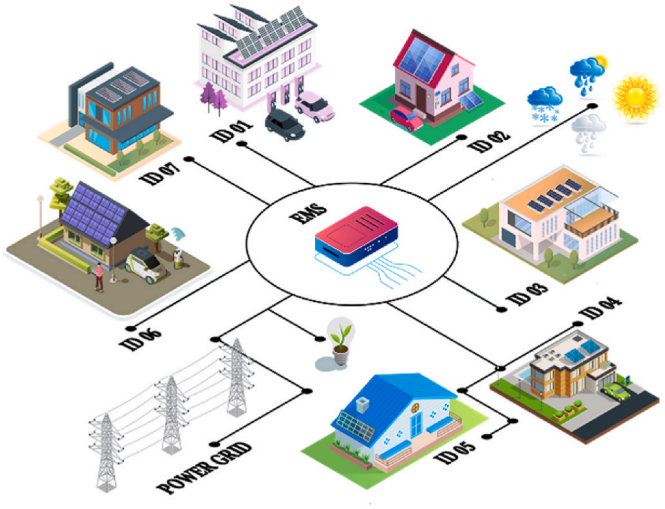


Fig. 1. Structural visualization of an SMG that is interconnected with smart buildings, a grid, a PV, ESSs, and V2B stations.

2.2. The proposed V2B model

In this study, EVs are considered as local mobile storage in the SMG, which contributes to increasing grid flexibility and on-site solar self-consumption. One of the key contributions is the development of a stochastic data-driven hybrid model for EVs and then integrating the model into the CityLearn environment.

To realistically simulate the behavior of EVs while the direct EV usage data is not available for the case of the study, we used a Norwegian national dataset published by Sørensen et al. (2024) [42]. This dataset contains more than 35,000 home charging sessions from 267 users in 12 locations, providing detailed records of EVs connection and disconnection times, battery capacity, charging power level, charged energy, and arrival SoC). Then, a stochastic distribution is applied to generate random behavior of EVs, which is used in several numbers of research [30,43–45]. To simulate the EV charging process dynamically, the SoC, the presence, and the capacity of the EVs are chosen randomly from a range of numbers from the mentioned real dataset. The number of charging stations is based on the real-world case study, but the availability of EVs is based on stochastic scheduling which demonstrates the number of EVs at the station in each time step. Arrival SoC is typically in the range of 0.3–0.5, but to have a wider range it is randomly selected from 0.2 to 0.6. The schedule of each EV is generated based on this approach and given to the CityLearn.

Moreover, the variety of each charging power is also considered based on the most common EV Level 2 residential chargers in Norway. Similarly, EV battery capacities are chosen randomly based on available most common EV models [46] with higher probabilities assigned to battery sizes of 60 kWh and 80 kWh [47]. Each building has a different number of chargers, and each EV has an arrival time domain from 7:00 to 10:00 and a departure time domain from 15:00 to 18:00. These hours are selected based on typical working time in Norway by considering the flexible delay period but with a higher probability of early arrivals and late departures. Although the dataset primarily reflects residential charging behavior, the patterns confirm that these timing assumptions are appropriate. The frequencies of these EV Availability Time and Arrival SoC are illustrated in Fig. 2.

To avoid unrealistic charging behavior and protect battery health, some operational assumptions are taken into account. Since EV battery degradation is strongly influenced by charge and discharge cycles and the price of EV batteries themselves is nearly half of the vehicle price [48], each EV is not allowed to discharge more than twice per day. At the charging stage, the EV can be charged only up to 80 % of the maximum battery capacity. But we should confess that there is a lack of knowledge regarding EV travel patterns and their impact on battery degradation while they are away from the building. To prevent a low partial load efficiency of power electronics, the minimum charging power is set at 10 % of the charging power rating [49]. This is enough in Norway because of having EV charging stations on all roads nearby. The influence of the self-discharge rate is ignored. All EV characteristics and EV charging point specifications are summarized in Table 3.

Fig. 3 represents the structure of the V2B model. At the beginning of

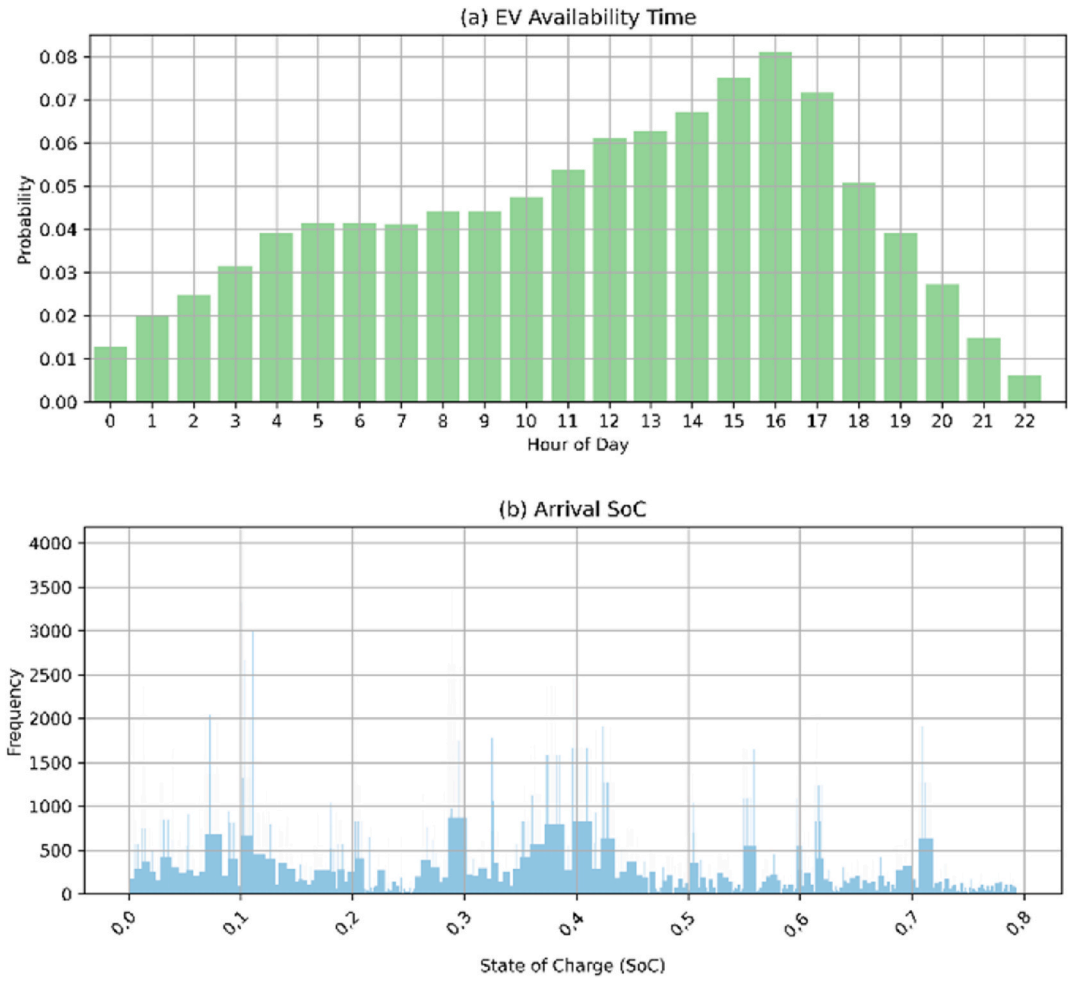


Fig. 2. Frequency of EV availability time and arrival SoC based on Norwegian dataset.

