The University of Tehran Press

Journal of Sustainable Energy Systems

Online ISSN: 2980-8693

Home Page: https://ses.ut.ac.ir

Reinforcement Learning (RL) in Energy Systems: A Review of Adaptive Optimization, Current Challenges, and Future Directions

Amirali Saifoddin[1*] | Ehsan Abdolvand[2] | Mohammadali Allahrabbi Shirazi[3]

1. Corresponding Author, Associate Professor, School of Energy Engineering and Sustainable Resources, Head of the Institute of Soft Technologies, College of Interdisciplinary Sciences and Technologies, University of Tehran, Tehran, Iran. Email: saifoddin@ut.ac.ir
2. PhD Student in Energy Systems Engineering, School of Energy Engineering and Sustainable Resources, College of Interdisciplinary Sciences and Technologies, University of Tehran, Tehran, Iran. Email: Ehsan.abdolvand@ut.ac.ir
3. PhD Student in Energy Systems Engineering, School of Energy Engineering and Sustainable Resources, College of Interdisciplinary Sciences and Technologies, University of Tehran, Tehran, Iran. Email: aliallahrabbi@ut.ac.ir

ARTICLE INFO

Article type:
Research Paper

Article History:
Received: 27 October 2025
Revised: 29 December 2025
Accepted: 25 February 2026
Published Online: 22 June 2026

Keywords:
Energy systems,
Adaptive optimization,
Reinforcement learning,
Energy system management.

ABSTRACT

With the increasing global demand for energy and the increasing complexity of energy systems, especially in the field of renewable resources and smart grids, the need for intelligent methods to optimize energy production, distribution, and consumption is increasingly felt. Reinforcement learning (RL), as an advanced branch of artificial intelligence, has provided new solutions for energy system management with the ability to learn optimal policies through dynamic interaction with the environment and adapt to uncertainties. This paper reviews the basic concepts of reinforcement learning, such as Markov decision processes and related algorithms, the advantages and disadvantages of this method, its practical applications in smart grid management, energy storage optimization, and electric vehicle management. RL is also compared with other optimization methods, such as supervised machine learning, evolutionary algorithms, and traditional mathematical models, and its future directions, including integration with new technologies such as the Internet of Things and blockchain, are reviewed. A special focus is placed on the potential of RL in solving Iran's endemic challenges, such as frequent blackouts and inefficient distribution networks, to propose solutions for energy sustainability at the national level.

Cite this article: Saifoddin, A.; Abdolvand, E. & Allahrabbi Shirazi, M. (2026). Reinforcement Learning (RL) in Energy Systems: A Review of Adaptive Optimization, Current Challenges, and Future Directions. *Journal of Sustainable Energy Systems*, 5 (3), 527-547. DOI: http//doi.org/10.22059/ses.2025.405984.1200

Publisher: University of Tehran Press.
DOI: http//doi.org/10.22059/ses.2025.405984.1200

Introduction

The contemporary global landscape is faced with the increasing demand for energy and the complexity of energy systems, especially in the field of renewable resources and smart grids. These developments highlight the need for intelligent and adaptive methods to optimize energy production, distribution, and consumption. Reinforcement learning (RL), as an advanced branch of artificial intelligence, has emerged as an effective approach to address complex and dynamic challenges in energy systems. RL enables agents to learn optimal policies and adapt to environmental uncertainties through interaction with the environment. This adaptive feature of RL is of particular importance in modern energy systems, which are inherently complex and uncertain. This paper reviews the fundamental concepts of reinforcement learning, its algorithms, its various applications in smart grid management, energy storage, and electric vehicles, and analyzes the challenges and benefits of this approach. This article

also examines the potential of RL in solving Iran's indigenous challenges, such as power generation fluctuations and frequent outages, and analyzes the future directions of this technique in terms of integration with new technologies such as the Internet of Things (IoT) and blockchain.

Methodology

The methodology of this study is a review and analysis that investigates and evaluates the applications of RL in energy system optimization. In this study, first, the theoretical foundations of RL and its fundamental principles such as Markov decision process and key algorithms such as Q-Learning and deep reinforcement learning (DRL) are introduced and analyzed. Then, the advantages and disadvantages of these algorithms in various energy applications, including smart grid management, energy consumption optimization in buildings, energy storage and electric vehicles, are examined. The paper also compares RL with other optimization methods such as evolutionary algorithms and supervised machine learning and critically analyzes its challenges and limitations, such as high computational intensity and scalability issues. This methodology is designed to provide an analytical framework to better understand the applications and future directions of RL in the field of energy optimization.

Result

The results of this study demonstrate the high potential of RL in energy system optimization. RL algorithms, especially in the form of DRL, have been able to effectively solve complex and dynamic problems in the energy domain. For example, in smart grid management, RL has been able to help optimize load distribution and reduce operating costs. Also, in energy storage systems, especially in microgrids and electric vehicles, RL has increased system efficiency by optimizing the charging and discharging timing of batteries, reducing the pressure on the power grid. In addition, the use of RL in buildings and HVAC systems has led to a reduction in energy consumption by up to 20%. However, this study also pointed out the limitations of RL, including the need for heavy computation, scalability challenges, and safety issues in practical applications.

Discussion and Conclusion

Reinforcement learning, as a leading method in artificial intelligence, has shown great potential in optimizing energy systems. This method is able to provide efficient solutions for smart grid management, energy storage optimization, and building energy consumption reduction by learning optimal policies through interaction with complex and dynamic environments. Although RL has been effective in reducing operating costs, improving battery efficiency, and reducing energy consumption, its implementation faces challenges such as the need for high computing power and the time-consuming nature of the training process. On the other hand, combining RL with new technologies such as the Internet of Things and blockchain can increase decision-making accuracy and create more sustainable and efficient energy systems. Especially in countries like Iran that face challenges such as frequent outages and distribution inefficiencies, the use of RL can help reduce network losses and improve energy sustainability. By overcoming existing challenges and utilizing new technologies, RL can become a global standard in energy optimization.

انتشارات دانشگاه تهران

فصلنامهٔ سیستم‌های انرژی پایدار

سایت نشریه: https://ses.ut.ac.ir

شاپا الکترونیکی: ۸۶۹۳-۲۹۸۰

یادگیری تقویتی (RL) در سامانه‌های انرژی: مروری بر بهینه‌سازی تطبیقی، چالش‌های کنونی و مسیرهای آینده

امیرعلی سیف‌الدین[۱*] | احسان عبدالوند[۲] | محمدعلی الهربی شیرازی[۳]

۱. نویسندهٔ مسئول، دانشیار، دانشکدهٔ مهندسی انرژی و منابع پایدار، رئیس مؤسسهٔ فناوری‌های نرم، دانشکدگان علوم و فناوری‌های میان‌رشته‌ای، دانشگاه تهران، تهران، ایران. رایانامه: saifoddin@ut.ac.ir

۲. دانشجوی دکتری مهندسی سیستم‌های انرژی، دانشکدهٔ مهندسی انرژی و منابع پایدار، دانشکدگان علوم و فناوری‌های میان‌رشته‌ای، دانشگاه تهران، تهران، ایران. رایانامه: Ehsan.abdolvand@ut.ac.ir

۳. دانشجوی دکتری مهندسی سیستم‌های انرژی، دانشکدهٔ مهندسی انرژی و منابع پایدار، دانشکدگان علوم و فناوری‌های میان‌رشته‌ای، دانشگاه تهران، تهران، ایران. رایانامه: aliallahrabbi@ut.ac.ir

اطلاعات مقاله

نوع مقاله:
پژوهشی

تاریخ‌های مقاله:
تاریخ دریافت: ۱۴۰۴/۰۸/۰۵
تاریخ بازنگری: ۱۴۰۴/۱۰/۰۸
تاریخ پذیرش: ۱۴۰۴/۱۲/۰۶
تاریخ انتشار: ۱۴۰۵/۰۴/۰۱

کلیدواژه:
سامانه‌های انرژی،
بهینه‌سازی تطبیقی،
یادگیری تقویتی،
مدیریت سیستم انرژی.

چکیده

صنعت با افزایش تقاضای جهانی برای انرژی و پیچیدگی روزافزون سیستم‌های انرژی، به‌ویژه در زمینهٔ منابع تجدیدپذیر و شبکه‌های هوشمند، نیاز به روش‌های هوشمند برای بهینه‌سازی تولید، توزیع و مصرف انرژی بیش از پیش احساس می‌شود. یادگیری تقویتی (RL)، به عنوان یکی از شاخه‌های پیشرفتهٔ هوش مصنوعی، با توانایی یادگیری سیاست‌های بهینه از طریق تعامل پویا با محیط و سازگاری با عدم قطعیت‌ها، راهکارهای نوینی برای مدیریت سیستم‌های انرژی ارائه کرده است. این مقاله به بررسی مفاهیم پایهٔ یادگیری تقویتی، مانند فرایندهای تصمیم‌گیری مارکوف و الگوریتم‌های مرتبط، مزایا و معایب این روش، کاربردهای عملی آن در مدیریت شبکه‌های هوشمند، بهینه‌سازی ذخیره‌سازی انرژی و مدیریت خودروهای الکتریکی می‌پردازد. همچنین، RL با سایر روش‌های بهینه‌سازی، نظیر یادگیری ماشین نظارت‌شده، الگوریتم‌های تکاملی و مدل‌های ریاضی سنتی مقایسه شده و جهت‌گیری‌های آیندهٔ آن، از جمله ادغام با فناوری‌های نوین مانند اینترنت اشیا و بلاک‌چین، بررسی می‌شود. تمرکز ویژه‌ای بر پتانسیل RL در حل چالش‌های بومی ایران، مانند خاموشی‌های مکرر و ناکارایی شبکه‌های توزیع، ارائه شده است تا راهکارهایی برای پایداری انرژی در سطح ملی پیشنهاد شود.

استناد: سیف‌الدین، امیرعلی؛ عبدالوند، احسان و الهربی شیرازی، محمدعلی (۱۴۰۵). یادگیری تقویتی (RL) در سامانه‌های انرژی: مروری بر بهینه‌سازی تطبیقی، چالش‌های کنونی و مسیرهای آینده. *فصلنامهٔ سیستم‌های انرژی پایدار*، ۵ (۳) ۵۲۷-۵۴۷.

DOI: http//doi.org/10.22059/ses.2025.405984.1200

DOI: http//doi.org/10.22059/ses.2025.405984.1200

ناشر: مؤسسه انتشارات دانشگاه تهران.

۱. مقدمه

چشم‌انداز جهانی معاصر بیش از پیش با افزایش پیچیدگی سامانه‌های انرژی و رشد روزافزون تقاضا برای انرژی شناخته می‌شود. این روند، در کنار ضرورت ارتقای پایداری زیست‌محیطی و مدیریت بهینهٔ منابع طبیعی محدود، توسعهٔ راهکارهای هوشمند و تطبیقی را برای بهینه‌سازی انرژی اجتناب‌ناپذیر ساخته است. ظهور زیرساخت‌های نوین و پیچیدهٔ انرژی ـ از جمله شبکه‌های هوشمند و ریزشبکه‌ها که با تمرکززدایی و پویایی فزاینده همراه هستند، نیاز فوری به پارادایم‌های نوین بهینه‌سازی را برجسته می‌سازد.

در این‌میان، یادگیری تقویتی (RL) به عنوان شاخه‌ای پیشرفته از هوش مصنوعی، به عنوان رویکردی محاسباتی قدرتمند برای مواجهه با چالش‌های چندبعدی سامانه‌های انرژی مطرح شده است. RL این امکان را فراهم می‌سازد که یک عامل هوشمند از طریق تعامل مستمر با محیط، سیاست‌های بهینه را بیاموزد و با عدم‌قطعیت‌های ذاتی سازگار شود؛ به گونه‌ای که بدون نیاز به دستورالعمل‌های ازپیش‌تعریف‌شده برای تمامی سناریوها، قادر به استنتاج راهبردهای کارآمد باشد. این ویژگی تطبیقی و مدل‌محور نبودن، به‌ویژه در سامانه‌های انرژی مدرن که به طور ذاتی پویا، پیچیده و نامطمئن هستند و روش‌های سنتی کنترل غالباً ناکارآمد جلوه می‌کنند، اهمیت حیاتی دارد.

این مقاله مروری جامع از کاربرد یادگیری تقویتی در بهینه‌سازی سامانه‌های انرژی ارائه می‌دهد. به این‌منظور، اصول بنیادین RL و الگوریتم‌های کلیدی آن، از جمله یادگیری تقویتی عمیق (DRL) و روش‌های بازیگر ـ منتقد، همراه با مزایا و محدودیت‌های هر یک بررسی می‌شوند. افزون بر این، مزایای بالقوهٔ RL نظیر قابلیت بهینه‌سازی تطبیقی در محیط‌های پویا تحلیل می‌شود و چالش‌های اساسی آن شامل شدت محاسباتی بالا، محدودیت مقیاس‌پذیری، کارایی نمونه و ملاحظات ایمنی به طور نقادانه مورد بحث قرار می‌گیرند.

