

Reinforcement Learning for Integrated MPPT and Battery Management in Photovoltaic Systems: A Systematic Review

Francisco Fidalgo and Ramiro Barbosa * 

Department of Electrical Engineering, ISEP, Polytechnic of Porto, Rua Dr. António Bernardino de Almeida, 431, 4249-015 Porto, Portugal

* Correspondence: rsb@isep.ipp.pt; Tel.: +351-228-340-500

Abstract

This work provides a systematic literature review on reinforcement learning (RL) for integrated maximum power point tracking (MPPT) and battery management in photovoltaic (PV) systems. As PV installations increasingly incorporate battery energy storage, the traditional objective of maximizing instantaneous power extraction is no longer sufficient on its own, since control decisions also affect battery state of charge, efficiency, degradation, and load support. Although RL has shown strong potential for sequential decision making in energy systems, most existing studies still treat MPPT and battery management as separate or only loosely coordinated problems. This review examines this issue by systematically examining how RL, particularly continuous-action methods, has been applied to coupled PV–battery control. The analysis highlights the shortcomings of discrete-action formulations in power-electronic systems and emphasizes the advantages and limitations of actor–critic approaches such as DDPG, TD3, PPO, and SAC for directly optimizing continuous-control variables. Across the reviewed literature, RL is found to be used predominantly at the supervisory energy management level, with PV generation often treated as exogenous rather than as an explicit control decision. This review therefore identifies a persistent structural separation between converter-level PV control and storage-aware energy management within a common learning and evaluation framework. It identifies continuous-action RL as a candidate formulation for unified PV–battery optimization while highlighting important challenges in constraint handling, state representation, sample efficiency, stability, and hardware validation.

Keywords: photovoltaic systems; MPPT; battery storage; continuous-action reinforcement learning; deep learning; actor–critic; SoC



Academic Editor: Marcin Kaminski

Received: 19 June 2026

Revised: 20 July 2026

Accepted: 4 August 2026

Published: 7 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The rapid expansion of PV systems, coupled with the increasing deployment of battery energy storage, has significantly transformed the control and optimization requirements of modern energy systems. While early research primarily focused on maximizing instantaneous power extraction through MPPT, contemporary PV–battery architectures demand coordinated control strategies that account for system-level dynamics, operational constraints, and long-term performance objectives.

Conventional MPPT techniques, such as Perturb and Observe (P&O) and Incremental Conductance (INC), remain widely used due to their simplicity and low computational cost. These methods iteratively adjust the operating point of the PV array to track the maximum power point under varying environmental conditions [1,2]. However, their reliance on

local perturbation mechanisms leads to well-known limitations, including steady-state oscillations, sensitivity to measurement noise, and degraded performance under rapidly changing irradiance [3,4]. Under partial shading, where multiple local maxima arise, these methods often converge to sub-optimal operating points [5,6]. Further, when considered as an isolated control objective, conventional MPPT generally focuses on instantaneous PV power extraction and does not by itself represent the effects of present operating-point decisions on future battery and system states.

To address these limitations, data-driven approaches based on artificial neural networks (ANNs) and deep learning (DL) have been introduced. These methods approximate the nonlinear relationship between PV measurements and optimal operating points, enabling improved tracking accuracy and reduced oscillations compared with classical algorithms [1,7]. Despite these advantages, many supervised learning-based MPPT formulations depend on offline training data and may exhibit limited adaptation when operating conditions differ from those represented during training. Moreover, such approaches typically learn static input–output mappings and do not explicitly account for sequential decision making or long-term system objectives.

Reinforcement learning has gained attention as an alternative approach by formulating MPPT as a sequential decision-making problem. Early RL-based approaches employ discrete-action formulations, where the control problem is modeled as a finite Markov Decision Process (MDP) and solved using algorithms such as Q-learning or SARSA. These methods demonstrate improved adaptability and reduced oscillatory behavior compared with perturbation-based techniques [8–10]. Extensions based on deep Q-networks (DQNs) further enhance scalability by approximating the action-value function with neural networks, enabling improved performance under complex operating conditions [11,12].

