Artificial Intelligence Review (2024) 57:88
https://doi.org/10.1007/s10462-024-10706-5

Reinforcement learning applications in environmental sustainability: a review

Maddalena Zuccotto[1] · **Alberto Castellini**[1] · **Davide La Torre**[2] · **Lapo Mola**[3,4] · **Alessandro Farinelli**[1]

Abstract

Environmental sustainability is a worldwide key challenge attracting increasing attention due to climate change, pollution, and biodiversity decline. Reinforcement learning, initially employed in gaming contexts, has been recently applied to real-world domains, including the environmental sustainability realm, where uncertainty challenges strategy learning and adaptation. In this work, we survey the literature to identify the main applications of reinforcement learning in environmental sustainability and the predominant methods employed to address these challenges. We analyzed 181 papers and answered seven research questions, e.g., "How many academic studies have been published from 2003 to 2023 about RL for environmental sustainability?" and "What were the application domains and the methodologies used?". Our analysis reveals an exponential growth in this field over the past two decades, with a rate of 0.42 in the number of publications (from 2 papers in 2007 to 53 in 2022), a strong interest in sustainability issues related to energy fields, and a preference for single-agent RL approaches to deal with sustainability. Finally, this work provides practitioners with a clear overview of the main challenges and open problems that should be tackled in future research.

Keywords Reinforcement learning · Environmental sustainability · Reinforcement learning applications · Sustainable development · Artificial intelligence

✉ Maddalena Zuccotto
maddalena.zuccotto@univr.it

[1] Department of Computer Science, University of Verona, Strada Le Grazie 15, 37134 Verona, Italy

[2] SKEMA Business School, Université Côte d'Azur, Sophia Antipolis, 60 Rue Fedor Dostoïevski, 06902 Valbonne, France

[3] Department of Management, University of Verona, Via Cantarane 24, 37129 Verona, Italy

[4] SKEMA Business School, Université Côte d'Azur (GREDEG), Sophia Antipolis, 60 Rue Fedor Dostoïevski, 06902 Valbonne, France

Published online: 12 March 2024

1 Introduction

Artificial Intelligence (AI) is taking an increasingly important role in industry and society. AI techniques have been recently introduced in autonomous driving, personalized shopping, and fraud prevention, just to make a few examples. A key challenge faced by today's society for which AI can bring an important advancement is environmental sustainability. Climate change, pollution, biodiversity decline, poor health, and poverty have led in the last years governments and companies to focus more and more their efforts and investments on solutions to environmental sustainability problems, which are usually characterized by an inefficient and increased use of resources. Environmental sustainability can be defined as a set of constraints regarding the use of renewable and nonrenewable resources on the one hand, pollution, and waste assimilation on the other (Goodland 1995). In this regard, in 2015, the United Nations published the "2030 Agenda for Sustainable Development" the centerpiece of which is 17 Sustainable Development Goals (United Nations 2015) to be fully achieved by 2030 to attain sustainable development in the economic, social, and environmental contexts, and eliminate all forms of poverty.

AI-based algorithms can control autonomous drones used in water monitoring (Steccanella et al. 2020; Marchesini et al. 2021; Bianchi et al. 2023), extract from acquired data new insight about environmental conditions (Castellini et al. 2020; Azzalini et al. 2020), improve the healthiness of indoor environments (Capuzzo et al. 2022), or demand forecast in district heating networks (Bianchi et al. 2019; Castellini et al. 2021, 2022). Several AI techniques have been employed to address various environmental sustainability challenges. These approaches enable the efficient management of distributed resources within smart grids (Roncalli et al. 2019; Orfanoudakis and Chalkiadakis 2023), improve the power flow for DC grids (Blij et al. 2020), increase the utilization of renewable resources for electric vehicle charging (Koufakis et al. 2020), and mitigate carbon emissions in urban transportation by fostering ridesharing and reducing traffic congestion (Bistaffa et al. 2021, 2017). Furthermore, a crucial aspect of climate change prevention involves optimizing the energy consumption associated with heating and cooling residential properties. To tackle this issue, AI-based approaches have been developed methods to enhance the efficiency of home systems (Panagopoulos et al. 2015; Auffenberg et al. 2017) and quantify the thermal efficiency of residences (Brown et al. 2021). Among the broad spectrum of AI techniques in this survey, we focus on Reinforcement Learning (RL) (Sutton and Barto 2018), which has recently obtained impressive success, achieving human-level performance in several tasks, such as in the context of games (Silver et al. 2016, 2017).

One of the most important and interesting challenges in today's RL research is the application of RL algorithms to real-world domains, where uncertainty makes strategy learning and adaptation much more complex than in game environments. In particular, the application of RL to environmental sustainability has achieved, in the last decade, a strong interest from both the computer science community and the communities of environmental sciences and business. Reducing carbon emissions requires increasing renewable resources usage, such as solar and wind power. While these resources are economically efficient, their stochastic and intermittent nature poses challenges in replacing nonrenewable energy sources within energy networks. RL, with a systematic trial-and-error interaction with dynamic environments, offers a promising approach for learning optimal policies that can adapt to changing system dynamics and effectively manage environmental uncertainty.

Thus, an RL agent is capable of handling variations in operating conditions, for instance, due to a change in resource availability or weather conditions.

This work surveys the recent use of RL to improve environmental sustainability. It provides a comprehensive overview of the different application domains where RL has been used, such as energy and water resource management, and traffic management. The goal is to show practitioners the state-of-the-art RL methods that are currently used to solve environmental sustainability problems in each of these domains. For each paper analyzed, we consider

- The problem tackled,
- The RL approach used,
- The challenges faced,
- The formalization of the RL problem (i.e., type of state/action space, type of transition model, type of RL method, performance measures used to evaluate the results).

The paper is structured as follows. Section 2 presents the surveys already available on topics close to RL and environmental sustainability. Section 3 presents the basic concepts of RL as well as a formalization of the main concepts. In Sect. 4, we present the research methodology used in our survey. Section 5 describes the results of our research, considering different levels of detail. In particular, in Sects. 5.1.1, 5.1.2 and 5.1.3, we provide a quantitative analysis of the state-of-the-art related to the application of RL in environmental sustainability over the last two decades. Then, Sect. 5.1.4 outlines domains where RL techniques are applied and the RL-based approaches employed to address environmental sustainability. In Sect. 5.2, our focus shifts to a subset of 35 main papers, for which we analyze the application domains of proposed RL techniques, provide technical insights into problem formalization, discuss the performance metrics used for evaluation, and consider the challenges addressed. Section 5.3 provides an in-depth analysis of each of these main papers. Finally, in Sect. 6 we discuss our findings, and in Sect. 7 we draw conclusions and summarize future directions.

2 Related work

The literature provides already some surveys on the application of RL to problems related to environmental sustainability, but all these works focus only on specific aspects of environmental sustainability or they consider also AI methods different from RL. For instance, Ma et al. (2020) focus on Energy-Harvesting Internet of Things (IoT) devices, offering insights into recent advancements addressing challenges in commercialization, standards development, context sensing, intermittent computing, and communication strategies. Charef et al. (2023) conduct a study considering various AI techniques, including RL, to enhance energy sustainability within IoT networks. They categorize studies based on the challenges they address, establishing connections between challenges and AI-based solutions while delineating the performance metrics used for evaluation. Within the domain of Architecture, Engineering, Construction, and Operation, Rampini and Re Cecconi (2022) concentrate on the application of AI techniques, including RL, in Asset Management. Their work reviews studies related to several aspects such as energy management, condition assessment, operations, risk, and project management, identifying key points for future development in this context. Alanne and

Sierla (2022) shift their focus to smart buildings, discussing the learning capabilities of intelligent buildings and categorizing learning application domains based on objectives. They also survey the application of RL and Deep Reinforcement Learning (DRL) in decision-making and energy management, encompassing aspects like control of heating and cooling systems and lighting systems. Within the context of smart buildings and smart grids, Mabina et al. (2021) examine the utilization of Machine Learning (ML), including RL, for optimizing energy consumption and electric water heater scheduling, emphasizing the advantages of these approaches in Demand Response (DR) due to their interaction with the environment. Himeur et al. (2022) investigate the integration of AI-big data analytics into various tasks such as load forecasting, water management, and indoor environmental quality monitoring, focusing on the role of RL and DRL in optimizing occupant comfort and energy consumption. Yang et al. (2020) focus on the application of RL and DRL techniques to sustainable energy and electric systems, addressing issues such as optimization, control, energy markets, cyber security, and electric vehicle management.

In the realm of transportation systems, Li et al. (2023) explore various topics, including cooperative mobility-on-demand systems, driver assistance systems, autonomous vehicles (AVs), and electric vehicles (EVs). Sabet and Farooq (2022) study the state-of-the-art in the context of Green Vehicle Routing Problems, which involve reducing greenhouse gas (GHG) emissions and addressing issues like charging activities, pickup and delivery operations, and energy consumption. Moreover, the authors note that most of the works leverage metaheuristics while using RL methods is uncommon. Chen et al. (2019) tackle sustainability concerns within the Internet of Vehicles, leveraging 5th generation mobile network (5G) technology, Mobile Edge Computing architecture, and DRL to optimize energy consumption and resource utilization. Rangel-Martinez et al. (2021) assess the application of ML techniques, including RL, in manufacturing, with a focus on energy-related fields impacting environmental sustainability. Sivamayil et al. (2023) explore a wide range of RL applications (e.g., Natural Language Processing, health care, etc.) emphasizing Energy Management Systems with an environmental sustainability perspective. Mischos et al. (2023) investigate Intelligent Energy Management Systems across diverse building environments, considering control types and optimization approaches, including ML, DL, and DRL. Yao et al. (2023) discuss the application of Agent-Based Modeling and Multi-Agent System modeling in the transition to Multi-Energy Systems, highlighting RL and suggesting future research directions in Multi-Agent Reinforcement Learning (MARL) for energy systems.

While these works address specific aspects of environmental sustainability using RL methods, our review takes a comprehensive approach, analyzing all contexts in which RL techniques have recently contributed to enhancing environmental sustainability. Our goal is to provide practitioners with insights into state-of-the-art methods for addressing environmental sustainability challenges across various application domains, including energy and water resource management and traffic management. In summary, the main contribution of this survey consists of offering an overview of RL application domains within the context of environmental sustainability.

3 Reinforcement learning: preliminaries and main definitions

In this section, we present the basic concepts of RL as well as a formalization of the main concepts. RL, a prominent machine learning paradigm, focuses on learning a policy that maximizes cumulative rewards, i.e., which action should be selected considering the environment configuration for achieving the best possible outcome. Key elements of RL are listed in the following:

- The *agent* is the entity that makes decisions and performs actions in the environment;
- The *environment* represents the system with which the agent interacts and provides the agent with feedback on the performed action;
- The *policy* is a function that defines the agent's behavior considering the environment configuration (i.e., a map between what the agent observes and what the agent should do);
- The *reward* is a numerical signal that provides feedback on the action performed by the agent;
- The *value function* specifies state values, namely, how valuable it is to reach a state, considering also future states reachable from it;
- The *model of the environment* (optional) is a stochastic function providing next state probability given current state and action, it allows simulating the behavior of the environment in response to the agent's actions.

RL methods (Sutton and Barto 2018) can be categorized into two main groups: model-free and model-based (Moerland et al. 2020). Over the past two decades, model-free methods have demonstrated significant success. Meanwhile, model-based approaches have become a focal point in current research due to their potential to enhance sample efficiency, which is a reduction in interactions with the environment. This efficiency is achieved by explicitly representing the model of the environment and incorporating relevant prior knowledge (Castellini et al. 2019; Zuccotto et al. 2022a, b). Additionally, model-based methods offer the advantage of addressing the risks associated with taking actions in partially observable environments (Mazzi et al. 2021, 2023; Simão et al. 2023) or partially known environment (Castellini et al. 2023).

A common framework to formalize the RL problem is by using Markov Decision Process (MDP) (Puterman 1994). An MDP is a tuple (S, A, T, R, γ) where S is a finite set of *states*, A is a finite set of *actions*, $T : S \times A \rightarrow \Pi(S)$ is the *transition model* where $\Pi(S)$ is the space of probability distribution over states, $R : S \times A \rightarrow \mathbb{R}$ is the *reward function* and $\gamma \in [0, 1)$ is the *discount factor*. The agent's goal is to maximize the *expected discounted return* $\mathbb{E}[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)]$ acting optimally, namely, choosing in each state s_t, at time t, the action a_t with the highest expected reward. The solution of an MDP is an optimal policy, namely, a function that optimally maps states into actions. A policy is optimal if it maximizes the expected discounted return. The discount factor γ reduces the weight of long-term rewards guaranteeing convergence. In the case of partially observable environments, an extension of the MDP framework, namely POMDP (Kaelbling et al. 1998), can be used. A POMDP is a tuple $(S, A, O, T, \Omega, R, \gamma)$ where the elements shared with MDP are augmented by Ω, a finite set of *observations*, and $O : S \times A \rightarrow \Pi(\Omega)$, the *observation model*. In contrast to MDPS, in POMDPs the agent is not able to directly observe the current state s_t but it maintains a probability distribution over states S, called *belief*, which updates at each timestep. The belief summarizes the agent's previous experiences, i.e. the sequence of actions and observations that the agent took from an initial belief b_0 to the belief b. The solution of a POMDP is an optimal *policy*, namely, a function that optimally maps belief states into actions. In the following, we will survey applications of RL to environmental sustainability, hence we will investigate how the elements described in this section (e.g., MDP modeling framework, RL algorithms, etc.) have been used so far to solve problems related to environmental sustainability.

4 Review methodology

In this section, we outline the research methodology we used for this study. It consists of 5 steps: (i) the definition of the research questions, (ii) paper collection process, (iii) the definition of inclusion and exclusion criteria, (iv) the identification of relevant studies based on inclusion and exclusion criteria, (v) data extraction and analysis.

Research questions. The first step involves defining the research questions we want to answer on the application of RL techniques for environmental sustainability. The goal of our questions is twofold: to offer a quantitative analysis of the state of the art related to the application of RL to environmental sustainability and to analyze the use of these techniques focusing on sustainability. Specifically, we aim to answer the following questions:

- RQ1: How many academic studies have been published from 2003 to 2023 about RL for environmental sustainability?
- RQ2: What were the most relevant publication channels used?
- RQ3: In which country were located the most active research centers?
- RQ4: What were the application domains and the methodologies used?
- RQ5: How was the RL problem formalized (i.e., type of state/action space, type of transition model, and type of dataset used)?
- RQ6: Which evaluation metrics were used to assess the performance?
- RQ7: What were the challenges addressed?

The databases we use to collect papers are those of the search engines Scopus and Web of Science. To limit the scope of research to the application of RL approaches for environmental sustainability, we define the following search strings:

- "reinforcement learning AND sustainable AND environment";
- "reinforcement learning AND environmental AND sustainability";
- "reinforcement learning AND environment AND sustainability";
- "reinforcement learning AND environmental AND sustainable".

The search on the two databases led to a total of 375 papers, 236 collected from Scopus and 139 from Web of Science.

Selection criteria for the initial set of (181) papers. To refine the results of the search, we outline the following inclusion and exclusion criteria.

Inclusion criteria. To determine studies eligible for inclusion in this work, we consider the following criteria:

- It is written in English;
- It is clearly focused on RL for environmental sustainability;
- In the case of duplicate articles, the most recent version is included.

Exclusion criteria. To further refine our search, we apply the following exclusion criteria: the study is an editorial, a conference review, or a book chapter.

Following these criteria, we found 181 papers (104 articles, 70 conference papers, and 7 reviews). We combine the information in the index keywords of these papers with their number of citations and the publication year. In particular, we compute the number of occurrences of each keyword to identify the application domains and methodologies most used in the literature. To this aim, we standardize the keywords to avoid spelling variations. Then, we combine these values with the number of citations and the publication year to identify the most recent and relevant studies. In cases where index keywords are missing, we use author keywords. For the only three papers that do not have author nor index keywords, we use the title as related keywords.

Selection criteria for the set of (35) main papers. To identify papers for the in-depth analysis, we applied the following criteria that consider the most important keyword occurrences (i.e., the most frequent keywords), the publication year, and the number of citations based on publication year.

- Presence of at least one keyword with no less than 10 occurrences;
- Publication year from 2013 to 2023;
- Number of citations:
  - Papers published in 2022–2021, at least 3 citations;
  - Papers published in 2020–2019, at least 10 citations;
  - Papers published in 2018–2013, at least 20 citations.

Following these criteria, we selected 35 studies that have been explored in-depth, and answers to the research questions defined above have been reported.

In the following sections, we first consider the initial 181 papers found using the search strings defined above and applying inclusion/exclusion criteria. In Sect. 5.1.1 we answer question RQ1 for those papers, in Sect. 5.1.2 we answer question RQ2, in Sect. 5.1.3 we answer question RQ3 and in Sect. 5.1.4 we answer question RQ4. Namely, we first analyze the number of papers that focus on RL for sustainability published in the last 20 years, then we identify the main international conferences, workshops, and journals used to disseminate research, subsequently, we find the research centers that are particularly active in this research/application topic, and finally, we analyze the application domains and RL methodologies used. From Sect. 5.2, we start focusing only on the main 35 papers identified using main papers selection criteria. In particular, we answer question RQ4 in Sect. 5.2.1, question RQ5 in Sect. 5.2.2, question RQ6 in Sect. 5.2.3, and question RQ7 in Sect. 5.2.4. Namely, for these main papers, we first analyze the application domains of RL techniques and the RL-based approaches used to tackle environmental sustainability; then we analyze the way in which the problem has been formalized; subsequently, we investigate the evaluation measures used; finally, we identify the main challenges addresses. Notice that questions RQ1, RQ2, and RQ3 have not been answered considering only the main 35 papers because these questions aim to provide a quantitative analysis of the state of the art as a whole, and this subset of articles is part of the 181 papers used to answer these three questions.

5 Results of the review

This section reports the results of the analysis provided in this survey, first for the initial set of 181 papers, then for the subset of the main 35 papers.

5.1 Analysis of the initial set of 181 papers

The initial set of papers, selected using the search strings of Sect. 4, is analyzed by answering questions RQ1, RQ2, RQ3, and RQ4.

5.1.1 RQ1: How many academic studies have been published from 2003 to 2023 about RL for environmental sustainability?

This research question aims to quantify the interest of the international scientific community in applying RL methods to environmental sustainability problems over the last 20 years. As shown in Fig. 1, the number of publications (pink dots) remained relatively low until 2018, with the number of publications each year less than five. Since 2019, there has been a rapid growth of up to 53 papers in 2022, showing the increasing interest in this topic during the last few years. It is important to notice that the data for the year 2023 are updated to April 2023 and do not represent a decrease in the number of studies published. Application of inclusion and exclusion criteria leads to no publication in the years 2004, 2005, 2010, and 2011. In Fig. 1, we also show that the increase in the number of publications fits an exponential pattern (green line) with a growth rate of 0.42 in the number of publications (from 2 papers in 2007 to 53 in 2022). To compute the regression model, we do not consider 2023 in the model since its information is partial.

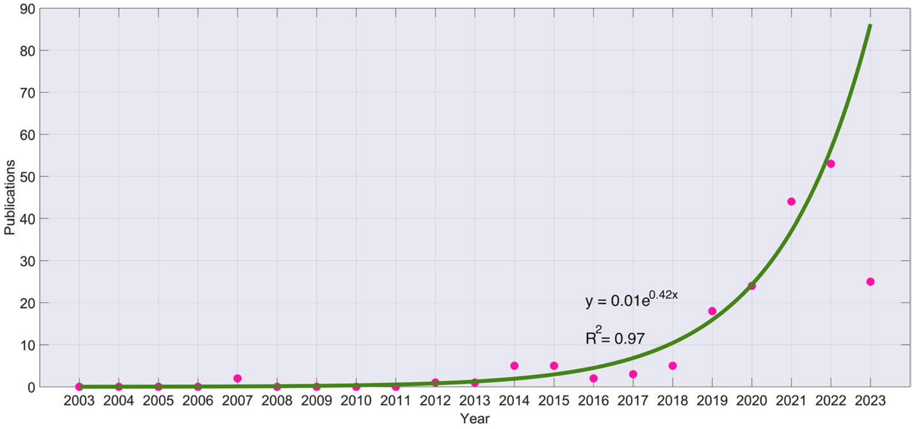

Fig. 1 Academic studies published from 2003 to 2023. Pink dots represent the number of publications per year used to compute the regression model represented by the green line

Table 1 Journals and conferences with at least two publications

	Publications	Scope
Conference		
Lecture notes in computer science (*)	5	CS
IEEE conference on intelligent transportation systems	3	CS + APP
International conference on autonomous agents and multiagent systems	2	CS
IEEE international conference on distributed computing systems	2	CS
International conference on mobility, sensing and networking	2	CS + APP
IOP conference series: earth and environmental science	2	APP
Land, water and environmental management: integrated systems for sustainability	2	APP
Journal		
IEEE access	7	APP
IEEE internet of things journal	5	CS + APP
Sustainability (Switzerland)	5	CS + APP
IEEE transactions on intelligent transportation systems	4	CS + APP
Sustainable cities and society	4	CS + APP
Energies	3	APP
IEEE transactions on green communications and networking	3	CS + APP
IEEE transactions on vehicular technology	3	CS + APP
Journal of cleaner production	3	APP
Applied energy	2	APP
Applied sciences (Switzerland)	2	APP
Electronics (Switzerland)	2	CS + APP
Energy and buildings	2	APP
IEEE sensors journal	2	APP
IEEE transactions on network and service management	2	CS + APP
IEEE wireless communications	2	APP
Journal of hydrology	2	APP
Resources, conservation and recycling	2	APP
Sensors	2	APP
Sustainable energy technologies and assessments	2	APP

In the "Scope" column, "CS" and "APP" indicate a technical/informatics or application-oriented perspective of the Conference/Journal, respectively, while "CS + APP" denotes a combination of them

*Including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics

5.1.2 RQ2: What were the most relevant publication channels used?

With this research question, we aim to show what are the main channels used to disseminate research in the application of RL techniques to environmental sustainability problems. In Table 1, we show the journals and conferences with at least 2 publications. As can be seen, the topics of the journals and conferences are very varied. In particular, some of these communication channels are specific for sustainability, e.g., "Sustainability (Switzerland)" and "Sustainable Cities and Society", and many are related to environmental aspects such as "IOP Conference Series: Earth and Environmental Science" and “IEEE Transactions on Green Communications and Networking”. Moreover, in the third column of Table 1, we provide an overview of the scope of the publication channels. To this aim, we analyze the information presented on the website of each conference and journal about its scope, indicating whether it has a technical/informatics or application-oriented perspective (“CS” or “APP”, respectively) or a combination of them (“CS + APP”). As can be seen, most of the publication channels are application-oriented (2 conferences + 12 journals), followed by those that present a combined scope (2 conferences + 8 journals), finally, a few of them (3 conferences) have a more technical/informatics perspective.