همچنین، نمونه‌های متنوعی از کاربرد RL در حوزه‌های مختلف انرژی معرفی می‌شود؛ از جمله مدیریت انرژی در شبکه‌های هوشمند، مدیریت کارای منابع در ریزشبکه‌ها، بهینه‌سازی بلادرنگ مصرف انرژی ساختمان‌ها، تصمیم‌گیری و کنترل در سامانه‌های قدرت، بهینه‌سازی مصرف انرژی در شبکه‌های توزیع و مدل‌سازی بازار برق. افزون بر آن، کاربرد مستقیم RL، به‌ویژه DRL، در کنترل پیشرفته سامانه‌های ذخیره‌سازی انرژی باتری ـ شامل زمان‌بندی شارژ/دشارژ و بهینه‌سازی فرایند شارژ خودروهای برقی (EV) تبیین می‌شود. به طور خاص، این مقاله به بررسی نقش RL در پاسخ‌گویی به چالش‌های انرژی ایران، همچون مدیریت خاموشی‌ها و نوسانات تولید برق، می‌پردازد و مسیرهای آینده را در راستای تلفیق RL با فناوری‌های نوظهور نظیر اینترنت اشیا (IoT) و بلاکچین مورد واکاوی قرار می‌دهد.

نوآوری این مطالعه در ارائهٔ مروری جامع بر کاربردهای یادگیری تقویتی در بهینه‌سازی سامانه‌های انرژی است. با توجه به پیچیدگی‌های فزایندهٔ سیستم‌های انرژی، به‌ویژه در زمینهٔ منابع تجدیدپذیر و شبکه‌های هوشمند، استفاده از روش‌های هوشمند برای مدیریت بهینهٔ تولید، توزیع و مصرف انرژی امری ضروری به نظر می‌رسد. این مقاله با هدف جمع‌آوری و تحلیل مطالعات پیشین، به ارزیابی مزایا، محدودیت‌ها و چالش‌های استفاده از یادگیری تقویتی در این حوزه می‌پردازد. همچنین، مطالعهٔ حاضر با مقایسهٔ این روش با سایر روش‌های بهینه‌سازی، مانند یادگیری ماشین نظارت‌شده و الگوریتم‌های تکاملی، سعی دارد دیدگاه‌های جدیدی در راستای بهبود پایداری و کارایی سامانه‌های انرژی ارائه دهد. این تحقیق به‌ویژه برای شناسایی راهکارهای مناسب برای حل مشکلات بومی ایران، نظیر نوسانات تولید برق و خاموشی‌های مکرر، اهمیت دارد.

هدف این مطالعهٔ مروری، بررسی و تحلیل کاربردهای یادگیری تقویتی در بهینه‌سازی سامانه‌های انرژی است. این تحقیق به‌ویژه به شناسایی چالش‌ها و مزایای استفاده از RL در زمینه‌هایی مانند مدیریت شبکه‌های هوشمند، بهینه‌سازی ذخیره‌سازی انرژی، و کاهش مصرف انرژی در ساختمان‌ها می‌پردازد. همچنین، مطالعهٔ حاضر روش‌های مختلف یادگیری تقویتی با سایر رویکردهای بهینه‌سازی، مانند الگوریتم‌های تکاملی و یادگیری ماشین نظارت‌شده را مقایسه می‌کند تا قوت‌ها و ضعف‌های این تکنیک‌ها در حل مسائل پیچیدهٔ سامانه‌های انرژی مشخص شود. در نهایت، هدف این است که با بررسی کاربردهای RL در

شرایط خاص بومی ایران، مانند چالش‌های مربوط به نوسانات تولید و خاموشی‌های مکرر، راهکارهایی برای ارتقای پایداری و بهینه‌سازی انرژی در سطح ملی ارائه شود.

۲. مروری بر ادبیات و مفاهیم یادگیری تقویتی

بهینه‌سازی انرژی در چشم‌انداز معاصر جهانی به عنوان یکی از دغدغه‌های اساسی مطرح است؛ دغدغه‌ای که از ضرورت دستیابی به کاهش چشمگیر هزینه‌ها، ارتقای پایداری زیست‌محیطی و مدیریت سنجیدهٔ منابع طبیعی محدود ناشی می‌شود. تلاش برای افزایش بهره‌وری انرژی و مدیریت مؤثر منابع نه‌تنها یک ملاحظه اقتصادی است، بلکه ضرورتی حیاتی برای جامعه و محیط زیست، به‌ویژه در سامانه‌های پویا و پیچیده‌ای همچون ریزشبکه‌ها، محسوب می‌شود [۱]. زیرساخت‌های نوین انرژی، که نمونهٔ بارز آن مفهوم در حال گسترش شبکه‌های هوشمند است، شاهد افزایش چشمگیر در پیچیدگی و غیرمتمرکز شدن بوده‌اند؛ موضوعی که توسعه و به‌کارگیری روش‌های پیشرفته کنترلی و تکنیک‌های پیچیدهٔ تصمیم‌گیری را الزامی می‌سازد [۲]. علاوه بر این، چالش‌های مرتبط با مسائل کنترلی کلان‌مقیاس، پیچیده و پویا در حوزه‌هایی همچون مدیریت انرژی و مواد ساختمان‌ها، نیاز فوری به پارادایم‌های نوآورانهٔ بهینه‌سازی و محیط زیستی را برجسته می‌سازد [۳].

در این چارچوب، یادگیری تقویتی به عنوان یک رویکرد محاسباتی قدرتمند مطرح می‌شود که در آن یک عامل هوشمند از طریق تعامل مستمر با محیط خود و دریافت بازخوردهای ارزشی در قالب پاداش یا تنبیه، سیاست‌های بهینه را فرا می‌گیرد. اساساً RLاین امکان را برای عامل فراهم می‌سازد تا به صورت خودکار دنباله‌ای از رفتارها یا اقدامات را کشف کند که در طولانی‌مدت به حداکثرسازی یک سیگنال پاداش عددی تجمعی منجر شود [۴]. این الگوی یادگیری از آن جهت متمایز است که به عامل اجازه می‌دهد از طریق فرایند آزمون و خطا، راهبردهای بهینه را استنتاج کند و با شرایط متغیر محیطی سازگار شود، بی‌آنکه نیازمند دستورالعمل‌های صریح و از پیش برنامه‌ریزی‌شده برای تمامی سناریوهای ممکن باشد [۲].

ساختار یک سامانهٔ متعارف یادگیری تقویتی بر پایهٔ چندین مؤلفهٔ اساسی بنا شده است که در کنار یکدیگر فرایند یادگیری را تسهیل می‌کنند:

- عامل (Agent): این بخش نمایانگر موجودیت یادگیرنده یا تصمیم‌گیرنده در چارچوب RL است. عامل وظیفه دارد محیط را ادراک کرده و اقدامات لازم را اجرا کند [۴].
- محیط (Environment): شامل تمامی عناصر بیرونی نسبت به عامل است. محیط بستر لازم برای کنش‌های عامل را فراهم می‌سازد و در پاسخ به آن‌ها با انتقال به حالت‌های جدید و ارائهٔ پاداش واکنش نشان می‌دهد [۴].
- حالت (State): بازنمایی جامعی از وضعیت یا شرایط کنونی محیط است که توسط عامل ادراک می‌شود. حالت، مبنای اصلی تصمیم‌گیری عامل را تشکیل می‌دهد [۴].
- اقدام (Action): انتخاب‌ها یا مداخلات گسسته یا پیوسته‌ای هستند که توسط عامل برای تأثیرگذاری بر محیط انجام می‌گیرند [۴].
- پاداش (Reward): یک سیگنال بازخوردی عددی و کلیدی است که میزان مطلوبیت یا نامطلوبیت فوریِ اقدام عامل در یک حالت خاص را کمی‌سازی می‌کند. هدف اصلی عامل، بیشینه‌سازی مجموع تجمعی این پاداش‌ها طی زمان است [۴].
- سیاست (Policy): راهبرد رفتاری اصلی عامل است که یک نگاشت از حالت‌های ادراک‌شدهٔ محیط به اقدامات متناظر را تعریف می‌کند. در اصل، سیاست رفتار آموخته‌شدهٔ عامل را تعیین می‌کند [۴].

اهمیت چشمگیر یادگیری تقویتی در حوزهٔ سامانه‌های انرژی ناشی از توانایی ذاتی آن در مدیریت مسائل پویا، بسیار پیچیده و ذاتاً نامطمئن است؛ مسائلی که ویژگی بارز زیرساخت‌های مدرن انرژی محسوب می‌شوند. به خلاف روش‌های متعارف کنترلی، الگوریتم‌های RL از ظرفیت منحصربه‌فردی برای یادگیری و بهبود راهبردهای کنترلی بهینه از طریق آزمایش تکرارشونده و سازگاری تدریجی برخوردار هستند [۲]. اهمیت تطبیقی یادگیری تقویتی آن را به‌ویژه برای طیف گسترده‌ای از کاربردها کارآمد می‌سازد؛ کاربردهایی که شامل موارد زیر بوده اما محدود به آن‌ها نمی‌شوند:

- مدیریت انرژی شبکه‌های هوشمند: الگوریتم‌های RL اثربخشی قابل توجهی در بهینه‌سازی جریان انرژی، هماهنگی پاسخ‌های سمت تقاضا و یکپارچه‌سازی بی‌وقفهٔ منابع انرژی تجدیدپذیر توزیع‌شده در شبکه‌های هوشمند، که ذاتاً پیچیده، پویا و با عدم قطعیت قابل توجه هستند، نشان داده‌اند [۲].
- مدیریت منابع انرژی بهینه در ریزشبکه‌ها: یادگیری تقویتی عمیق، یکی از شاخه‌های برجستهٔ RL، با موفقیت برای دستیابی به مدیریت منابع انرژی با بهره‌وری بالا در ریزشبکه‌ها به کار گرفته شده است؛ ریزشبکه‌ها به عنوان مؤلفه‌ای حیاتی از سامانه‌های انرژی غیرمتمرکز آینده شناخته می‌شوند [۱].
- بهینه‌سازی آنلاین انرژی ساختمان‌ها: به طور ویژه برای مقابله با چالش‌های بزرگ، پیچیده و پویا در سیستم‌های انرژی ساختمان‌ها مناسب است، که مستقیم به بهبود بهره‌وری عملیاتی انرژی می‌انجامد [۳].
- تصمیم‌گیری و کنترل در سامانه‌های قدرت الکتریکی: تکنیک‌های RL به طور گسترده برای طیف وسیعی از وظایف تصمیم‌گیری و کنترل در سامانه‌های گستردهٔ برق مورد بررسی قرار گرفته‌اند و از قابلیت یادگیری سیاست‌های بهینه در محیط‌های عملیاتی بسیار پیچیده و پویا بهره می‌برند [۵].
- بهینه‌سازی مصرف انرژی در شبکه‌های توزیع برق: الگوریتم‌های RL به طور مؤثری برای بهینه‌سازی مصرف انرژی در شبکه‌های توزیع برق خاص به کار گرفته شده‌اند، نمونه‌ای از آن کاربردها در ایران مشاهده شده است [۶].
- مدل‌سازی بازار برق: می‌تواند به طور استراتژیک در مدل‌های مبتنی بر عامل برای شبیه‌سازی و بهینه‌سازی رفتار شرکت‌کنندگان در بازارهای عمده‌فروشی برق، مانند بازار برق عمده‌فروشی ایران، مورد استفاده قرار گیرد [۷].

بنابراین، هدف اصلی این مقالهٔ علمی، انجام یک مرور جامع بر کاربردهای رو به رشد یادگیری تقویتی در جنبه‌های مختلف بهینه‌سازی انرژی است. این مرور به طور دقیق مزایا و معایب بالقوهٔ مرتبط با به‌کارگیری RL در این حوزهٔ حیاتی را تحلیل خواهد کرد و همچنین، یک ارزیابی مقایسه‌ای با روش‌های متعارف بهینه‌سازی موجود ارائه خواهد داد.

۳. مبانی نظری یادگیری تقویتی و نتایج

یادگیری تقویتی یک پارادایم از یادگیری ماشین است که در آن یک عامل با تعامل مستمر با محیط، یاد می‌گیرد تصمیم‌های متوالی اتخاذ کند. هدف عامل، بیشینه‌سازی یک سیگنال پاداش عددی طی زمان است که بر اساس اقداماتش از محیط دریافت می‌کند. این فرایند تکراری شامل مشاهدهٔ حالت فعلی، انجام یک اقدام، دریافت پاداش و انتقال به حالت جدید است. مجموع پاداش‌ها، عامل را هدایت می‌کند تا یک سیاست بهینه کشف کند؛ سیاستی که بهترین اقدام را در هر حالت مشخص تعیین می‌کند [۸].

بنیان ریاضی اکثر مسائل یادگیری تقویتی بر فرایند تصمیم‌گیری مارکوف (MDP) استوار است. یک MDP به طور رسمی اجزای یک مسئلهٔ RL را تعریف می‌کند که شامل مجموعه‌ای از حالت‌ها (States)، مجموعه‌ای از اقدامات، تابع انتقال که نحوهٔ تغییر حالت محیط توسط اقدامات را توصیف می‌کند، و تابع پاداش است. یکی از ویژگی‌های اساسی MDP، خاصیت مارکوف است که بیان می‌کند حالت آینده تنها به حالت فعلی و اقدام انجام‌شده بستگی دارد و مستقل از توالی رویدادهایی است که به حالت فعلی منتهی شده‌اند. این خاصیت، پیچیدگی فرایند یادگیری را کاهش می‌دهد، زیرا هر نقطه تصمیم‌گیری را از نظر تاریخچهٔ فوری خود مستقل می‌سازد [۸].