Table 3
EV characteristics.

Variables	Value (Unit)
Battery capacities	20.0, 60.0, 80.0, 100.0 (kWh)
Charging powers	7.4, 11 (kW)
Efficiency	0.9 (–)
Arrival hour	7 to 10
Departure hour	15 to 18
Arrival SoC	0.2 to 0.6
Minimum SoC limit	0.1
Maximum SoC limit	0.8
Discharge count limit	2

each day, PV generations, electrical demand, EES, and EV states are updated. EV states consist of EVs' schedules and the SoC of each EV in the buildings. We defined the time step as an hour to cope with the charging interval term. Electricity from the PV or the grid can be consumed by the buildings, stored in EES, or used to charge the EVs. Also, if the electricity is not enough, firstly EES and secondly EVs can be discharged to help supply the electricity of the building. After receiving the power from the PV, the ESS and V2B model updates the SoCs, considering the capacity. Rewards are given for the position in the next step of the EV. If EVs leave the next step, they must receive 10 % more than the initial SoC, as a guarantee that they will have enough charge to reach another charging station. After each EV receives the rewards, all the EV states are updated.

2.3. Modeling in CityLearn

The proposed EV model is developed in Python and is integrated into CityLearn, a Gym environment for benchmarking advanced control algorithms, including RL in urban energy management [7], to evaluate the model and implement the case study as a microgrid controlled by an RL-based EMS. This integration enables coordinated control of EV charging alongside other energy systems to optimize the overall performance of an SMG. CityLearn is explored in several studies. These studies have shown promising results in improving EMS using RL techniques and highlight the potential for real-world applications [16, 50,51]. As an example, MARLISA [52], a controller combining cooperative multi-agent RL with an iterative sequential action selection algorithm, demonstrated significant reductions in peak load and improved load factor compared to manually optimized rule-based, and independent RL controllers. Another example called MERLIN [41] addressed challenges in RL training, evaluation, deployment, and transfer of control policies for DERs, showing performance improvements and cost reductions through transfer learning. Coordinated energy management using RL showcased the ability to flatten load profiles and optimize energy consumption in a cluster of buildings [11]. All these features make CityLearn suitable for use in this research. It is important to have a communication interface for simulation and hybrid models and it is necessary to define transformation paths for all sectors to analyze the development and changes of neighborhoods over time. Hence, we customize the Citylearn environment by integrating a V2B model which is described in section 2.2. Moreover, Fig. 4 shows how the CityLearn environment is used by compiling the V2B model and real-world case

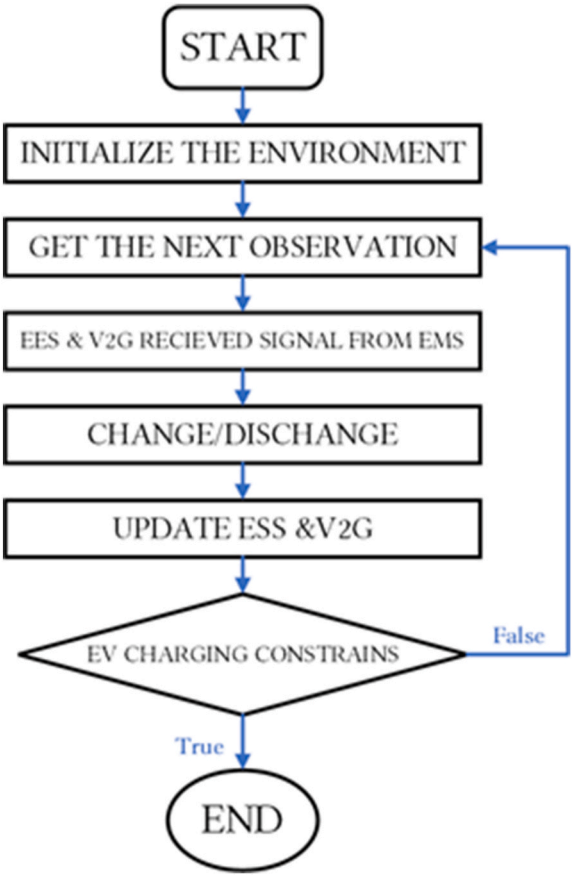


Fig. 3. Structure of V2B model.

study to reach the goal of increasing the penetration of solar energy in the proposed SMG.

The EMS is developed based on an RL algorithm tailored for energy management in SMG. In our study, the SAC (Soft Actor-Critic) algorithm is used as an alternative for RBC and MPC in battery and EV management, where each building's battery and EV charging spots are controlled independently by an agent. SAC is a model-free off-policy RL algorithm that allows for reusing experience and learning from fewer samples. It incorporates an actor-critic architecture, off-policy updates, and entropy maximization to facilitate efficient exploration and stable training. SAC learns from three functions: the actor (policy), the critic (soft Q-function), and the value function (V) [53](See Fig. 5). The steps

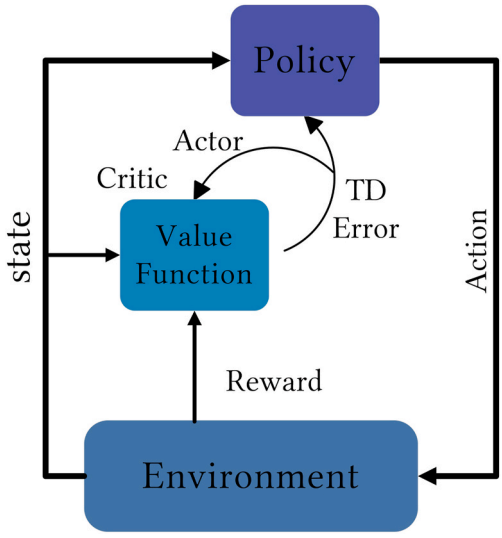


Fig. 5. Soft actor-critic reinforcement learning algorithm.