5.1.3 RQ3: In which country were located the most active research centers?

This research question aims to show which countries whose research centers are most concerned with the application of RL methods to environmental sustainability issues. With this in mind, we leverage the information in the Scopus and Web of Science databases about the 181 papers that were not excluded by the application of inclusion and exclusion criteria. In Fig. 2, we show only the countries with at least 5 publications and, as we can see, the highest number of papers comes from research centers located in China (33 papers), followed by the United States (29 papers), and the United Kingdom (17 papers). It is important to note that most of these works are developed in collaboration between research centers in multiple countries, so we count the paper for each collaborating country. To show co-author relationships, in Fig. 3, we represent only countries with at least 5 occurrences among analyzed documents. Each country is depicted as a circle, a link between 2 circles represents a co-authorship relation, and the line weight is proportional to the number of papers in the co-authorship relationship. As we can see, the countries with

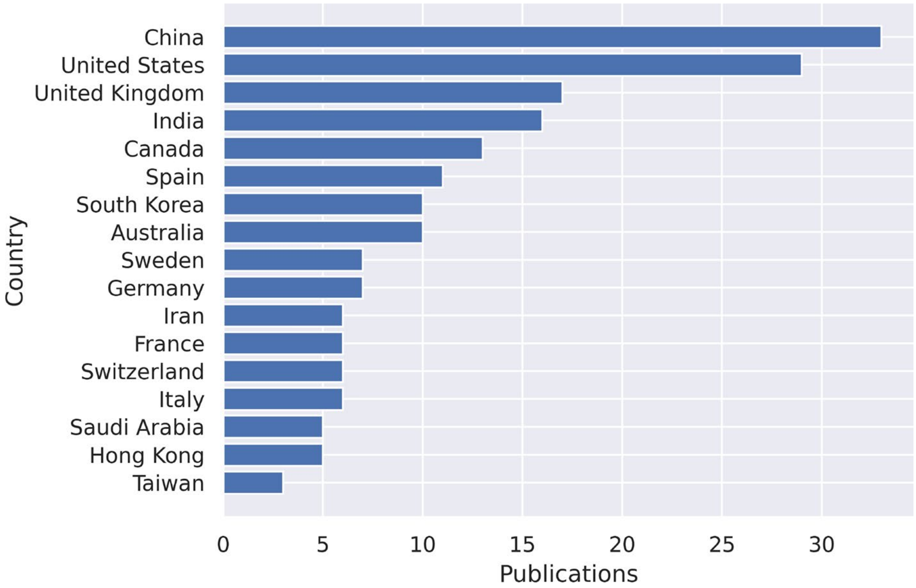

Fig. 2 Number of publications per Country on RL approaches for environmental sustainability

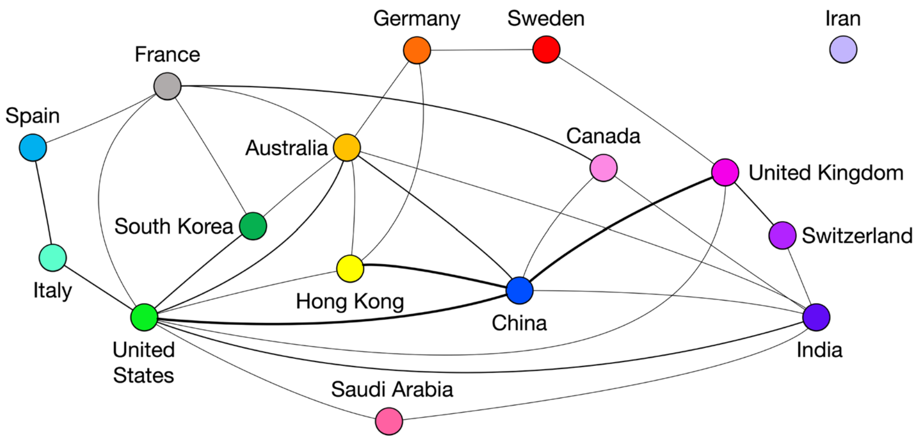

Fig. 3 Co-author relationships with Country as a unit of analysis. Nodes represent states, and links depict co-authorship relationships. The thickness of the link is proportional to the number of papers in the co-authorship relationship

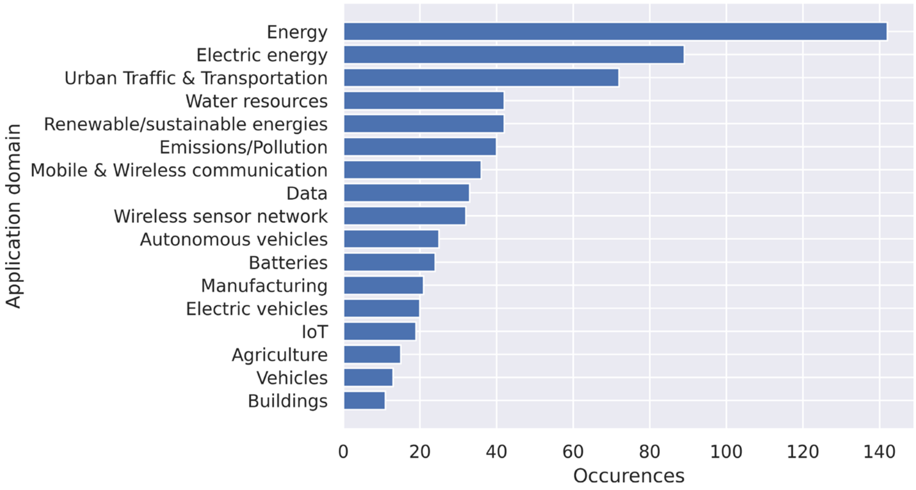

Fig. 4 Overview of application domains. For each application domain (y-axis), we show the number of occurrences of keywords belonging to its macro-area (x-axis)

more links are the United States (9 links), followed by Australia (7 links), China, and India (6 links).

5.1.4 RQ4: What were the application domains and the methodologies used?

This research question aims to analyze the application domains and the RL methodologies used for tackling issues related to environmental sustainability. To this aim, we analyze the index keywords of the 181 papers that were not excluded by applying the inclusion and exclusion criteria and the authors' keywords for works with no index keywords. In Fig. 4,

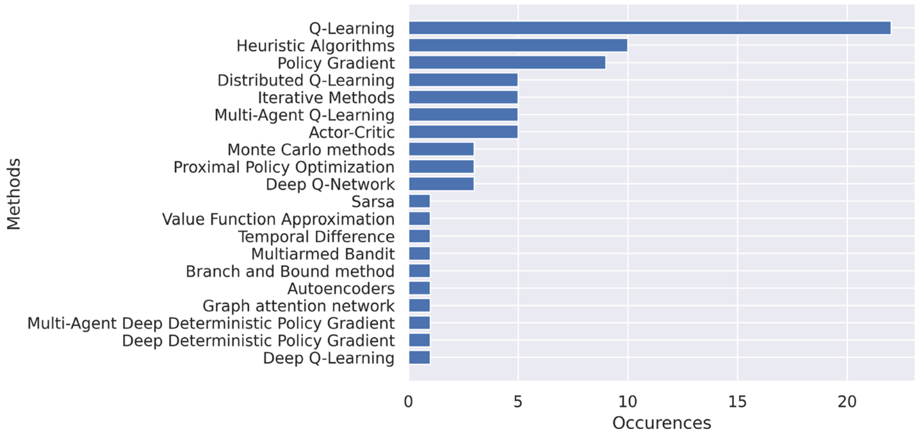

Fig. 5 Overview of RL methods used. For each RL method (y-axis), we show the number of occurrences of corresponding keywords (x-axis)

we show the application domains with more than 10 index keyword occurrences. We group the keywords into macro areas, such as in "Energy" we include keywords like "energy", "energy conservation", "energy consumption", etc., in "Electric energy" we group keywords such as "electric energy storage", "electric load dispatching", "smart grid", etc. (see Appendix for details). The image clearly shows that there is a wide variety of application domains, but most of the applications deal with sustainability issues related to energy fields.

Regarding the proposed approaches, we follow the same procedure as previously described for application domains grouping keywords that refer to the same method. For example, in "Actor-Critic" we group keywords such as "actor critic", "advantage actor-critic (A2C)", and "soft actor critic". As we can see in Fig. 5, the most widely used RL method for dealing with environmental sustainability in different application domains is a state-of-the-art model-free algorithm, namely Q-Learning (Watkins 1989). It is important to note that, in the image, we show only RL approaches, but there are also index keywords related to other approaches like "genetic algorithm", "simulated annealing", etc.

Moreover, we perform a bibliometrics analysis on the co-occurrence of index keywords by using VOSviewer (Perianes-Rodriguez et al. 2016). Having a co-occurrence means that 2 keywords occur in the same work. After a data cleaning process, VOSviewer detects 17 clusters by considering keywords with 3 occurrences at least. In Fig. 6, each cluster corresponds to a color, and each element of the cluster, namely a keyword, is depicted by a circle in the cluster color. For instance, the blue cluster is made of several blue nodes, each of which contains a keyword (e.g., electric vehicles, charging (batteries)) belonging cluster. The size of the circle and the circle label depend on the number of occurrences of the related keyword. Lines between items depict co-occurrences of keywords in a paper. Each cluster groups keywords identifying an application domain and/or the approaches used to tackle related environmental sustainability issues. For example, cluster 1 (red colored on the top-right) is somewhat related to traffic signals control for traffic management through the application of control strategies. Cluster 2 (green colored on the left) is related to power management and energy harvesting in wireless sensor networks.

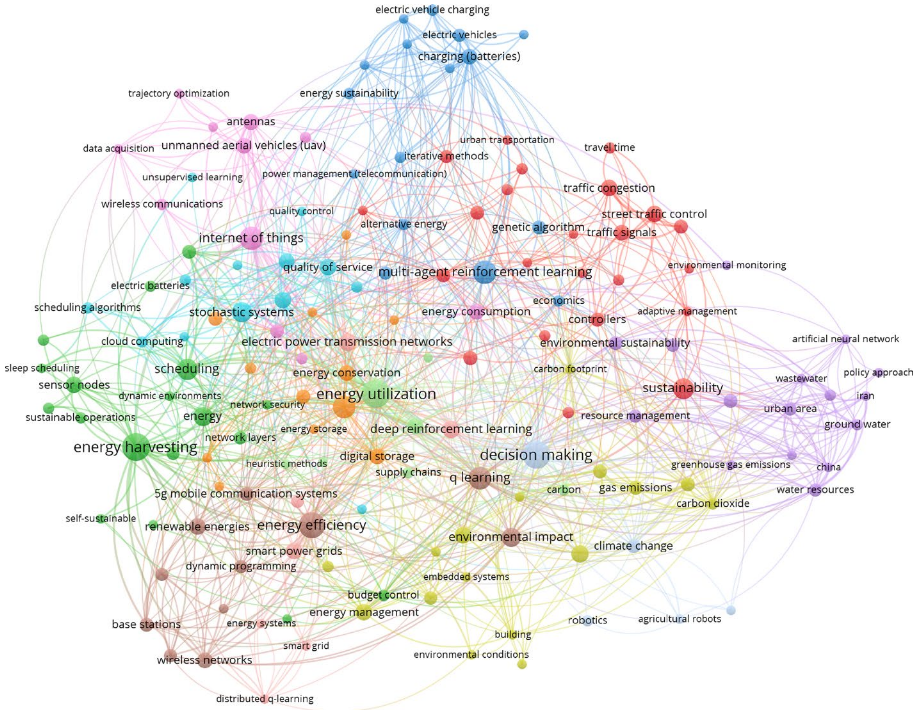

Fig. 6 Bibliometrics analysis on the co-occurrence of index keywords. Each color outlines a cluster, and each circle of the cluster color represents a keyword, while edges represent co-occurrences of keywords in the same work

5.2 Analysis of the 35 main papers

In this section, we focus on the 35 papers chosen using the selection criteria for the main papers (see Sect. 4). First, we provide a high-level analysis of the application domain and the RL approaches used to address environmental sustainability issues (research question RQ4). Then, we give an overview of the RL problem formalization (i.e., type of state/action space, type of transition model, type of RL method) (research question RQ5). Subsequently, we analyze the performance measures used to evaluate the results (research question RQ6). Finally, we evaluate the main challenges faced (research question RQ7).

5.2.1 RQ4: What were the application domains and the methodologies used?

In Table 2, we summarize the application domains and the RL approaches used in the selected works. First, we group the 35 main works according to their main related application domains (first column). It is important to note that application domains may overlap consequently, we report all application domains common to all papers in the same group. Then, we indicate for each paper (second column) the method behind the proposed technique (third column). The selected papers tackle environmental sustainability issues in the application domains shown in Fig. 4. In particular, the most relevant

Table 2 Technical information about the selected works. In the third column, we report the RL methodology used

Application domain	Paper	Method	State	Action	Tr. model	Dataset
Electric vehicles, batteries, energy	Sultanuddin et al. (2023)	DDQN	N/A	D	St.	S
	Zhang et al. (2021a)	MA AC	N/A	D*	Det.*	R
IoT	Ajao and Apeh (2023)	Q-Learning	D*	D*	St.	S
	Zhang et al. (2021b)	Adaptive regression	D*	C*	Det.*	S
	Han et al. (2020)	MAB	N/A	D*	St.	S
Water resources	Emamjomehzadeh et al. (2023)	Q-Learning	N/A	N/A	N/A	S
	Skardi et al. (2020)	Q-Learning	N/A	N/A	Det., St.	R + S
Emissions/pollution	Chen et al. (2021)	MADDPG	C	C	N/A	S
	Huo et al. (2023)	Multi-agent Q-learning	D*	D	N/A	S
Agriculture	Elavarasan and Durairaj Vincent (2020)	DRQN	D*	N/A	N/A	R + S
Data, energy	Shaw et al. (2022)	Q-Learning, SARSA	D*	D*	St.	S (R based)
	Venkataswamy et al. (2023)	AC	D*	D	N/A	S, R
Urban traffic and transportation	Ounoughi et al. (2022)	DQN	N/A	D*	N/A	R
	Alizadeh Shabestray and Abdulhai (2019)	DQN	D*	D	St.	S
	Aziz et al. (2018)	RMART	D*	D*	St.*	S
	Khalid et al. (2023)	DQN	D*	D*	Det.*	S
Buildings, energy	Kathirgamanathan et al. (2021)	SAC	C	N/A	N/A	R + S
	De Gracia et al. (2015)	SARSA(λ)	D*	D*	Det.	S
Manufacturing	Wang and Wang (2022)	Policy network	D*	D*	Det.*	S*
	Leng et al. (2021)	Q-Learning*	N/A	D*	St.*	S (R based)
Mobile and wireless communication, energy, renewable/sustainable energies	Liu et al. (2021)	DDPG	C	C	N/A	S
	Miozzo et al. (2015)	Distributed Q-Learning	D*	D	N/A	S
	Miozzo et al. (2017)	Distributed Q-Learning	D*	D	N/A	S
	Giri and Majumder (2022)	Deep Q-Learning	C*	D*	St.*	S
Mobile and wireless communication	Al-Jawad et al. (2021)	Q-Learning	D*	D*	N/A	S

Table 2 (continued)

Application domain	Paper	Method	State	Action	Tr. model	Dataset
Energy, electric energy	Sheikhi et al. (2016)	Q-Learning	N/A	N/A	St.	S
	Harrold et al. (2022)	Rainbow DQN	C*	D	N/A	R + S
	Jendoubi and Bouffard (2022)	MADDPG	N/A	C	St.*	S*
Energy	Gao et al. (2023)	Q-Learning	N/A	D*	St.*	R
Wireless sensor network, energy, renewable/sustainable energies	Hsu et al. (2014)	Q-Learning*	D*	D*	N/A	S
	Chen et al. (2016)	Q-Learning	D*	D*	St.*	R + S
	Feng et al. (2023)	DDPG	C	C	St.*	S
Autonomous vehicles	Bouhamed et al. (2020)	DDPG	C*	C	N/A	S
		Q-Learning	D*	D*		
	Sacco et al. (2021)	AC	C*	D*	N/A	R + S
	Gu et al. (2023)	Policy gradient	C*	D	St*	S

In the fourth and fifth columns, we indicate with “C” and “D” continuous and discrete state/action spaces, respectively. In the sixth column, “Det.” and “St.” denote deterministic and stochastic transition models respectively. In the seventh column “S” and “R” indicate synthetic and real-world datasets, respectively application domain correlates to the macro area of “Energy”. Indeed, it involves more than half of the papers in the table, considering both the works in which it represents the main application domain and those in which it is related to the main application domain. In Table 2, we also show that 16 out of the 35 selected papers use DRL approaches such as Deep Q-Network (DQN) (Mnih et al. 2015) and Double Deep Q-Network (DDQN) (van Hasselt et al. 2016), and another 2 rely on DRL techniques in multi-agent contexts, such as Multi-Agent Deep Deterministic Policy Gradient (MADDPG) (Lowe et al. 2017). RL techniques are used by 10 articles, here the most used method is Q-Learning, and 7 apply RL approaches to a multi-agent context. Finally, only 1 paper adopts a Genetic Algorithm-based RL (GARL) approach.

5.2.2 RQ5: How was the RL problem formalized (i.e., type of state/action space, type of transition model, and type of dataset used)?

This research question deals with a technical point of view, which we think may be helpful for practitioners to get an overview of the environments considered by the authors in developing the proposed methods. In Table 2, we summarize the information related to problem formulation that we found in the selected papers. For each paper, we point out if the state and action spaces are continuous or discrete and whether the transition model is deterministic or stochastic. Finally, we provide information on the dataset used in the experiments, outlining whether real-world or synthetic data are used. It is important to note that not all papers explicitly provide this information. Thus, we mark with “*” all information inferred from reading the article. On the other hand, “N/A” specifies that the information available was not enough to infer the required data.

In the selected papers, most of the spaces of states and actions are discrete. Indeed, only 9 approaches use a continuous space state (third column of Table 2), and 6 use a continuous action space (fourth column of Table 2). Regarding the transition model, we can see that the model is stochastic in most cases where the information is available. In De Gracia et al. (2015), “Det.” and “St.” are both reported because the authors test the proposed methodology on both models. Finally, in the last column, we note that most of the experiments are performed on synthetic datasets. In fact, only 9 papers use real-world data, 6 of which combine them with synthetic data (“R + S” in the table), while 2 others use the real data to generate larger data sets from them (“S (R based)” in the table). Only Venkataswamy et al. (2023) test the proposed approach on both dataset types (“R, S” in the table).