الگوریتم‌های یادگیری تقویتی را می‌توان به طور کلی بر اساس رویکردشان در یادگیری سیاست بهینه دسته‌بندی کرد:

از شاخص‌ترین این الگوریتم‌ها، Q-Learning است که به عنوان یک الگوریتم RL بدون مدل شناخته می‌شود و قادر است سیاست‌های بهینه را بدون نیاز به مدل دینامیک محیط یاد بگیرد. این الگوریتم با یادگیری تابع ارزش ـ اقدام یا همان تابع Q کار می‌کند که پاداش‌های آیندهٔ مورد انتظار را برای انجام یک اقدام خاص در یک حالت مشخص تخمین می‌زند. عامل با توجه به پاداش‌های مشاهده‌شده و مقادیر Q حالت‌های بعدی، مقادیر Q خود را به صورت تکراری به‌روزرسانی می‌کند، به گونه‌ای که درنهایت به مقادیر Q بهینه برسد و بهترین سیاست را تعریف کند [۸].

یادگیری تقویتی عمیق از ترکیب شبکه‌های عصبی عمیق با الگوریتم‌های سنتی RL، برای مسائل پیچیده‌تر که شامل فضای حالت و اقدام با ابعاد بالا هستند، بهره می‌برد. شبکه‌های عصبی عمیق این توانایی را دارند که نمایش‌ها و الگوهای پیچیده را مستقیم از داده‌های خام ورودی یاد بگیرند و به طور مؤثر به عنوان تقریب‌زننده‌های قدرتمند تابع Q یا سیاست عمل کنند. این ترکیب، به عوامل DRL امکان می‌دهد تا با چالش‌هایی مانند حجم عظیم داده‌های حسگر در ساختمان‌های هوشمند یا الگوهای پیچیدهٔ تقاضا در شبکه‌های هوشمند به‌خوبی مقابله کنند [۹ و ۱۰].

یادگیری تقویتی عمیق علاوه بر استفاده از شبکه‌های عصبی برای مسائل پیچیده، به‌ویژه در زمینه‌هایی مانند سیستم‌های انرژی و ساختمان‌های هوشمند، امکان بهره‌برداری از الگوریتم‌های یادگیری چندعاملی را نیز فراهم می‌کند. این رویکرد اجازه می‌دهد که چندین عامل به طور هم‌زمان با یکدیگر تعامل داشته باشند و تصمیمات بهینه را در شرایطی که نیاز به همکاری و هماهنگی دارند، اتخاذ کنند. این ویژگی، به‌ویژه در شبکه‌های هوشمند و مدیریت مصرف انرژی، بسیار حائز اهمیت است، زیرا می‌تواند به بهینه‌سازی مصرف انرژی و مدیریت تقاضا در مقیاس بزرگ‌تر کمک کند. در این راستا، DRL به عنوان ابزاری برای بهبود فرایندهای تصمیم‌گیری در مواجهه با عدم قطعیت و پیچیدگی‌های محیطی عمل می‌کند و نتایج بهینه‌تری نسبت به الگوریتم‌های سنتی ارائه می‌دهد.

Deep RL برای مدیریت شبکه‌های هوشمند پیچیده ضروری است، زیرا توانایی پردازش حجم بالای داده‌های پویا و پیچیده را دارد و قادر است از تجربیات گذشته برای بهینه‌سازی تصمیم‌گیری در محیط‌های غیرقطعی و در حال تغییر استفاده کند. به خلاف روش‌های سنتی که نیازمند مدل‌های ثابت و دقیق هستند، Deep RL به طور انعطاف‌پذیر با چالش‌های ناشی از تغییرات پویا و عدم قطعیت در شبکه‌های هوشمند سازگار می‌شود، و راه‌حل‌های بهینه‌ای در زمان واقعی ارائه می‌دهد.

روش‌های Actor-Critic ترکیبی از یادگیری مبتنی بر سیاست و مبتنی بر ارزش را ارائه می‌دهند. یک عامل Actor-Critic شامل دو مؤلفهٔ اصلی است Actor و Critic است. نقش Actor یادگیری و به‌روزرسانی سیاست است و تعیین می‌کند در هر حالت مشخص کدام اقدام انجام شود. از سوی دیگر، Critic یک تابع ارزش (مانند تابع ارزش حالت یا تابع ارزش ـ اقدام) را یاد می‌گیرد که اقدامات انجام‌شده توسط Actor را ارزیابی می‌کند. ارزیابی Critic سپس برای به‌روزرسانی و بهبود سیاست Actor استفاده می‌شود، که این فرایند به یادگیری پایدارتر و مؤثرتر نسبت به روش‌های صرفاً مبتنی بر سیاست یا مبتنی بر ارزش منجر می‌شود [۸].

یادگیری تقویتی چندعامله (MARL) الگوریتم‌های یادگیری تقویتی را به سیستم‌های چندعامله (MASs) اعمال می‌کند و به ایجاد سامانه‌های توزیع‌شده که در آن چندین عامل به طور هم‌زمان تعامل دارند، کمک می‌کند. به خلاف یادگیری تقویتی تک‌عامله که معمولاً با استفاده از فرایند تصمیم‌گیری مارکوف مدل می‌شود، MARL از بازی مارکوف یا بازی تصادفی استفاده می‌کند تا تأثیرات متقابل عوامل و محیط‌های غیرایستا و غیرقطعی را در نظر بگیرد. در این چارچوب، عوامل وضعیت‌ها و اعمال خود را ارزیابی کرده و به طور پویا رفتار خود را تنظیم می‌کنند تا پاداش‌ها را به حداکثر برسانند. الگوریتم‌های MARL اغلب از مکانیزم‌هایی مانند یادگیری مشارکتی برای بهینه‌سازی سراسری استفاده می‌کنند، که در آن عوامل اطلاعات خود را به اشتراک می‌گذارند. این الگوریتم‌ها به دسته‌های کاملاً مشارکتی، کاملاً رقابتی یا ترکیبی تقسیم می‌شوند [۱۱].

MARL همچنین یک چارچوب مدل‌بی‌نیاز برای مدیریت انرژی غیرمتمرکز و پاسخگویی به تقاضا در ریزشبکه‌ها فراهم می‌آورد. در این سیستم، عامل ارائه‌دهندهٔ خدمات (SPA) به طور پویا قیمت خرید برق را تعیین می‌کند، در حالی که عوامل تولید ـ مصرف‌کننده (PAs) به این قیمت‌گذاری پاسخ می‌دهند و با تنظیم شارژ/دشارژ باتری‌های خود، مدیریت تقاضا را بهینه می‌کنند. این تصمیم‌گیری جمعی، به بهره‌برداری از ذخیره‌سازی باتری‌های پراکنده در ساعت‌های اوج مصرف کمک می‌کند، که در نتیجه هم پایداری شبکه را تقویت و هم مزایای اقتصادی برای تولید ـ مصرف‌کنندگان فراهم می‌آورد [۱۲].

۳-۱. ارتباط یادگیری تقویتی با بهینه‌سازی انرژی

یادگیری تقویتی و شکل‌های مختلف الگوریتمی آن به عنوان ابزاری قدرتمند برای بهینه‌سازی جنبه‌های متنوع مدیریت انرژی مطرح شده است، چرا که قادر است استراتژی‌های تصمیم‌گیری بهینه را در محیط‌های پویا و نامطمئن بیاموزد [۱۳ و ۱۴]. در

ماهیت خود، RL یک مسئلهٔ بهینه‌سازی است، جایی که هدف عامل یافتن بهترین توالی اقدامات (سیاست بهینه) برای بیشینه‌سازی پاداش تجمعی یا کمینه‌سازی یک تابع هزینه طی زمان است [۸]. این تمرکز ذاتی بر بهینگی، RL را به گزینه‌ای بسیار مناسب برای سیستم‌های انرژی تبدیل می‌کند، جایی که اهدافی مانند کاهش مصرف انرژی، کمینه‌سازی هزینه‌های عملیاتی، بیشینه‌سازی بهره‌وری یا تضمین ثبات شبکه از اهمیت بالایی برخوردارند [۱۵ و ۱۶].

یادگیری تقویتی را می‌توان به طور مؤثر در مدیریت بار در شبکه‌های هوشمند به کار گرفت [۱۰]. به عنوان مثال، یک چارچوب مبتنی بر RL می‌تواند برای مدیریت بهینهٔ انرژی در خانه‌ها استفاده شود تا تصمیم‌گیری دربارهٔ زمان مصرف، ذخیره یا تولید انرژی به منظور کاهش هزینه‌ها انجام گیرد [۱۵]. به طور مشابه، رویکردهای مبتنی بر RL در مدیریت بهینهٔ انرژی در سیستم‌های هیبریدی نیز نقش مؤثری دارند، که اغلب شامل متعادل‌سازی میان منابع مختلف انرژی و نیازهای مصرف‌کنندگان است [۱۴].

یکی از نمونه‌های شاخص کاربرد یادگیری تقویتی در بهینه‌سازی انرژی، کنترل ذخیره‌سازی باتری در سیستم‌های انرژی پیچیده است. در ریزشبکه‌ها، جایی که منابع انرژی و بارها نوسان دارند، یادگیری تقویتی عمیق می‌تواند برنامه‌های شارژ و دشارژ سیستم‌های ذخیره‌سازی انرژی باتری را به منظور افزایش بهره‌وری، یکپارچه‌سازی انرژی‌های تجدیدپذیر و تضمین ثبات شبکه بهینه‌سازی کند [۱۳]. کاربرد مستقیم RL، از جمله DRL، در سیستم‌های مدیریت باتری (BMS) امکان کنترل پیشرفته بر سلامت باتری، طول عمر و توزیع انرژی را فراهم می‌کند و به طور پویا با شرایط زمان واقعی سازگار می‌شود [۱۷]. توانایی DRL در مدیریت فضاهای حالت با ابعاد بالا آن را برای کنترل بهینه و زمان واقعی جریان انرژی به سمت باتری‌ها و از آن‌ها ایده‌آل می‌سازد. این کاربردها همچنین شامل برنامه‌ریزی شارژ خودروهای الکتریکی (EV) می‌شود، جایی که الگوریتم‌های RL می‌توانند زمان‌بندی شارژ را بهینه کنند، هزینه‌ها را کاهش دهند و فشار بر شبکه را کم کنند [۱۶].

۳-۲. مزایای یادگیری تقویتی

یادگیری تقویتی چندین مزیت چشمگیر را ارائه می‌دهد که از جمله مهم‌ترین آن‌ها ماهیت تطبیقی ذاتی و توانایی یادگیری بدون مدل است. این ویژگی به عامل‌ها امکان می‌دهد حتی در محیط‌های پویا و نامطمئن، بدون نیاز به مدل ازپیش‌تعریف‌شدهٔ سیستم، سیاست‌های بهینه را کشف کنند. چنین قابلیتی به‌ویژه در کاربردهایی مانند پاسخگویی به تقاضا در سامانه‌های انرژی اهمیت دارد، جایی که شرایط می‌توانند به‌سرعت و به طور غیرقابل پیش‌بینی تغییر کنند. عامل‌های RL قادر هستند به صورت خودکار یاد بگیرند چگونه به این شرایط متغیر واکنش نشان دهند و در نتیجه، راهبردهای کنترلی مقاوم و بسیار کارآمدی ایجاد کنند که طراحی دستی آن‌ها یا اتکا به قواعد ایستا دشوار خواهد بود. توانایی RL در مدیریت عدم قطعیت‌ها و دینامیک‌های پیچیده، بدون نیاز به مدل‌سازی صریح سیستم، آن را به ابزاری قدرتمند برای توسعهٔ راهکارهای تاب‌آور و سازگارپذیر تبدیل می‌سازد [۱۸].

علاوه بر این، RL کارآمدی عملی خود را از طریق پیاده‌سازی‌های موفق در دنیای واقعی در حوزه‌های پیچیده گوناگون نشان داده و توانایی خود را در مدیریت پویایی‌های واقعی و پیچیده اثبات کرده است. یکی از نمونه‌های شاخص در این زمینه، بهینه‌سازی شارژ خودروهای الکتریکی است، جایی که عامل‌های RL با موفقیت به کار گرفته شده‌اند تا با پیچیدگی‌های ناشی از رفتارهای پویا و متغیر کاربران و همچنین، محدودیت‌های شبکه سازگار شوند. چنین کاربردهایی در دنیای واقعی، ظرفیت بالای RL را در ارائهٔ مزایای ملموس از طریق مدیریت هوشمند سامانه‌هایی نشان می‌دهند در آن‌ها شهود انسانی یا الگوریتم‌های ایستا ممکن است ناکافی باشند. توانایی RL در یادگیری مستقیم از طریق تعامل با محیط واقعی بدون نیاز به یک مدل کامل و بی‌نقص از آن محیط، یکی از قوت‌های کلیدی این رویکرد محسوب می‌شود [۱۹].

یادگیری تقویتی عمیق، که تلفیقی از RL و شبکه‌های عصبی عمیق است، مزیت قابل توجهی در حل مسائل کنترل پیوسته و مدیریت فضاهای اقدام با ابعاد بالا ارائه می‌دهد. الگوریتم‌های DRL با بهره‌گیری از قدرت بازنمایی پیشرفتهٔ شبکه‌های عصبی عمیق، قادر هستند سیاست‌ها و توابع ارزش پیچیده را تقریب بزنند و به عامل‌ها امکان دهند اقدام‌های کنترلی دقیق و ظریف را در محیط‌هایی اجرا کنند که در آن‌ها انتخاب‌های گسسته کافی نیستند یا تعداد اقدامات گسسته ممکن به طور بازدارنده‌ای بزرگ است. این قابلیت به‌ویژه برای وظایفی که نیازمند تنظیمات دقیق هستند، مانند کنترل رباتیک، بازی‌های پیچیده، یا تنظیم پیوسته

جریان توان در سامانه‌های انرژی، اهمیت دارد و دامنهٔ کاربرد RL را فراتر از وظایف ساده‌تر تصمیم‌گیری گسسته گسترش می‌دهد. توانایی DRL در یادگیری نگاشت‌های پیچیده از مشاهدات به اقدامات پیوستهٔ آن را برای طیف وسیعی از کاربردهای پیشرفته کنترلی مناسب می‌سازد [۲۰].