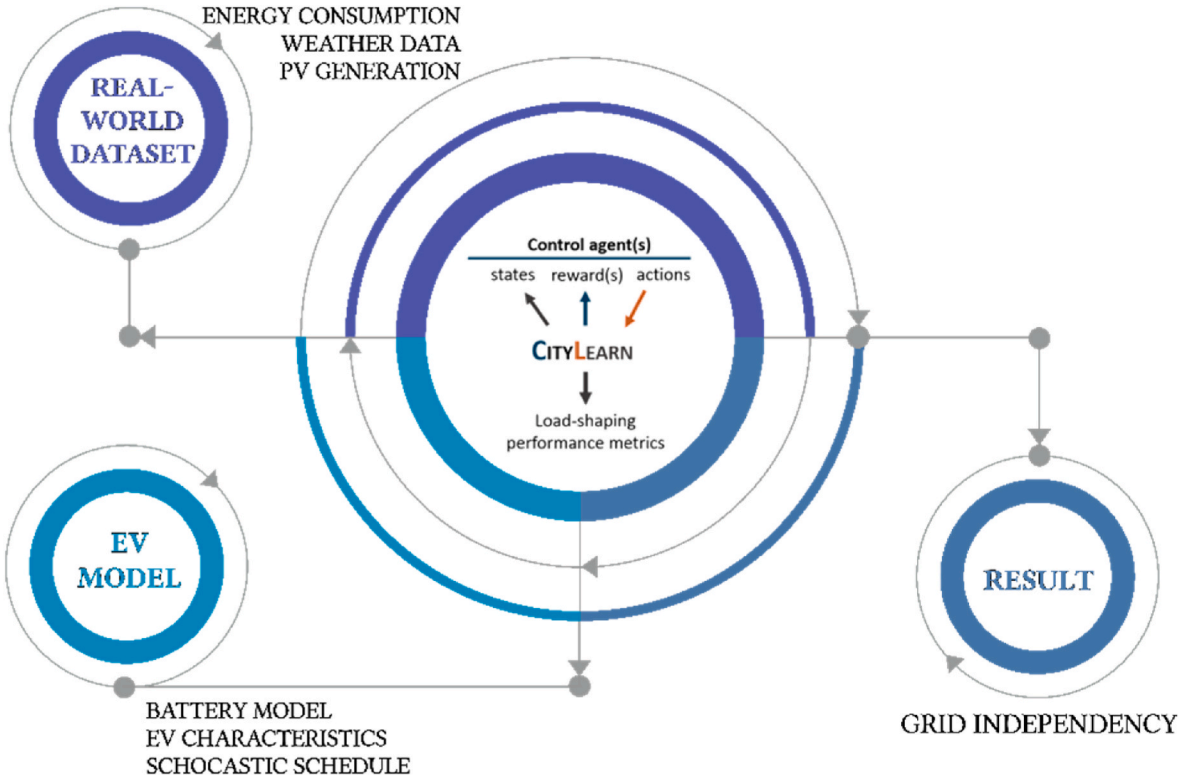


Fig. 4. Conceptual graph of what is added to Citylearn.

of implementation of RL-EMS Algorithm using SAC are summarized in Table 4.

For comparing the performance of the SAC policy, an RBC is used as a reference control policy which is validated in Ref. [44]. RBC is a method of controlling systems by defining a set of predetermined rules or conditions. These rules are typically based on expert knowledge outlining how the system should behave in different situations. Instead of learning from data or optimizing parameters, RBC relies on instructions or guidelines to make decisions and take action. The strategy is such that the agent charges the controlled storage system(s) by 9.1 % of its maximum capacity every hour between 10:00 p.m. and 08:00 a.m., and discharges 8.0 % of its maximum capacity at every other hour. This strategy minimized the error between simulated and measured battery electricity consumption.

2.4. Reward function design

Designing an efficient reward function is one of the most important aspects of using RL in complex energy systems. The reward directly defines the learning objective and shapes the agent's behavior. In this work, the reward function is formulated with a clear energy efficiency objective: to maximize the self-consumption of locally generated solar energy while preserving the health of the battery and EV and minimizing unnecessary energy exchanges with the grid.

As defined in Eq. (1), the total system reward at each time step is the sum of individual building rewards in the SMG of $n_{Building}$ that are designed to maximize self-consumption of solar generation, E. It encourages net-zero energy use by penalizing grid load satisfaction when there is energy in a building's battery or available EVs out of n_{EV} visiting EVs $SoC > 0.0$ as well as penalizing net export when these storage devices are not fully charged $SoC < 1.0$ through the penalty term, p defined in Eq. (2). There is no penalty nor reward given when the battery and EVs are fully charged during net export to the grid. Whereas the penalty is maximized when there is a net import from the grid when the storage systems are charged to capacity. Note that only net import is penalized for a building with no PV generation. In particular, the reward indirectly supports economic efficiency and system robustness by incentivizing storage during surplus periods (e.g., high PV or low electricity prices) and discharge during deficit periods (e.g., high demand or high grid costs), without explicitly encoding time-of-use prices, thus keeping the approach generalizable to other SMG setups. Overall, this reward function captures both physical constraints and strategic energy management goals, ensuring that the SAC agent learns policies that are not only optimal in simulation, but also meaningful and transferable in real-world deployments.

$$r = \sum_{i=0}^{n_{Building}-1} -p^i . \left|E^i\right| \tag{1}$$

$$p^i = 1 + n^i_{EV} + \frac{E^i}{\left|E^i\right|} \cdot \left(SoC^i_{Battery} + \sum_{j=0}^{n^i_{EV}-1} SoC^i_j \right) \tag{2}$$

2.5. Training parameters

The proposed SAC algorithm is trained using the measured data from the case study explained in section 3, and its performance is evaluated based on some KPIs defined in section 2.6. Initially, the raw time series data is preprocessed to remove inconsistencies, adjust time steps, and normalize important features. This ensures that the SAC agent is exposed to consistent and meaningful data inputs, facilitating stable learning dynamics. Next, the cleared time series data is forwarded to the proposed model for training that intelligently learns discriminative features. Finally, the model is tested and evaluated using numerous KPIs.

In this research, the agents are trained for 10 episodes, and each episode has 8760 timesteps. The results from the sensitive analysis of the SAC hyperparameter grid search in Ref. [41] show the values of the decay rate (τ), discount factor (λ), learning rate (α), and temperature (T) that maximize the cumulative reward sum for each building's SAC agent. Therefore, the hyperparameter values used in the SAC agents for this work are: $\tau = 0.05$, $\lambda = 0.90$, $\alpha = 0.005$, and $T = 0.5$. These values maximized the cumulative reward sum across agents controlling each building and their respective storage systems. The relatively high discount factor ensures that long-term energy management goals are prioritized, such as maximizing PV self-consumption over time.

This configuration allows the SAC agent to efficiently coordinate the operation of building batteries and EVs, and dynamically respond to changes in load, PV generation, and storage state in real time.

2.6. Key performance indicators

The performance is evaluated based on five energy-related Key Performance Indicators (KPIs). These KPIs aim to assess both grid interaction and system efficiency and are designed to be minimized. The KPIs include self-consumption ratio (*SCR*), zero net energy (*ZNE*), average daily peak (*ADP*), ramping (*R*), and 1 - Load Factor (1-LF) as formulated in Eqs. (3)–(7). The evaluation is carried out over a total of n time steps in an episode. The KPIs *ADP*, *R*, and (1-LF) are calculated at the SMG level using the aggregated district-level hourly net electricity consumption. The two other KPIs are computed at the building level using the building-level hourly net electricity consumption and then reported as the average of the building-level values at the SMG level. SCR is defined as the sum of the electricity received from the grid divided by total consumption values. As the objective is to maximize self-consumption, this KPI is considered -*SCR*. The definition of each of the KPIs is well described in Ref. [41] as follows:

Table 4
Steps of RL-EMS Algorithm using SAC.