5.2.3 RQ6: Which evaluation metrics were used to assess the performance?

This research question aims to provide an overview of the authors’ performance measure choices to evaluate the proposed approaches in the 35 selected papers. In the second column of Table 3, we report information about the metrics found in the articles, which are also indicated in the in-depth analysis of each paper in Section 5.3. As we can see in Table 3, the performance measures vary widely depending on the application domain and the goal of the method proposed in each paper. For example, reward is used as a metric in 9 articles but is computed differently depending on the context. Concerning electric vehicles, in Sultanuddin et al. (2023), the reward corresponds to a penalty function considering the cost of charging and a departure incentive. Instead, in wastewater treatment plants (WWTPs), Chen et al. (2021) use a reward function that takes into account the operational

Table 3 We summarize the performance measures used by the authors to evaluate the proposed approaches in the second column and the challenges they address in the third column

Paper	Performance measures	Challenges
Sultanuddin et al. (2023)	Reward, voltage levels, load curves, charging/discharging curves	Grid overload prevention, driving pattern uncertainty, dimensionality
Zhang et al. (2021a)	MCWT, MCP, TSF, CFR	Dimensionality, coordination and cooperation among agents, charging requests competitiveness, joint optimization of multiple optimization objectives
Ajao and Apeh (2023)	Detection accuracy, recall, precision, specificity, F-measure	Security threat to sustainability functionality
Zhang et al. (2021b)	Operational logs, power wastage, power requirements, average failure ratio	Energy requirements management, smart power allocation
Han et al. (2020)	Number of ready nodes, throughput	Spatial uncertainty
Emamjomehzadeh et al. (2023)	Water table level, nitrate concentration, energy usage, GHG emissions	WEF nexus modeling and management for an urban area integrated water resource management
Skardi et al. (2020)	Water and wastewater allocation, water and groundwater level, nitrate concentration	Social attachments quantification and consideration, cooperation among agents
Chen et al. (2021)	Reward, Q-values, influents, inflow rate, DO and dosage values, energy consumption, cost, EP, and GHG emissions	WWTP impact optimization
Huo et al. (2023)	Productivity, operational mistakes, GHG emissions, queuing time	Operational randomness and uncertainties
Elavarasan and Durairaj Vincent (2020)	R2, MAE, MSE, RMSE, MedAE, MSLE, MAPE, PDF, explained variance score, accuracy	Mapping between raw data and crop yield values, effectiveness dependence on extracted features quality
Shaw et al. (2022)	Energy consumption, SLAV, number of migrations, ESV	Energy awareness, slow convergence to optimal policy
Venkataswamy et al. (2023)	Monetary job value	Intermittent power, environment nonuniformity, system design and configuration effects, learning and improving available heuristic policies
Ounoughi et al. (2022)	MSE, MAE, noise levels, CO_2 emission, fuel consumption	Sustainability and proactivity integration
Alizadeh Shabestray and Abdulhai (2019)	Average of intersection travel time, queue time, and network travel time, weighted average intersection person travel time	Regular and transit vehicles consideration
Aziz et al. (2018)	Average delay, stopped delay, number of stops, and network-wide delay, GHG emissions	Traffic congestion information sharing, reward function dynamic adaptation

Table 3 (continued)

Paper	Performance measures	Challenges
Khalid et al. (2023)	Execution time, reward, path planned, distance	Quality of experience assurance, optimal order user serving, distance minimization
Kathirgamanathan et al. (2021)	Energy purchased and cost, discomfort, reward, temperature, power demand	Robustness, scalability, lack of well-established environments
De Gracia et al. (2015)	Electrical energy saving	Energy saving maximization, thermal energy storage optimization
Wang and Wang (2022)	Overall Nondominated Vector Generation, C Metric, Hyper volume, $D1_R$	Energy awareness with simultaneous makespan and energy minimization
Leng et al. (2021)	MSE, MSLE, RMSE, R2, unit and total profit, acceptance rate	Demand uncertainty, order customization
Liu et al. (2021)	Achievable rate, transmission power	Effective communication, IRS phase optimization
Miozzo et al. (2015)	*Throughput gain, traffic drop rate, energy efficiency, energy efficient improvement*, traffic demand, harvested energy, battery level, policy, normalized load at the macro bs, total energy spent, average load, average cell load for the macro bs battery outage, Jain's fairness index	Energy harvesting
Miozzo et al. (2017)	Switch-off rate, battery level, excess energy	
Giri and Majumder (2022)	Reward, capacity, network lifetime, average delay	Dimensionality, efficient use of collected energy for QoS
Al-Jawad et al. (2021)	Throughput, packet loss, rejected flows, PSNR, MOS	Sustainable QoS
Sheikhi et al. (2016)	Storage charge level, operational cost, primary energy involved	Energy system parameter variability or stochasticity, smart grid architecture issues
Harrold et al. (2022)	MAPE, energy cost savings, relative savings, episodic rewards, and value distribution	Energy arbitrage, renewables usage improvement, limited data availability
Jendoubi and Bouffard (2022)	Annual total cost, daily operation cost, PV production, aggregated demand, power to be charged/discharged, generator provided power, provider delivered electricity, PAR	Energy dispatch
Gao et al. (2023)	Working and non-working energy consumption, the ratio between working and non-working energy consumption	Energy consumption minimization

Table 3 (continued)

Paper	Performance measures	Challenges
Hsu et al. (2014)	RBE, EDC, OTRT, ToD achievability	Simultaneous achievement of throughput on demand satisfaction and power consumption reduction
Chen et al. (2016)	Nodes potential energy, network lifetime, area coverage ratio, number of residual alive nodes versus the network lifetime, recharging cycle	Simultaneous area coverage and energy balancing
Feng et al. (2023)	Distribution of the ratio between the expected per-slot harvested energy and the MS-to-sink distance within the network, moving trajectories and steps, reward, actor-loss, battery level, accuracy, convergence, training time	Lack of energy-related information, energy harvesting and data transmission trade-off
Bouhamed et al. (2020)	Path followed, reward, battery level, completion time of the tour against the ground unit transmission power	Limited battery capacity, obstacle awareness
Sacco et al. (2021)	Task completion time, utility, average node-antenna distance, energy consumption, task completion time against the average computing workload, CDF and utility evolution	Task completion time reduction, energy efficiency
Gu et al. (2023)	Energy loss, collision loss, reward, lane changes	Efficient response to environmental observations

The papers are grouped according to the main related application domains, as in Table 2. Performance measures written in italics are used in both Miozzo et al. (2015) and Miozzo et al. (2017)

cost, consisting of multiple components, such as energy cost and biogas price, and several indicators, like energy consumed by the aeration and sludge treatment processes and GHG emissions. Another performance measure common to multiple application domains is, for example, energy consumption. Indeed, it is used in contexts such as water resources management (Emamjomehzadeh et al. 2023), WWTPs (Chen et al. 2021), data centers (Shaw et al. 2022), and AVs (Sacco et al. 2021). Even approaches related to the same application domain may differ in terms of performance measures depending on their objective. Considering, for example, the water resources context, both Emamjomehzadeh et al. (2023) and Skardi et al. (2020) evaluate their proposed approaches using resource level and nitrate concentration. However, in (Emamjomehzadeh et al. 2023), energy consumption and GHG emissions are also considered, while in (Skardi et al. 2020), resource allocation is used.

5.2.4 RQ7: What were the challenges addressed?

This research question aims to offer an overview of the issues that the authors have tackled within the 35 selected papers. In the third column of Table 3, we summarize information about the challenges addressed in the articles, which are also indicated in the in-depth analysis of each paper in Section 5.3. As with the performance measures, we can see in Table 3 that the challenges faced vary greatly depending on the application context and the goal of the method proposed in each paper. As an example, considering the domain of electric vehicles, Sultanuddin et al. (2023) address several challenges, like avoiding network energy overload at peak times, considering the uncertainty of driving patterns, and managing large state spaces. On the other hand, in addition to the challenge related to dimensionality, Zhang et al. (2021a) also address issues related to coordination and collaboration among agents, the competitiveness of charging demands, and joint optimization of multiple objective functions. However, although not explicitly stated by the authors, the challenge that unites these papers is the development of approaches capable of adapting to changes in a dynamic environment and managing the uncertainty associated with the environment that, in many cases, arises from the use of renewable resource sources, which have a stochastic and intermittent nature whose management adds further complexity to the problem.

5.3 Analysis of single papers (grouped by application domain)

In this section, we group the 35 main papers by application domain and analyze each single paper answering research questions RQ4, RQ5, RQ6, and RQ7. This provides the reader interested in a specific application domain with a deep knowledge of the main features of these papers. Notice that in answering RQ5, we use the information available in Table 2 and report in the text a “(*)” for all information inferred from reading the article.

5.3.1 Electric vehicles, Batteries, Energy

The transportation system is characterized by an increasing presence of EVs due to their eco-friendly features. In Sultanuddin et al. (2023), it is proposed a DDQN-based approach to provide a smart scalable charging strategy for EV fleets that ensures all cars have sufficient charging for their trips without exceeding the maximum energy threshold of the power grid. The charging management system combines information on the current state of the network and vehicle with historical data, being able to schedule charging at least 24 hours in advance. In developing the proposed approach, it is considered an environment with discrete actions and a stochastic transition model. The experimental evaluation is performed on a synthetic dataset by using as metrics the reward, the voltage levels, the load curves, and the charging/discharging curves. The rapid growth in the popularity of EVs subjects the power grid infrastructure to challenges, such as preventing grid overload at peak times. Moreover, the authors address issues related to driving pattern uncertainty and handling large state spaces.

Zhang et al. (2021a) propose a framework for charging recommendations based on MARL, called Multi-Agent Spatio-Temporal Reinforcement Learning (MASTER). By leveraging a multi-agent actor-critic framework with Centralized Training and Decentralized Execution (CTDE), the proposed approach increases the collaboration and cooperation among agents, and it can make use of information about possible future charging competition through the use of a delayed access strategy. The framework is further extended to multi-critics for addressing multiple objective optimizations. MASTER works in environments characterized by discrete actions (*) and a deterministic transition model (*), and it has been tested on a real-world dataset. To evaluate its performance, the Mean Charging Wait Time (MCWT), Mean Charging Price (MCP), Total Saving Fee (TSF), and Charging Failure Rate (CFR) are used as performance measures. In the development of the proposed charging recommendation approach, the authors face several challenges such as dealing with large state and action space, coordination and cooperation among agents in a large-scale system, potential competitiveness of future charging requests, and the joint optimization of multiple optimization objectives.

5.3.2 IoT

Recent years have seen rapid advances in IoT technology enabling the development of smart services such as smart cities, buildings, and oceans. Regarding smart cities, Ajao and Apeh (2023) consider the Industrial Internet of Things and present a framework for edge computing vulnerabilities. Indeed, edge computing security threatens the sustainability functionality of urban infrastructure with various attacks, such as Man-in-the-Middle and denial of service. In particular, to tackle authentication and privacy violation problems, this work proposes a secure framework modeling in Petri Net, namely Secure Trust-Aware Philosopher Privacy and Authentication (STAPPA), on which a Distributed Authorization Algorithm is implemented. Moreover, a GARL approach is developed to optimize the network during learning, detect anomalies, and optimize routing. This work regards an environment characterized by discrete state and action spaces (*), and a stochastic transition model. The authors test the proposed approach on a synthetic dataset and assess the performance in anomaly detection and detection accuracy by using the popular detection accuracy, recall, precision, specificity, and F-measure. Ajao and Apeh (2023) deal with security challenges, in particular authentication and privacy violation problems.

Zhang et al. (2021b) propose an IoT-based Smart Green Energy (IoT-SGE) management system for improving the energy management of power grids allowed by DRL. The proposed approach is able to balance power availability and demand by keeping grid states steady, thus reducing power wastage. In developing IoT-SGE, it is considered an environment with discrete states (*), continuous actions (*), and a deterministic transition model (*). The proposed approach has been evaluated on a synthetic dataset by the use of operational logs, power wastage and requirement, and average failure ratio as metrics. The authors address an energy sustainability issue, in particular, they aim to manage energy requirements and allocate smart power systems.

In the context of smart ocean systems, Han et al. (2020) present an analytical model to evaluate the performance of an Internet of Underwater Things network with energy harvesting capabilities. The goal of this work is the maximization of IoT nodes throughput by optimally selecting the window size. To this aim, the authors propose an RL approach and leverage the Branch and Bound method to solve the optimization problem by autonomously adapting random access parameters through interaction with the network environment. Considering a realistic scenario, it is proposed a MARL approach to deal with the lack of network information. In this case, random access parameters autonomously adapt by using a distributed Multi-Armed Bandit (MAB)-based algorithm for each node. The environment considered in this work is characterized by deterministic actions (*) and a stochastic transition model. The authors test the proposed approach on a synthetic dataset, evaluating its performance in channel access regulation in relation to the number of ready nodes per time slot and throughput. Finally, this work addresses a fairness issue due to spatial uncertainty in underwater acoustic communication, to deal with which the authors formalize an optimization problem for maximizing the IoT network nodes throughput.

5.3.3 Water resources

Water resource management is a key aspect of sustainable development and usually does not include social aspects. Emamjomehzadeh et al. (2023) propose a novel urban water metabolism model that combines urban metabolism with the Water, Energy, and Food (WEF) (Radini et al. 2021) nexus and thus it can consider interconnections among water, energy, food, material, and GHG emissions. Moreover, this work proposes a physical-behavioral model that relates the proposed approach to a MARL agent-based model neither fully cooperative nor fully competitive developed using Q-Learning. In this case, the only technical information available concerns the use of a synthetic dataset. The proposed approach is evaluated in terms of water table level, nitrate density, energy usage, and GHG emissions. Considering water resource management challenges related to sustainability, the authors aim to model and manage the WEF nexus for Integrated Water Resource Management in an urban area, taking into account stakeholders' characteristics.

Skardi et al. (2020) propose, instead, an approach for quantifying and including social attachments in water and wastewater allocation tasks. This work proposes a paired physical-behavioral model, and the authors leverage Q-Learning to include social and behavioral aspects in the decision-making process. Specifically, it uses the approach proposed by Bazzan et al. (2011) to integrate Social Analysis in Q-Learning, and they choose between individual or social behavior through the use of specific reward functions. In developing the proposed method both a deterministic and a stochastic transition model is considered. Tests are performed on a dataset that combines real-world and synthetic data, and the performance evaluation is conducted considering water and treated wastewater allocation to the agents, water and groundwater level, and the concentration of nitrates to measure groundwater quality. Using Social Network Analysis, the authors tackle a key challenge in common resource management, i.e., the cooperation among agents. Also, they aim to quantify and include social attachments in water resource management.

5.3.4 Emissions/pollution

The development of WWTPs has a positive impact on environmental protection by reducing pollution but, at the same time, they consume resources and produce GHG emissions as well as residual sludge. With this in mind, Chen et al. (2021) propose an approach based on MADDPG to control Dissolved Oxygen (DO) and chemical dosage at once and improve sustainability accordingly. Specifically, the proposed approach uses two agents, one to control DO and one to control chemical dosage. Moreover, a reward function is designed based on life cycle cost and various Life Cycle Assessment mid-point indicators respectively. The proposed approach is developed considering an environment with continuous state and action spaces and tested on a synthetic dataset. To evaluate the training process, the reward and the Q-values determined by trained critic networks are used as metrics, while to analyze the variation of the influents and control parameters, the authors leverage the influents (COD, TN, TP, and NH_3-N), inflow rate, DO and dosage values. Finally, to assess the impact of the proposed approach are used energy consumption, cost, Eutrophication Potential (EP), and GHG emissions. WWTPs have a positive impact on environmental protection since they reduce contaminants and environmental pollution. However, at the same time, WWTPs consume resources and produce GHG emissions as well as residual sludge, thus the authors seek to optimize their impact on environmental sustainability.

Intelligent fleet management is crucial in mitigating direct GHG emissions in open-pit mining operations. In this context, Huo et al. (2023) propose a MARL-based dispatching system for reducing GHG emissions. To this aim, this work presents an environment for haulage simulation that integrates a component for real-time computing of GHG emissions. Then, Q-Learning is leveraged to improve fleet productivity and reduce trucks' emissions by decreasing their waiting time. In the development of the proposed approach, an environment characterized by discrete state (*) and action spaces is considered. Tests are performed on a synthetic dataset and productivity, number of operational mistakes, GHG emissions, and time spent in queue are used as evaluation metrics. In this work, the authors tackle operational randomness and uncertainties in fleet management for reducing haul trucks' GHG emissions in open-pit mining operations

5.3.5 Agriculture

In the context of sustainable agriculture, one of the key aspects of food security is crop yield prediction. Elavarasan and Durairaj Vincent (2020) tackle this problem by using a DRL approach, specifically a Deep Recurrent Q-Network (DRQN) (Hausknecht and Stone 2015) model. It consists of a Recurrent Neural Network (RNN) (Rumelhart et al. 1986) on top of the DQN. The proposed approach sequentially stacks the RNN layers, feeds the network with pre-trained parameters, and adds a linear layer to map the RNN output into Q-values. The Q-Learning network builds a crop yield prediction environment as a 'yield prediction game' that leverages both parametric feature combinations and thresholds useful in agricultural production. The authors consider an environment with discrete states (*) and test their approach on a dataset combining real-world and synthetic data, evaluating the performance by using the following metrics: Determination Coefficient (R2), Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Median Absolute Error (MedAE), Mean Squared Logarithmic Error (MSLE), Mean Absolute Percentage Error (MAPE), Probability Density function (PDF), Explained Variance Score, and accuracy. Finally, Elavarasan and Durairaj Vincent (2020) address issues related to the application of Deep Learning (DL) methods to crop yield prediction for increasing food production. Specifically, the authors tackle the incapability of DL approaches to directly map, linearly or non-linearly, raw data with crop yield values and the strong dependence of their effectiveness on the quality of features extracted from data.

5.3.6 Data, energy

Data centers are among the largest consumers of energy. In Shaw et al. (2022), an RL-based Virtual Machine (VM) consolidation algorithm named Advanced Reinforcement Learning Consolidation Agent (ARLCA). Its aim consists of simultaneously improving energy efficiency and delivery service guarantees. In this work, a global resource manager constantly monitors the state of the system and identifies hosts that may be overloaded due to the resource demand change over time. The proposed approach rebalances the VM distribution and avoids the rapid overloading of hosts while ensuring efficient operation. This work presents two implementations of ARLCA based on two RL methods, i.e., Q-Learning and SARSA, and it tests two different approaches to balance the exploration-exploration tradeoff, namely ϵ-greedy, and softmax. Finally, the authors leverage the Potential Based Reward Shaping (Ng et al. 1999) technique to include domain knowledge in the reward structure and speed up the learning process. ARLCA works in an environment with discrete state and action spaces (*) and a stochastic transition model. Its performance is evaluated on a synthetic dataset (real-world-based). To evaluate the proposed VM consolidation algorithms, energy consumption, Service Level Agreement Violations (SLAV), number of migrations, and Energy Service Level Agreement Violations (ESV) are used as performance measures. In this work, the authors tackle a key challenge for cloud computing services, namely energy awareness. Further, they also face the slow convergence to the optimal policy of conventional RL algorithms.

Renewable energy Aware Resource management (RARE), a DRL approach for job scheduling in a green data center, is presented in Venkataswamy et al. (2023). This work proposes a customized actor-critic method in which the authors use three Deep Neural Networks (DNNs): the encoder, the actor, and the critic. The encoder summarizes information about the state of the environment into a compact representation of it, used as input for both the action and the critic. The actor returns the probability of choosing each scheduling action, while the critic estimates, for each action, the total expected value achieved by starting in the current state and applying a specific action. Moreover, since DRL requires a significant amount of interactions with the environment to explore it and then to adapt a randomly initialized DNN policy, the authors leverage an offline learning algorithm, namely, Behavioral Cloning, to learn a policy based on existing heuristic policy data used as prior experience. In particular, the actor network is trained to imitate the action selection process of data within the replay memory. In developing RARE, it is considered an environment characterized by discrete states (*) and actions and tested the performance on both synthetic and real-world datasets by using the total job economic value as metrics. In this work, the authors tackle several challenges related to the application of RL techniques to the context of green datacenters. The first issue relates to the environment. The dynamic of green data center environments makes the scheduling process difficult as it has to consider and manage the intermittent and variable nature of renewable energy sources. Moreover, the lack of uniformity in the environments makes it challenging to compare different approaches. The second challenge highlights the absence of discussion regarding the systems design choices effect (e.g., the planning horizon size). Such lack does not help to clarify the reasons for the better performance of the RL scheduler over heuristic policies. Furthermore, the authors discuss employing RL schedulers as a black box, without considering different configurations, such as the size of the neural network, which can lead to improved performance. Finally, the last challenge highlights that existing RL schedulers do not focus on learning and improving available heuristic policies.

5.3.7 Urban traffic and transportation

In recent years, the traffic congestion level has increased significantly with a consequent negative impact on the environment. Ounoughi et al. (2022) present EcoLight, an approach for controlling traffic signals based on DRL, which aims to reduce noise pollution, CO_2 emissions, and fuel consumption. The proposed method combines the Sequence to Sequence Long Short Term Memory (SeqtoSeq-LSTM) prediction model with the DQN algorithm. SeqtoSeq-LSTM is used to forecast the traffic noise level that is part of the traffic information given as input to the DQN to determine the action to perform. EcoLight works in environments with discrete actions (*) and has been tested on a real-world dataset. The performance of EcoLight is evaluated by using the MSE, MAE, noise levels, CO_2 emission, and fuel consumption as metrics. In this work, the authors tackle the issue of developing a control method that considers not only mobility and current traffic conditions but also integrates sustainability and proactivity.

On the other hand, Alizadeh Shabestray and Abdulhai (2019) present Multimodal iNtelligent Deep (MiND), a DRL-based traffic signal controller that considers both regular vehicles and public transit and leverages sensors' information, like occupancy, position, and speed, to optimize the flow of people through an intersection by using DQN. In developing MiND, the authors regard an environment characterized by discrete states (*) and action and a stochastic transition method and test the proposed approach on a synthetic dataset. To assess the performance of the proposed approach the following measures are used: average intersection travel time, average in queue time, average network travel time, and weighted average intersection person travel time. In this work, the authors have to fulfill some important requirements to develop a real-time adaptive traffic signal controller. Indeed, the controller has to consider both regular vehicles and public transit traffic, and leverage sensors' data on vehicle speed, position, and occupancy, moreover, the decision-making process should be fast.