RL همچنین کاربردپذیری گسترده‌ای در بخش‌های حیاتی مختلف نشان داده و نقش چشمگیری در ارتقای بهره‌وری و بهینه‌سازی در حوزه‌های متنوع ایفا می‌کند [۲۱]. به عنوان نمونه، مطالعات متعددی به بررسی نقش RL در بهینه‌سازی استفاده از انرژی‌های تجدیدپذیر در ساختمان‌ها پرداخته‌اند و ظرفیت آن را در مدیریت مؤثر منابع ناپایدار مانند انرژی خورشیدی و بادی و همچنین، سامانه‌های ذخیره‌سازی انرژی نشان داده‌اند [۲۲]. به همین‌ترتیب، راهبردهای کنترلی مبتنی بر RL به طور فزاینده‌ای در ارتقای بهره‌وری انرژی در شبکه‌های هوشمند مورد توجه قرار گرفته‌اند و از طریق اتخاذ تصمیم‌های هوشمندانه دربارهٔ جریان و مصرف انرژی، به پایداری و تاب‌آوری بیشتر شبکه‌های توزیع برق کمک می‌کنند [۲۳]. مطالعات مقایسه‌ای RL با سایر روش‌های کنترلی، مانند کنترل پیش‌بین مدل نیز بیانگر اهمیت آن در مدیریت انرژی ساختمان‌ها است [۲۴].

فراتر از سامانه‌های فیزیکی، یادگیری تقویتی در حوزهٔ تصمیم‌گیری‌های پیشرفته مالی، به‌ویژه در مالی انرژی، نیز توجه روزافزونی را به خود جلب کرده است. توانایی RL در یادگیری از طریق تعامل و سازگاری با شرایط متغیر بازار، آن را به ابزاری نویدبخش برای وظایفی همچون معاملات بهینه، مدیریت ریسک و بهینه‌سازی پرتفوی در بازارهای پرنوسان انرژی تبدیل کرده است. این گسترهٔ کاربردی وسیع، بر انعطاف‌پذیری و چندوجهی بودن RL در دستیابی به پایداری انرژی، تعالی عملیاتی و تصمیم‌گیری‌های راهبردی مالی تأکید می‌کند [۲۱].

۳-۳. معایب یادگیری تقویتی

با وجود قابلیت‌های چشمگیر، RL، به‌ویژه DRL، با چالش‌هایی در زمینهٔ بهره‌وری نمونه و نیاز به کاوش گسترده مواجه است. در محیط‌هایی با فضای اقدام گسستهٔ بزرگ، DRL اغلب به تعداد بسیار زیادی از تعاملات نیاز دارد تا عامل بتواند به طور مؤثر یاد بگیرد. این میزان بالای کاوش می‌تواند بسیار زمان‌بر و از نظر محاسباتی پرهزینه باشد، زیرا عامل باید طیف وسیعی از اقدامات مختلف را امتحان کند تا مشخص شود کدام‌یک به دریافت پاداش منجر می‌شوند. این مسئله به‌ویژه در کاربردهای دنیای واقعی که جمع‌آوری داده می‌تواند پرهزینه، منابع‌بر یا حتی خطرناک باشد، مانند سامانه‌های کنترل صنعتی یا تنظیمات آزمایشی حساس، به مانعی جدی تبدیل می‌شود؛ چرا که شکست در فرایند کاوش می‌تواند پیامدهای زیان‌باری داشته باشد. نیاز به حجم بالای نمونه‌ها، به معنای افزایش زمان آموزش و افزایش تقاضای محاسباتی برای دستیابی به عملکرد رضایت‌بخش است [۲۵].

مقیاس‌پذیری یکی دیگر از موانع اساسی در به‌کارگیری یادگیری تقویتی در زیرساخت‌های کلان‌مقیاس، به‌ویژه در شبکه‌های قدرت مدرن محسوب می‌شود. با گسترش ابعاد و پیچیدگی این سامانه‌ها، ابعاد فضای حالت و فضای اقدام به صورت نمایی افزایش می‌یابد و به پدیدهٔ شناخته‌شدهٔ نفرین ابعاد منجر می‌شود. این امر فرایند کاوش کارآمد و یادگیری سیاست‌های بهینه را در شبکه‌ای وسیع از اجزای در تعامل، به طور جدی دشوار می‌سازد؛ چرا که تعداد حالت‌ها و اقدامات ممکن به‌سرعت غیرقابل مدیریت می‌شود. غلبه بر این محدودیت‌های ذاتی مستلزم طراحی‌های معماری نوآورانه و پارادایم‌های محاسباتی پیشرفته است که توانایی پردازش مقیاس عظیم سامانه‌های پیچیدهٔ دنیای واقعی مانند شبکه‌های قدرت را داشته باشند. در غیاب راه‌حل‌های مؤثر برای مسئلهٔ مقیاس‌پذیری، استقرار عملی RL در چنین زیرساخت‌های کلان و حیاتی همچنان یک چالش عمده باقی می‌ماند [۲۶].

یکی از نگرانی‌های اساسی در استقرار یادگیری تقویتی در زیرساخت‌های حیاتی، نظیر شبکه‌های هوشمند برق، تضمین ایمنی و قابلیت اطمینان سیاست‌های یادگرفته‌شده است. اجرای اقدامات ناایمن در هر یک از مراحل یادگیری (کاوش) یا عملیاتی می‌تواند به اختلالات گسترده، خسارت‌های اقتصادی، یا حتی بی‌ثباتی‌های خطرناک در سامانه منجر شود. به عنوان نمونه، یک عامل RL که کنترل یک شبکه قدرت را به عهده دارد، ممکن است تصمیمی اتخاذ کند که به خاموشی گسترده یا آسیب به تجهیزات منجر شود، در صورتی که قیود ایمنی به‌درستی لحاظ نشده باشند. از این‌رو، در سال‌های اخیر توجه فزاینده‌ای به حوزهٔ یادگیری تقویتی ایمن معطوف شده است؛ حوزه‌ای که فراتر از بیشینه‌سازی صرف پاداش، بر ادغام قیود صریح ایمنی با هدف

تضمین پایداری سامانه و جلوگیری از بروز رفتارهای نامطلوب تمرکز دارد. تحقق قابلیت اطمینان در سامانه‌های خودمختار، به‌ویژه آن‌هایی که به طور پویا یاد می‌گیرند و سازگار می‌شوند، نه تنها برای عملکرد ایمن بلکه برای پذیرش اجتماعی و اعتماد عمومی به این فناوری‌ها نقشی تعیین‌کننده ایفا می‌کند [۲۷].

افزون بر چالش‌های پیشین، شدت محاسباتی مورد نیاز برای آموزش عامل‌های یادگیری تقویتی یک ضعف عملی مهم محسوب می‌شود. توسعه و استقرار راهکارهای مؤثر RL، به‌ویژه در حوزه‌هایی همچون مدیریت انرژی در ساختمان‌ها، مستلزم بهره‌گیری از منابع محاسباتی گسترده و فرایندهای بهینه‌سازی پیچیده است. این نیاز بالا به توان پردازشی و حافظه اغلب به زمان‌های طولانی آموزش و همچنین، مصرف بالای انرژی منجر می‌شود که خود می‌تواند مانعی برای به‌کارگیری گستردهٔ این فناوری به شمار آید، به‌ویژه برای سازمان‌هایی که دسترسی محدودی به زیرساخت‌های محاسباتی قدرتمند دارند. ذات تکرارشوندهٔ فرایند آموزش در RL، در ترکیب با معماری‌های پیچیدهٔ شبکه‌های عصبی عمیق، نیازمند قدرت محاسباتی چشمگیری است تا عامل بتواند به طور کامل فضای تصمیم‌گیری را کاوش کند و در نهایت به سیاست‌های کارآمد همگرا شود [۲۴]. با مطالعهٔ منابع مختلف، مزایا و معایب مدل یادگیری تقویتی، در جدول ۱ بیان شده است.

جدول ۱. مزایا و معایب یادگیری تقویتی

دسته‌بندی	مزایا	معایب
قابلیت‌های کلی یادگیری تقویتی	• الگوریتم‌های یادگیری تقویتی می‌توانند بدون نیاز به مدل‌های صریح از محیط، سیاست‌های کنترلی بهینه را فرا بگیرند. • این الگوریتم‌ها برای مسائل پیچیده تصمیم‌گیری مناسب هستند. • یادگیری تقویتی توانایی سازگاری با محیط‌های پویا و نامطمئن را دارد. • این روش قادر است فضاهای کنش بزرگ و پیچیده، چه پیوسته و چه گسسته، را مدیریت کند.	• الگوریتم‌های یادگیری تقویتی اغلب برای یادگیری مؤثر به حجم زیادی از داده‌ها یا تعاملات با محیط نیاز دارند که این امر می‌تواند زمان‌بر و منابع‌بر باشد. • قابلیت تعمیم‌پذیری می‌تواند مسئله‌ساز باشد؛ به طوری که یک عامل آموزش‌دیده در یک محیط ممکن است در محیطی اندکی متفاوت بدون بازآموزی عملکرد مطلوبی نداشته باشد. • ایمنی و قابلیت اطمینان از نگرانی‌های اساسی هستند، به‌ویژه در پیاده‌سازی‌های واقعی که در آن اقدامات اکتشافی می‌توانند به وضعیت‌های ناایمن یا نامطلوب منجر شوند. • تفسیرپذیری سیاست‌های یادگرفته‌شده می‌تواند پایین باشد، که درک دلایل تصمیم‌گیری‌های یک عامل یادگیری تقویتی را دشوار می‌سازد.
کاربرد در سامانه‌های انرژی و شبکه‌های هوشمند	• بهینه‌سازی بهره‌وری انرژی در شبکه‌های هوشمند. • ارتقای بهره‌برداری از انرژی‌های تجدیدپذیر در ساختمان‌ها. • تسهیل پیاده‌سازی واقعی برای شارژ خودروهای برقی. • نشان دادن ظرفیت امیدبخش در مدیریت انرژی ساختمان. • کاربرد در پاسخگویی به تقاضا از طریق بهینه‌سازی الگوهای مصرف انرژی. • قابلیت استفاده در حوزهٔ مالی انرژی، شامل معاملات و مدیریت پرتفوی. • پتانسیل بالاتر برای مقاومت در برابر عدم قطعیت‌های مدل و شرایط متغیر در مقایسه با کنترل پیش‌بینانه مدل‌محور (MPC) در مدیریت انرژی ساختمان.	• چالش‌های مقیاس‌پذیری در سامانه‌های قدرت در مقیاس بزرگ به دلیل ابعاد بالای فضاهای حالت و عمل وجود دارد که به افزایش پیچیدگی محاسباتی و کندی فرایند یادگیری منجر می‌شود. • ایمنی در شبکه‌های هوشمند اهمیت اساسی دارد؛ تضمین اینکه عامل‌های RL موجب ناپایداری یا شکست سامانه نشوند، یک چالش عمده محسوب می‌شود. • یکپارچه‌سازی با زیرساخت‌های موجود شبکه و محدودیت‌های عملیاتی بلادرنگ می‌تواند پیچیده باشد. • دسترسی و کیفیت داده برای آموزش در سامانه‌های انرژی واقعی ممکن است محدود یا مالکیتی باشد. • بار محاسباتی می‌تواند قابل توجه باشد، به‌ویژه برای کنترل بلادرنگ در سامانه‌های بزرگ. • فقدان معیارهای استاندارد در برخی کاربردهای انرژی، مقایسه عملکرد را دشوار می‌سازد.

در نهایت، پیچیدگی پیاده‌سازی و تنظیم دقیق الگوریتم‌های یادگیری تقویتی نیز از دیگر معایب اساسی این رویکرد به شمار می‌آید. معماری پیچیدهٔ شبکه‌های عصبی عمیق [۲۰] همراه با ماهیت تکرارشوندهٔ فرایند آموزش RL و نیاز به روش‌های پیشرفتهٔ بهینه‌سازی [۲۴]، سبب می‌شود که طراحی و استقرار راهکارهای مبتنی بر RL فرایندی ساده و سرراست نباشد. دستیابی به عملکرد مطلوب اغلب مستلزم دانش تخصصی در زمینهٔ تنظیم ابرپارامترها و طراحی معماری مدل است. افزون بر این، حساسیت بالای الگوریتم‌های RL نسبت به انتخاب این پارامترها باعث می‌شود که فرایند آموزش نیازمند آزمون و خطاهای گسترده و تسلط کارشناسانه باشد. این مسئله نه تنها بار محاسباتی را افزایش می‌دهد، بلکه تقاضای قابل توجهی برای منابع انسانی متخصص نیز ایجاد می‌کند و چالش‌های عملیاتی را در فراتر از بُعد محاسباتی تشدید می‌سازد [۲۵].