RL-EMS Algorithm using SAC.
1: Initializes the environment parameters such as maximum PV power, maximum battery capacity, and maximum EV capacity.
2: Initializes the Q-table based on the number of PV states, battery states, EV states, and actions.
3: Defines the hyperparameters such as learning rate (alpha), discount factor (gamma), epsilon-greedy parameter (epsilon), and number of episodes.
4: Defines the reward function that calculates the net power based on the PV power level, battery charge, and EV charge.
5: Defines the epsilon-greedy policy that selects an action based on the current state and epsilon value.
6: for the specified number of episodes:
7: Runs the Q-learning algorithm by determining a possible action set.
8: Updates the Q-values based on the current state, action, next state, and reward.
9: Updates the episode variables and checks if the episode is complete.
10: End for
11: Determine the optimal policy based on the Q-values.

$$-SCR = -\frac{\sum_{h=0}^{n-1} E_h^{Building} - E_h^{Solar}}{\sum_{h=0}^{n-1} E_h^{Building}} \quad (3)$$

$$ZNE = \sum_{h=0}^{n-1} E_h^{Building} \quad (4)$$

$$ADP = \frac{\sum_{d=0}^{364}\sum_{h=0}^{23} max\left(E_{24d+h}^{district}, \ldots, E_{24d+23}^{district}\right)}{365} \quad (5)$$

$$R = \sum_{h=0}^{n-1} \left|E_h^{district} - E_{h-1}^{district}\right| \quad (6)$$

$$1 - LF = \left(\sum_{m=0}^{11} 1 - \frac{\left(\sum_{h=0}^{729} E_{730m+h}^{district}\right)\Big/730}{max\left(E_{730m}^{district}, \ldots, E_{730m+729}^{district}\right)}\right)\Big/12 \quad (7)$$

These KPIs provide a multidimensional picture of system performance, capturing aspects of energy autonomy, peak load handling, steady load, and operational efficiency – all of which are crucial for evaluating advanced EMSs in smart microgrids. The KPIs are reported as normalized values with respect to the baseline outcome (Eq. (8)) where the baseline outcome is when buildings are not equipped with any battery storage i.e., no control.

$$\text{KPI} = \frac{KPI_{control}}{KPI_{baseline(no\ storage)}} \quad (8)$$

3. The selected case study

3.1. Real-world scenario

The seven pilot buildings in Norway are investigated as real-world case study in this research, representing northern Europe with cold winters and mild humid summers. There are various building types, including two healthcare centers, two schools, one medical center, and one sports arena (See Table 5) [54]. The training dataset consists of real-world hourly data including the electrical demand of the buildings, solar power generation, as well as weather data. Energy consumption readings, environmental conditions, and solar power generation are measured by Elhub (which is a centralized information exchange platform for the Norwegian electricity market, facilitating efficient data management and communication among market participants [55]). Historical data is used for weather [56]. The data measured on an hourly scale includes the main parameters necessary for energy simulation (e.g., air temperature, relative humidity, as well as direct and scattered solar radiation) and related to the period of 2022. The average daily electricity load before battery flexibility (i.e., solid black line) and solar generation (i.e., dashed blue line) profiles are presented in Fig. 6, where the share of the electrical demands is denoted for all buildings, and then the total electrical demand and the PV generation of the microgrid are illustrated in one chart. While EV usage data was not available for this case study, a data-driven stochastic EV model was developed using national-scale EV charging data from Ref. [42], which is explained in section 2.2. This hybrid data approach enables the simulation to remain in real behavior patterns while preserving the integrity of the actual case study. Thereafter, this case study is used to evaluate the EV model and SAC-EMS. It is a real-world case, and we do not change the system capacities and specifications. It is obvious that the capacity of the current PV systems is not enough for sharing among all buildings. Therefore, we decided to design a PV system for each building in the microgrid and also a battery storage system, explained in section 3.2.

3.2. PV and battery design scenario

To evaluate the impact of solar penetration and storage capacity, we have designed a battery package and PV system for each building based on its electrical demand. Thus, each building has its own PV generation for afterward. The PV is designed based on the maximum available roof area and looking at the maximum electrical demand, so it is kind of the potential to have a PV system for each building. The solar radiation and all other information used in this design are similar to Building ID. 1 specifications. The model is validated by real data from Building ID. 1. Moreover, the battery system is sized based on the maximum daily energy storage requirement in 2022, ensuring sufficient capacity to shift or store surplus solar energy. This scenario represents a decentralized energy-autonomous configuration, allowing each building to operate more independently while supporting self-consumption and demand-side flexibility. The load profiles are illustrated in Fig. 7.

3.3. PV sharing scenario

To assess how solar energy penetration can affect the KPIs and help the microgrid have a better performance, two more scenarios are selected for PV generation. The first case applied to the model is a real-world situation in which only one building has a PV plant for itself (see Fig. 6 and Table 5) and it does not share it with other buildings. In the second scenario, each building has PV and battery. Furthermore, the third case or scenario is the situation where the available PV energy generation can be shared between all buildings in the SMG. With the regulatory changes that the Energy Ministry of Norway has made, buildings can contribute to each other by sharing self-produced renewable electricity on the same property [57]. In this regard, the PV production is shared between the buildings based on their electrical demand in Table 5. Then, it can be seen that PV generation has been more fit to the demand of each building over time. This third scenario represents a future-ready, district-level optimization strategy that showcases the potential of collaborative EMS and resource sharing for achieving higher RES self-consumption and operational efficiency. The

Table 5
An overview of the case study (Buildings).

Building ID. No.	Building type	floor area (m^2)	Year built	Annual Electrical Demand (kWh/m^2)	Annual PV Generation (kWh/m^2)	EV Charging Numbers	Operation Time (Hour)
ID. 1	Health care center	7039	2015	88	145	5	24
ID. 2	Sports arena	2610	2015	40	N/A	3	8
ID. 3	Elementary school	4477	1984	95	N/A	N/A	8
ID. 4	Health care center	5980	2012	111	N/A	5	24
ID. 5	Medical center	2700	1979	170	N/A	N/A	8
ID. 6	High school	9700	2015	59	N/A	7	8
ID. 7	Elementary school	2490	1959 (renovated in 2002)	115	N/A	N/A	8

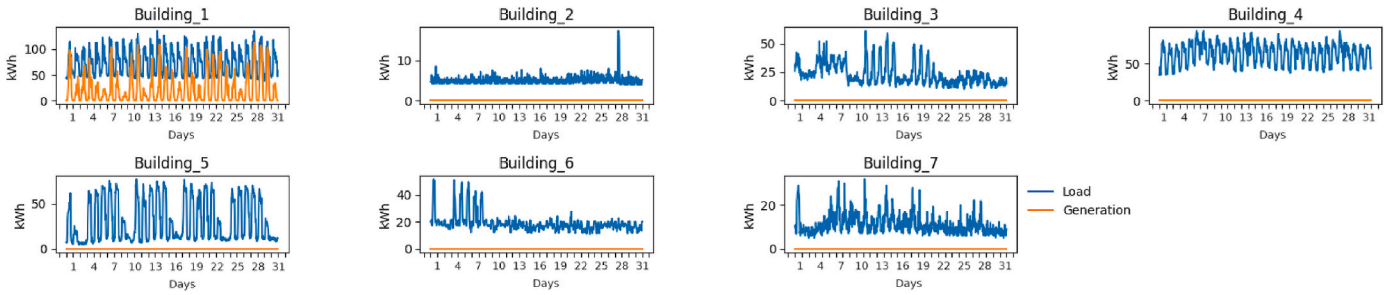


Fig. 6. Load Profiles of Buildings and the Microgrid (Real-World scenario).