Aziz et al. (2018) present an RL-based approach to control traffic signals in connected vehicle environments for reducing travel delays and GHG emissions. The proposed method, the R-Markov Average Reward Technique (RMART), leverages congestion information sharing among neighbor signal controllers and a multi-reward structure that can dynamically adapt the reward function according to the level of congestion at intersections. The considered environment presents discrete state (*) and action spaces (*) and a stochastic (*) transition model. The authors test RMART on a synthetic dataset and to evaluate its performance they use as metrics the average delay, stopped delay, number of stops, and network-wide delay, while to assess the performance from a sustainability point of view they leverage GHG emissions, i.e., CO, CO_2, NOX, VOC, PM10. Finally, this work deals with the traffic signal control problem to reduce travel delays and GHG emissions by addressing the following issues: the sharing of congestion information among neighbor signal controllers and the dynamic adaptation of the reward function on the base of congestion level.

Reducing the number of drivers who commute in search of car parking in urban centers has a positive impact on environmental sustainability. In this context, Khalid et al. (2023) propose a Long-range Autonomous Valet Parking framework that optimizes the path planning of AVs to minimize distance while serving all users by picking them up and dropping them off at their required spots. The authors propose two learning-based solutions: Double-Layer Ant Colony Optimization (DL-ACO) and DQN-based algorithms. DL-ACO can be applied in new or unfamiliar environments, while DQN can be used in familiar environments to make efficient and fast decisions since it is pre-trainable. The DL-ACO approach determines the most efficient path between pairs of spots and subsequently establishes the optimal order in which users can be served. To deal with dynamic environments, it is proposed a DQN-based algorithm in which the agent learns to solve the task by interacting with the environment and using memory experience replay and the target network. The proposed techniques aim to improve the carpool and parking experience while reducing the congestion rate. In this work, the environment considered is characterized by discrete states (*) and actions (*), and a deterministic (*) transition model. The proposed approach is tested on a synthetic dataset and execution time, reward, path planned, and distance are used as performance measures. In this work, the authors deal with path planning problems in dynamic environments while ensuring the quality of experience for each user, optimizing the order of user pick-up and drop-off, and finally minimizing the overall distance.

5.3.8 Buildings

Buildings are interesting from a DR and Demand Side Management point of view. In this context, Kathirgamanathan et al. (2021) leverage a DRL algorithm, namely Soft Actor-Critic (SAC), intending to automatize energy management and harness energy flexibility by controlling the cooling set point in a commercial building environment. In developing the proposed approach, the authors regard an environment with continuous states, and they evaluate the performance on a dataset that combines real-world and synthetic data using as evaluation metrics the energy purchased, energy cost, discomfort, total reward, temperature evolution, and power demand. Kathirgamanathan et al. (2021) tackle the application of DRL methods to automatize DR without the need for a specific building model and their robustness to different operating environments and scalability. Moreover, the authors point out that the lack of well-established environments makes it challenging to compare RL algorithms over different buildings.

De Gracia et al. (2015) instead consider Thermal Energy Storage, and in particular latent heat, techniques to maximize energy savings by leveraging a Ventilated Double Skin Facade (VDSF) with Phase Change Material (PCM) used as a cold energy storage system. By using an RL approach, i.e., SARSA(λ), the authors control the VDSF to optimally schedule the solidification of PCM through mechanical ventilation during nighttime and the stored cold release into the indoor environment at peak demand time, considering weather and indoor conditions. The environment considered in this work presents discrete states (*), discrete actions (*), and a deterministic transition model. Moreover, the proposed approach is evaluated on a synthetic dataset considering electrical energy savings. This work aims to maximize energy savings by considering both the benefit of VDSF and the energy used in the solidification process. Therefore, it is crucial to determine the best time for the charging process to solidify the PCM and store coldness.

5.3.9 Manufacturing

Manufacturing industries are among the largest energy consumers, so it is crucial to develop approaches that make them more energy efficient. In this regard, Wang and Wang (2022) tackle the Energy-Aware Distributed Hybrid Flow-shop Scheduling Problem (EADHFSP). The goal of this work consists of simultaneously minimizing two conflicting objectives, makespan, and Total Energy Consumption. To this aim, the authors formulate a mixed-integer linear programming model of the EADHFSP and combine a Cooperative Memetic Algorithm with an RL-based agent to solve the problem. The authors combine two heuristics to initialize the population with various solutions and finally propose an improvement scheme in which solutions are refined by using the appropriate operator determined by a policy agent, while the solution selection is performed through the use of a decomposition strategy for balancing convergence and diversity. In this work, it is considered an environment characterized by discrete state (*) and action (*) spaces, a deterministic (*) transition model, and the performance of the presented approach is tested on a synthetic dataset considering the Overall Nondominated Vector Generation, C Metric, Hyper volume, and $D1_R$ as evaluation metrics. This work addresses the EADHFSP with the minimization of makespan and total energy consumption, a challenging problem due to the simultaneous optimization of two conflicting objectives.

Leng et al. (2021) focus on Printed Circuit Board (PCB) manufacturing and propose a Loosely-Coupled Deep Reinforcement Learning (LCDRL) model for energy-efficient order acceptance decisions. The authors leverage DL, specifically a Convolutional Neural Network (LeCun 1989), to obtain an accurate prediction of the production cost, makespan, and carbon consumption of each order by considering historical order labeled data. Then, the proposed approach combines the forecasted data with order features to decide whether to accept the order and determine the optimal acceptance sequence by using a reinforcement learning approach based on Q-Learning. The authors regard an environment with discrete actions (*) and a stochastic transition model, and they test the proposed method on a synthetic dataset (real-world-based). As performance measures, the metrics MSE, MSLE, RMSE, and R2 are used to evaluate the prediction accuracy of LCDRL, while the performance of the approach is assessed in terms of unit profit, total profit, and acceptance rate. This work tackles the problem of order acceptance in PCB manufacturing to achieve energy efficiency, reduce carbon emissions, and improve material usage. Two critical aspects of PCB manufacturing are demand uncertainty and order customization which can lead to different profits, energy consumption, and carbon emissions. These two factors have to be considered in production planning under production constraints.

5.3.10 Mobile and wireless communication

Sustainable energy infrastructures need high-quality communication systems to connect user facilities and power plants for providing information interaction. In this context, Liu et al. (2021) propose the use of a 6G network and Intelligent Reflective Surface (IRS) technology to create a wireless networking platform and suggest a DRL method to optimize the phase shift of IRS and therefore improve the communication quality. Combining the 6G Network with the IRS technology, the authors provide high-quality coverage while gaining energy-saving benefits. In particular, this work proposes the application of the Deep Deterministic Policy Gradient (DDPG) (Lillicrap et al. 2016) algorithm to configure the IRS phase shift for enhancing system coverage. The authors consider an environment characterized by continuous state and action spaces. The performance of the proposed approach is assessed on a synthetic dataset using two reflection units as metrics: the achievable rate to measure the service quality and the transmission power. Developing sustainable energy infrastructure is challenging from several points of view. Liu et al. (2021) tackle the need for an effective global covering communication system using the IRS technology, whose phase shift configuration is challenging itself.

In the context of two-tier urban Heterogeneous Networks (HetNets), Miozzo et al. (2015) model the Small Cell (SC) network as a decentralized multi-agent system. The authors' goal consists of improving system performance and self-sustainability of the SCs in terms of energy consumption. To this aim, they leverage the distributed Q-Learning algorithm so that every agent learns an appropriate Radio Resource Management (RRM) policy. Miozzo et al. (2015) is extended in Miozzo et al. (2017). Here, it is proposed to train offline the algorithm to compute Q-values with which initialize the Q-tables of the SCs that will be used in the online method. In both approaches, the environment presents discrete states (*) and actions (*) and the dataset used is synthetic. In both works, the authors evaluate the proposed approaches in terms of network performance by using the throughput gain and traffic drop rate and their energy performance in terms of energy efficiency and energy efficiency improvement. Moreover, Miozzo et al. (2015) analyze the behavior of the HetNet considering traffic demand, harvested energy, battery level, policy, and normalized load at the macro Base Station (BS). Also, the authors consider as performance metrics the total amount of energy the system spent, the average load, the average cell load for the macro BS battery outage, and Jain's fairness index to assess the Quality of Service (QoS) improvement. Finally, Miozzo et al. (2017) assess the computed policy by leveraging the switch-off rate as a performance measure and use the battery level to analyze the convergence of the online algorithm and evaluate the excess energy over storage capacity. Both works address the problem of introducing energy harvesting into the computation of sleeping strategies to achieve energy efficiency. This is challenging due to the irregular and intermittent nature of renewable energies.

Giri and Majumder (2022), instead, leverage a Deep Q-Learning algorithm for optimizing resource allocation in energy-harvested cognitive radio networks, where primary users networks share channel resources with secondary users and nodes can harvest energy from the environment, such as solar or wind. The proposed approach addresses the dynamic allocation of resources to achieve optimal network and throughput capacity, considering QoS, energy constraints, and interference limitations. Moreover, the authors utilize both linear and non-linear energy-harvested models, proposing a novel reward function that incorporates the non-linear one/model. The proposed approach works in environments characterized by continuous states (*), discrete actions (*), and a stochastic transition model (*) and it has been tested on a synthetic dataset by using reward, capacity, network lifetime, and average delay as performance measures. Giri and Majumder (2022) address the limitations of Q-Learning-based allocation methods, then allow dealing with high-dimensional problems, improving convergence performance, and efficiently harnessing the collected energy to meet the network's QoS requirements.

Internet traffic has increased in recent years, and in the development of next-generation networks, it is important to address the QoS issue sustainably. In this context, Al-Jawad et al. (2021) propose an RL-based algorithm to solve routing problems in a Software Defined Network (SDN) environment, named Reinforcement lEarning-based Dynamic rOuting (REDO). Indeed, the proposed approach leverages Q-Learning to handle traffic flows by determining the most appropriate routing strategy among a set of conventional routing algorithms with the aim to maximize flows meeting the Service Level Agreement as to throughput, packet loss, and rejection rate. In developing REDO, the authors consider an environment with discrete state (*) and action spaces (*). The performance of the proposed approach is evaluated on a synthetic dataset in terms of throughput, packet loss, rejected flows, PSNR, and Mean Opinion Score (MOS). In the development of next-generation networks like SDN, Al-Jawad et al. (2021) address the problem of providing QoS sustainably through the solution of a traffic flow routing problem.

5.3.11 Electric energy

One way to increase environmental sustainability is to improve the energy efficiency of smart hubs. To this aim, Sheikhi et al. (2016) present the new Smart Energy Hub framework for modeling distinct energy infrastructures under a single framework. The authors' goal consists of optimizing the electrical and natural gas consumption of a residential customer through the use of Q-Learning. Moreover, to improve and support information management among users and utility service providers, the proposed framework leverages Cloud Computing based systems. In this case the only technical information available concern the use of a stochastic transition model and a synthetic dataset. To evaluate the performance of the proposed approach, the metrics used are the storage charge level, the operational cost, and the primary energy involved. As regards dynamic load management in smart hubs, the authors tackle two issues. The first one is related to energy system parameters which are often assumed to be constant but can vary with time or be stochastic in practice. The second one is, instead, related to the conventional smart grid architecture, which has several reported issues, including exposure to cyber-attacks, single failure problems, limited memory and storage capacity in the energy management system, and difficulties in implementing real-time early warning systems due to limited energy and bandwidth resources.

In the context of smart energy networks, Harrold et al. (2022) consider a microgrid environment and leverage DRL to control a battery for energy arbitrage and increased use of renewable energies, namely solar and wind energy. Specifically, the authors apply the Rainbow Deep Q-Network (Hessel et al. 2018) algorithm and add predicted values for demand, Renewable Energy Source (RES), and energy price to agents' information by leveraging an Artificial Neural Network. In this work, the environment considered is characterized by continuous states (*) and discrete actions. The authors test the proposed approach on a dataset that considers both real-world and synthetic data and assess the prediction accuracy using the MAPE. Also, they evaluate the performance of the proposed approach through energy cost savings, relative savings, episodic rewards, and value distribution. This work tackles the problem of controlling an Energy Storage System in a microgrid with its demand, RES, and dynamic energy pricing to perform energy arbitrage and improve the use of RES leading to reduced energy cost. Finally, the authors point out the limited availability of data that requires an efficient algorithm training procedure.

A key aspect of sustainability and cost-effectiveness in grid operation is optimal energy dispatch. Jendoubi and Bouffard (2022) address a multi-dimensional power dispatch problem within a power system by leveraging MARL, specifically the MADDPG algorithm. The proposed control framework performs CTDE to improve the coordination among dispatchable units without communication needed, and thus it mitigates data privacy and communication issues. In developing the presented approach, it is considered an environment with continuous actions and a stochastic transition model (*). The dataset used to evaluate the performance is synthetic and the proposed method is evaluated in terms of the annual total cost, variation of daily operation cost, photovoltaics (PV) production, aggregated demand, amount of power to be charged/discharged, amount of power provided by a diesel generator, amount of electricity delivered by the electricity provider, the difference in the amount of electricity delivered by the electricity provider between two consecutive time steps and Peak-to-Average Ratio (PAR). The authors address the energy dispatch aspects related to the development of distributed energy resources control strategies in grid operation to simultaneously reduce costs and delays and allow local coordination among energy resources.

5.3.12 Energy

In recent years, international trade and container handling at port terminals have increased greatly. Improving sustainability in port operations closely relates to the energy consumption at Automated Container Terminals, where Automatic Stacking Cranes (ASCs) are used to load, unload, and pile containers. In this context, Gao et al. (2023) propose a digital twin-based approach for container yard management. Specifically, this work focuses on determining the optimal allocation of container tasks and scheduling of ASCs to reduce the energy consumption of ASCs while maintaining efficient loading and unloading operations. The proposed approach leverages a virtual container yard to simulate the operating plan and mixed integer programming model to optimize the scheduling problem taking into account the energy consumption. Finally, the authors use the Q-Learning algorithm to determine the optimal scheduling plan and minimize energy consumption. The environment considered in this work presents discrete actions (*) and a stochastic (*) transition model. The performance of the proposed approach is evaluated on a real-world dataset using working and non-working energy consumption and the ratio between them as metrics. To improve the sustainability of port operations, in this work the problem of optimizing container yard operations to minimize energy consumption is addressed. Indeed, several factors can introduce randomness and uncertainty into these operations, and incorrect distribution of tasks can lead to suboptimal utilization of ASCs.

5.3.13 Wireless sensor network

Regarding embedded systems powered by a renewable energy source like an Energy Harvesting Wireless Sensor Node (EHWSN), Hsu et al. (2014) present a method called Reinforcement Learning-based throughput on-demand provisioning dynamic power management (RLTDPM). By leveraging the Q-Learning algorithm, the proposed approach allows the EHWSN to adapt the operational duty cycle to satisfy both the energy neutrality condition and the throughput on-demand (ToD) requirement, ensuring perpetual operation. In developing RLTDPM, the authors regard an environment characterized by discrete state (*) and actions (*) and evaluate the performance on a synthetic dataset by considering the residual battery energy (RBE), exercised duty cycle (EDC), offset to the required ToD (OTRT), and ToD achievability. In this work, the authors address the problem of simultaneously achieving two mutually conflicting goals i.e., satisfying ToD and reducing power consumption.

Energy-Harvesting Wireless Sensor Networks (WSNs) are widely used in energy-constrained operation problems. In particular, Chen et al. (2016) focus on Solar-Powered Wireless Sensor Networks (SPWSNs) and present an RL-based Sleep Scheduling for Coverage algorithm to improve the sustainability of SPWSN's operations. The proposed approach leverages a precedence operator in the group formation algorithm to prioritize sensors in sparsely covered areas ensuring the desired coverage distribution. Then, the authors propose a multi-sensor cooperation Q-Learning group model to properly choose nodes' working modes by leveraging the developed learning and action selection strategies. The whole group learns the sleep schedule by changing the role of the active node. The environment considered in this work presents discrete state (*) and action (*) spaces and a stochastic transition model. The proposed approach is tested on a dataset that combines real-world and synthetic data, and its performance is evaluated in terms of energy balancing between group members by using the potential energy of nodes as metrics, network lifetime, area coverage ratio, number of residual alive nodes versus the network lifetime, and the recharging cycle. In this work, the authors tackle a sleep scheduling problem to simultaneously achieve the desired area coverage and energy balance between group nodes to extend the network lifetime.

On the other hand, Feng et al. (2023) propose an RL-based approach to maximize data throughput in self-sustainable WSNs. The authors consider a Mobile Sensor (MS) that collects and transmits data to a fixed sink while moving within the network and harvesting energy from the environment. By leveraging DDPG, the MS can determine the optimal trajectory to optimize the EH performance and data transmission dealing with unknown energy supply dynamics. The environment considered in this work is characterized by continuous states and actions, and a stochastic transition model (*). Moreover, the performance of the proposed approach is assessed on a synthetic dataset considering as evaluation metrics the distribution of the ratio between the expected per-slot harvested energy and the MS-to-sink distance within the network, moving trajectories of the MS, reward, actor-loss, battery level, accuracy, moving steps, convergence, and training time. The authors tackle two main challenges concerning the MS's trajectory optimization to maximize data throughput. The first relates to the lack of energy-related information such as the energy sources' placement, future energy harvesting potential, and statistical parameters like the average energy harvesting rate, which makes the problem challenging. The second consists of the tradeoff between energy harvesting and data transmission. Indeed, moving closer to energy sources allows the MS to increase the energy harvesting amount. However, this may lead to decreasing data transmission power due to a possible increase in the distance between the MS and the sink.

5.3.14 Autonomous vehicles

In the last decade, Unmanned Aerial Vehicles (UAVs), i.e., drones, have been used in various scenarios such as rapid disaster response, Search-And-Rescue, environmental monitoring, etc., where humans are unable to operate in a timely and efficient manner, for example, due to the presence of physical obstacles. Bouhamed et al. (2020) consider the application of UAVs as mobile data collection units in delay-tolerant WSNs. The authors propose a mechanism that exploits two RL techniques, namely DDPG and Q-Learning algorithms. The proposed approach uses DDPG to determine the best trajectory for a UAV to reach the target destination while avoiding obstacles in the environment. Q-Learning, on the other hand, is used to schedule the best order of visiting nodes to minimize the time needed to collect data within a predefined time limit. In this work, the environment presents continuous state (*) and action spaces for the DDPG-based part of the approach while discrete states (*) and actions (*) for the Q-Learning-based one. The proposed mechanism is tested on a synthetic dataset and to evaluate its obstacle avoidance and scheduling performance, the authors analyze the path followed by the UAV, the reward collected, the UAV's battery level, and the completion time of the tour against the ground unit transmission power.

This work addresses issues related to the limited battery capacity of UAVs and challenges related to navigating in obstacle-prone environments to enable communication between the UAV and low transmission power sensors.

Sacco et al. (2021) propose a MARL approach based on the actor-critic framework to tackle task offloading problems from UAV swarms in edge computing environments to simultaneously reduce task completion time and improve energy efficiency. The proposed approach determines a distributed decision strategy through the collaboration among the system's mobile nodes that share information about the overall system state. This information is then used by the agents to decide whether to compute a task locally or offload it to the edge cloud and in this case, the proposed technique chooses the best transmission technology among Wi-Fi access points and mobile network. In developing the proposed approach, the environment considered presents continuous states (*) and discrete actions (*), and the dataset used for testing combines real-world and synthetic data. The performance of the presented techniques is assessed in terms of task completion time and utility against a varying number of agents, and average node-antenna distance. Then, the authors evaluate the energy consumption necessary to complete the task by varying the average node-antenna distance and computing workload and assess the task completion time against the average computing workload. In addition, the cumulative distribution function (CDF) and utility evolution through episodes are considered to analyze the variability of performance among nodes and convergence performance, respectively. Finally, in this work, the authors tackle the problem of reducing the time necessary for task completion of UAV swarms by leveraging task offloading to the edge cloud while improving energy efficiency.

In the context of autonomous driving, Gu et al. (2023) tackle the application of RL methods focusing on energy-saving and environmentally friendly driving strategies within a cooperative adaptive cruise control platoon. More precisely, the goal of this work consists of training platoon member vehicles to react effectively when the leading vehicle faces a severe collision. The authors leverage the Policy Gradient algorithm for training an RL agent to minimize the energy consumption in inter-vehicle communication for decision-making while avoiding collisions or minimizing the resulting damage. To this aim, two different loss functions are used, i.e., collision loss and energy loss. Moreover, utilizing a specific reward function can both ensure the vehicle's safety and consider the fuel consumption resulting from the action performed by the vehicle. This work considers an environment characterized by continuous states (*), discrete actions, and a stochastic transition model (*). The proposed approach has been tested on a synthetic dataset using energy loss, collision loss, reward, and lane changes as metrics. A key challenge of green autonomous driving addressed in this work is the development of effective strategies that can respond to environmental observations by automatically generating appropriate control signals.

6 Discussion

The analysis of the literature performed in this work shows that most of the works about RL for environmental sustainability concern the energy application domain, followed by urban traffic and transportation. The main RL technique used in the reviewed manuscripts is Q-Learning. Concerning the 35 selected articles, we observe that energy-related issues involve most of the papers, and about half of them leverage DRL approaches, such as DQN and DDQN. In developing the proposed methods, the authors mainly consider domains with discrete state and action spaces, and stochastic transition models, using synthetic datasets to evaluate the performance.