۳-۴. راهکارهای نوین غلبه بر چالش‌های یادگیری تقویتی در سامانه‌های انرژی

چالش اصلی به‌کارگیری روش‌های استاندارد یادگیری تقویتی در سامانه‌های قدرت، تأمین ایمنی است؛ زیرا کوچک‌ترین خطا می‌تواند به پیامدهای فاجعه‌باری مانند آسیب به تجهیزات یا بروز خاموشی‌های گسترده منجر شود. یادگیری تقویتی ایمن (Safe RL) چارچوبی بنیادی فراهم می‌کند که در آن پایداری و عملکرد قابل‌اعتماد شبکه قدرت هم‌زمان با بهینه‌سازی کارکرد سیستم تضمین می‌شود. راهکارهای موجود در این حوزه به طور کلی در سه دسته اصلی قابل طبقه‌بندی هستند:

- الگوریتم‌های مبتنی بر فرم‌بندی مقید محیط (Constrained Markov Decision Processes: CMDPs) که در آن‌ها قیود ایمنی و عملیاتی مستقیم در ساختار مسئلهٔ تصمیم‌گیری لحاظ می‌شود.
- رویکردهای مبتنی بر ادغام مدل‌های دینامیکی سیستم (Model-Based RL) که با استفاده از دانش قبلی نسبت به رفتار سامانه، کارایی نمونه را افزایش داده و نیاز به اکتشاف پرریسک در دنیای واقعی را کاهش می‌دهند.
- معماری‌های تخصصی مبتنی بر لایه یا فیلتر ایمنی (Safety Layer/Filter Architectures) که به عنوان یک سازوکار حفاظتی، از نقض قیود فیزیکی و بهره‌برداری در حین فرایند یادگیری یا اجرای سیاست جلوگیری می‌کنند [۲۸].

۳-۴-۱. الگوریتم‌های مبتنی بر فرم‌بندی مقید

فرایند تصمیم‌گیری مارکوف مقید (CMDP) در اصل همان مدل استاندارد تصمیم‌گیری مارکوف است که با مجموعه‌ای از قیود تکمیل شده است. این چارچوب، محدودیت‌هایی را بر سیاست‌ها (قواعد بهره‌برداری) که عامل یادگیرندهٔ مجاز به پیروی از آن‌هاست، اعمال می‌کند. برای اعمال این محدودیت‌ها، در CMDPها مجموعه‌ای از توابع هزینه تعریف می‌شود که هنگام انتقال سیستم از یک حالت به حالت بعدی، بسته به عمل انجام‌شده، مقدار مشخصی هزینه به آن اختصاص می‌دهند. هدف اصلی آن است که سیاستی انتخاب شود که ضمن بیشینه‌سازی عملکرد کلی (پاداش)، به طور سخت‌گیرانه به مجموعه قیود تعریف‌شده پایبند باشد [۲۹]. مزایای به‌کارگیری CMDPها، به‌ویژه در کاربردهای حساس به ایمنی مانند سامانه‌های انرژی، شامل موارد زیر است:

- تضمین ایمنی: CMDPها قیود را به صورت صریح اعمال می‌کنند و از این‌رو تضمین‌های ایمنی تأمین می‌شود؛ در حالی که روش‌های یادگیری متعارف در صورت نبود نظارت ممکن است حدود بهره‌برداری را نقض کنند.
- اولویت‌دهی به ایمنی: ساختار CMDP به گونه‌ای است که ایمنی را بر بیشینه‌سازی صرف پاداش مقدم می‌دارد.
- افزایش پایداری: سازوکارهای ایمنیِ یکپارچه‌شده سبب بهبود پایداری عملکرد طی فرایند آموزش می‌شوند.
- تناسب بالا با محیط‌های واقعی: این چارچوب برای محیط‌های پرریسک و حساس به ایمنی بسیار مناسب است، زیرا در تمام مراحل تصمیم‌گیری، رعایت قیود را تضمین می‌کند [۲۹].

۳-۴-۲. رویکردهای مبتنی بر ادغام مدل‌های دینامیکی سیستم

بهره‌گیری از دانش پیشین از طریق تکنیک‌های کنترل مبتنی بر نظریهٔ مجموعه‌ها (Set-Theoretic Control) از مزیت بالایی برخوردار است، زیرا به طراحی یک سیاست کنترلی مبتنی بر شبکهٔ عصبی منجر می‌شود که به طور تضمین‌شده قیود ایمنی حالت‌های بحرانی سیستم را ارضا می‌کند و در نتیجه، ریسک انجام اعمال ناایمن در مرحلهٔ اکتشاف حذف می‌شود. این رویکرد از

نظر محاسباتی نیز کارآمد است، زیرا با بهره‌برداری از ویژگی‌های مجموعه‌های ناوردای مقاوم کنترل‌شده، یک «فیلتر ایمنی» نوین با فرم بسته ایجاد می‌کند که خروجی عامل یادگیرنده را به صورت آنی به مجموعه اعمال ایمن نگاشت می‌دهد، بی‌آنکه نیازی به حل مسائل کنترل پیش‌بین مدل‌محور (MPC) یا مسائل فرافکنی (Projection) به صورت برخط وجود داشته باشد.

در نهایت، مزیت اصلی این روش آن است که سیاست یادگرفته‌شده در مقایسه با تکنیک‌های کنترل مقاوم مبتنی بر بازخور خطی، هزینهٔ کمتر دارد و در عین حال، همان سطح سخت‌گیرانه از تضمین ایمنی را حفظ می‌کند [۳۰].

۳-۴-۳. معماری‌های تخصصی مبتنی بر لایه یا فیلتر ایمنی (Safety Layer/Filter Architectures)

معماری‌های تخصصی مبتنی بر لایه یا فیلتر ایمنی (Safety Layer/Filter Architectures) به عنوان سازوکارهای حفاظتی کلیدی شناخته می‌شوند که در آن‌ها صورت‌بندی قیود ایمنی از هستهٔ الگوریتم یادگیری تقویتی به طور کامل تفکیک می‌شود. این تفکیک به لایه ایمنی امکان می‌دهد تا پیش از اجرای هر عمل، امکان‌پذیری آن را بر اساس توابع قید ازپیش‌تعریف‌شده ارزیابی کرده و به این‌ترتیب، رعایت قیود سخت را به صورت تضمین‌شده برقرار سازد؛ به گونه‌ای که از بروز تخطی‌های فیزیکی هم در مرحلهٔ آموزش اکتشافی و هم در مرحلهٔ بهره‌برداری از سیاست کنترلی جلوگیری شود.

مزیت بنیادین این معماری، کارایی محاسباتی بالا است، زیرا تضمین ایمنی بدون نیاز به حل یک برنامه‌ریزی ریاضی به صورت برخط (Real-Time) حاصل می‌شود. این ساختار همچنین اجازه می‌دهد که هر الگوریتم بهینه‌سازی مبتنی بر یادگیری تقویتی بتواند در یک چارچوب ایمن مورد استفاده قرار گیرد. افزون بر این، روش‌هایی که دانش خبره را در قالب‌هایی نظیر سیاست پشتیبان یکپارچه‌سازی می‌کنند، در مقایسه با روش‌های بدون قید، از سودمندی اولیه به‌مراتب بالاتری برخوردار هستند [۳۱].

۴. نمونه‌های کاربردی یادگیری تقویتی در بهینه‌سازی انرژی

RL به عنوان یکی از روش‌های پیشرفته هوش مصنوعی، در حوزهٔ بهینه‌سازی انرژی کاربردهای عملی گسترده‌ای یافته است. این روش با تمرکز بر تصمیم‌گیری پویا و یادگیری از تعاملات محیطی، امکان حل مسائل پیچیده‌ای را فراهم می‌کند که روش‌های سنتی مانند برنامه‌ریزی ریاضی در آن‌ها ناکارآمد هستند. در ادامه، نمونه‌های کاربردی RL در زمینه‌های مختلف بهینه‌سازی انرژی بررسی می‌شود، با تأکید بر مثال‌های عملی و نتایج کمی. این کاربردها نشان‌دهندهٔ پتانسیل RL در کاهش هزینه‌ها، افزایش کارایی و مدیریت پایدار منابع انرژی هستند. تمرکز بر مثال‌های مبتنی بر داده‌های واقعی یا شبیه‌سازی‌های پیشرفته است تا جنبه‌های عملی برجسته شود [۳۲].

۴-۱. مدیریت شبکه‌های هوشمند

برای شبکه‌های هوشمند به عنوان سیستم‌های توزیع انرژی پویا، با چالش‌هایی مانند نوسانات بار، تلفات انرژی و نیاز به تعادل عرضه و تقاضا مواجه هستند. یادگیری تقویتی، به‌ویژه یادگیری تقویتی عمیق، برای تنظیم دینامیک بار و مدیریت فعال شبکه‌ها استفاده می‌شود.

برای مثال، در مطالعه‌ای از DRL برای مدیریت فعال شبکه‌های توزیع استفاده شده است. این رویکرد با مدل‌سازی شبکه به عنوان یک محیط مارکوف و بهره‌گیری از الگوریتم‌های DRL مانند DDPG، توانسته است تنظیم بار را به صورت پویا انجام دهد و تلفات انرژی را کاهش دهد. در این روش، عامل RL با دریافت پاداش بر اساس کاهش تلفات و پایداری ولتاژ، سیاست‌های بهینه را می‌آموزد. نتایج شبیه‌سازی روی شبکه‌های واقعی نشان‌دهندهٔ کاهش ۱۰ ـ ۱۵ درصد در هزینه‌های عملیاتی شبکه است، که این کاهش عمدتاً از طریق بهینه‌سازی توزیع بار و جلوگیری از اضافه‌بار خطوط حاصل می‌شود. این مثال نشان می‌دهد RL نه تنها در محیط‌های شبیه‌سازی‌شده، بلکه در سیستم‌های واقعی با داده‌های زمانی، کارایی بالایی دارد و می‌تواند با تغییرات ناگهانی تقاضا سازگار شود [۳۳].

۴–۲. بهینه‌سازی ذخیره‌سازی انرژی

ذخیره‌سازی انرژی، به‌ویژه در سیستم‌های مبتنی بر انرژی‌های تجدیدپذیر مانند خورشیدی، نیاز به کنترل هوشمند شارژ و دشارژ باتری‌ها دارد تا کارایی سیستم افزایش یابد و وابستگی به شبکهٔ اصلی کاهش یابد [۲۸]. الگوریتم‌های RL مانند Proximal Policy Optimization (PPO) برای این‌منظور مناسب هستند، زیرا می‌توانند با عدم قطعیت‌های محیطی مانند تغییرات شدت تابش خورشیدی کنار بیایند.

یک مثال کاربردی در این زمینه، استفاده از PPO در مدیریت باتری‌های ذخیره‌سازی در سیستم‌های خورشیدی است. در مطالعه‌ای که در سال ۲۰۲۴ منتشر شد، یک رویکرد DRL برای کنترل باتری در ساختمان‌های مسکونی با پنل‌های خورشیدی پیشنهاد شده است. این مدل با تعریف حالت‌های سیستم (مانند سطح شارژ باتری، پیش‌بینی تولید خورشیدی و تقاضای بار) و پاداش بر اساس کاهش واردات برق از شبکه، سیاست‌های بهینه شارژ/دشارژ را می‌آموزد. نتایج بر اساس داده‌های واقعی نشان‌دهندهٔ بهبود ۱/۶۲ درصد در کاهش واردات برق نسبت به روش‌های سنتی مانند Q-Learning است، که این بهبود به دلیل پایداری بالاتر PPO در آموزش است. این کاربرد نه تنها کارایی ذخیره‌سازی را افزایش می‌دهد، بلکه به پایداری سیستم‌های انرژی تجدیدپذیر کمک می‌کند و می‌تواند در مقیاس بزرگ‌تر، مانند میکروشبکه‌ها، گسترش یابد [۳۴].

۴–۳. مدیریت انرژی در خودروهای الکتریکی

خودروهای الکتریکی با چالش‌هایی مانند زمان‌بندی شارژ بهینه در ایستگاه‌ها مواجه هستند تا هزینه‌ها کاهش یابد و فشار روی شبکهٔ برق کم شود. Q-Learning به عنوان یک الگوریتم ساده و مؤثر RL، برای این مسائل استفاده می‌شود، زیرا می‌تواند سیاست‌های بهینه را بدون نیاز به مدل دقیق محیط بیاموزد.

برای نمونه، در یک طرح زمان‌بندی شارژ مبتنی بر Q-Learning، تمرکز بر بهینه‌سازی عملیات ایستگاه‌های شارژ است. مطالعه‌ای در سال ۲۰۱۹ نشان می‌دهد این رویکرد با مدل‌سازی وضعیت باتری EV، قیمت برق پویا و ظرفیت ایستگاه به عنوان حالت‌ها، توانسته است سود عملیاتی را افزایش دهد. عامل RL با دریافت پاداش بر اساس کاهش هزینهٔ شارژ و افزایش رضایت کاربر، جدول Q را به‌روزرسانی می‌کند و زمان‌بندی بهینه را پیشنهاد می‌دهد. نتایج شبیه‌سازی روی داده‌های واقعی ایستگاه‌های شارژ نشان‌دهندهٔ بهبود سودآوری تا ۲۰ درصد نسبت به روش‌های ایستا است. این مثال تأکید می‌کند که Q-Learning در مسائل با فضای حالت محدود، مانند زمان‌بندی EV، کارایی بالایی دارد و می‌تواند با ادغام داده‌های اینترنت اشیا (IoT) گسترش یابد [۳۵].

۴–۴. بهینه‌سازی مصرف در ساختمان‌ها

ساختمان‌های تجاری و مسکونی بخش عمده‌ای از مصرف انرژی را تشکیل می‌دهند، و سیستم‌های گرمایش، تهویه و تهویه مطبوع (HVAC) از اصلی‌ترین مصرف‌کنندگان هستند. RL برای تنظیم هوشمند HVAC استفاده می‌شود تا مصرف انرژی کاهش یابد در حالی که راحتی کاربران حفظ شود.