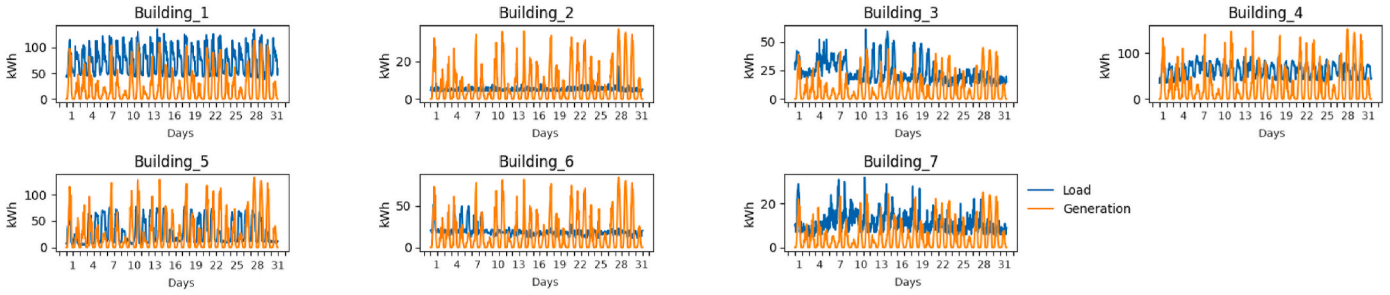


Fig. 7. Load Profiles of Buildings and the Microgrid (Designed _PV scenario).

load profiles of the third scenario are illustrated in Fig. 8.

4. Results and discussions

The SAC-RL agent is trained for 10 episodes, where each episode represents a full year of operation (8760 time steps). The model performance is validated using five key energy management KPIs: *SCR*, *ADP*, *ZNE*, *R*, and 1-LF as defined in section 2.6. These metrics assess the system's ability to optimize the use of RES, reduce peak loads, and maintain stable grid interaction. SAC is compared to both a baseline scenario without storage control and a validated RBC policy used in previous CityLearn studies [41].The plots in Fig. 9 illustrate the daily net-electricity consumption profiles of the SMG with storage and PV (i.e., blue line) compared to the status with no active storage (i.e., red line). No active storage means that there are no battery systems and the EVs are not considered as a storage system; they are only consumers.

It shows that SAC RL EMS can flatten load curves, reduce peaks, and shift energy consumption to periods of higher solar energy production or availability of EVs. It is obvious that a higher capacity of PV and the number of |EVs would also be effective in this progress. For the buildings with 24-h operation, the performance is much better (e.g., Buildings ID. 1 and 4), because there are some nighttime EVs for exchanging energy during the period with no production. Also, buildings with more EV charging stations or larger storage capacities are reducing energy consumption even more. The SAC policy slightly improves load shapes in most of the buildings. The SAC profiles have lower peaks compared to the baseline, exhibiting smoother profiles with no rapid ramping. In contrast, some buildings (e.g., Building ID. 3, 5, and 7), show similar profiles to the baseline, not only in this period but also in the KPIs for these buildings, which are 0.99. This indicates that the controller for electrical storage has little effect on energy flexibility in those buildings because they have limited participation in the V2B model. It highlights that storage flexibility, not just a control strategy, is key to achieving optimal outcomes. Interestingly, in some buildings, EV could decrease the peak, but in some cases, it causes a minor jump down in load which makes *R* worse or not better. In this situation, if the price were considered, it would be a profit for the system.

Moreover, Fig. 10 displays the results of the evaluation after training the SAC policy for 10 episodes, compared to the baseline model (i.e., no storage) and the RBC policy in all three scenarios. In Fig. 10 (i), as a real-world scenario, the SAC policy outperforms both the baseline and RBC policies in minimizing all KPIs. However, it is performing slightly better in *R*, Ramping. The figure indicates that the proposed EMS has the potential to enhance KPIs in comparison to both the baseline and RBC, even with minor improvements due to an under-designed PV system. In the second scenario (ii), when the PV and battery are designed realistically per building, it can be seen that *SCR*, *ADP*, and *ZNE* are improving significantly, showing better load-shifting and energy self-sufficiency. However, *R* increases, caused by sharper transitions from EV discharging. The only reason can be for having less smoothness of the load

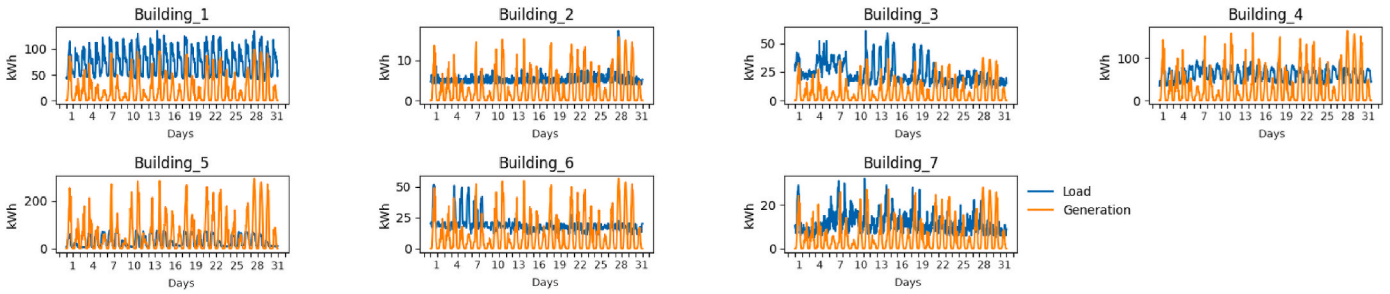


Fig. 8. Load Profiles of Buildings and the Microgrid (Shared_PV scenario).