Problems related to environmental sustainability were traditionally tackled with optimization techniques in which the concept of adaptability has to be introduced explicitly. In contrast, one of the strengths of RL is its natural way of dealing with adaptability to changing or different environments, a crucial feature in environmental sustainability problems since in this context the agent has to handle variations in operating conditions due to, for example, changes in resource availability or weather conditions. For instance, Chen et al. (2016) introduce a RL-based Sleep Scheduling for Coverage (RLSSC) approach to ensure sustainable time-slotted operations in solar-powered wireless sensor networks. This algorithm is compared to LEACH (Heinzelman et al. 2002), a high-energy-efficient hierarchical routing protocol, wherein the node chosen to be active in the current round is ineligible for selection in the subsequent round, and a random algorithm that randomly determines active nodes within a group. Among the various aspects considered, a crucial criterion for evaluating algorithm effectiveness lies in maintaining equilibrium in energy levels, as significant disparities in current residual energy arise when a node receives an energy supplement. RLSSC initially exhibits fluctuations but eventually converges through iterative learning, exhibiting slight oscillations up and down in response to varying solar strength throughout the day. Moreover, the proposed approach demonstrates real-time energy balancing among sensor nodes. In contrast, non-RL-based methods lack the capacity to adapt to the dynamic environment. Another aspect to consider is network lifetime, where RLSSC excels in adapting to uncertainties associated with harvesting time and the amount of acquired energy. This adaptability enables RLSSC to dynamically adjust its scheme in real-time, effectively extending the overall network lifetime. This is only one of several examples showing that RL can provide a strong advantage in solving problems related to environmental sustainability because of its natural capability to deal with uncertainty and adaptation in sequential decision-making.

However, we identify several open problems in the application of RL techniques to environmental sustainability. These concern scalability, data efficiency, and the necessity to deal with large data volumes, often posing cost challenges. In future developments, it is crucial to improve pre-training methods that allow the generation of initial policies by simulation and leverage knowledge acquired by solving a related task. RL methods are also sensitive to reward function therefore reward engineering is important to avoid a negative impact on performance. Moreover, in dealing with environmental sustainability problems in specific contexts like IoT, it is of particular importance to consider the presence of computational limitations and then optimize the computational complexity of the method. Finally, we note that most of the approaches involve single-agent systems. Extending the proposed approaches to the multi-agent context would allow the cooperative computation of optimal policies accounting for common performance objectives to improve shared resources management and environmental sustainability.

7 Conclusions

This review focuses on the application of RL techniques to address environmental sustainability challenges, a topic of increasing interest in the international scientific community. We have examined several contexts where RL techniques have been recently used to enhance environmental sustainability, offering practitioners insights into state-of-the-art methodologies across diverse application domains. RL has found practical application in environmental sustainability because the inherent uncertainty of this domain poses challenges to strategy learning and adaptation that can be naturally tackled by RL. The review of the literature performed in this survey has identified the most common applications of RL in environmental sustainability and the most popular methods used to address these challenges in the last two decades. We have first provided a quantitative analysis of the state-of-the-art related to the application of RL in environmental sustainability and then analyzed the use of these techniques, focusing on sustainability concerns. In particular, we have provided an overview of the application domains of the proposed RL techniques and the approaches used for environmental sustainability issues. Moreover, we have narrowed our attention to 35 selected papers and provided technical information on the formalization of the RL problem, the performance measures adopted for evaluation, and the challenges addressed.

Keywords mapping into macro areas

See Tables 4, 5, 6, 7 , 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20.

Table 4 Keywords grouped in the "Energy" macro area

Energy		
Energy	Energy aware	Energy conservation
Energy consumption	Energy distributions	Energy efficiency
Energy harvesting	Energy management	Energy management systems
Energy resource	Energy source	Energy storage
Energy sustainability	Energy systems	Energy utilization
Distributed energies	Dynamic energy	Energy allocations
Energy arbitrages	Energy availability	Energy consumption balances
Energy flexibility	Energy infrastructures	Energy management strategy
Energy market	Energy neutrality	Energy optimal scheduling
Energy routing	Energy savings	Energy storage system
Energy theft	Energy transfer	Energy usage
Energy use	Energy-awareness	Energy-constrained networks
Energy-saving strategies	Harvesting energies	Intelligent energy management
Non-renewable energy	Total energy consumption	Wireless energy transfers
Energy trading		

Table 5 Keywords grouped in the "Electric energy" macro area

Electric energy	
Electric energy storage	Electric load dispatching
Electric power transmission networks	Electric power utilization
Dynamic energy managements	Dynamic loads
Dynamic power management	Electric load management
Electric loads	Electric power system control
Electric power transmission	Electrical networks
Electricity demands	Electricity loss
Micro grid	Smart grid
Smart power grids	Electricity grids
Grid resilience	Power grids
Electricity grids	Smart grid communications
Grid resilience	Distributed power generation
Power management	Power supply
Low power electronics	Dynamic power management
Electric power system control	Electric power transmission
inductive power transmission	Low power
Low power and lossy network (lln)	Low power networks
Low-power consumption	Low-power devices
Power	Power allocations
Power dispatch	Power grids
Power harvesting	Power limitations
Power system simulator for engineering	Power system simulators
Power traces	Power control
Power transmission systems	Reactive power
Reactive power output	Power management (telecommunication)

Table 6 Keywords grouped in the "Urban Traffic and Transportation" macro area

Urban traffic and transportation		
street traffic control	Sustainable mobility	Traffic congestion
Traffic emission	Traffic flow	Traffic light control
Traffic management	Traffic signal control	Traffic signals
Adaptive traffic signal control	Intelligent traffic controls	Optimal traffic control
Traffic	Traffic conditions	Traffic environment
Traffic light	Traffic management strategies	Traffic scheduling
Urban traffic	Highway administration	Motor transportation
Transportation	Transportation system	Urban transportation
Bus bunching	Bus transportation	Sustainable transportation
Intelligent transportation	Transport systems	Transportation network
Transportation planning		

Table 7 Keywords grouped in the "Water resources" macro area

Water resources		
Aquifer	Ground water	Wastewater
Wastewater treatment	Water management	Water quality
Water resources	Water resources management	Water supply
Water treatment	Surface water	Sustainable wastewater treatments
Waste water management	Waste water recycling	Wastewater treatment plant
Water metabolism	Water purification	Water resources systems
Water treatment plants	Groundwater resources	Water level
Water quantity	Watersheds	

Table 8 Keywords grouped in the "Renewable/sustainable energies" macro area

Renewable/sustainable energies		
Renewable energies	Renewable energy resources	Renewable energy source
Renewable resource	Renewables	Alternative energy
Solar energy	Green energy	Smart renewable energy
Use of renewable energies	Sustainable energy	Solar power generation
Renewable power generation	Tidal power	Wind power
Hydroelectric power plants	Hydropower	

Table 9 Keywords grouped in the “Emissions/Pollution” macro area

Emissions/Pollution		
Carbon emission	Carbon footprint	Emission control
Gas emissions	Greenhouse gas	Greenhouse gas emissions
Acoustic noise	Air pollution monitoring	Atmospheric pollution
Carbon abatement strategy	Carbon sequestration	Co2 emissions
Greenhouse emissions	Greenhouse gas emission reduction	Groundwater pollution
Noise pollution	Pollution control	Vehicular emission
Water pollution	Air quality	

Table 10 Keywords grouped in the “Mobile and Wireless communication” macro area

Mobile and wireless communication	
5G mobile communication systems	Mobile telecommunication systems
6g mobile communication	Mobile communications
Base stations	Small cells
Wireless communication links	Wireless communications
Wireless telecommunication systems	Sustainable wireless communication network
Wireless communications networks	Communication network
Heterogeneous networks	Wireless powered communication network

Table 11 Keywords grouped in the “Data” macro area

Data		
Data acquisition	Data handling	Data transfer
Digital storage	Big data	Data aggregation
Data aggregation and fusion	Data analytics	Data distribution
Data logger	Data mining	Data sensing
Data-driven approach	Data-driven design	Database technology
Datacenter	Distributed database	Next generation data centers
Data transfer	Data-communication	Real-time data

Table 12 Keywords grouped in the “Wireless sensor network” macro area

Wireless sensor network	
Wireless sensor network	Wireless sensor node
Heterogeneous wireless sensor networks	Solar-powered wireless sensor networks
Rechargeable sensor networks	Wireless smart sensors
Sensor nodes	Integrating sensors
Sensor networks	Wireless smart sensors
Sensor networks	Smart sensors
Adaptive sensor selections	Mobile sensors
Sensor	Sensor payloads
Sleep scheduling	

Table 13 Keywords grouped in the “Autonomous vehicles” macro area

Autonomous vehicles	
Autonomous driving	Autonomous navigation
Autonomous vehicles	Unmanned aerial vehicles (uav)
Autonomous unmanned aerial vehicles	Uav networks
Unmanned aerial vehicle	Uav positioning
Autonomous vehicle control	Autonomous ship
Auto-navigation	Automated vehicles

Table 14 Keywords grouped in the “Batteries” macro area

Batteries		
Battery energy storage systems	Battery management systems	Battery storage
Charging (batteries)	Secondary batteries	Battery operation
Electric batteries	Battery capacity	Lead acid batteries
Lithium batteries	Residual battery	

Table 15 Keywords grouped in the “Manufacturing” macro area

Manufacturing	
Manufacture	Manufacturing
Production control	Supply chains
Sustainable manufacturing	Distributed manufacturing systems
Manufacturing environments	Large scale manufacturing systems
Manufacturing industries	Manufacturing sector
Printed circuit board manufacturing	Process manufacturing
Sustainable manufacturing engineering and resource-efficient production	

Table 16 Keywords grouped in the "Electric vehicles" macro area

Electric vehicles	
charging station	Electric vehicle charging
Electric vehicle charging station	Electric vehicles
Electric vehicle charging station recommendation	E mobilities
Fuel cell hybrid electric vehicles	Vehicle-to-grid

Table 17 Keywords grouped in the "IoT" macro area

IoT		
Internet of things	Internet of underwater things	Green internet of thing
Industrial internet of thing		

Table 18 Keywords grouped in the "Agriculture" macro area

Agriculture		
Agricultural robots	Agriculture	Crops
Agricultural productions	Agricultural products	Crop productivity
Smart agricultures	Sustainable agricultural	Sustainable agricultural system
Sustainable agriculture		

Table 19 Keywords grouped in the "Vehicles" macro area

Vehicles				
Intelligent vehicle highway systems	Vehicles	Transit vehicles	Vehicle dispatch	Vehicle fleets

Table 20 Keywords grouped in the "Buildings" macro area

Buildings		
Building	Building coverage ratios	Building energy
Building energy flexibility	Building energy management systems	Building energy managements
Building stocks	College buildings	Intelligent buildings
Office buildings		

Database search results

In the following section, we report the results of the search performed on the databases used, namely Scopus and Web of Science. In the following table, for each paper, we indicate the authors, title, source title, and publication year in addition to the database in which a corresponding record is present. More specifically, in the "Database" column, we use the letters "S" and "W" to indicate the presence of the paper on the Scopus and W databases, respectively, while "S, W" indicates the paper's presence on both databases.

Authors	Title	Source title	Year	Database
Sultanuddin S.J., Vibin R., Rajesh Kumar A., Behera N.R., Pasha M.J., Baseer K.K.	Development of improved reinforcement learning smart charging strategy for electric vehicle fleet	Journal of Energy Storage	2023	S, W
Ajao L.A., Apeh S.T.	Secure edge computing vulnerabilities in smart cities sustainability using petri net and genetic algorithm-based reinforcement learning	Intelligent Systems with Applications	2023	S
Wang J., Sun L.	Robust dynamic bus control: a distributional multi-agent reinforcement learning approach	IEEE Transactions on Intelligent Transportation Systems	2023	S, W
Szoke L., Aradi S., Bécsi T.	Traffic signal control with successor feature-based deep reinforcement learning agent	Electronics (Switzerland)	2023	S, W
Ali M.Y., Alsaeedi A., Shah S.A.A., Yafooz W.M.S., Malik A.W.	Energy efficient data dissemination for large-scale smart farming using reinforcement learning	Electronics (Switzerland)	2023	S, W
Yao R., Hu Y., Varga L.	Applications of agent-based methods in multi-energy systems-a systematic literature review	Energies	2023	S, W
Kazemeini A., Swei O.	Identifying environmentally sustainable pavement management strategies via deep reinforcement learning	Journal of Cleaner Production	2023	S, W
Emamjomehzadeh O., Kerachian R., Emami-Skardi M.J., Momeni M.	Combining urban metabolism and reinforcement learning concepts for sustainable water resources management: A nexus approach	Journal of Environmental Management	2023	S
Charef N., Ben Mnaouer A., Aloqaily M., Bouachir O., Guizani M.	Artificial intelligence implication on energy sustainability in Internet of Things: A survey	Information Processing and Management	2023	S, W
Savazzi S., Rampa V., Kianoush S., Bennis M.	An energy and carbon footprint analysis of distributed and federated learning	IEEE Transactions on Green Communications and Networking	2023	S, W
Naseer F., Khan M.N., Altalbe A.	Telepresence robot with DRL assisted delay compensation in iot-enabled sustainable healthcare environment	Sustainability (Switzerland)	2023	S
Kolat M., Kővári B., Bécsi T., Aradi S.	Multi-Agent reinforcement learning for traffic signal control: a cooperative approach	Sustainability (Switzerland)	2023	S, W

Authors	Title	Source title	Year	Database
Sivamayil K., Rajasekar E., Aljafari B., Nikolovski S., Vairavasundaram S., Vairavasundaram I.	A systematic study on reinforcement learning based applications	Energies	2023	S, W
Khalid M., Wang L., Wang K., Aslam N., Pan C., Cao Y.	Deep reinforcement learning-based long-range autonomous valet parking for smart cities	Sustainable Cities and Society	2023	S
Gao Y., Chang D., Chen C.-H.	A digital twin-based approach for optimizing operation energy consumption at automated container terminals	Journal of Cleaner Production	2023	S
Badakhshan S., Jacob R.A., Li B., Zhang J.	Reinforcement learning for intentional islanding in resilient power transmission systems	2023 IEEE Texas Power and Energy Conference, TPEC 2023	2023	S
Zhang W., Valencia A., Chang N.	Fingerprint networked reinforcement learning via multiagent modeling for improving decision making in an urban food-energy-water nexus	IEEE Transactions on Systems, Man, and Cybernetics: Systems	2023	S, W
Li C., Bai L., Yao L., Waller S.T., Liu W.	A bibliometric analysis and review on reinforcement learning for transportation applications	Transportmetrica B	2023	S, W
Venkataswamy V., Grigsby J., Grimshaw A., Qi Y.	RARE: renewable energy aware resource management in datacenters	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	2023	S, W
No author name available [Conference Review]	9th International Conference on Sustainable Design and Manufacturing, SDM 2022	Smart Innovation, Systems and Technologies	2023	S
Koch L., Picerno M., Badalian K., Lee S.-Y., Andert J.	Automated function development for emission control with deep reinforcement learning	Engineering Applications of Artificial Intelligence	2023	S, W
Huo D., Sari Y.A., Kealey R., Zhang Q.	Reinforcement learning-based fleet dispatching for greenhouse gas emission reduction in open-pit mining operations	Resources, Conservation and Recycling	2023	S
Feng Y., Zhang X., Jia R., Lin F., Lu J., Zheng Z., Li M.	Intelligent trajectory design for mobile energy harvesting and data transmission	IEEE Internet of Things Journal	2023	S, W
Chen M., Li Y., Zhang X., Liao R., Wang C., Bi X.	Optimization of river environmental management based on reinforcement learning algorithm: a case study of the Yellow River in China	Environmental Science and Pollution Research	2023	S, W

Authors	Title	Source title	Year	Database
Gu Z., Liu Z., Wang Q., Mao Q., Shuai Z., Ma Z.	Reinforcement learning-based approach for minimizing energy loss of driving platoon decisions	Sensors	2023	W
Baba-Nalikant M., Syed-Mohamad S. M., Husin M. H., Abdullah N. A., Saleh M. S. M., Rahim A. A.	A zero-waste campus framework: perceptions and practices of university campus community in Malaysia	Recycling	2023	W
Daradkeh M.	Lurkers versus contributors: an empirical investigation of knowledge contribution behavior in open innovation communities	Journal of Open Innovation: Technology, Market, and Complexity	2022	S
Jendoubi I., Bouffard F.	Data-driven sustainable distributed energy resources' control based on multi-agent deep reinforcement learning	Sustainable Energy, Grids and Networks	2022	S
Tomin N., Shakirov V., Kurbatsky V., Muzychuk R., Popova E., Sidorov D., Kozlov A., Yang D.	A multi-criteria approach to designing and managing a renewable energy community	Renewable Energy	2022	S, W
Adetunji K.E., Hofsajer I.W., Abu-Mahfouz A.M., Cheng L.	A novel dynamic planning mechanism for allocating electric vehicle charging stations considering distributed generation and electronic units	Energy Reports	2022	S
Li R., Zhang X., Jiang L., Yang Z., Guo W.	An adaptive heuristic algorithm based on reinforcement learning for ship scheduling optimization problem	Ocean and Coastal Management	2022	S, W
Zhang W., Xie M., Scott C., Pan C.	Sparsity-aware intelligent spatiotemporal data sensing for energy harvesting IoT system	IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems	2022	S, W
Yao L., Leng Z., Jiang J., Ni F.	Large-scale maintenance and rehabilitation optimization for multi-lane highway asphalt pavement: a reinforcement learning approach	IEEE Transactions on Intelligent Transportation Systems	2022	S, W
Adetunji K.E., Hofsajer I.W., Abu-Mahfouz A.M., Cheng L.	An optimization planning framework for allocating multiple distributed energy resources and electric vehicle charging stations in distribution networks	Applied Energy	2022	S, W

Authors	Title	Source title	Year	Database
Mahmud S., Abbasi A., Chakrabortty R.K., Ryan M.J.	A self-adaptive hyper-heuristic based multi-objective optimisation approach for integrated supply chain scheduling problems	Knowledge-Based Systems	2022	S, W
Musaddiq A., Ali R., Kim S.W., Kim D.-S.	Learning-based resource management for low-power and lossy IoT networks	IEEE Internet of Things Journal	2022	S
Giri M.K., Majumder S.	Deep Q-learning based optimal resource allocation method for energy harvested cognitive radio networks	Physical Communication	2022	S
Selukar M., Jain P., Kumar T.	Inventory control of multiple perishable goods using deep reinforcement learning for sustainable environment	Sustainable Energy Technologies and Assessments	2022	S, W
No author name available [Conference Review]	IFAC Workshop on Control for Smart Cities, CSC 2022 - Proceedings	IFAC-PapersOnLine	2022	S
Alibabaei K., Gaspar P.D., Assunção E., Alirezazadeh S., Lima T.M., Soares V.N.G.J., Caldeira J.M.L.P.	Comparison of on-policy deep reinforcementlearning A2C with off-policy DQN in irrigation optimization: a case study at a site in portugal	Computers	2022	S, W
Raza A., Shah M.A., Khattak H.A., Maple C., Al-Turjman F., Rauf H.T.	Collaborative multi-agents in dynamic industrial internet of things using deep reinforcement learning	Environment, Development and Sustainability	2022	S, W
Shaw R., Howley E., Barrett E.	Applying reinforcement learning towards automating energy efficient virtual machine consolidation in cloud data centers	Information Systems	2022	S, W
Jurj S.L., Werner T., Grundt D., Hagemann W., Möhlmann E.	Towards safe and sustainable autonomous vehicles using environmentally-friendly criticality metrics	Sustainability (Switzerland)	2022	S
Oubbati O.S., Atiquzzaman M., Lim H., Rachedi A., Lakas A.	Synchronizing UAV teams for timely data collection and energy transfer by deep reinforcement learning	IEEE Transactions on Vehicular Technology	2022	S, W
Xu G., Guo F.	Sustainability-oriented maintenance management of highway bridge networks based on Q-learning	Sustainable Cities and Society	2022	S

Authors	Title	Source title	Year	Database
Wang J.-J., Wang L.	A cooperative memetic algorithm with learning-based agent for energy-aware distributed hybrid flow-shop scheduling	IEEE Transactions on Evolutionary Computation	2022	S, W
Zhang T., Gou Y., Liu J., Yang T., Cui J.-H.	UDARMF: An underwater distributed and adaptive resource management framework	IEEE Internet of Things Journal	2022	S, W
Zhang M., Lu Y., Hu Y., Amaitik N., Xu Y.	Dynamic scheduling method for job-shop manufacturing systems by deep reinforcement learning with proximal policy optimization	Sustainability (Switzerland)	2022	S, W
Ma Y., Kassler A., Ahmed B.S., Krakhmalev P., Thore A., Toyser A., Lindbäck H.	Using deep reinforcement learning for zero defect smart forging	Advances in Transdisciplinary Engineering	2022	S
Jang J., Yang H.J.	Deep learning-aided user association and power control with renewable energy sources	IEEE Transactions on Communications	2022	S, W
Danassis P., Erden Z.D., Faltings B.	Exploiting environmental signals to enable policy correlation in large-scale decentralized systems	Autonomous Agents and Multi-Agent Systems	2022	S
Manchella K., Haliem M., Aggarwal V., Bhargava B.	PassGoodPool: joint passengers and goods fleet management with reinforcement learning aided pricing, matching, and route planning	IEEE Transactions on Intelligent Transportation Systems	2022	S, W
Yk S., Wu J., Song S.	Research on autonomous driving decision based on improved deep deterministic policy algorithm	SAE Technical Papers	2022	S
Neumann M., Palkovits D.S.	Reinforcement learning approaches for the optimization of the partial oxidation reaction of methane	Industrial and Engineering Chemistry Research	2022	S, W
Luo M., Du B., Klemmer K., Zhu H., Wen H.	Deployment optimization for shared e-mobility systems with multi-agent deep neural search	IEEE Transactions on Intelligent Transportation Systems	2022	S, W
Dey S., Saha S., Singh A.K., McDonald-Maier K.	SmartNoshWaste: using blockchain, machine learning, cloud computing and QR code to reduce food waste in decentralized web 3.0 enabled smart cities	Smart Cities	2022	S