یک مثال برجسته، استفاده از RL برای کنترل HVAC در ساختمان‌های هوشمند است. در مطالعه‌ای که در سال ۲۰۲۵ منتشر شد، یک رویکرد RL مبتنی بر دانش متخصصان برای تسریع یادگیری آنلاین پیشنهاد شده است. این مدل با ترکیب قوانین اولیه HVAC (مانند تنظیم دما بر اساس اشغال) و یادگیری تقویتی، توانسته است مصرف انرژی را تا ۲۰ درصد کاهش دهد. عامل RL با پاداش بر اساس تعادل بین مصرف انرژی و سطح راحتی (بر اساس سنسورهای دما و رطوبت)، سیاست‌های بهینه را می‌آموزد. نتایج روی ساختمان‌های واقعی نشان‌دهندهٔ کاهش قابل توجه مصرف بدون تأثیر منفی بر کاربران است. این کاربرد نشان می‌دهد RL می‌تواند با داده‌های سنسورهای هوشمند ادغام شود و در ساختمان‌های بزرگ مقیاس‌پذیر باشد [۳۶].

۴–۵. کاربردهای ایرانی

در ایران، سیستم‌های برق با چالش‌هایی مانند خاموشی‌های مکرر، نوسانات تولید (به دلیل وابستگی به منابع فسیلی و تجدیدپذیر

ناپایدار) و ناکارایی در توزیع مواجه هستند. یادگیری تقویتی می‌تواند در مدیریت شبکه‌های توزیع کمک کند، به‌ویژه در شرایط عدم قطعیت بالا مانند پیک مصرف تابستانی یا اختلالات ناشی از تحریم‌ها.

برای مثال، در مطالعه‌ای که توسط مؤسسهٔ تحقیقات نیرو در ایران انجام شده، از Q-Learning برای مدیریت توزیع انرژی در میکروشبکه‌ها استفاده شده است. این رویکرد با مدل‌سازی عوامل متعدد (مانند تولید محلی و تقاضا) به عنوان سیستم‌های چندعاملی، توانسته است تعادل بار را بهبود بخشد و خاموشی‌ها را کاهش دهد. نتایج شبیه‌سازی روی شبکه‌های توزیع ایران نشان‌دهندهٔ کاهش تلفات تا ۱۲ درصد است، که این می‌تواند به سیاست‌گذاران کمک کند تا RL را در برنامه‌های ملی انرژی ادغام کنند. پیشنهاد می‌شود که RL در مدیریت شبکه‌های توزیع ایران، مانند ادغام با سیستم‌های SCADA، برای پیش‌بینی و جلوگیری از خاموشی‌ها استفاده شود، که این امر می‌تواند پایداری انرژی را افزایش دهد و وابستگی به واردات را کم کند [۳۷].

در مجموع، این نمونه‌ها نشان‌دهندهٔ تنوع کاربردهای RL در بهینه‌سازی انرژی هستند و تأکید می‌کنند که این روش می‌تواند چالش‌های جهانی و بومی را حل کند. تحقیقات آینده می‌تواند بر ترکیب RL با فناوری‌های نوین مانند بلاک‌چین برای امنیت بیشتر تمرکز کند.

بخش انرژی در ایران با چالش‌های مزمن و بومی در زمینهٔ حفظ پایداری و بهره‌وری مواجه است که ضرورت استفاده از پیشرفت‌های فناورانه مانند یادگیری تقویتی را اجتناب‌ناپذیر می‌سازد. در این‌میان، نیاز به ارتقای تاب‌آوری سامانه‌ها در برابر تهدیدهای کم‌احتمال اما پرشدت، همچون زمین‌لرزه، سیلاب، توفان و بارش‌های سنگین برف، به‌ویژه در زیرساخت‌های توزیع، به‌وضوح قابل مشاهده است. این رخدادها موجب خاموشی‌های گسترده و آسیب‌های جدی به شبکه‌های توزیع می‌شوند. علاوه بر این، تمرکز پژوهش‌های داخلی هم‌زمان بر تقویت شبکهٔ فیزیکی در برابر نارسایی‌ها و خطرات خارجی و مقابله با تلفات غیر فنی، نظیر سرقت برق، است. این چالش‌ها ناکارآمدی‌های سیستماتیک در سازوکارهای توزیع و نگهداشت انرژی را برجسته می‌کنند [۳۸].

کاربرد روش‌های پیشرفتهٔ RL در چارچوب محدودیت‌های زیرساخت انرژی ایران، به‌ویژه در بازار عمده‌فروشی برق، در پژوهش‌های داخلی اثبات شده است. مطالعات مدل‌سازی عامل‌محور نشان داده‌اند واحدهای تولید برق می‌توانند از RL برای تعیین بهینهٔ قیمت‌های پیشنهادی بهره‌برداری کنند که به افزایش سود بلندمدت منجر می‌شود. این پژوهش کارایی RL را در مدیریت محیط‌های اقتصادی پیچیده تأیید می‌کند و زمینه را برای به‌کارگیری الگوریتم‌های هوشمند در مسائل پیچیدهٔ بهینه‌سازی انرژی فراهم می‌آورد [۳۹].

چالش‌های ناشی از ادغام منابع انرژی تجدیدپذیر و منابع انرژی توزیع‌شده، نظیر مدیریت تولید ناپایدار آن‌ها و تضمین امنیت سامانهٔ توزیع، از طریق بازارهای انرژی تراکنشی مدرن و یادگیری تقویتی عمیق قابل مدیریت است. الگوریتم‌های DRL، به‌ویژه Soft Actor-Critic به تولید ـ مصرف‌کنندگان این امکان را می‌دهند که راهبردهای بهینه برای مدیریت تولید و مصرف خود در بازارهای رقابتی بیابند، و این‌گونه پایداری شبکه حفظ شود [۴۰].

یادگیری تقویتی عمیق چندعامله چارچوبی قدرتمند برای حل مسائل کلیدی انرژی کشور از جمله مدیریت پیک بار و هماهنگی تولید پراکنده فراهم می‌آورد. در این چارچوب، عامل‌های DRL شامل شرکت‌های خدمات‌دهنده و تولید ـ مصرف‌کنندگان منفرد به طور پویا با یکدیگر تعامل دارند تا در ساعت‌های اوج بار، انرژی ذخیره‌شده به شبکه تزریق شود. این رویکرد، که از الگوریتم‌هایی نظیر Deep Q-Network بهره می‌برد، نیاز به مصرف توان رزرو پرهزینه را کاهش می‌دهد و تقاضای انرژی در دوره‌های اوج را به حداقل می‌رساند [۱۲].

در سطح خانگی، یادگیری تقویتی عمیق می‌تواند در سامانه‌های خودکار مدیریت انرژی خانگی بهینه‌سازی مصرف انرژی را در لبهٔ شبکه تسهیل کند. این الگوریتم‌ها، مانند زمان‌بندی روشن‌سازی سیستم‌های گرمایشی و بهینه‌سازی استفاده از تولید محلی فتوولتائیک، مصرف انرژی وارداتی از شبکه را کاهش می‌دهند و امکان جابه‌جایی بار را فراهم می‌آورند [۴۱].

در نهایت، استفاده از یادگیری تقویتی برای حل چالش‌های مزمن انرژی در ایران مستلزم بازنگری در زیرساخت‌های نظارتی و کنترلی است. راهبرد ملی بر تسریع خودکارسازی و هوشمندسازی شبکه‌های توزیع تأکید دارد و پیاده‌سازی سیستم‌های یکپارچهٔ مدیریت توزیع، مدیریت پاسخگویی بار و مدیریت منابع انرژی توزیع‌شده را ضروری می‌سازد. این زیرساخت‌ها زمینه‌ساز

استقرار سیستم‌های مبتنی بر RL خواهند بود که امکان پیش‌بینی پیشرفته و کنترل خودکار برای مقابله با نوسانات تولید و ناکارآمدی‌های توزیع را فراهم می‌آورند [۴۲].

۴-۶. مقایسهٔ یادگیری تقویتی با سایر روش‌های مدل‌سازی

۴-۶-۱. مقایسه با یادگیری ماشین نظارت‌شده

یادگیری ماشین نظارت‌شده، مانند شبکه‌های عصبی مصنوعی (ANN) و مدل‌های رگرسیون، در مسائل انرژی که داده‌های برچسب‌دار در دسترس هستند، عملکرد بالایی دارد. این روش‌ها برای پیش‌بینی مصرف یا تولید انرژی با دقت بالا مناسب هستند، اما در محیط‌های پویا که نیاز به تصمیم‌گیری در زمان واقعی است، محدودیت دارند، زیرا به داده‌های ازپیش‌تعیین‌شده وابسته‌اند. در مقابل، RL نیازی به داده‌های برچسب‌دار ندارد و با یادگیری از طریق آزمون‌وخطا، می‌تواند سیاست‌های بهینه را در شرایط متغیر، مانند مدیریت بار در شبکه‌های هوشمند، ایجاد کند. با این‌حال، پیچیدگی محاسباتی RL و نیاز به شبیه‌سازی‌های گسترده، آن را در مسائل ساده‌تر پرهزینه‌تر از روش‌های نظارت‌شده می‌کند. برای مثال، ANN می‌تواند مصرف انرژی ساختمان‌ها را با دقت بالا پیش‌بینی کند، اما RL در مدیریت پویا بار و واکنش به تغییرات ناگهانی کارآمدتر است [۴۳].

۴-۶-۲. مقایسه با مدل‌های ریاضی سنتی

مدل‌های ریاضی سنتی، مانند برنامه‌ریزی خطی (LP) و برنامه‌ریزی پویا، به دلیل سادگی و سرعت محاسباتی بالا، در مسائل انرژی با روابط خطی و مشخص، مانند تخصیص منابع در سیستم‌های پایدار، کاربرد دارند. این روش‌ها در شرایط شناخته‌شده کارایی بالایی دارند، اما در مسائل غیرخطی و پویا، مانند نوسانات تولید انرژی بادی یا خورشیدی، ناکارآمد هستند. RL با توانایی مدیریت سیستم‌های غیرخطی و متغیر، انعطاف‌پذیری بیشتری ارائه می‌دهد و می‌تواند سیاست‌های بهینه را در محیط‌های پیچیده بیاموزد. با این‌حال، نیاز به توان محاسباتی بالا و طراحی محیط‌های شبیه‌سازی‌شده، RL را در مقایسه با روش‌های ریاضی ساده پرهزینه‌تر می‌کند. برای مثال، مدل‌های ریاضی سنتی می‌توانند تخصیص منابع را با سرعت بالا انجام دهند، اما RL در بهینه‌سازی بلندمدت و مدیریت عدم قطعیت‌ها کارآمدتر است [۴۴].

۴-۶-۳. مقایسه با یادگیری ماشین نظارت‌شده

الگوریتم‌های تکاملی، مانند الگوریتم ژنتیک (GA) و بهینه‌سازی ازدحام ذرات (PSO)، برای حل مسائل بهینه‌سازی چندمنظوره در سیستم‌های انرژی، مانند طراحی سیستم‌های ترکیبی خورشیدی و بادی، مناسب هستند. این روش‌ها با جست‌وجوی گسترده در فضای راه‌حل، جواب‌های نزدیک به بهینه را در مسائل ایستا ارائه می‌دهند. با این‌حال، در مسائل پویا که نیاز به واکنش سریع به تغییرات محیطی (مانند نوسانات تولید انرژی تجدیدپذیر) دارند، سرعت و انعطاف‌پذیری کمتری دارند. RL، به‌ویژه یادگیری تقویتی عمیق (Deep RL)، با یادگیری مداوم و سازگاری با تغییرات، در مدیریت زمان واقعی سیستم‌های انرژی برتری دارد. با این‌وجود، زمان آموزش طولانی‌تر RL و نیاز به منابع محاسباتی بیشتر، در مقایسه با GA یا PSO که پیاده‌سازی ساده‌تری دارند، یک محدودیت است. برای مثال، PSO می‌تواند طراحی سیستم‌های انرژی ترکیبی را سریع‌تر انجام دهد، اما RL در مدیریت زمان واقعی میکروشبکه‌ها قوی‌تر عمل می‌کند [۴۵].

این مقایسه نشان می‌دهد RL در مسائل پویا و غیرخطی انرژی، مانند مدیریت شبکه‌های هوشمند یا ذخیره‌سازی باتری، برتری دارد، اما در مسائل ساده‌تر یا با داده‌های برچسب‌دار، روش‌های نظارت‌شده یا ریاضی سنتی ممکن است کارآمدتر باشند. انتخاب روش مناسب به نوع مسئله، منابع محاسباتی در دسترس، و نیاز به سازگاری با تغییرات محیطی بستگی دارد. جدول ۲ این مقایسه را از بعدهای مختلف به‌خوبی نشان می‌دهد.