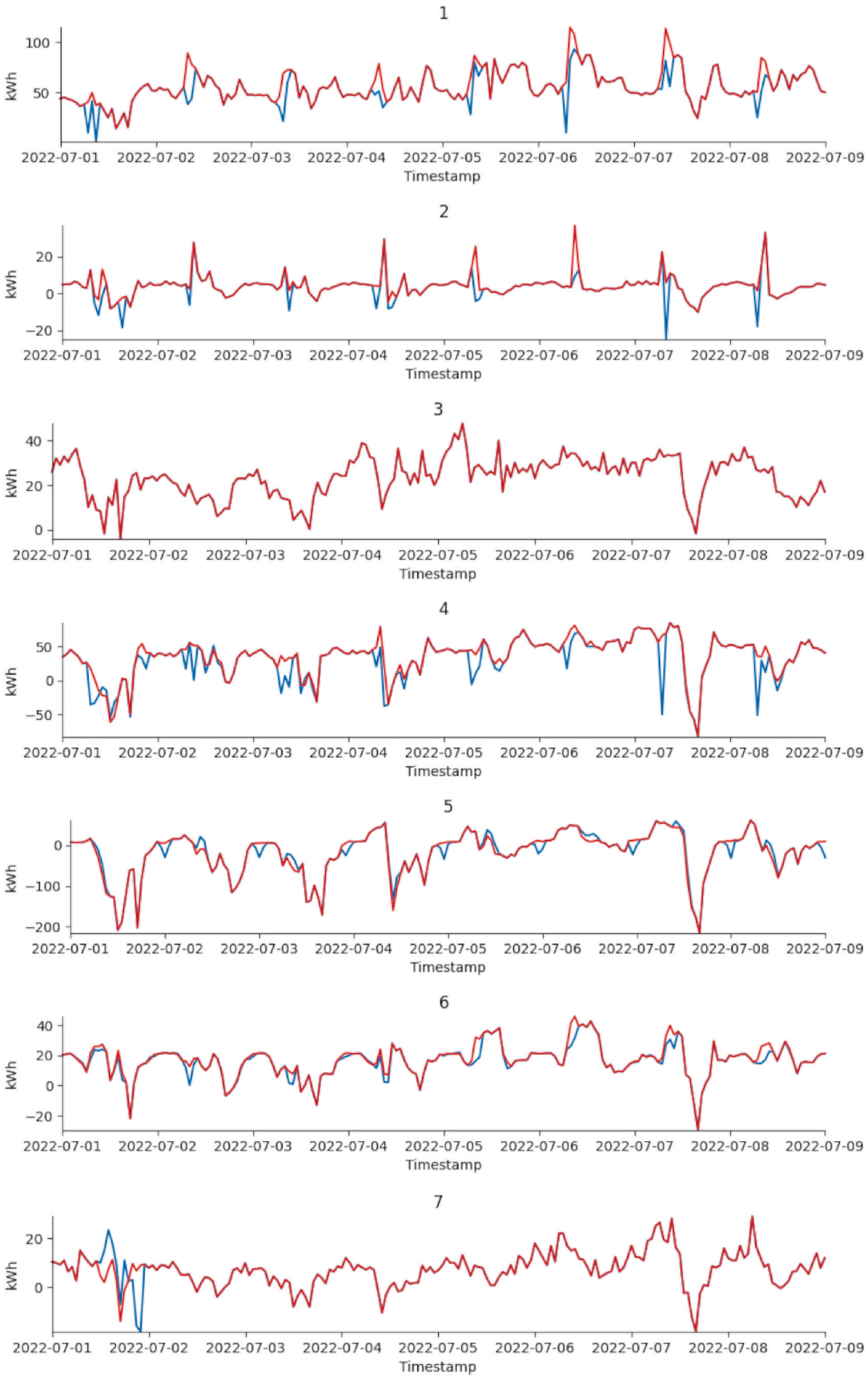


Fig. 9. A comparison of microgrid energy consumption for with storage and PV (blue line) and without any storage (red line).

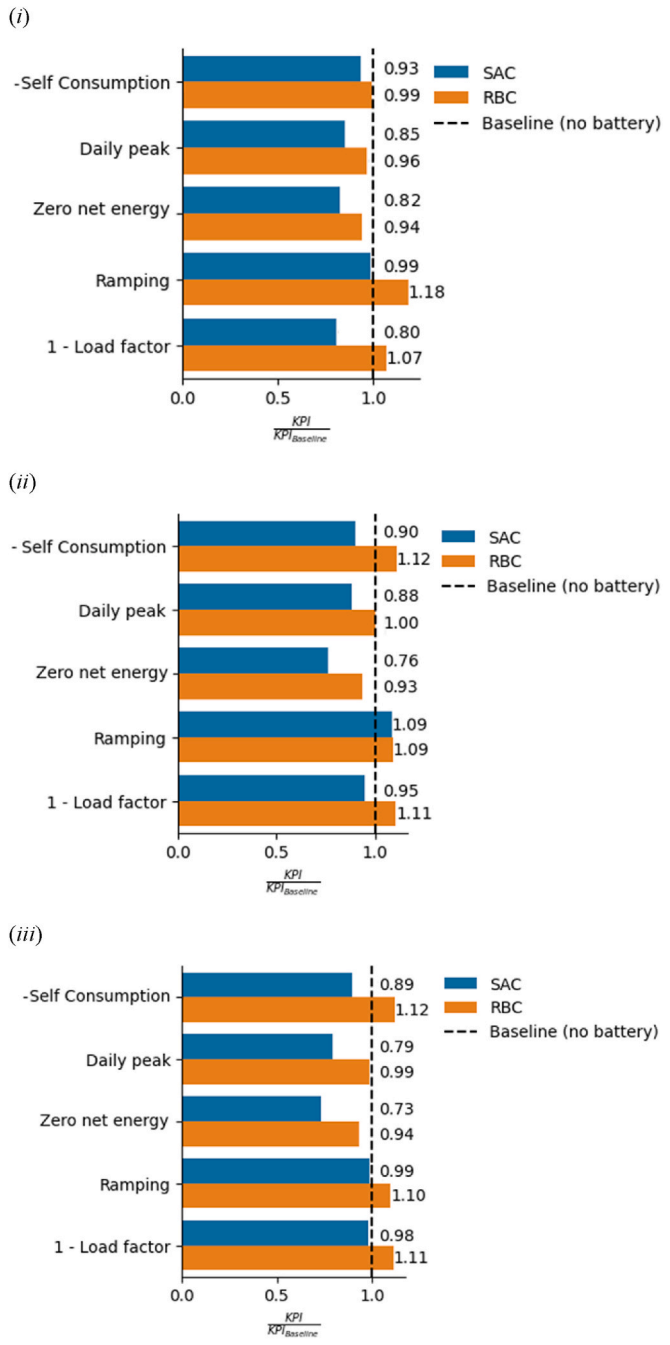


Fig. 10. District-Level KPIs (*i*: Real World scenario), (*ii*: Designed PV scenario), and (*iii*: Shared PV scenario).

profile, where high *R* means an abrupt change in grid load. Here, what makes some abrupt is a jump down because of EV discharging. However, the best KPIs belong to the third scenario (iii), in which every building shares its potential PV capacity to supply a high percentage of the district's consumption. In this case, all KPIs improve, especially *ADP* and *ZNE* are significantly decreased, due to the EMS leveraging district-wide flexibility. This highlights that solar PV can reduce peak electricity consumption and grid load even if self-generation is unavailable. Less 1-LF means the efficiency of electricity consumption is higher than (i) and (ii). The 1-LF improvement suggests better load factor, indicating more stable, efficient grid usage.

Fig. 10 analysis highlights how the SAC-based EMS effectively improves grid interaction and energy efficiency. It is worth noting that the scenario (iii) achieves the best performance in *SCR* (i.e., 0.89), *ADP* (i.e., 0.79), and *ZNE* (i.e., 0.73). These KPIs show approximately 20–23 % improvements compared to the RBC. This reflects the significant benefit of buildings with PV generation, which increases the overall utilization of RESs and reduces grid dependency. On the other hand, scenario (i) shows the most significant improvements in *R* with 19 % reduction and 27 % in *1–LF*, indicating smoother load profiles and better load consistency even under limited PV availability. These results confirm that while distributed PV capacity improves renewable integration, the SAC itself is very effective in handling load flexibility and reducing peak loads. Across all KPIs and scenarios, the SAC consistently outperforms the RBC, demonstrating its suitability for intelligent and dynamic control of smart microgrids.

The centralized SAC-RL agent demonstrates promising potential in finding efficient energy management strategies to improve the KPIs of a real-world application. After just 10 training episodes on the dataset for a specific MG, the agent achieves an improved reward function compared to the manually predefined RBC and baseline. These results consider the agent's adaptability over various buildings and climates and the limited information required for its state space. The RL agent shows potential for more significant improvements in MG-DSM when fine-tuned and trained on larger datasets, making it a valuable approach for energy management in grid-interactive communities and smart MG.

Moreover, this confirms findings from recent literature that RL methods, especially SAC, are effective in complex, dynamic environments such as smart microgrids. While RBC relies on static rules, SAC learns adaptive policies through interaction, allowing it to handle variability in demand, PV generation, and EV behavior. These improvements are particularly notable in the PV sharing scenario, where SAC achieves up to 25 % better load factor performance and 20 % higher self-consumption rate than RBC.