Authors	Title	Source title	Year	Database
Kim J., Park J., Cho K.	Continuous autonomous ship learning framework for human policies on simulation	Applied Sciences (Switzerland)	2022	S, W
Tuli S., Gill S.S., Xu M., Garraghan P., Bahsoon R., Dustdar S., Sakellariou R., Rana O., Buyya R., Casale G., Jennings N.R.	HUNTER: AI based holistic resource management for sustainable cloud computing	Journal of Systems and Software	2022	S, W
Wang T., Liu L., Ding T.	Optimized sustainable strategy in aerial terrestrial IoT network	Proceedings - 2022 18th International Conference on Mobility, Sensing and Networking, MSN 2022	2022	S, W
Hoover W., Guerra-Zubiaga D.A., Banta J., Wandene K., Key K., Gonzalez-Badillo G.	Industry 4.0 trends in intelligent manufacturing automation exploring machine learning	ASME International Mechanical Engineering Congress and Exposition, Proceedings (IMECE)	2022	S
No author name available [Conference Review]	Proceedings - 22nd IEEE International Conference on Data Mining Workshops, ICDMW 2022	IEEE International Conference on Data Mining Workshops, ICDMW	2022	S
Heo S., Mayer P., Magno M.	Predictive energy-aware adaptive sampling with deep reinforcement learning	ICECS 2022 - 29th IEEE International Conference on Electronics, Circuits and Systems, Proceedings	2022	S, W
No author name available [Conference Review]	20th international conference on service-oriented computing, ICSOC 2022	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	2022	S
Rampini L., Re Cecconi F.	Artificial intelligence in construction asset management: a review of present status, challenges and future opportunities	Journal of Information Technology in Construction	2022	S, W
Wang K., Yang R., Liu C., Samarasinghalage T., Zang Y.	Extracting electricity patterns from high-dimensional data: a comparison of K-Means and DBSCAN algorithms	IOP Conference Series: Earth and Environmental Science	2022	S
No author name available [Conference Review]	Proceedings - 2022 IEEE international conference on autonomic computing and self-organizing systems companion, ACSOS-C 2022	Proceedings - 2022 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion, ACSOS-C 2022	2022	S
Dusparic I.	Reinforcement learning for sustainability: adapting in large-scale heterogeneous dynamic environments	Proceedings - 2022 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion, ACSOS-C 2022	2022	S, W

Authors	Title	Source title	Year	Database
Amadi K.W., Iyalla I., Radhakrishna P., Al Saba M.T., Waly M.M.	Continuous dynamic drill-off test whilst drilling using reinforcement learning in autonomous rotary drilling system	Society of Petroleum Engineers - ADIPEC 2022	2022	S
No author name available [Conference Review]	Proceedings - 2022 IEEE 5th international conference on artificial intelligence and knowledge engineering, AIKE 2022	Proceedings - 2022 IEEE 5th International Conference on Artificial Intelligence and Knowledge Engineering, AIKE 2022	2022	S
Wu L., Guo S., Liu Y., Hong Z., Zhan Y., Xu W.	Sustainable federated learning with long-term online VCG Auction Mechanism	Proceedings - International Conference on Distributed Computing Systems	2022	S, W
Baumgart U., Burger M.	Optimal control of traffic flow based on reinforcement learning	Communications in Computer and Information Science	2022	S
Sabet S., Farooq B.	Green vehicle routing problem: state of the art and future directions	IEEE Access	2022	S, W
No author name available [Conference Review]	Proceedings of International Conference on Computing, Communication, Security and Intelligent Systems, IC3SIS 2022	Proceedings of International Conference on Computing, Communication, Security and Intelligent Systems, IC3SIS 2022	2022	S
Andersen P.-A., Goodwin M., Granmo O.-C.	CaiRL: a high-performance reinforcement learning environment toolkit	IEEE Conference on Computatonal Intelligence and Games, CIG	2022	S
Korecki M., Helbing D.	Analytically guided reinforcement learning for green it and fluent traffic	IEEE Access	2022	S
No author name available [Conference Review]	3rd international conference on resources and environmental research, ICRER 2021	IOP Conference Series: Earth and Environmental Science	2022	S
Ounoughi C., Touibi G., Yahia S.B.	EcoLight: eco-friendly traffic signal control driven by urban noise prediction	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	2022	S, W
Tang Y., Deng X., Yi L., Xia Y., Yang L.T., Tang A.X.	Collaborative intelligent confident information coverage node sleep scheduling for 6G-empowered green IoT	IEEE Transactions on Green Communications and Networking	2022	S, W
Paul S., Chowdhury S.	A graph-based reinforcement learning framework for urban air mobility fleet scheduling	AIAA Aviation 2022 Forum	2022	S
No author name available [Conference Review]	7th international scientific-technical conference, MANUFACTURING 2022	Lecture Notes in Mechanical Engineering	2022	S

Authors	Title	Source title	Year	Database
No author name available [Conference Review]	2nd conference of innovative product design and intelligent manufacturing system, IPDIMS 2020	Lecture Notes in Mechanical Engineering	2022	S
No author name available [Conference Review]	7th international scientific-technical conference, MANUFACTURING 2022	Lecture Notes in Mechanical Engineering	2022	S
Isufaj R., Sebastia D.A., Piera M.A.	Toward conflict resolution with deep multi-agent reinforcement learning	Journal of Air Transportation	2022	S
Yuan M., Pun M., Wang D.	Rényi state entropy maximization for exploration acceleration in reinforcement learning	IEEE Transactions on Artificial Intelligence	2022	S
Eriksson K., Ramasamy S., Zhang X., Wang Z., Danielsson F.	Conceptual framework of scheduling applying discrete event simulation as an environment for deep reinforcement learning	Procedia CIRP	2022	S
Zhang W., Liu H., Xiong H., Xu T., Wang F., Xin H., Wu H.	RLCharge: imitative multi-agent spatiotemporal reinforcement learning for electric vehicle charging station recommendation	IEEE Transactions on Knowledge and Data Engineering	2022	S, W
Zhang W., Zhang J., Xie M., Liu T., Wang W., Pan C.	M2M-routing: environmental adaptive multi-agent reinforcement learning based multi-hop routing policy for self-powered IoT systems	Proceedings of the 2022 Design, Automation and Test in Europe Conference and Exhibition, DATE 2022	2022	S, W
No author name available [Conference Review]	7th international scientific-technical conference, MANUFACTURING 2022	Lecture Notes in Mechanical Engineering	2022	S
No author name available [Conference Review]	7th international scientific-technical conference, MANUFACTURING 2022	Lecture Notes in Mechanical Engineering	2022	S
Mao Z., Fang Z., Li M., Fan Y.	EvadeRL: evading PDF malware classifiers with deep reinforcement learning	Security and Communication Networks	2022	S
Lin B., Duan J., Han M., Cai L.X.	Decentralized reinforcement learning-based access control for energy sustainable underwater acoustic sub-network of MWCN	Wireless Networks (United Kingdom)	2022	S
No author name available [Conference Review]	7th international scientific-technical conference, MANUFACTURING 2022	Lecture Notes in Mechanical Engineering	2022	S

Authors	Title	Source title	Year	Database
Maree C., Omlin C.W.	Balancing profit, risk, and sustainability for portfolio management	2022 IEEE Symposium on Computational Intelligence for Financial Engineering and Economics, CIFEr 2022 - Proceedings	2022	S, W
Lee J., Sun Y.G., Sim I., Kim S.H., Kim D.I., Kim J.Y.	Non-technical loss detection using deep reinforcement learning for feature cost efficiency and imbalanced dataset	IEEE Access	2022	S, W
Liu Y., Yang M., Guo Z.	Reinforcement learning based optimal decision making towards product lifecycle sustainability	International Journal of Computer Integrated Manufacturing	2022	S, W
Alanne K., Sierla S.	An overview of machine learning applications for smart buildings	Sustainable Cities and Society	2022	S
Samuel O., Javaid N., Alghamdi T.A., Kumar N.	Towards sustainable smart cities: a secure and scalable trading system for residential homes using blockchain and artificial intelligence	Sustainable Cities and Society	2022	S, W
Harrold D.J.B., Cao J., Fan Z.	Data-driven battery operation for energy arbitrage using rainbow deep reinforcement learning	Energy	2022	S, W
No author name available [Conference Review]	Sustainable smart cities and territories international conference, SSCT 2021	Lecture Notes in Networks and Systems	2022	S
Dhiman S., Lallotra B.,	Replacing of steel with bamboo as reinforcement with addition of sisal fiber		2022	W
Bekdas G., Yucel M., Nigdeli S. M.	Generation of eco-friendly design for post-tensioned axially symmetric reinforced concrete cylindrical walls by minimizing of CO2 emission	Structural Design of Tall and Special Buildings	2022	W
Vandaele M., Stalhammar S.	Hope dies, action begins? The role of hope for proactive sustainability engagement among university students	International Journal of Sustainability in Higher Education	2022	W
Sagar K. V., Jerald J.	Real-time automated guided vehicles scheduling with markov decision process and double Q-Learning algorithm	Materials Today-Proceedings	2022	W
Kalinin M., Ovasapyan T., Poltavtseva M.	Application of the learning automaton model for ensuring cyber resiliency	Symmetry-Basel	2022	W

Authors	Title	Source title	Year	Database
Liu Q., Sun S., Rong B., Kadoch M.	Intelligent reflective surface based 6G communications for sustainable energy infrastructure	IEEE Wireless Communications	2021	S, W
Muhammad G., Hossain M.S.	Deep-reinforcement-learning-based sustainable energy distribution for wireless communication	IEEE Wireless Communications	2021	S
Grzelczak M., Duch P.	Deep reinforcement learning algorithms for path planning domain in grid-like environment	Applied Sciences (Switzerland)	2021	S
Gao A., Wang Q., Liang W., Ding Z.	Game combined multi-agent reinforcement learning approach for UAV assisted offloading	IEEE Transactions on Vehicular Technology	2021	S, W
Atli İ., Ozturk M., Valastro G.C., Asghar M.Z.	Multi-objective uav positioning mechanism for sustainable wireless connectivity in environments with forbidden flying zones	Algorithms	2021	S
Guo L., Li Z., Outbib R.	Reinforcement learning based energy management for fuel cell hybrid electric vehicles	IECON Proceedings (Industrial Electronics Conference)	2021	S, W
Kővári B., Szőke L., Bécsi T., Aradi S., Gáspár P.	Traffic signal control via reinforcement learning for reducing global vehicle emission	Sustainability (Switzerland)	2021	S, W
Rangel-Martinez D., Nigam K.D.P., Ricardez-Sandoval L.A.	Machine learning on sustainable energy: A review and outlook on renewable energy systems, catalysis, smart grid and energy storage	Chemical Engineering Research and Design	2021	S, W
Choi J.-H., Yang B., Yu C.W.	Artificial intelligence as an agent to transform research paradigms in building science and technology	Indoor and Built Environment	2021	S
Shani P., Chau S., Swei O.	All roads lead to sustainability: opportunities to reduce the life-cycle cost and global warming impact of U.S. roadways	Resources, Conservation and Recycling	2021	S
Jia R., Zhang X., Feng Y., Wang T., Lu J., Zheng Z., Li M.	Long-term energy collection in self-sustainable sensor networks: a deep Q-learning approach	IEEE Internet of Things Journal	2021	S, W

Authors	Title	Source title	Year	Database
Zhao J., Rodriguez M.A., Buyya R.	A deep reinforcement learning approach to resource management in hybrid clouds harnessing renewable energy and task scheduling	IEEE International Conference on Cloud Computing, CLOUD	2021	S, W
Kathirgamanathan A., Mangina E., Finn D.P.	Development of a soft actor critic deep reinforcement learning approach for harnessing energy flexibility in a large office building	Energy and AI	2021	S, W
Mabina P., Mukoma P., Booysen M.J.	Sustainability matchmaking: linking renewable sources to electric water heating through machine learning	Energy and Buildings	2021	S
Chen K., Wang H., Valverde-Pérez B., Zhai S., Vezzaro L., Wang A.	Optimal control towards sustainable wastewater treatment plants based on multi-agent reinforcement learning	Chemosphere	2021	S, W
Sacco A., Flocco M., Esposito F., Marchetto G.	Supporting sustainable virtual network mutations with mystique	IEEE Transactions on Network and Service Management	2021	S
Munir M.S., Tran N.H., Saad W., Hong C.S.	Multi-agent meta-reinforcement learning for self-powered and sustainable edge computing systems	IEEE Transactions on Network and Service Management	2021	S
Pérez-Pons M.E., Alonso R.S., García O., Marreiros G., Corchado J.M.	Deep q-learning and preference based multi-agent system for sustainable agricultural market	Sensors	2021	S
Zhang X., Manogaran G., Muthu B.	IoT enabled integrated system for green energy into smart cities	Sustainable Energy Technologies and Assessments	2021	S, W
Emami-Skardi M.J., Momenzadeh N., Kerachian R.	Social learning diffusion and influential stakeholders identification in socio-hydrological environments	Journal of Hydrology	2021	S, W
Almalki A.J., Alsofyani M., Alghuried A., Wocjan P., Wang L.	Model-based variational autoencoders with autoregressive flows	Proceedings of the 2021 5th World Conference on Smart Trends in Systems Security and Sustainability, WorldS4 2021	2021	S
Liu B., Han W., Wang E., Ma X., Xiong S., Qiao C., Wang J.	An efficient message dissemination scheme for cooperative drivings via multi-agent hierarchical attention reinforcement learning	Proceedings - International Conference on Distributed Computing Systems	2021	S, W
Ghosh S., De S., Chatterjee S., Portmann M.	Learning-based adaptive sensor selection framework for multi-sensing WSN	IEEE Sensors Journal	2021	S

Authors	Title	Source title	Year	Database
Li L., Luo Y., Pu L.	Q-learning enabled intelligent energy attack in sustainable wireless communication networks	IEEE International Conference on Communications	2021	S, W
Sacco A., Esposito F., Marchetto G., Montuschi P.	Sustainable task offloading in UAV networks via multi-agent reinforcement learning	IEEE Transactions on Vehicular Technology	2021	S
Razack A.J., Ajith V., Gupta R.	A deep reinforcement learning approach to traffic signal control	2021 IEEE Conference on Technologies for Sustainability, SusTech 2021	2021	S
Zhang W., Liu H., Wang F., Xu T., Xin H., Dou D., Xiong H.	Intelligent electric vehicle charging recommendation based on multi-agent reinforcement learning	The Web Conference 2021 - Proceedings of the World Wide Web Conference, WWW 2021	2021	S, W
Park J., Lee J., Kim T., Ahn I., Park J.	Co-evolution of predator-prey ecosystems by reinforcement learning agents	Entropy	2021	S, W
Eyni A., Skardi M.J.E., Kerachian R.	A regret-based behavioral model for shared water resources management: Application of the correlated equilibrium concept	Science of the Total Environment	2021	S, W
Raeisi M., Mahboob A.S.	Intelligent control of urban intersection traffic light based on reinforcement learning algorithm	26th International Computer Conference, Computer Society of Iran, CSICC 2021	2021	S, W
Piovesan N., Lopez-Perez D., Miozzo M., Dini P.	Joint load control and energy sharing for renewable powered small base stations: a machine learning approach	IEEE Transactions on Green Communications and Networking	2021	S
Yin H., Wei J., Zhao H., Xiong J., Mei K., Zhang L., Ren B., Ma D.	An intelligent adaptative architecture for wireless communication in complex scenarios	Scientia Sinica Informationis	2021	S
Leng J., Ruan G., Song Y., Liu Q., Fu Y.,Ding K., Chen X.	A loosely-coupled deep reinforcement learning approach for order acceptance decision of mass-individualized printed circuit board manufacturing in industry 4.0	Journal of Cleaner Production	2021	S, W
Baumgart U., Burger M.	A reinforcement learning approach for traffic control	International Conference on Vehicle Technology and Intelligent Transport Systems, VEHITS - Proceedings	2021	S, W
Dinh T.H.L., Kaneko M., Wakao K., Kawamura K., Moriyama T., Takatori Y.	Towards an energy-efficient DQN-based user association in Sub6GHz/mm wave integrated networks	Proceedings - 2021 17th International Conference on Mobility, Sensing and Networking, MSN 2021	2021	S

Authors	Title	Source title	Year	Database
Barth A., Zhang L., Ma O.	Cooperation of a team of heterogeneous swarm robots for space exploration	Proceedings of the International Astronautical Congress, IAC	2021	S
Al-Jawad A., Comsa I.-S., Shah P., Gemikonakli O., Trestian R.	REDO: a reinforcement learning-based dynamic routing algorithm selection method for SDN	2021 IEEE Conference on Network Function Virtualization and Software Defined Networks, NFV-SDN 2021 - Proceedings	2021	S, W
Cloud J.M., Nieves R.J., Duke A.K., Muller T.J., Janmohamed N.A., Buckles B.C., Dupuis M.A.	Towards autonomous lunar resource excavation via deep reinforcement learning	Accelerating Space Commerce, Exploration, and New Discovery conference, ASCEND 2021	2021	S
Isufaj R., Sebastia D.A., Piera M.A.	Towards conflict resolution with deep multi-agent reinforcement learning	14th USA/Europe Air Traffic Management Research and Development Seminar, ATM 2021	2021	S
No author name available [Conference Review]	7th International Conference on Life System Modeling and Simulation, LSMS 2021, and the 7th International Conference on Intelligent Computing for Sustainable Energy and Environment, ICSEE 2021	Communications in Computer and Information Science	2021	S
No author name available [Conference Review]	18th International Conference on Mobile Systems and Pervasive Computing, MobiSPC 2021, the 16th International Conference on Future Networks and Communications, FNC 2021 and the 11th International Conference on Sustainable Energy Information Technology, SEIT 2021	Procedia Computer Science	2021	S
Gambin A.F., Angelats E., Gonzalez J.S., Miozzo M., DIni P.	Sustainable marine ecosystems: deep learning for water quality assessment and forecasting	IEEE Access	2021	S, W
No author name available [Conference Review]	AHFE conferences on human factors in software and systems engineering, artificial intelligence and social computing, and energy, 2021	Lecture Notes in Networks and Systems	2021	S
Serrano J.C., Mula J., Poler R.	Digital twin for supply chain master planning in zero-defect manufacturing	IFIP Advances in Information and Communication Technology	2021	S

Authors	Title	Source title	Year	Database
Chaudhuri R., Mukherjee K., Narayanam R., Vallam R.D.	Collaborative reinforcement learning framework to model evolution of cooperation in sequential social dilemmas	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	2021	S, W
Danassis P., Erden Z.D., Faltings B.	Improved cooperation by exploiting a common signal	Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS	2021	S
Daneshvar M., Asadi S., Mohammadi-Ivatloo B.	Energy trading possibilities in the modern multi-carrier energy networks	Power Systems	2021	S
Zheng Z., Yan P., Chen Y., Cai J., Zhu F.	Increasing crop yield using agriculture sensing data in smart plant factory	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	2021	S
Musaddiq A., Ali R., Choi J.-G., Kim B.-S., Kim S.-W.	Collision observation-based optimization of low-power and lossy IoT network using reinforcement learning	Computers, Materials and Continua	2021	S, W
Ballis H., Dimitriou L.	Evaluating the performance of reinforcement learning signalling strategies for sustainable urban road networks	Advances in Intelligent Systems and Computing	2021	S
Surovik D., Wang K., Vespignani M., Bruce J., Bekris K.E.	Adaptive tensegrity locomotion: Controlling a compliant icosahedron with symmetry-reduced reinforcement learning	International Journal of Robotics Research	2021	S, W
Mandhare P., Yadav J., Kharat V., Patil C. Y.	Control and coordination of self-adaptive traffic signal using deep reinforcement learning	International Journal of Next-Generation Computing	2021	W
Hamutoglu N. B., Unveren-Bilgic E. N., Salar H. C., Sahin Y. L.	The effect of E-learning experience on readiness, attitude, and self-control/self-management	Journal of Information Technology Education-Innovations in Practice	2021	W
de Oliveira A. M. L., Marques C. V., Field's K. A. P.	Application of mathematical modeling in the construction of the ecological house and its perspectives in teaching	Cadernos Educacao Tecnologia e Sociedade	2021	W
Tiwari T., Shastry N., Nandi A.	Deep learning based lateral control system	Proceedings - 2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security, iSSSC 2020	2020	S