در قالب جمع‌بندی کمی، مطابق شکل ۱ و بر پایهٔ اطلاعات جدول ۳، به‌کارگیری یادگیری تقویتی RL در حوزه‌های مختلف سیستم‌های انرژی موجب بهبودهای مشخصی شده است. همان‌گونه که مشاهده می‌شود، روش‌های مبتنی بر RL توانسته‌اند ۱۰ درصد کاهش در هزینه‌های عملیاتی شبکه‌های توزیع ایجاد کنند. همچنین، هزینه‌های انرژی تجهیزات گرمایشی ۹ درصد و

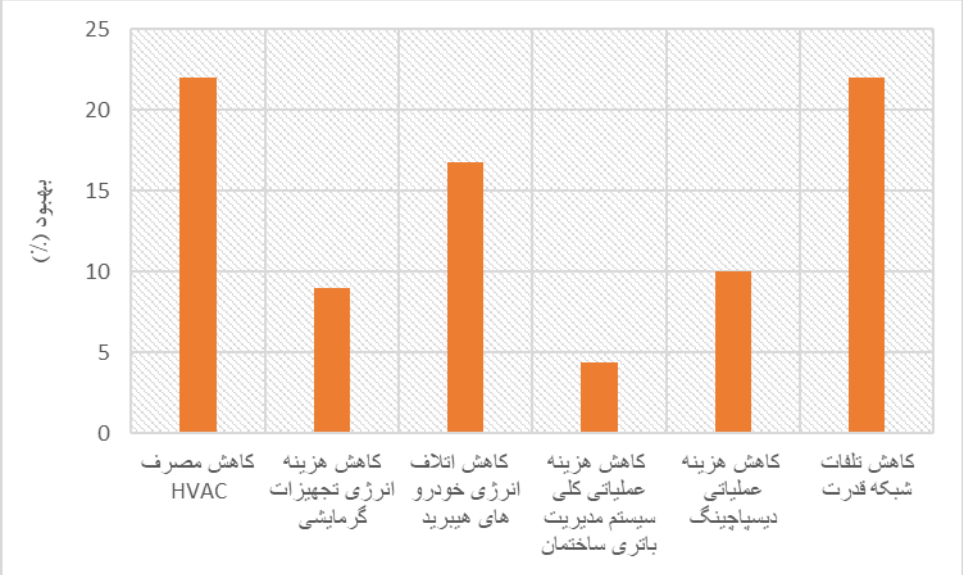
مصرف انرژی سیستم‌های HVAC تا ۲۰ درصد کاهش یافته است. در بخش حمل‌ونقل الکتریکی نیز یادگیری تقویتی سبب بهبود ۱۶/۸ درصدی اتلاف انرژی سیستم ذخیرهٔ انرژی هیبریدی (HESS) در یک خودروی هیبریدی برقی پلاگین (PHEV) شده است. افزون بر این، در شبکه‌های قدرت نیز کاهش هزینه‌های عملیاتی تا ۲۲ درصد را باعث شده.

جدول ۲. مقایسهٔ روش RL با دیگر الگوریتم‌ها

کاربرد در سیستم‌های انرژی	سازگاری با مسائل پویا	پیچیدگی پیاده‌سازی	سرعت محاسباتی	دقت پیش‌بینی	روش
مدیریت پویای شبکه‌های هوشمند، کنترل باتری	بسیار بالا	بالا	متوسط تا پایین	متوسط تا بالا	یادگیری تقویتی (RL)
پیش‌بینی تقاضای انرژی	پایین	متوسط	بالا	بالا	یادگیری نظارت‌شده
طراحی سیستم‌های انرژی ترکیبی	متوسط	متوسط	متوسط	متوسط	الگوریتم‌های تکاملی
تخصیص منابع در سیستم‌های پایدار	پایین	پایین	بالا	بالا (در مسائل خطی)	مدل‌های ریاضی سنتی

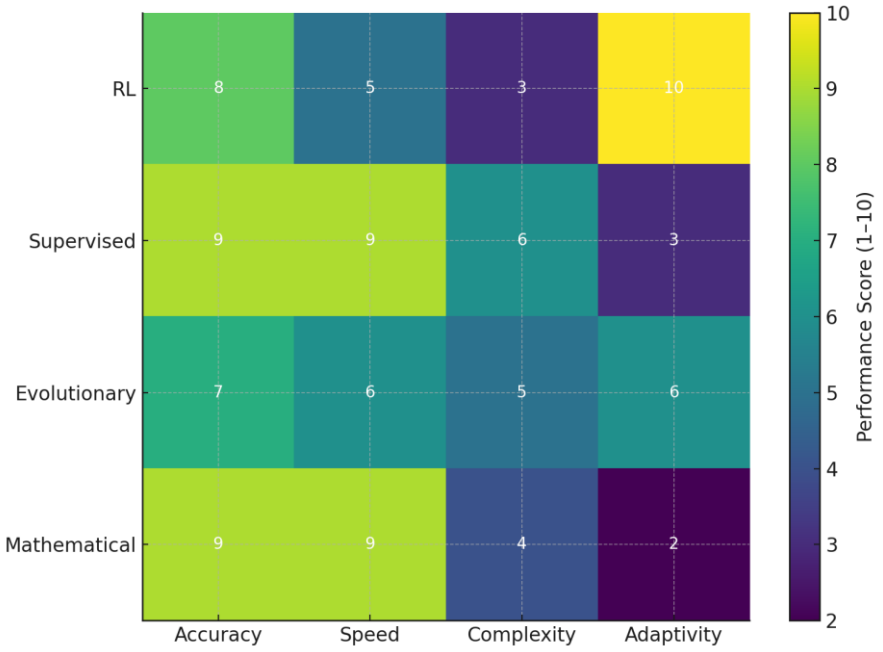
جدول ۳. میزان بهبود در سیستم‌های انرژی با استفاده از روش RL

حوزهٔ کاربرد	مقدار بهبود	نوع الگوریتم RL	نوع مطالعه (شبیه‌سازی/واقعی)	منبع
سیستم‌های تهویهٔ مطبوع مسکونی	۲۲ درصد صرفه‌جویی انرژی	Ibex-RL (Gnu-RL اصلاح شده)	واقعی	[۴۶]
کنترل ساختمان برای بهینه‌سازی مصرف انرژی	۹ درصد کاهش هزینهٔ انرژی تجهیزات گرمایشی	PhysQ	شبیه‌سازی	[۴۷]
سیستم ذخیرهٔ انرژی هیبریدی در یک خودروی هیبریدی برقی پلاگین	۱۶/۸ درصد کاهش نسبی کل اتلاف انرژی	Q-learning	شبیه‌سازی	[۴۸]
توسعهٔ سیستم دومرحله‌ای مدیریت باتری برای کاهش هزینهٔ عملیاتی برای ساختمان‌های با ساختارهای پیچیدهٔ تعرفه برق	۴/۴۳ درصد کاهش هزینهٔ عملیاتی کلی	PPO DQN	شبیه‌سازی	[۴۹]
بهینه‌سازی دیسپاچینگ سیستم‌های قدرت	۲۲ درصد کاهش تلفات شبکه	DRL	شبیه‌سازی	[۵۰]



شکل ۱. میزان بهبود حاصل از اعمال یادگیری تقویتی در سیستم‌های انرژی

همچنین، شکل ۲ مقایسهٔ بصری چهار رویکرد مدل‌سازی رایج در سیستم‌های انرژی شامل یادگیری تقویتی، یادگیری نظارت‌شده، الگوریتم‌های تکاملی و مدل‌های ریاضی سنتی نشان می‌دهد هر روش قوت‌ها و محدودیت‌های متفاوتی دارد. یادگیری تقویتی بالاترین امتیاز را از نظر سازگاری و تطبیق با شرایط پویا کسب کرده، اما از نظر سرعت و پیچیدگی پیاده‌سازی در رتبهٔ پایین‌تری قرار دارد. روش‌های نظارت‌شده در دقت و سرعت بسیار قوی هستند، اما در محیط‌های پویا و غیرخطی عملکرد محدودی دارند. الگوریتم‌های تکاملی تعادل نسبی میان دقت، سرعت و سازگاری ارائه می‌دهند، در حالی که مدل‌های ریاضی سنتی در سرعت و سادگی پیاده‌سازی برتر هستند، ولی در مواجهه با مسائل غیرخطی و متغیر، انعطاف‌پذیری پایینی دارند. این مقایسه نشان می‌دهد انتخاب روش مناسب باید بر اساس ماهیت مسئله، پیچیدگی محیط، نیاز به تطابق‌پذیری و هدف بهینه‌سازی انجام گیرد.



شکل ۲. مقایسهٔ نقشهٔ حرارتی رویکردهای مدل‌سازی در سیستم‌های انرژی

در مقایسه چهار روش رایج در مدل‌سازی سیستم‌های انرژی، شامل یادگیری تقویتی (RL)، یادگیری نظارت‌شده، الگوریتم‌های تکاملی (EA) و مدل‌های ریاضی سنتی، معیارهای مختلفی برای ارزیابی هر روش استفاده شده است [۵۱]. بر اساس مطالعات تجربی، یادگیری تقویتی (DRL) بالاترین عملکرد را در شرایط پویا و پیچیده نشان داده است و قادر است صرفه‌جویی‌های قابل توجهی، مانند ۱۵ ـ ۲۰ درصد کاهش هزینه در سیستم‌های انرژی خانگی، ایجاد کند. با این‌حال، پیچیدگی پیاده‌سازی و نیاز به داده‌های زیاد از محدودیت‌های آن هستند. الگوریتم‌های تکاملی به‌ویژه در مسائل با فضای جست‌وجوی وسیع عملکرد خوبی دارند و در برخی موارد می‌توانند بهتر از روش‌های مبتنی بر RL عمل کنند، به‌ویژه در مسائل چندهدفه و غیرخطی. مدل‌های ریاضی سنتی، اگرچه ساده‌تر و سریع‌تر در پیاده‌سازی هستند، در مواجهه با عدم قطعیت و سیستم‌های غیرخطی انعطاف‌پذیری کمتری دارند. در نهایت، انتخاب روش مناسب به پیچیدگی سیستم و نیاز به دقت و سرعت بستگی دارد [۵۱].

کاربرد الگوریتم‌ها و روش‌های مدل‌سازی، چه در سطح رگرسیونی و چه در زمینه‌های هوش مصنوعی و یادگیری ماشین، در سیستم‌های انرژی و مهندسی بسیار گسترده و حیاتی است [۵۲]. این روش‌ها به تحلیل و بهینه‌سازی عملکرد سیستم‌ها کمک کرده و زمینه‌ساز توسعهٔ راهکارهای هوشمند در این حوزه هستند. در تحقیقات آینده، می‌توان بر مقایسهٔ دقیق‌تر این رویکردها تمرکز بیشتری داشت تا با ارزیابی ویژگی‌ها و محدودیت‌های هر یک، دیدگاه جامع‌تری نسبت به انتخاب مدل بهینه بر اساس معیارهایی همچون سرعت پردازش و کاربردپذیری در شرایط مختلف به دست آید. این مقایسه‌ها می‌توانند به انتخاب بهترین

رویکرد برای سیستم‌های پیچیدهٔ انرژی کمک کنند. در کنار استفاده از این مدل‌ها، می‌توان در سیستم‌های انرژی، از جمله منابع تجدیدپذیر مانند انرژی خورشیدی، به مسائل زیست‌محیطی نیز توجه ویژه‌ای داشت. ترکیب بهینهٔ انرژی و محیط زیست نه تنها به کاهش مصرف سوخت‌های فسیلی و انتشار گازهای گلخانه‌ای کمک می‌کند، بلکه راه‌حل‌هایی هوشمندانه برای حفظ پایداری منابع طبیعی و بهبود کیفیت محیط زیست در آینده فراهم می‌آورد. مطالعات مختلفی به جنبه‌های نوین محیط زیستی اشاره کرده‌اند که می‌توانند به رویکردهای بهینه‌سازی اضافه شوند تا بهبود بهره‌وری انرژی همراه با کاهش اثرات منفی زیست‌محیطی تحقق یابد [۵۳ و ۵۴].

۵. نتیجه‌گیری و بحث تحلیلی

یادگیری تقویتی به عنوان یک روش پیشرو در هوش مصنوعی، پتانسیل بالایی در بهینه‌سازی سیستم‌های انرژی نشان داده است. این روش با توانایی یادگیری سیاست‌های بهینه از طریق تعامل با محیط‌های پویا، راه‌حل‌هایی کارآمد برای مدیریت شبکه‌های هوشمند، بهینه‌سازی ذخیره‌سازی انرژی، و کاهش مصرف در ساختمان‌ها ارائه کرده است. مطالعات نشان داده‌اند RL می‌تواند هزینه‌های عملیاتی شبکه‌های توزیع را تا ۱۰ ـ ۱۵ درصد کاهش دهد، کارایی باتری‌ها را در سیستم‌های تجدیدپذیر بهبود بخشد، و مصرف انرژی سیستم‌های HVAC را تا ۲۰ درصد کم کند. برتری RL نسبت به روش‌های سنتی، مانند برنامه‌ریزی خطی یا الگوریتم‌های تکاملی، در انعطاف‌پذیری آن برای مدیریت مسائل غیرخطی و متغیر نهفته است. این ویژگی، RL را به ابزاری نوآورانه برای سیستم‌های انرژی با عدم قطعیت بالا، مانند نوسانات تولید انرژی خورشیدی و بادی، تبدیل کرده است [۵۵].

پیاده‌سازی RL در سیستم‌های انرژی با چالش‌هایی مواجه است. نیاز به توان محاسباتی بالا، به‌ویژه در یادگیری تقویتی عمیق (Deep RL) و زمان‌بر بودن فرایند آموزش، محدودیت‌های اصلی هستند. طراحی محیط‌های شبیه‌سازی‌شده دقیق برای آموزش عامل RL، به‌خصوص در سیستم‌های پیچیده مانند میکروشبکه‌ها، نیازمند داده‌های باکیفیت و زیرساخت‌های محاسباتی قوی است. علاوه بر این، ناپایداری در فرایند آموزش، مانند مشکل تعادل بین اکتشاف و بهره‌برداری، می‌تواند کارایی RL را در کاربردهای عملی کاهش دهد. این چالش‌ها نشان‌دهندهٔ نیاز به توسعهٔ الگوریتم‌های کارآمدتر و روش‌های آموزشی با مصرف منابع کمتر است [۵۶].