5. Conclusion and recommendation

This study has proposed an energy management system (EMS) based on reinforcement learning (RL) for smart microgrids (SMGs), focusing on coordinated control of stationary batteries and bidirectional charging of electric vehicles (EVs). Among all RL strategies, the Soft Actor-Critic (SAC) algorithm is designed as the controller due to its robust performance in high-dimensional, continuous action spaces typical of energy systems. The reward function is designed to maximize the self-consumption of locally generated solar energy while avoiding degradation of the battery and EV by minimizing unnecessary energy exchanges with the grid. Another key contribution of this study is the development of a stochastic data-driven EV model derived from over 35,000 real charging sessions from Norwegian users. The model generates probabilistic EV arrival/departure times, state of charge (SoC), and charging behavior in accordance with observed national patterns. This EV model has been integrated into the CityLearn environment. The SAC-based EMS is trained using the measured dataset that consists of real-world hourly data from a Norwegian case study, including the electrical demand of the buildings, solar power generation, as well as weather data.

Then, it is evaluated across three different scenarios: (i) Real World scenario, (ii) Designed PV scenario, and (iii) Shared PV scenario. Simulation results show that the SAC-based EMS system significantly improves system performance across all key performance indicators (KPIs), namely self-consumption ratio (*SCR*), zero net energy (*ZNE*), average daily peak (*ADP*), ramping (*R*) and 1–load factor (*1–LF*). Notably, SAC achieved its best performance in Scenario (iii), where PV energy is shared across all buildings, achieving a *ZNE* of 0.73, which reflects improvements of up to 21 % over RBC and 27 % over the baseline. In Scenario (i) (i.e., the real-world setting with only one PV system), SAC reduced Ramping by 19 % and 1–LF by 27 % compared to RBC, highlighting its effectiveness even under limited generation capacity. These results point out the importance of combining advanced control strategies with flexible resources such as EVs and storage

systems to improve grid interaction, smooth load profiles, and increase renewable self-consumption.

Although the study focused on the development of EMS algorithms rather than system design optimization, the results indicate that better-designed PV and battery configurations can open up even greater flexibility and sustainability. Future work will investigate optimal sizing strategies, multi-agent coordination, considering other storage systems, and dynamic pricing schemes to further advance the effective and sustainable operation of SMGs.

CRediT authorship contribution statement

Parisa Hajialigol: Writing – review & editing, Writing – original draft, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Kingsley Nweye:** Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Conceptualization. **Mohammadreza Aghaei:** Writing – review & editing, Supervision, Resources, Project administration, Conceptualization. **Behzad Najafi:** Supervision, Methodology, Formal analysis. **Amin Moazami:** Supervision, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Zoltan Nagy:** Supervision, Software, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the writingprocess

During the preparation of this work the authors used ChatGPT (OpenAI) to improve language and readability. After using this tool/service, the authors reviewed and edited the content as needed and took full responsibility for the content of the publication.

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.

Acknowledgments

This paper is supported by the European Union's Horizon 2020 research and innovation program under the grant agreement for the COLLECTiEF (Collective Intelligence for Energy Flexibility) project (grant agreement ID: 101033683).

Data availability

The authors do not have permission to share data.

References

[1] DNV. The energy transition outlook 2022. DNV; 2022 [Online]. Available: Available: https://www.dnv.com/energy-transition-outlook/.
[2] IEA. World energy outlook 2022 [Online]. Available: https://www.iea.org/reports/world-energy-outlook-2022; 2022.
[3] Lasseter RH. Microgrids. In: 2002 IEEE power engineering society winter meeting. Conference proceedings (Cat. No. 02CH37309). IEEE; 2002. p. 305–8.
[4] Kaur A, Kaushal J, Basak P. A review on microgrid central controller. Renew Sustain Energy Rev 2016;55:338–45.
[5] IEA. World energy outlook 2023. International Energy Agency; 2022 [Online]. Available: https://www.iea.org/reports/world-energy-outlook-2023.
[6] Nagy Z, et al. Ten questions concerning reinforcement learning for building energy management. Build Environ 2023:110435.
[7] Vázquez-Canteli JR, Dey S, Henze G, Nagy Z. CityLearn: standardizing research in multi-agent reinforcement learning for demand response and urban energy management. ArXiv Prepr ArXiv201210504 2020.
[8] Hajialigol P, Moazami A, Aghaei M. Comparative analysis of simulation tools for developing, testing, and benchmarking advanced control algorithms in building energy management systems. Front Energy Effic 2025;3:1546824.
[9] Kim S, Lim H. Reinforcement learning based energy management algorithm for smart energy buildings. Energies 2018;11(8):2010.
[10] Cao D, et al. Reinforcement learning and its applications in modern power and energy systems: a review. J Mod Power Syst Clean Energy 2020;8(6):1029–42.
[11] Pinto G, Piscitelli MS, Vázquez-Canteli JR, Nagy Z, Capozzoli A. Coordinated energy management for a cluster of buildings through deep reinforcement learning. Energy 2021;229. https://doi.org/10.1016/j.energy.2021.120725.
[12] Foruzan E, Soh L-K, Asgarpoor S. Reinforcement learning approach for optimal distributed energy management in a microgrid. IEEE Trans Power Syst 2018;33(5):5749–58.
[13] Mbuwir BV, Massip E, Thoelen K, Deconinck G. Applying reinforcement learning to maximise photovoltaic self-consumption for electric vehicle charging. In: CIRED 2020 Berlin workshop (CIRED 2020). IET; 2020. p. 285–8.
[14] Kumar NM, et al. Distributed energy resources and the application of AI, IoT, and blockchain in smart grids. Energies 2020;13(21):5739.
[15] Mathankumar M, Gunapriya B, Guru RR, Singaravelan A, Sanjeevikumar P. AI and ML powered IoT applications for energy management in electric vehicles. Wirel Pers Commun 2022;126(2):1223–39.
[16] Nweye K, Liu B, Stone P, Nagy Z. Real-world challenges for multi-agent reinforcement learning in grid-interactive buildings. Energy AI 2022;10:100202.
[17] Hussain A, Musilek P. Energy management of buildings with energy storage and solar photovoltaic: a diversity in experience approach for deep reinforcement learning agents. Energy AI 2024;15:100313.
[18] Hosseini M, Erba S, Hajialigol P, Aghaei M, Moazami A, Nik VM. Enhancing climate resilience in buildings using Collective Intelligence: a pilot study on a Norwegian elderly care center. Energy Build 2024;308:114030. https://doi.org/10.1016/j.enbuild.2024.114030.
[19] Kofinas P, Dounis AI, Vouros GA. Fuzzy Q-Learning for multi-agent decentralized energy management in microgrids. Appl Energy 2018;219:53–67. https://doi.org/10.1016/j.apenergy.2018.03.017.
[20] Ahrarinouri M, Rastegar M, Seifi AR. Multiagent reinforcement learning for energy management in residential buildings. IEEE Trans Ind Inf 2020;17(1):659–66.
[21] Nakabi TA, Toivanen P. Deep reinforcement learning for energy management in a microgrid with flexible demand. Sustain Energy Grids Netw 2021;25:100413.
[22] Fang X, Zhao Q, Wang J, Han Y, Li Y. Multi-agent deep reinforcement learning for distributed energy management and strategy optimization of microgrid market. Sustain Cities Soc 2021;74:103163.
[23] Guo C, Wang X, Zheng Y, Zhang F. Real-time optimal energy management of microgrid with uncertainties based on deep reinforcement learning. Energy 2022;238:121873.
[24] Dorokhova M, Martinson Y, Ballif C, Wyrsch N. Deep reinforcement learning control of electric vehicle charging in the presence of photovoltaic generation. Appl Energy 2021;301:117504.
[25] Aljohani TM, Mohammed O. A real-time energy consumption minimization framework for electric vehicles routing optimization based on SARSA reinforcement learning. Vehicles 2022;4(4):1176–94.
[26] Zhang J, et al. Optimal operation of energy storage system in photovoltaic-storage charging station based on intelligent reinforcement learning. Energy Build 2023;299:113570. https://doi.org/10.1016/j.enbuild.2023.113570.
[27] Hosseini S, Hajialigol P, Aghaei M, Erba S, Nik V, Moazami A. Improving climate resilience and thermal comfort in a complex building through enhanced flexibility of the energy system. In: 2022 international conference on smart energy systems and technologies (SEST); 2022. p. 1–6. https://doi.org/10.1109/SEST53650.2022.9898453.
[28] Polimeni S, Moretti L, Martelli E, Leva S, Manzolini G. A novel stochastic model for flexible unit commitment of off-grid microgrids. Appl Energy 2023;331:120228.
[29] Hirschburger R, Weidlich A. Profitability of photovoltaic and battery systems on municipal buildings. Renew Energy 2020;153:1163–73.
[30] Carlucci S, Causone F, Biandrate S, Ferrando M, Moazami A, Erba S. On the impact of stochastic modeling of occupant behavior on the energy use of office buildings. Energy Build 2021;246:111049.
[31] Han L, Yang K, Ma T, Yang N, Liu H, Guo L. Battery life constrained real-time energy management strategy for hybrid electric vehicles based on reinforcement learning. Energy 2022;259:124986.
[32] Sørensen ÅL, Westad MC, Delgado BM, Lindberg KB. Stochastic load profile generator for residential EV charging. In: E3S web of conferences. EDP Sciences; 2022, 03005.
[33] Sørensen ÅL, Morsund BB, Andresen I, Sartori I, Lindberg KB. Energy profiles and electricity flexibility potential in apartment buildings with electric vehicles–A Norwegian case study. Energy Build 2024;305:113878.
[34] Guo G, Gong Y. Energy management of intelligent solar parking lot with EV charging and FCEV refueling based on deep reinforcement learning. Int J Electr Power Energy Syst 2022;140:108061.
[35] Park K, Moon I. Multi-agent deep reinforcement learning approach for EV charging scheduling in a smart grid. Appl Energy 2022;328:120111.
[36] Viziteu A, et al. Smart scheduling of electric vehicles based on reinforcement learning. Sensors 2022;22(10):3718.
[37] Wang Y, Qiu D, Strbac G. Multi-agent deep reinforcement learning for resilience-driven routing and scheduling of mobile energy storage systems. Appl Energy 2022;310:118575.
[38] Blum D, Wetter M. MPCPy: an open-source software platform for model predictive control in buildings. Berkeley, CA (United States): Lawrence Berkeley National Lab. (LBNL); 2019.
[39] Blum D, et al. Building optimization testing framework (BOPTEST) for simulation-based benchmarking of control strategies in buildings. J Build Perform Simul 2021;14(5):586–610.
[40] Scharnhorst P, et al. Energym: a building model library for controller benchmarking. Appl Sci 2021;11(8):3518.