Even so, discrete-action RL can introduce limitations when inherently real-valued control variables are represented using a finite action set. Coarse discretization may restrict control resolution and can produce quantization effects or oscillatory behavior, whereas finer discretization increases the size of the state–action space and may aggravate the curse of dimensionality [13]. These trade-offs are particularly relevant when the RL action represents a real-valued converter command, such as a duty-ratio, voltage, current, or power reference. However, discrete actions remain appropriate for operating-mode selection, switching decisions, and explicitly enumerated feasible actions.

It is also important to distinguish the RL action representation from the physical switching process. A duty ratio may be represented as a continuous policy output, while its implementation through PWM remains a discrete high-frequency switching process.

Continuous-action RL avoids explicit discretization by directly parameterizing real-valued control commands. Actor–critic methods, including deep deterministic policy gradient (DDPG), Twin Delayed DDPG (TD3), Proximal Policy Optimization (PPO), and Soft Actor–Critic (SAC), have reported favorable and, in some studies, superior performance relative to the selected baselines in MPPT applications, including faster convergence, reduced steady-state oscillations, and high tracking efficiency under the tested dynamic conditions [14–17]. Unlike static supervised mappings, these methods optimize a cumulative reward through interaction with the modeled environment and can therefore support adaptation to nonlinear and time-varying dynamics. These reported advantages are dependent on the action definition, reward design, training conditions, constraints, and comparison baseline and do not establish general superiority over discrete, mixed-action, or conventional control approaches.

Despite these advances, the majority of existing RL-based MPPT studies focus on standalone PV systems. In practical deployment, PV arrays are increasingly integrated with battery storage and variable loads, where control decisions influence not only instan-

taneous power extraction but also the evolution of battery state of charge (SoC), system efficiency, and component lifetime. In such systems, the traditional objective of maximizing instantaneous PV power may conflict with broader operational goals, such as preventing battery over-charge, minimizing degradation, or satisfying load demand. Consequently, the MPPT objective or its resulting operating reference may need to be coordinated with battery and system-level objectives, particularly when storage or converter constraints become active. Such coordination may be implemented through hierarchical, cascaded, supervisory, or joint-decision architectures.

This interaction motivates an examination of how PV operating-point control and battery-aware objectives are represented across the existing literature. Within the corpus reviewed in this article, many RL formulations optimize supervisory battery, grid, or load decisions while leaving MPPT to a separate controller. Relatively few reviewed studies include converter-level PV operating-point decisions and battery-aware long-horizon objectives within the same learning and evaluation framework.

To examine these questions, this article presents a structured review of RL methods related to MPPT and battery management. This review maps the reported architectures, action variables, control layers, constraint treatments, and validation approaches and summarizes the study-specific evidence reported for discrete-, continuous-, and mixed-action spaces. Particular attention is given to the assumptions, limitations, and validation requirements associated with each formulation.

The contribution of this work is twofold:

- A critical analysis of discrete- and continuous-action RL approaches, highlighting their limitations and suitability for power-electronic control;
- A structured review of integrated PV–battery control strategies, identifying open challenges and future research directions for continuous-action RL.

The remainder of this paper is structured as follows: Section 2 examines the relationship between MPPT and battery management in PV–battery systems and motivates the need for integrated control formulations. Section 3 describes the PRISMA-guided review methodology adopted in this study. Section 4 presents the results of the systematic review and synthesizes the literature in relation to the research questions. Section 5 discusses the major research gaps and future opportunities for continuous-action RL in integrated PV–battery control. Finally, Section 6 concludes the paper.

2. Integrated MPPT and Battery Management

The integration of battery energy storage fundamentally alters the objective of MPPT in photovoltaic systems. While conventional MPPT maximizes instantaneous PV power, PV–battery systems introduce additional constraints related to SoC, charging limits, efficiency, and lifetime. Consequently, MPPT may no longer be treated as an isolated control task but ought to be embedded within a broader system-level optimization framework.