جهت‌گیری‌های آینده RL در بهینه‌سازی انرژی می‌تواند بر ادغام با فناوری‌های نوین متمرکز شود. ترکیب RL با اینترنت اشیا امکان جمع‌آوری داده‌های بی‌درنگ از سنسورهای انرژی را فراهم می‌کند، که می‌تواند دقت تصمیم‌گیری را بهبود بخشد. همچنین، استفاده از بلاک‌چین برای مدیریت غیرمتمرکز انرژی، مانند معاملات در بازارهای محلی، می‌تواند با RL ادغام شود تا سیستم‌های امن و شفاف ایجاد کند. توسعهٔ الگوریتم‌های RL کم‌مصرف، مانند یادگیری آفلاین یا مدل‌های فشرده‌شده، نیز می‌تواند ردپای محاسباتی را کاهش دهد و کاربرد RL را در سیستم‌های انرژی پایدار گسترش دهد [۵۷].

در مجموع، یادگیری تقویتی با ارائهٔ راه‌حل‌های هوشمند و انعطاف‌پذیر، آینده‌ای روشن در بهینه‌سازی انرژی دارد. با غلبه بر چالش‌های محاسباتی و بهره‌گیری از فناوری‌های نوین، RL می‌تواند به یک استاندارد در مدیریت انرژی پایدار تبدیل شود، به‌ویژه در کشورهایی مانند ایران که نیاز به راه‌حل‌های نوآورانه برای مدیریت انرژی دارند.

منابع

[1] Almutairi A, Abiyo S, Hyejian J. Secured and Smart System for Energy Management in Microgrids Using Deep Reinforcement Learning. IEEE Transactions on Consumer Electronics. 2025 Jun 4.

[2] Vamvakas D, Michailidis P, Korkas C, Kosmatopoulos E. Review and evaluation of reinforcement learning frameworks on smart grid applications. Energies. 2023 Jul 12;16(14):5326.

[3] Moaven, M., Allahrabbi Shirazi, M., & Ghodusinejad, M. H. (2025). Systematic Analysis of the Impact of Life Cycle Assessment of Materials on Urban Heat Island Mitigation: A Path Toward Sustainable Urban Policy Using the PRISMA Method. Urban Development Policy Making, 2(1), 95-110.

[4] Sutton RS, Barto AG. Reinforcement learning: An introduction. Cambridge: MIT press; 1998 Mar 1.

[5] Glavic M, Fonteneau R, Ernst D. Reinforcement learning for electric power system decision and control: Past considerations and perspectives. IFAC-PapersOnLine. 2017 Jul 1;50(1):6918-27.

[6] Jahani MT, Nazarian P, Safari A, Haghifam MR. Multi-objective optimization model for optimal reconfiguration of distribution networks with demand response services. Sustainable Cities and Society. 2019 May 1;47:101514.

[7] Maeda I, DeGraw D, Kitano M, Matsushima H, Sakaji H, Izumi K, Kato A. Deep reinforcement learning in agent based financial market simulation. Journal of Risk and Financial Management. 2020 Apr 11;13(4):71.

[8] Russell SJ, Norvig P. Artificial intelligence: A modern approach;[the intelligent agent book]. Prentice hall; 1995.

[9] Gupta A, Badr Y, Negahban A, Qiu RG. Energy-efficient heating control for smart buildings with deep reinforcement learning. Journal of Building Engineering. 2021 Feb 1;34:101739.

[10] Kumar A, Maulik A, Chinmaya KA. Energy Management Strategies for Active Distribution Networks and Microgrids–A Comprehensive Survey. IETE Technical Review. 2025 Jun 28:1-40.

[11] Liang, Jiaxin, et al. "A review of multi-agent reinforcement learning algorithms." Electronics 14.4 (2025): 820.

[12] Shojaeighadikolaei, Amin, et al. "Distributed Energy Management and Demand Response in Smart Grids: A Multi-Agent Deep Reinforcement Learning Framework." arXiv preprint arXiv:2211.15858 (2022).

[13] Arwa EO, Folly KA. Reinforcement learning techniques for optimal power control in grid-connected microgrids: A comprehensive review. Ieee Access. 2020 Nov 17;8:208992-9007.

[14] Duan J, Yi Z, Shi D, Lin C, Lu X, Wang Z. Reinforcement-learning-based optimal control of hybrid energy storage systems in hybrid AC–DC microgrids. IEEE Transactions on Industrial Informatics. 2019 Jan 31;15(9):5355-64.

[15] Latoń D, Grela J, Ożadowicz A. Applications of Deep Reinforcement Learning for Home Energy Management Systems: A Review. Energies. 2024 Dec 20;17(24):6420.

[16] Saifoddin A, Mirzaei N, Allahrabbi Shirazi M, Yousefi H. Comparative Applications of Supervised and Unsupervised Machine Learning Models in Energy Systems. Journal of Energy Management and Technology. 2025 Dec 1;9(4):284-90.

[17] Subramanya R, Sierla SA, Vyatkin V. Exploiting battery storages with reinforcement learning: a review for energy professionals. IEEE Access. 2022 May 18;10:54484-506.

[18] Vázquez-Canteli JR, Nagy Z. Reinforcement learning for demand response: A review of algorithms and modeling techniques. Applied energy. 2019 Feb 1;235:1072-89.

[19] Michailidis P, Michailidis I, Kosmatopoulos E. Reinforcement Learning for Electric Vehicle Charging Management: Theory and Applications. Energies. 2025 Oct 1;18(19):5225.

[20] Duan Y, Chen X, Houthooft R, Schulman J, Abbeel P. Benchmarking deep reinforcement learning for continuous control. InInternational conference on machine learning 2016 Jun 11 (pp. 1329-1338). PMLR.

[21] Giannelos S. Reinforcement Learning in Energy Finance: A Comprehensive Review. Energies (19961073). 2025 Jun 1;18(11).

[22] Mohammaddini S, Yousefi H, Abdoos M, Shirazi MA, Hajinezhad A. Techno-economic simulation of solar flat plate collector systems for building hot water demand supply. Energy. 2025 Oct 16:138933.
[23] Singh D, Shah OA, Arora S. Adaptive control strategies for effective integration of solar power into smart grids using reinforcement learning. Energy Storage and Saving. 2024 Dec 1;3(4):327-40.
[24] Wang D, Zheng W, Wang Z, Wang Y, Pang X, Wang W. Comparison of reinforcement learning and model predictive control for building energy system optimization. Applied Thermal Engineering. 2023 Jun 25;228:120430.
[25] Dulac-Arnold G, Evans R, van Hasselt H, Sunehag P, Lillicrap T, Hunt J, Mann T, Weber T, Degris T, Coppin B. Deep reinforcement learning in large discrete action spaces. arXiv preprint arXiv:1512.07679. 2015 Dec 24.
[26] Stavrev S, Ginchev D. Reinforcement learning techniques in optimizing energy systems. Electronics. 2024 Apr 12;13(8):1459.
[27] Bui VH, Das S, Hussain A, Hollweg GV, Su W. A critical review of safe reinforcement learning techniques in smart grid applications. arXiv preprint arXiv:2409.16256. 2024 Sep 24.
[28] Jin, Ming. "Reinforcement Learning Meets the Power Grid: A Contemporary Survey with Emphasis on Safety and Multi-agent Challenges." Foundations and Trends® in Electric Energy Systems 8.3-4 (2025): 169-316.
[29] Bui, Van-Hai, et al. "A critical review of safe reinforcement learning strategies in power and energy systems." Engineering Applications of Artificial Intelligence 143 (2025): 110091.
[30] Tabas, Daniel, and Baosen Zhang. "Computationally efficient safe reinforcement learning for power systems." 2022 American Control Conference (ACC). IEEE, 2022.
[31] Ceusters, Glenn, et al. "Safe reinforcement learning for multi-energy management systems with known constraint functions." Energy and AI 12 (2023): 100227.
[32] Zhang H, Sun X, Lee MH, Moon J. Deep reinforcement learning-based active network management and emergency load-shedding control for power systems. IEEE Transactions on Smart Grid. 2023 Aug 8;15(2):1423-37.
[33] Guan Y, Ma W, Che L, Shahidehpour M. Model-Based Safe Reinforcement Learning for Active Distribution Network Scheduling. IEEE Transactions on Smart Grid. 2025 Mar 17.
[34] Allahrabbi Shirazi, M. A., Goldoust, A., Khatami, M., Abedi, E., & Janfeshan, M. H. (2024). Modeling and simulation of a solar tracker with bifacial panel: a case study of Tehran city. Journal of Sustainable Energy Systems, 3(3), 271-287.
[35] Dang Q, Wu D, Boulet B. A q-learning based charging scheduling scheme for electric vehicles. In2019 IEEE Transportation Electrification Conference and Expo (ITEC) 2019 Jun 19 (pp. 1-5). IEEE.
[36] Xu S, Fu Y, Wang Y, Yang Z, Huang C, O'Neill Z, Wang Z, Zhu Q. Efficient and assured reinforcement learning-based building HVAC control with heterogeneous expert-guided training. Scientific reports. 2025 Mar 5;15(1):7677.
[37] Samadi E, Badri A, Ebrahimpour R. Q-Learning-Oriented Distributed Energy Management of Grid-Connected Microgrid. In2021 29th Iranian Conference on Electrical Engineering (ICEE) 2021 May 18 (pp. 318-322). IEEE.
[38] Ministry of Energy / Tavanir Company. Research Priorities of Iran's Ministry of Energy in 2021. Tehran. 2021; Available: https://gsme.semnan.ac.ir/uploads/23/2021/Aug/21/tavanir_1.docx (In persia).
[39] Asghari Oskoei M, Fallahi F, Doostizadeh M, Moshiri S. Reinforcement Learning Applied to Multi Agent Modelling, the Case of the Iranian Power Market. Iranian Energy Economics. 2017;7(25):1-40.
[40] Mousavi Ziabari Z, Azmi R. Demand Forecasting for Dynamic Pricing in Smart Electricity Markets. 10th Conference on Information Technology and Knowledge (IKT2019). Tehran. 2019.
[41] Mohsenzadeh-Yazdi, Hossein, Hamed Kebriaei, and Farrokh Aminifar. "Multi-agent reinforcement learning in a new transactive energy mechanism." IET Generation, Transmission & Distribution 18.18 (2024): 2943-2955
[42] Lissa P, Deane C, Schukat M, Seri F, Keane M, Barrett E. Deep reinforcement learning for home energy management system control. Energy and AI. 2021 Mar 1;3:100043.

[43] Salmi C, Senoussi NE, Chaal D, Boudjadi M. An in-depth Investigation Into the Application of Flash Memory In a Business Intelligence Database Environment.
[44] Su Y, Yue S, Qiu L, Chen J, Wang R, Tan M. Energy management for scalable battery swapping stations: A deep reinforcement learning and mathematical optimization cascade approach. Applied Energy. 2024 Jul 1;365:123212.
[45] Li P, Hao J, Tang H, Fu X, Zhen Y, Tang K. Bridging evolutionary algorithms and reinforcement learning: A comprehensive survey on hybrid algorithms. IEEE Transactions on evolutionary computation. 2024 Aug 14.
[46] Mulayim, Ozan Baris, et al. "Comparative Field Deployment of Reinforcement Learning and Model Predictive Control for Residential HVAC." arXiv preprint arXiv:2510.01475 (2025).
[47] Gokhale, Gargya, Bert Claessens, and Chris Develder. "PhysQ: a physics informed reinforcement learning framework for building control." arXiv preprint arXiv:2211.11830 (2022).
[48] Xiong, Rui, Jiayi Cao, and Quanqing Yu. "Reinforcement learning-based real-time power management for hybrid energy storage system in the plug-in hybrid electric vehicle." Applied energy 211 (2018): 538-548.
[49] Im, Jaedong, et al. "Reinforcement learning-based energy management system in the complex electric tariff environment." International Journal of Electrical Power & Energy Systems 172 (2025): 111038.
[50] Zhang, Haifeng, et al. "Resilient dispatching optimization of power system driven by deep reinforcement learning model." Discover Artificial Intelligence 5.1 (2025): 189.
[51] Xu Y, Li Y, Gao W. Comparative Analysis of Reinforcement Learning Approaches for Multi-Objective Optimization in Residential Hybrid Energy Systems. Buildings. 2024;14(9):2645. doi: 10.3390/buildings14092645.
[52] Naziri A, Tahavvor AR, Shirazi MA, Zahedi R. Feasibility study on heat recovery from gas turbine exhaust for absorption chiller operation and efficiency enhancement using neural networks. Thermal Science and Engineering Progress. 2025;104210.
[53] Shirazi MA, Zahedi R, Yousefi H, Aslani A. Environmental and damage assessment of electric vehicles compared to internal combustion engine vehicles under various ambient temperature scenarios using the LCA approach. Energy Nexus. 2025 Nov 19;100606.
[54] Allah Rabbi Shirazi MA. Idea generation and examination of environmental challenges of floating solar photovoltaic power plants on wetlands and its economic advantage for local communities. Journal of Sustainable Energy Systems. 2024;3(1):39-51.
[55] Pallonetto F, De Rosa M, Milano F, Finn DP. Demand response algorithms for smart-grid ready residential buildings using machine learning models. Applied energy. 2019 Apr 1;239:1265-82.
[56] Zhang Z, Zhang D, Qiu RC. Deep reinforcement learning for power system applications: An overview. CSEE Journal of Power and Energy Systems. 2019 Oct 7;6(1):213-25.
[57] Ozan AT, Kamalaruban P. Applications of reinforcement learning in energy systems. Renewable and Sustainable Energy Reviews. 2021 Mar 1;137:110618.