[41] Nweye K, Sankaranarayanan S, Nagy Z. MERLIN: multi-agent offline and transfer learning for occupant-centric operation of grid-interactive communities. Appl Energy 2023;346:121323.
[42] Sørensen ÅL, Sartori I, Lindberg KB, Andresen I. Electric vehicle charging dataset with 35,000 charging sessions from 12 residential locations in Norway. Data Brief 2024;57:110883.
[43] Zhang B, Hu W, Xu X, Li T, Zhang Z, Chen Z. Physical-model-free intelligent energy management for a grid-connected hybrid wind-microturbine-PV-EV energy system via deep reinforcement learning approach. Renew Energy 2022;200:433–48.
[44] Abdullah HM, Gastli A, Ben-Brahim L. Reinforcement learning based EV charging management systems–a review. IEEE Access 2021;9:41506–31.
[45] Ottesen SO, Tomasgard A. A stochastic model for scheduling energy flexibility in buildings. Energy 2015;88:364–76.
[46] S. Dyhr, "Now 3 of 4 Norwegians buy an electric car," Norsk elbilforening. [Online]. Available: https://elbil.no/now-3-of-4-norwegians-buy-an-electric-car/%0A.
[47] Sørensen ÅL, Sartori I, Lindberg KB, Andresen I. A method for generating complete EV charging datasets and analysis of residential charging behaviour in a large Norwegian case study. Sustain Energy Grids Netw 2023;36:101195.
[48] Liu K, Hu X, Yang Z, Xie Y, Feng S. Lithium-ion battery charging management considering economic costs of electrical energy loss and battery degradation. Energy Convers Manag 2019;195:167–79.
[49] Van Roy J, Leemput N, Geth F, Büscher J, Salenbien R, Driesen J. Electric vehicle charging in an office building microgrid with distributed energy resources. IEEE Trans Sustain Energy 2014;5(4):1389–96.
[50] Glatt R, Soper B, Goldhahn R. Collaborative energy demand response with centralized actor and decentralized critic. Livermore, CA (United States): Lawrence Livermore National Lab.(LLNL); 2021.
[51] Kathirgamanathan A, Twardowski K, Mangina E, Finn DP. A centralised soft actor critic deep reinforcement learning approach to district demand side management through CityLearn. In: Proceedings of the 1st international workshop on reinforcement learning for energy management in buildings \& cities; 2020. p. 11–4.
[52] Vazquez-Canteli JR, Henze G, Nagy Z. MARLISA: multi-agent reinforcement learning with iterative sequential action selection for load shaping of grid-interactive connected buildings. In: *Proceedings of the 7th ACM international Conference on Systems for energy-efficient buildings, cities, and transportation*, in BuildSys '20. New York, NY, USA: Association for Computing Machinery; 2020. p. 170–9. https://doi.org/10.1145/3408308.3427604.
[53] Haarnoja T, Zhou A, Abbeel P, Levine S. Soft actor-critic: off-policy maximum entropy deep reinforcement learning with a stochastic actor. In: International conference on machine learning. PMLR; 2018. p. 1861–70.
[54] COLLECTiEF. COLLECTiEF. 2020.
[55] elhub, "Elhub AS." 2020.
[56] NASA Langley Research Center, "Prediction of worldwide energy resources (POWER).".
[57] Ministry of Petroleum and Energy, "Establishes regulatory changes for the sharing of self-produced renewable electricity on the same property," Regjeringen.no. [Online]. Available: https://www.regjeringen.no/no/aktuelt/fastsetter-forskrift sendringer-for-deling-av-egenprodusert-fornybar-strom-pa-samme-eiendom/id2975877/.