2.1. Coupling Between MPPT and Battery Dynamics

PV–battery systems may be implemented using DC-coupled or AC-coupled architectures, but the coupling between MPPT and storage dynamics is particularly strong in DC-coupled configurations, where the PV array and battery share a common DC bus [18]. In such systems, the selection of the PV operating point directly affects battery charging behavior, so enforcing strict MPPT without accounting for storage constraints may lead to excessive charging currents, unnecessary curtailment, inefficient energy utilization, or accelerated battery degradation [19,20].

Conventional control strategies typically adopt a hierarchical structure in which MPPT maximizes PV power at a lower level, while a supervisory controller independently man-

ages battery charging and load supply. Although this decoupled architecture simplifies implementation, it implicitly assumes that maximizing instantaneous PV power is consistent with broader system-level objectives. In practice, this assumption may fail, particularly when the battery approaches its operational limits, leading to sub-optimal energy routing, unnecessary cycling, or increased stress on storage components [21].

To mitigate these limitations, several coordinated architectures and supervisory strategies have been proposed. An example of such architecture is shown in Figure 1, illustrating a representative DC-coupled PV–battery system in which the PV array and battery energy storage system exchange power through a shared DC bus. Although the PV-side and battery-side converters are often controlled independently, their operation remains physically coupled through the common DC link. As a result, the operating point selected for the PV array directly influences battery charging power, state-of-charge evolution, and the amount of energy available for load support. Consequently, decisions traditionally associated with MPPT can have significant implications for battery utilization, efficiency, and degradation.

While the topology shown employs separate DC–DC converters for the PV array and battery, similar interactions exist in alternative architectures, including non-inverting four-switch buck–boost converters, bidirectional buck–boost interfaces, and integrated multiport converter topologies. Likewise, though the coupling of these subsystems is particularly evident in DC-coupled configurations, similar dependencies arise in many alternative architectures, including AC-coupled ones. Although these configurations differ in terms of power-flow flexibility and converter integration, they share the fundamental characteristic that PV generation and battery behavior cannot be considered entirely independent control problems.

From a control perspective, this architecture highlights a structural mismatch that persists throughout much of the literature. The physical system is inherently coupled, yet MPPT and battery management are frequently formulated as separate optimization tasks, even if sometimes coordinated. This observation motivates integrated control approaches in which the PV operating point is treated as an explicit decision variable alongside battery charging and discharging actions, enabling system-level objectives such as efficiency, battery lifetime, and load support to be considered jointly rather than sequentially.

Multiport converters, such as three-port DC–DC converters, enable unified coordination of PV, battery, and load interfaces, reducing redundant conversion stages and improving energy routing flexibility [22]. Likewise, SoC-aware supervisory strategies can dynamically adjust power references to prevent over-charge and over-discharge while reducing cycling stress [21,23]. However, even in these settings, many studies still optimize battery scheduling while assuming that MPPT operates independently. For example, Voigt et al. [24] improved storage scheduling through machine learning-based load forecasting and optimization, yet the PV operating point was not included among the storage-scheduling decision variables.

With that, it can be seen that the intrinsic coupling of PV operating-point selection, future SoC evolution, and load demand remains underexploited in many existing control frameworks. These limitations motivate integrated control strategies that explicitly account for the interaction among PV generation, storage dynamics, and system-level objectives.