Authors	Title	Source title	Year	Database
No author name available [Conference Review]	Proceedings - 2020 International Conference on Pervasive Artificial Intelligence, ICPAI 2020	Proceedings - 2020 International Conference on Pervasive Artificial Intelligence, ICPAI 2020	2020	S
Pang G., Zhu X., Lu K., Peng Z., Deng W.	A simulator for reinforcement learning training in the recommendation field	Proceedings - 2020 IEEE International Symposium on Parallel and Distributed Processing with Applications, 2020 IEEE International Conference on Big Data and Cloud Computing, 2020 IEEE International Symposium on Social Computing and Networking and 2020 IEEE International Conference on Sustainable Computing and Communications, ISPA-BDCloud-SocialCom-SustainCom 2020	2020	S
Chemingui Y., Gastli A., Ellabban O.	Reinforcement learning-based school energy management system	Energies	2020	S, W
Krejci S.E., Ramroop-Butts S., Torres H.N., Isokpehi R.D.	Visual literacy intervention for improving undergraduate student critical thinking of global sustainability issues	Sustainability (Switzerland)	2020	S, W
Ballis H., Dimitriou L.	Evaluation of reinforcement learning traffic signalling strategies for alternative objectives: implementation in the network of nicosia, cyprus	Transport and Telecommunication	2020	S, W
Guliyev H.B., Tomin N.V., Ibrahimov F.S.	Methods of intelligent protection from asymmetrical conditions in electric networks	E3S Web of Conferences	2020	S
Wölfle D., Vishwanath A., Schmeck H.	A guide for the design of benchmark environments for building energy optimization	BuildSys 2020 - Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation	2020	S
Liu H., Zhang C., Guo Q.	Data-driven robust voltage/var control using PV inverters in active distribution networks	Proceedings - 2020 International Conference on Smart Grids and Energy Systems, SGES 2020	2020	S, W
Nakamoto Y., Kumalija E., Zhang M.	Toward autonomous adaptive embedded systems for sustainable services using reinforcement learning (WiP report)	Proceedings - 2020 8th International Symposium on Computing and Networking Workshops, CANDARW 2020	2020	S

Authors	Title	Source title	Year	Database
No author name available [Conference Review]	Proceedings - 2020 8th International Symposium on Computing and Networking Workshops, CANDARW 2020	Proceedings - 2020 8th International Symposium on Computing and Networking Workshops, CANDARW 2020	2020	S
Han M., Del Castillo L.A., Khairy S., Chen X., Cai L.X., Lin B., Hou F.	Multi-agent reinforcement learning for green energy powered IoT networks with random access	IEEE Vehicular Technology Conference	2020	S, W
Henderson P., Hu J., Romoff J., Brunskill E., Jurafsky D., Pineau J.	Towards the systematic reporting of the energy and carbon footprints of machine learning	Journal of Machine Learning Research	2020	S
Tan Z., Karakose M.	Comparative study for deep reinforcement learning with CNN, RNN, and LSTM in autonomous navigation	2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy, ICDABI 2020	2020	S
Lee S., Cho Y., Lee Y.H.	Injection mold production sustainable scheduling using deepreinforcement learning	Sustainability (Switzerland)	2020	S, W
Han M., Duan J., Khairy S., Cai L.X.	Enabling sustainable underwater IoT networks with energy harvesting: a decentralized reinforcement learning approach	IEEE Internet of Things Journal	2020	S, W
Opalic S.M., Goodwin M., Jiao L., Nielsen H.K., Lal Kolhe M.	A deep reinforcement learning scheme for battery energy management	2020 5th International Conference on Smart and Sustainable Technologies, SpliTech 2020	2020	S
Dawn S., Saraogi U., Thakur U.S.	Agent-based learning for auto-navigation within the virtual city	2020 International Conference on Computational Performance Evaluation, ComPE 2020	2020	S, W
Xi F., Ruan X.	Influence of intelligent environmental art based on reinforcement learning on the regionality of architectural design	Proceedings of the International Conference on Electronics and Sustainable Communication Systems, ICESC 2020	2020	S
Skardi M.J.E., Kerachian R., Abdolhay A.	Water and treated wastewater allocation in urban areas considering social attachments	Journal of Hydrology	2020	S, W
Banerjee P.S., Mandal S.N., De D., Maiti B.	RL-sleep: temperature adaptive sleep scheduling using reinforcement learning for sustainable connectivity in wireless sensor networks	Sustainable Computing: Informatics and Systems	2020	S, W

Authors	Title	Source title	Year	Database
Piovesan N., Miozzo M., Dini P.	Modeling the environment in deep reinforcement learning: The case of energy harvesting base stations	ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings	2020	S, W
Radenkovic M., Ha Huynh V.S.	Energy-aware opportunistic charging and energy distribution for sustainable vehicular edge and fog networks	2020 5th International Conference on Fog and Mobile Edge Computing, FMEC 2020	2020	S, W
Ma D., Lan G., Hassan M., Hu W., Das S.K.	Sensing, computing, and communications for energy harvesting IoTs: a survey	IEEE Communications Surveys and Tutorials	2020	S, W
Miozzo M., Piovesan N., Dini P.	Coordinated load control of renewable powered small base stations through layered learning	IEEE Transactions on Green Communications and Networking	2020	S
No author name available [Conference Review]	Proceedings of the 19th international conference on autonomous agents and multiagent systems, AAMAS 2020	Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS	2020	S
Yu K.-H., Jaimes E., Wang C.-C.	Ai based energy optimization in association with class environment	ASME 2020 14th International Conference on Energy Sustainability, ES 2020	2020	S
Kazmi H., Driesen J.	Automated demand side management in buildings	Artificial Intelligence Techniques for a Scalable Energy Transition: Advanced Methods, Digital Technologies, Decision Support Tools, and Applications	2020	S
No author name available [Conference Review]	18th international conference on practical applications of agents and multi-agent systems, PAAMS 2020	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	2020	S
Bouhamed O., Ghazzai H., Besbes H., Massoud Y.	A UAV-assisted data collection for wireless sensor networks: autonomous navigation and scheduling	IEEE Access	2020	S, W
Elavarasan D., Durairaj Vincent P.M.	Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications	IEEE Access	2020	S
Yang T., Zhao L., Li W., Zomaya A.Y.	Reinforcement learning in sustainable energy and electric systems: a survey	Annual Reviews in Control	2020	S, W

Authors	Title	Source title	Year	Database
Perera-Villalba J. J., Martinez-Borreguero G., Naranjo-Correa F. L., Mateos-Nunez M.	Validation of a didactic intervention based on video games for the teaching of sustainability contents in secondary education	14th International Technology, Education and Development Conference (INTED2020)	2020	W
Worum H., Lillekroken D., Ahlsen B., Roaldsen K. S., BerglandA.	Otago exercise programme-from evidence to practice: a qualitative study of physiotherapists' perceptions of the importance of organisational factors of leadership, context and culture for knowledge translation in Norway	BMC Health Services Research	2020	W
Temesgene D.A., Miozzo M., Dini P.	Dynamic control of functional splits for energy harvesting virtual small cells: A distributed reinforcement learning approach	Computer Communications	2019	S, W
Lin K., Lin B., Chen X., Lu Y., Huang Z., Mo Y.	A time-driven workflow scheduling strategy for reasoning tasks of autonomous driving in edge environment	Proceedings - 2019 IEEE Intl Conf on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking, ISPA/BDCloud/SustainCom/SocialCom 2019	2019	S
Bhargavi K., Sathish Babu B.	Load balancing scheme for the public cloud using reinforcement learning with raven roosting optimization policy (RROP)	CSITSS 2019 - 2019 4th International Conference on Computational Systems and Information Technology for Sustainable Solution, Proceedings	2019	S
Strnad F.M., Barfuss W., Donges J.F., Heitzig J.	Deep reinforcement learning in World-Earth system models to discover sustainable management strategies	Chaos	2019	S, W
Firdausiyah N., Taniguchi E., Qureshi A.G.	Impacts of urban consolidation centres for sustainable city logistics using adaptive dynamic programming based multi-agent simulation	IOP Conference Series: Earth and Environmental Science	2019	S
Xu T., Wang N., Lin H., Sun Z.	UAV autonomous reconnaissance route planning based on deep reinforcement learning	Proceedings of the 2019 IEEE International Conference on Unmanned Systems, ICUS 2019	2019	S

Authors	Title	Source title	Year	Database
Alizadeh Shabestray S.M., Abdulhai B.	Multimodal iNtelligent Deep (MiND) traffic signal controller	2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019	2019	S, W
Ebell N., Gütlein M., Pruckner M.	Sharing of energy among cooperative households using distributed multi-agent reinforcement learning	Proceedings of 2019 IEEE PES Innovative Smart Grid Technologies Europe, ISGT-Europe 2019	2019	S, W
Vo N.N.Y., He X., Liu S., Xu G.	Deep learning for decision making and the optimization of socially responsible investments and portfolio	Decision Support Systems	2019	S, W
Chang S., Saha N., Castro-Lacouture D., Yang P.P.-J.	Multivariate relationships between campus design parameters and energy performance using reinforcement learning and parametric modeling	Applied Energy	2019	S, W
Qin F.-B., Xu D.	Review of robot manipulation skill models	Zidonghua Xuebao/Acta Automatica Sinica	2019	S
Shabana Anjum S., Md Noor R., Ahmedy I., Anisi M.H.	Energy optimization of sustainable Internet of Things (IoT) systems using an energy harvesting medium access protocol	IOP Conference Series: Earth and Environmental Science	2019	S
Liu Q., Liu Z., Xu W., Tang Q., Zhou Z., Pham D.T.	Human-robot collaboration in disassembly for sustainable manufacturing	International Journal of Production Research	2019	S, W
Shi B., Yuan H., Shi R.	Pricing cloud resource based on multi-agent reinforcement learning in the competing environment	Proceedings - 16th IEEE International Symposium on Parallel and Distributed Processing with Applications, 17th IEEE International Conference on Ubiquitous Computing and Communications, 8th IEEE International Conference on Big Data and Cloud Computing, 11th IEEE International Conference on Social Computing and Networking and 8th IEEE International Conference on Sustainable Computing and Communications, ISPA/IUCC/BDCloud/SocialCom/SustainCom 2018	2019	S

Authors	Title	Source title	Year	Database
Do Q.V., Koo I.	Dynamic bandwidth allocation scheme for wireless networks with energy harvesting using actor-critic deep reinforcement learning	1st International Conference on Artificial Intelligence in Information and Communication, ICAIIC 2019	2019	S, W
Blad C., Koch S., Ganeswarathas S., Kallesøe C.S., Bøgh S.	Control of HVAC-systems with slow thermodynamic using reinforcement learning	Procedia Manufacturing	2019	S
Mikhail M., Yacout S., Ouali M.-S.	Optimal preventive maintenance strategy using reinforcement learning	Proceedings of the International Conference on Industrial Engineering and Operations Management	2019	S
Chen H., Zhao T., Li C., Guo Y.	Green internet of vehicles: architecture, enabling technologies, and applications	IEEE Access	2019	S, W
Chaudhuri R., Vallam R.D., Garg S., Mukherjee K., Kumar A., Singh S., Narayanam R., Mathur A., Parija G.	Collaborative reinforcement learning model for sustainability of cooperation in sequential social dilemmas	Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS	2019	S, W
Park J.Y., Nagy Z.	The influence of building design, sensor placement, and occupant preferences on occupant centered lighting control	Computing in Civil Engineering 2019: Smart Cities, Sustainability, and Resilience - Selected Papers from the ASCE International Conference on Computing in Civil Engineering 2019	2019	S
No author name available [Conference Review]	2nd International Conference on Intelligent Human Systems Integration, IHSI 2019	Advances in Intelligent Systems and Computing	2019	S
McLauchlan A., Joao E.	Recognising learning' as an uncertain source of SEA effectiveness	Impact Assessment and Project Appraisal	2019	W
Halim D. A., Karyanto P., Sarwono	Education for sustainable development: student's biophilia and the emome model as an alternative efforts of enhancement in the perspectives of education	2nd International Conference on Science, Mathematics, Environment, and Education, 2019	2019	W
Saifuddin M. R. B., Logenthiran T., Naayagi R. T., Woo W. L.	A nano-biased energy management using reinforced learning multi-agent on layered coalition model: consumer sovereignty	IEEE Access	2019	W

Authors	Title	Source title	Year	Database
Temesgene D.A., Miozzo M., Dini P.	Dynamic functional split selection in energy harvesting virtual small cells using temporal difference learning	IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC	2018	S, W
No author name available [Conference Review]	6th IEEE international conference on advanced logistics and transport, ICALT 2017 - proceedings	6th IEEE International Conference on Advanced Logistics and Transport, ICALT 2017 - Proceedings	2018	S
Prauzek M., Mourcet N.R.A., Hlavica J., Musilek P.	Q-learning algorithm for energy management in solar powered embedded monitoring systems	2018 IEEE Congress on Evolutionary Computation, CEC 2018 - Proceedings	2018	S, W
Ghanshala K.K., Sharma S., Mohan S., Nautiyal L., Mishra P., Joshi R.C.	Self-organizing sustainable spectrum management methodology in cognitive radio vehicular Adhoc network (CRAVENET) environment: a reinforcement learning approach	ICSCCC 2018 - 1st International Conference on Secure Cyber Computing and Communications	2018	S, W
Aziz H.M.A., Zhu F., Ukkusuri S.V.	Learning-based traffic signal control algorithms with neighborhood information sharing: an application for sustainable mobility	Journal of Intelligent Transportation Systems: Technology, Planning, and Operations	2018	S
Laubis K., Knöll F., Zeidler V., Simko V.	Crowdsensing-based road condition monitoring service: an assessment of its managerial implications to road authorities	Lecture Notes in Business Information Processing	2018	S
Hwangbo S., Yoo C.	A methodology of a hybrid hydrogen supply network (HHSN) under alternative energy resources (AERs) of hydrogen footprint constraint for sustainable energy production (SEP)	Computer Aided Chemical Engineering	2018	S
Ganapathi Subramanian S., Crowley M.	Combining MCTS and A3C for prediction of spatially spreading processes in forest wildfire settings	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	2018	S
Serrat R., Alcala M., Delgado-Aguilar M., Tarres J., Oliver-Ortega H., Mutje P.	Case study: development of biodegradable hybrid materials as a substitute for glass fiber reinforced composites	EDULEARN18: 10th International Conference on Education and New Learning Technologies	2018	W
Huang J., Gao Y., Lu S., Zhao X. B., Deng Y. D., Gu M.	Energy-efficient automatic train driving by learning driving patterns		2018	W

Authors	Title	Source title	Year	Database
Miozzo M., Giupponi L., Rossi M., Dini P.	Switch-On/off policies for energy harvesting small cells through distributed Q-learning	2017 IEEE Wireless Communications and Networking Conference Workshops, WCNCW 2017	2017	S, W
Lindkvist E., Ekeberg Ö., Norberg J.	Strategies for sustainable management of renewable resources during environmental change	Proceedings of the Royal Society B: Biological Sciences	2017	S
Perolat J., Leibo J.Z., Zambaldi V., Beattie C., Tuyls K., Graepel T.	A multi-agent reinforcement learning model of common-pool resource appropriation	Advances in Neural Information Processing Systems	2017	S, W
Dos Santos Mignon A., De Azevedo Da Rocha R.L.	An adaptive implementation of ϵ-greedy in reinforcement learning	Procedia Computer Science	2017	S
Ciric D., Todorovic V., Lalic B.	Boosting student entrepreneurship through idealab concept in Western Balkan countries	9th International Conference on Education and New Learning Technologies (EDULEARN17)	2017	W
Sheikhi A., Rayati M., Ranjbar A.M.	Dynamic load management for a residential customer; reinforcement learning approach	Sustainable Cities and Society	2016	S, W
Chen H., Li X., Zhao F.	A reinforcement learning-based sleep scheduling algorithm for desired area coverage in solar-powered wireless sensor networks	IEEE Sensors Journal	2016	S, W
Mathlouthi S., Trabelsi F.B.F., Zribi C.B.O.	A novel approach based on reinforcement learning for anaphora resolution in Arabic texts	Proceedings of the 28th International Business Information Management Association Conference - Vision 2020: Innovation Management, Development Sustainability, and Competitive Economic Growth	2016	S
Kaiser, A., Kragulj, F., Grisold, T.	Identifying human needs in organizations to develop sustainable intellectual capital - reflections on best practices	Proceedings of the 8th European Conference on Intellectual Capital (ECIC 2016)	2016	W
Atesok K., Satava R. M., Van Heest A., Hogan M. V., Pedowitz R. A., Fu F. H., Sitnikov I., Marsh J. L., Hurwitz S. R.	Retention of skills after simulation-based training in orthopaedic surgery	Journal of the American Academy of Orthopaedic Surgeons	2016	W
De Gracia A., Fernández C., Castell A., Mateu C., Cabeza L.F.	Control of a PCM ventilated facade using reinforcement learning techniques	Energy and Buildings	2015	S
Soares I.B., De Hauwere Y.-M., Januarius K., Brys T., Salvant T., Nowe A.	Departure MANagement with a reinforcement learning approach: respecting CFMU slots	IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC	2015	S

Authors	Title	Source title	Year	Database
Jin J., Ma X.	Adaptive group-based signal control using reinforcement learning with eligibility traces	IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC	2015	S
Hsu R.C., Lin T.-H., Chen S.-M., Liu C.-T.	Dynamic energy management of energy harvesting wireless sensor nodes using fuzzy inference system with reinforcement learning	Proceeding - 2015 IEEE International Conference on Industrial Informatics, INDIN 2015	2015	S, W
Miozzo M., Giupponi L., Rossi M., Dini P.	Distributed Q-learning for energy harvesting heterogeneous networks	2015 IEEE International Conference on Communication Workshop, ICCW 2015	2015	S
Vankov P., Vankova D.,	Sustainable Educational & Emotional Model - An Experience from Bulgaria	EDULEARN15: 7th International Conference on Education and New Learning Technologies	2015	W
Morales R. C., Sotomayor J. J., Hochstetter J., Figueroa D.	REFCOMTIC Interactive Manual for The Strengthen of Transversal Teaching Competences	INTED2015: 9th International Technology, Education and Development Conference	2015	W
Geller E. S.	Seven Life Lessons From Humanistic Behaviorism: How to Bring the Best Out of Yourself and Others	Journal of Organizational Behavior Management	2015	W
Comşa I.S., Aydin M., Zhang S., Kuonen P., Wagen J.-F., Lu Y.	Scheduling policies based on dynamic throughput and fairness tradeoff control in LTE-A networks	Proceedings - Conference on Local Computer Networks, LCN	2014	S, W
Hsu R.C., Liu C.-T., Wang H.-L.	A reinforcement learning-based ToD provisioning dynamic power management for sustainable operation of energy harvesting wireless sensor node	IEEE Transactions on Emerging Topics in Computing	2014	S
Urieli D., Stone P.	TacTex'13: A champion adaptive power trading agent	AAMAS 2014 Workshop on Adaptive and Learning Agents, ALA 2014	2014	S, W
Urieli D., Stone P.	TacTex'13: A champion adaptive power trading agent	13th International Conference on Autonomous Agents and Multiagent Systems, AAMAS 2014	2014	S
Urieli D., Stone P.	TacTex'13: A champion adaptive power trading agent	Proceedings of the National Conference on Artificial Intelligence	2014	S, W
Lindkvist E., Norberg J.	Modeling experiential learning: The challenges posed by threshold dynamics for sustainable renewable resource management	Ecological Economics	2014	S

Authors	Title	Source title	Year	Database
Crowley M.	Using equilibrium policy gradients for spatiotemporal planning in forest ecosystem management	IEEE Transactions on Computers	2014	S
Bielskis A.A., Guseinoviene E., Zutautas L., Drungilas D., Dzemydiene D., Gricius G.	Modeling of Ambient Comfort Affect Reward based on multi-agents in cloud interconnection environment for developing the sustainable home controller	2013 8th International Conference and Exhibition on Ecological Vehicles and Renewable Energies, EVER 2013	2013	S
Bielskis A.A., Guseinoviene E., Dzemydiene D., Drungilas D., Gricius G.	Ambient lighting controller based on reinforcement learning components of multi-agents	Elektronika ir Elektrotechnika	2012	S, W
No author name available [Conference Review]	APBITM 2011 - Proceedings 2011 IEEE International Summer Conference of Asia Pacific Business Innovation and Technology Management	APBITM 2011 - Proceedings2011 IEEE International Summer Conference of Asia Pacific Business Innovation and Technology Management	2011	S
No author name available [Conference Review]	IEEE 2011 EnergyTech, ENERGYTECH 2011	IEEE 2011 EnergyTech, ENERGYTECH 2011	2011	S
Anyanwu L.O., Keengwe J., Arome G.A.	Scalable intrusion detection with recurrent neural networks	ITNG2010 - 7th International Conference on Information Technology: New Generations	2010	S
Sabbadin R., Spring D., Bergonnier E.	A reinforcement-learning application to biodiversity conservation in Costa-Rican forest	MODSIM07 - Land, Water and Environmental Management: Integrated Systems for Sustainability, Proceedings	2007	S
Chadès I., Martin T.G., Curtis J.M.R., Barreto C.	Managing interacting species: A reinforcement learning decision theoretic approach	MODSIM07 - Land, Water and Environmental Management: Integrated Systems for Sustainability, Proceedings	2007	S
Bielskis A. A., Denisovas V., Ramasauskas O.	Ambient Intelligence of e-possibilities perception for sustainable development	4th International Conference Citizens and Governance for Sustainable Development	2006	W
Salden A.H., Kempen Ma.	Sustainable cybernetics systems: Backbones of ambient intelligent environments	Ambient Intelligence: A Novel Paradigm	2005	S

Authors	Title	Source title	Year	Database
Chen L., Evans T., Anand S., Boufford J., Brown H., Chowdhury M., Cueto M., Dare L., Dussault G., Elzinga G., Fee E., Habte D., Hanvoravongchai P., Jacobs M., Kurowski C., Michael S., Pablos-Mendez A., Sewankambo N., Solimano G., Stilwell B., de Waal A., Wibulpolprasert S.	Human resources for health: overcoming the crisis	LANCET	2004	W
Salden A., de Heer J.	Natural anticipation and selection of attention within sustainable intelligent multimodal systems by collective intelligent agents	8th World Multi-Conference on Systemics, Cybernetics and Informatics, Vol IX, Proceedings: Computer Science and Engineering: I	2004	W

Author contributions MZ: Conceptualization, Investigation, Methodology, Writing—Original Draft. AC, DLT and AF: Supervision, Conceptualization, Investigation, Validation, Writing—Review and Editing. LM: Conceptualization,Validation, Writing—Review and Editing.

Funding Open access funding provided by Università degli Studi di Verona within the CRUI-CARE Agreement.

Declarations

Competing interests The authors declare no competing interests.

References

Ajao L, Apeh S (2023) Secure edge computing vulnerabilities in smart cities sustainability using petri net and genetic algorithm-based reinforcement learning. Intell Syst Appl. https://doi.org/10.1016/j.iswa.2023.200216

Al-Jawad A, Comşa I, Shah P, et al (2021) REDO: a reinforcement learning-based dynamic routing algorithm selection method for SDN. In: IEEE conference on network function virtualization and software defined networks (NFV-SDN), pp 54–59, https://doi.org/10.1109/NFV-SDN53031.2021.9665140

Alanne K, Sierla S (2022) An overview of machine learning applications for smart buildings. Sustain Cities Soc. https://doi.org/10.1016/j.scs.2021.103445

Alizadeh Shabestray SM, Abdulhai B (2019) Multimodal iNtelligent Deep (MiND) traffic signal controller. In: IEEE intelligent transportation systems conference (ITSC), pp 4532–4539, https://doi.org/10.1109/ITSC.2019.8917493

Auffenberg F, Snow S, Stein S et al (2017) A comfort-based approach to smart heating and air conditioning. ACM Trans Intell Syst Technol. https://doi.org/10.1145/3057730

Aziz H, Zhu F, Ukkusuri S (2018) Learning-based traffic signal control algorithms with neighborhood information sharing: an application for sustainable mobility. J Intell Trans Syst Technol Plan Operat. https://doi.org/10.1080/15472450.2017.1387546

Azzalini D, Castellini A, Luperto M, et al (2020) HMMs for anomaly detection in autonomous robots. In: Proceedings of the 2020 international conference on autonomous agents and multiagent systems, AAMAS, p 105–113, https://doi.org/10.5555/3398761.3398779

Bazzan ALC, Peleteiro-Ramallo A, Burguillo-Rial JC (2011) Learning to cooperate in the iterated prisoner's dilemma by means of social attachments. J Braz Comput Soc 17(3):163–174. https://doi.org/10.1007/s13173-011-0038-2

Bianchi F, Castellini A, Tarocco P, et al (2019) Load forecasting in district heating networks: Model comparison on a real-world case study. In: Machine learning, optimization, and data science: 5th international conference, LOD 2019, proceedings. Springer-Verlag, p 553–565, https://doi.org/10.1007/978-3-030-37599-7_46

Bianchi F, Corsi D, Marzari L, et al (2023) Safe and efficient reinforcement learning for environmental monitoring. In: Proceedings of Ital-IA 2023: 3rd National Conference on Artificial Intelligence, CEUR Workshop Proceedings, vol 3486. CEUR-WS.org, pp 2610–615

Bistaffa F, Farinelli A, Chalkiadakis G et al (2017) A cooperative game-theoretic approach to the social ridesharing problem. Artif Intell 246:86–117. https://doi.org/10.1016/j.artint.2017.02.004

Bistaffa F, Blum C, Cerquides J et al (2021) A computational approach to quantify the benefits of ridesharing for policy makers and travellers. IEEE Trans Intell Transport Syst 22(1):119–130. https://doi.org/10.1109/TITS.2019.2954982

Blij NHVD, Chaifouroosh D, Cañizares CA, et al (2020) Improved power flow methods for DC grids. In: 29th IEEE international symposium on industrial electronics, ISIE. IEEE, pp 1135–1140, https://doi.org/10.1109/ISIE45063.2020.9152570

Bouhamed O, Ghazzai H, Besbes H et al (2020) A UAV-assisted data collection for wireless sensor networks: Autonomous navigation and scheduling. IEEE Access. https://doi.org/10.1109/ACCESS.2020.3002538

Brown J, Abate A, Rogers A (2021) QUILT: quantify, infer and label the thermal efficiency of heating and cooling residential homes. In: BuildSys '21: The 8th ACM international conference on systems for energy-efficient buildings, cities, and transportation. ACM, pp 51–60, https://doi.org/10.1145/3486611.3486653

Capuzzo M, Zanella A, Zuccotto M, et al (2022) IoT systems for healthy and safe life environments. In: IEEE forum on research and technologies for society and industry innovation (RTSI), pp 31–37, https://doi.org/10.1109/RTSI55261.2022.9905193

Castellini A, Chalkiadakis G, Farinelli A (2019) Influence of state-variable constraints on partially observable monte carlo planning. In: Proceedings of the twenty-eighth international joint conference on artificial intelligence, IJCAI 2019. International Joint Conferences on Artificial Intelligence Organization, pp 5540–5546, https://doi.org/10.24963/ijcai.2019/769

Castellini A, Bicego M, Masillo F et al (2020) Time series segmentation for state-model generation of autonomous aquatic drones: a systematic framework. Eng Appl Artif Intell. https://doi.org/10.1016/j.engappai.2020.103499

Castellini A, Bianchi F, Farinelli A (2021) Predictive model generation for load forecasting in district heating networks. IEEE Intell Syst 36(4):86–95. https://doi.org/10.1109/MIS.2020.3005903

Castellini A, Bianchi F, Farinelli A (2022) Generation and interpretation of parsimonious predictive models for load forecasting in smart heating networks. Appl Intell 52(9):9621–9637. https://doi.org/10.1007/s10489-021-02949-4

Castellini A, Bianchi F, Zorzi E, et al (2023) Scalable safe policy improvement via Monte Carlo tree search. In: Proceedings of the 40th international conference on machine learning, proceedings of machine learning research, vol 202. PMLR, pp 3732–3756

Charef N, Ben Mnaouer A, Aloqaily M et al (2023) Artificial intelligence implication on energy sustainability in internet of things: a survey. Info Process Manag. https://doi.org/10.1016/j.ipm.2022.103212

Chen H, Li X, Zhao F (2016) A reinforcement learning-based sleep scheduling algorithm for desired area coverage in solar-powered wireless sensor networks. IEEE Sensors Journal. https://doi.org/10.1109/JSEN.2016.2517084

Chen H, Zhao T, Li C et al (2019) Green internet of vehicles: Architecture, enabling technologies, and applications. IEEE Access. https://doi.org/10.1109/ACCESS.2019.2958175

Chen K, Wang H, Valverde-Pérez B et al (2021) Optimal control towards sustainable wastewater treatment plants based on multi-agent reinforcement learning. Chemosphere. https://doi.org/10.1016/j.chemosphere.2021.130498

De Gracia A, Fernández C, Castell A et al (2015) Control of a PCM ventilated facade using reinforcement learning techniques. Energy Build. https://doi.org/10.1016/j.enbuild.2015.06.045

Elavarasan D, Durairaj Vincent P (2020) Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. IEEE Access. https://doi.org/10.1109/ACCESS.2020.2992480

Emamjomehzadeh O, Kerachian R, Emami-Skardi M et al (2023) Combining urban metabolism and reinforcement learning concepts for sustainable water resources management: a nexus approach. J Environ Manag. https://doi.org/10.1016/j.jenvman.2022.117046

Feng Y, Zhang X, Jia R et al (2023) Intelligent trajectory design for mobile energy harvesting and data transmission. IEEE Internet Things J. https://doi.org/10.1109/JIOT.2022.3202252

Gao Y, Chang D, Chen CH (2023) A digital twin-based approach for optimizing operation energy consumption at automated container terminals. J Clean Prod. https://doi.org/10.1016/j.jclepro.2022.135782

Giri MK, Majumder S (2022) Deep Q-learning based optimal resource allocation method for energy harvested cognitive radio networks. Phys Commun. https://doi.org/10.1016/j.phycom.2022.101766

Goodland R (1995) The concept of environmental sustainability. Ann Rev Ecol Syst 26(1):1–24. https://doi.org/10.1146/annurev.es.26.110195.000245

Gu Z, Liu Z, Wang Q et al (2023) Reinforcement learning-based approach for minimizing energy loss of driving platoon decisions. Sensors. https://doi.org/10.3390/s23084176

Han M, Duan J, Khairy S et al (2020) Enabling sustainable underwater IoT networks with energy harvesting: a decentralized reinforcement learning approach. IEEE Internet Things J. https://doi.org/10.1109/JIOT.2020.2990733

Harrold D, Cao J, Fan Z (2022) Data-driven battery operation for energy arbitrage using rainbow deep reinforcement learning. Energy. https://doi.org/10.1016/j.energy.2021.121958

Hausknecht M, Stone P (2015) Deep recurrent Q-learning for partially observable MDPs. Preprint at https://arxiv.org/abs/1507.06527

Heinzelman W, Chandrakasan A, Balakrishnan H (2002) An application-specific protocol architecture for wireless microsensor networks. IEEE Trans Wireless Commun 1(4):660–670. https://doi.org/10.1109/TWC.2002.804190

Hessel M, Modayil J, van Hasselt H, et al (2018) Rainbow: combining improvements in deep reinforcement learning. In: Proceedings of the AAAI conference on artificial intelligence, pp 3215–3222

Himeur Y, Elnour M, Fadli F et al (2022) Ai-big data analytics for building automation and management systems: a survey, actual challenges and future perspectives. Artif Intell Rev 56(6):4929–5021. https://doi.org/10.1007/s10462-022-10286-2

Hsu R, Liu CT, Wang HL (2014) A reinforcement learning-based ToD provisioning dynamic power management for sustainable operation of energy harvesting wireless sensor node. IEEE Trans Emerg Topics Comput. https://doi.org/10.1109/TETC.2014.2316518

Huo D, Sari Y, Kealey R et al (2023) Reinforcement learning-based fleet dispatching for greenhouse gas emission reduction in open-pit mining operations. Resour Conserv Recycl. https://doi.org/10.1016/j.resconrec.2022.106664

Jendoubi I, Bouffard F (2022) Data-driven sustainable distributed energy resources' control based on multi-agent deep reinforcement learning. Sustain Energy Grids Netw. https://doi.org/10.1016/j.segan.2022.100919

Kaelbling LP, Littman ML, Cassandra AR (1998) Planning and acting in partially observable stochastic domains. Artif Intell 101(1–2):99–134. https://doi.org/10.1016/S0004-3702(98)00023-X

Kathirgamanathan A, Mangina E, Finn D (2021) Development of a soft actor critic deep reinforcement learning approach for harnessing energy flexibility in a large office building. Energy AI. https://doi.org/10.1016/j.egyai.2021.100101

Khalid M, Wang L, Wang K et al (2023) Deep reinforcement learning-based long-range autonomous valet parking for smart cities. Sustain Cities Soc. https://doi.org/10.1016/j.scs.2022.104311

Koufakis AM, Rigas ES, Bassiliades N et al (2020) Offline and online electric vehicle charging scheduling with V2V energy transfer. IEEE Trans Intell Transport Syst 21(5):2128–2138. https://doi.org/10.1109/TITS.2019.2914087

LeCun Y (1989) Generalization and network design strategies. Connect Perspect 19(143–155):18

Leng J, Ruan G, Song Y et al (2021) A loosely-coupled deep reinforcement learning approach for order acceptance decision of mass-individualized printed circuit board manufacturing in industry 4.0. J Clean Prod. https://doi.org/10.1016/j.jclepro.2020.124405

Li C, Bai L, Yao L et al (2023) A bibliometric analysis and review on reinforcement learning for transportation applications. Transportmetrica B. https://doi.org/10.1080/21680566.2023.2179461

Lillicrap TP, Hunt JJ, Pritzel A, et al (2016) Continuous control with deep reinforcement learning. In: International conference on learning representations, ICLR

Liu Q, Sun S, Rong B et al (2021) Intelligent reflective surface based 6G communications for sustainable energy infrastructure. IEEE Wireless Commun. https://doi.org/10.1109/MWC.016.2100179

Lowe R, Wu Y, Tamar A, et al (2017) Multi-agent actor-critic for mixed cooperative-competitive environments. In: Proceedings of the international conference on neural information processing systems, NIPS, p 6382-6393

Ma D, Lan G, Hassan M et al (2020) Sensing, computing, and communications for energy harvesting IoTs: a survey. IEEE Commun Surv Tutor 22(2):1222–1250. https://doi.org/10.1109/COMST.2019.2962526

Mabina P, Mukoma P, Booysen M (2021) Sustainability matchmaking: linking renewable sources to electric water heating through machine learning. Energy Build. https://doi.org/10.1016/j.enbuild.2021.111085

Marchesini E, Corsi D, Farinelli A (2021) Benchmarking safe deep reinforcement learning in aquatic navigation. In: IEEE/RSJ international conference on intelligent robots and systems, IROS. IEEE, pp 5590–5595, https://doi.org/10.1109/IROS51168.2021.9635925

Mazzi G, Castellini A, Farinelli A (2021) Rule-based shielding for partially observable monte-carlo planning. In: Proceedings of the international conference on automated planning and scheduling pp 243–251. https://doi.org/10.1609/icaps.v31i1.15968

Mazzi G, Castellini A, Farinelli A (2023) Risk-aware shielding of partially observable monte carlo planning policies. Artif Intell 324:103987

Miozzo M, Giupponi L, Rossi M, et al (2015) Distributed Q-learning for energy harvesting heterogeneous networks. In: IEEE international conference on communication workshop (ICCW), pp 2006–2011, https://doi.org/10.1109/ICCW.2015.7247475

Miozzo M, Giupponi L, Rossi M, et al (2017) Switch-on/off policies for energy harvesting small cells through distributed Q-learning. In: IEEE wireless communications and networking conference workshops (WCNCW), pp 1–6, https://doi.org/10.1109/WCNCW.2017.7919075

Mischos S, Dalagdi E, Vrakas D (2023) Intelligent energy management systems: a review. Artif Intell Rev. https://doi.org/10.1007/s10462-023-10441-3

Mnih V, Kavukcuoglu K, Silver D et al (2015) Human-level control through deep reinforcement learning. Nature 518(7540):529–533. https://doi.org/10.1038/nature14236

Moerland TM, Broekens J, Plaat A, et al (2020) Model-based reinforcement learning: a survey. arXiv abs/2206.09328. https://doi.org/10.48550/ARXIV.2206.09328

Ng A, Harada D, Russell SJ (1999) Policy invariance under reward transformations: theory and application to reward shaping. In: Proceedings of the international conference on machine learning, ICML, p 278–287

Orfanoudakis S, Chalkiadakis G (2023) A novel aggregation framework for the efficient integration of distributed energy resources in the smart grid. In: Proceedings of the 2023 international conference on autonomous agents and multiagent systems. AAMAS. ACM, pp 2514–2516, https://doi.org/10.5555/3545946.3598986

Ounoughi C, Touibi G, Yahia S (2022) EcoLight: eco-friendly traffic signal control driven by urban noise prediction. Lecture Notes Comput Sci. https://doi.org/10.1007/978-3-031-12423-5_16

Panagopoulos AA, Alam M, Rogers A, et al (2015) AdaHeat: a general adaptive intelligent agent for domestic heating control. In: Proceedings of the 2015 international conference on autonomous agents and multiagent systems, AAMAS. ACM, pp 1295–1303

Perianes-Rodriguez A, Waltman L, van Eck NJ (2016) Constructing bibliometric networks: a comparison between full and fractional counting. J Info 10(4):1178–1195. https://doi.org/10.1016/j.joi.2016.10.006

Puterman ML (1994) Markov Decision Processes: Discrete Stochastic Dynamic Programming. Wiley, Hoboken

Radini S, Marinelli E, Akyol Çağrı et al (2021) Urban water-energy-food-climate nexus in integrated wastewater and reuse systems: cyber-physical framework and innovations. Appl Energy 298:117268

Rampini L, Re Cecconi F (2022) Artificial intelligence in construction asset management: A review of present status, challenges and future opportunities. J Info Technol Construct. https://doi.org/10.36680/j.itcon.2022.043

Rangel-Martinez D, Nigam K, Ricardez-Sandoval L (2021) Machine learning on sustainable energy: a review and outlook on renewable energy systems, catalysis, smart grid and energy storage. Chem Eng Res Design. https://doi.org/10.1016/j.cherd.2021.08.013

Roncalli M, Bistaffa F, Farinelli A (2019) Decentralized power distribution in the smart grid with ancillary lines. Mobile Netw Appl 24(5):1654–1662. https://doi.org/10.1007/s11036-017-0893-y

Rumelhart DE, Hinton GE, Williams RJ (1986) Learning representations by back-propagating errors. Nature 323(6088):533–536. https://doi.org/10.1038/323533a0

Sabet S, Farooq B (2022) Green vehicle routing problem: State of the art and future directions. IEEE Access 10:101622–101642. https://doi.org/10.1109/ACCESS.2022.3208899

Sacco A, Esposito F, Marchetto G et al (2021) Sustainable task offloading in UAV networks via multi-agent reinforcement learning. IEEE Trans Vehicul Technol. https://doi.org/10.1109/TVT.2021.3074304

Shaw R, Howley E, Barrett E (2022) Applying reinforcement learning towards automating energy efficient virtual machine consolidation in cloud data centers. Info Syst. https://doi.org/10.1016/j.is.2021.101722

Sheikhi A, Rayati M, Ranjbar A (2016) Dynamic load management for a residential customer; reinforcement learning approach. Sustain Cities Soc. https://doi.org/10.1016/j.scs.2016.04.001

Silver D, Huang A, Maddison CJ et al (2016) Mastering the game of Go with deep neural networks and tree search. Nature. https://doi.org/10.1038/nature16961

Silver D, Schrittwieser J, Simonyan K et al (2017) Mastering the game of go without human knowledge. Nature. https://doi.org/10.1038/nature24270

Simão TD, Suilen M, Jansen N (2023) Safe policy improvement for POMDPs via finite-state controllers. Proc AAAI Conf Artif Intell 37(12):15109–15117. https://doi.org/10.1609/aaai.v37i12.26763

Sivamayil K, Rajasekar E, Aljafari B et al (2023) A systematic study on reinforcement learning based applications. Energies. https://doi.org/10.3390/en16031512

Skardi M, Kerachian R, Abdolhay A (2020) Water and treated wastewater allocation in urban areas considering social attachments. J Hydrol. https://doi.org/10.1016/j.jhydrol.2020.124757

Steccanella L, Bloisi D, Castellini A et al (2020) Waterline and obstacle detection in images from low-cost autonomous boats for environmental monitoring. Robot Auton Syst 124:103346

Sultanuddin S, Vibin R, Rajesh Kumar A et al (2023) Development of improved reinforcement learning smart charging strategy for electric vehicle fleet. J Energy Storage. https://doi.org/10.1016/j.est.2023.106987

Sutton RS, Barto AG (2018) Reinforcement learning: an introduction. A Bradford Book, Denver

United Nations (2015) Transforming our world: the 2030 agenda for sustainable development

van Hasselt H, Guez A, Silver D (2016) Deep reinforcement learning with double Q-earning. In: Proceedings of the AAAI conference on artificial intelligence, pp 2094–2100, https://doi.org/10.1609/aaai.v30i1.10295

Venkataswamy V, Grigsby J, Grimshaw A et al (2023) RARE: renewable energy aware resource management in datacenters. Lecture Notes Comput Sci. https://doi.org/10.1007/978-3-031-22698-4_6

Wang JJ, Wang L (2022) A cooperative memetic algorithm with learning-based agent for energy-aware distributed hybrid flow-shop scheduling. IEEE Trans Evol Comput. https://doi.org/10.1109/TEVC.2021.3106168

Watkins CJCH (1989) Learning from delayed rewards. King's College, Oxford

Yang T, Zhao L, Li W et al (2020) Reinforcement learning in sustainable energy and electric systems: a survey. Ann Rev Control 49:145–163. https://doi.org/10.1016/j.arcontrol.2020.03.001

Yao R, Hu Y, Varga L (2023) Applications of agent-based methods in multi-energy systems—a systematic literature review. Energies. https://doi.org/10.3390/en16052456

Zhang W, Liu H, Wang F, et al (2021a) Intelligent electric vehicle charging recommendation based on multi-agent reinforcement learning. In: Proceedings of the web conference, WWW, p 1856–1867, https://doi.org/10.1145/3442381.3449934

Zhang X, Manogaran G, Muthu B (2021b) IoT enabled integrated system for green energy into smart cities. Sustain Energy Technol Assess. https://doi.org/10.1016/j.seta.2021.101208

Zuccotto M, Castellini A, Farinelli A (2022a) Learning state-variable relationships for improving POMCP performance. In: Proceedings of the 37th ACM/SIGAPP symposium on applied computing. Association for Computing Machinery, SAC, p 739–747

Zuccotto M, Piccinelli M, Castellini A et al (2022b) Learning state-variable relationships in POMCP: a framework for mobile robots. Front Robotics AI 2022:183

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.