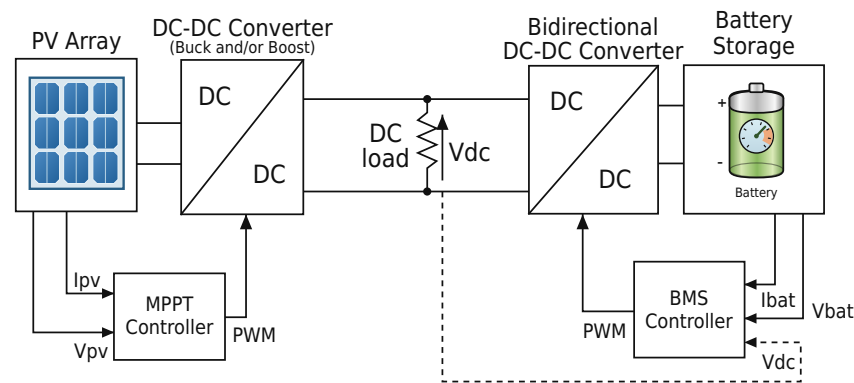


Figure 1. Representative DC-coupled PV–battery microgrid architecture (adapted from [25]). The PV array is interfaced through a DC–DC converter implementing MPPT control (using a given algorithm like P&O), while the battery is connected via a bidirectional DC–DC converter to a shared DC bus. The shared DC link illustrates the coupling between PV operating-point decisions and battery charging dynamics.

2.2. Limitations of Decoupled MPPT and Energy Management

Established PV–battery control systems frequently distribute MPPT, battery regulation, DC-bus control, and energy management functions across hierarchical or cascaded layers. Such structures simplify implementation, support time-scale separation, and can provide effective coordination when supervisory logic modifies operating references or control modes in response to SoC, load, or power-balance conditions.

In typical cascaded implementations, an MPPT or supervisory controller generates the operating reference, while faster inner voltage and current loops track that reference and regulate the converter and DC bus. A limitation arises when the MPPT and battery-management objectives are specified independently, such that maximizing instantaneous PV power is assumed to remain consistent with broader system-level objectives. This assumption may fail near battery-SoC limits or under variable load conditions, potentially leading to unnecessary curtailment, inefficient cycling, or increased battery stress [21]. Because present PV operating-point and storage decisions influence subsequent SoC evolution and constraint activation, PV–battery systems exhibit temporal interdependence. This motivates explicit coordination between MPPT and storage management, whether through hierarchical, cascaded, supervisory, or joint-decision control formulations.

Although RL is especially well suited to sequential PV–battery optimization, earlier ANN-based extensions also attempted to incorporate storage-aware behavior into MPPT-related control. Such approaches reported improvements in local dynamic performance, including smoother charging behavior and improved DC-link regulation in standalone or grid-integrated settings [25–27]. However, these methods remained fundamentally supervised and largely static, typically coordinating battery regulation and PV control without explicitly formulating the long-horizon coupling of PV operating-point selection, future SoC evolution, and system-level objectives. As a result, they only partially address the limitations of decoupled MPPT and energy management.

A representative non-RL example illustrating converter-level joint regulation is the DG–PV–BES islanded system proposed by Jha et al. [28]. In that architecture, the bidirectional buck–boost converter performs the dual function of PV MPPT and battery charging/discharging current regulation, thereby avoiding a fully decoupled arrangement in which MPPT and storage control are handled by separate converters or independent supervisory loops. The proposed scheme is particularly relevant because it demonstrates that PV operating-point regulation and battery protection can be coordinated within a shared electrical control structure while also being validated on an experimental prototype in variable-irradiance and -fault scenarios. Although the method is not learning-based,

it provides an important benchmark for this review by showing that unified PV–battery coordination is technically achievable at the converter layer and need not be restricted to supervisory scheduling formulations.

Figure 2 conceptually summarizes the changes from decoupled MPPT and battery-management architectures toward increasingly integrated control formulations. Most conventional PV–battery systems remain close to the left side of the spectrum, where MPPT and energy management are treated as distinct tasks. Coordinated approaches introduce supervisory layers that exchange information between subsystems but still preserve separate optimization objectives. In contrast, fully integrated formulations treat PV operating-point selection, battery charging/discharging decisions, and grid interaction as components of a single control problem. This distinction is particularly important because the degree of integration determines whether long-term interactions between energy harvesting, storage utilization, and future system states can be explicitly considered during decision making. Even so, these categories describe how objectives and decision variables are allocated; they do not represent a ranking of technical maturity or performance. Coordinated multi-layer architectures preserve dedicated inner control loops while exchanging references, measurements, and constraints across supervisory and converter levels. Joint-decision architectures instead determine multiple PV- and battery-side variables within a shared optimization or control framework. The appropriate structure depends on the application, time scales, safety requirements, model availability, and implementation constraints.

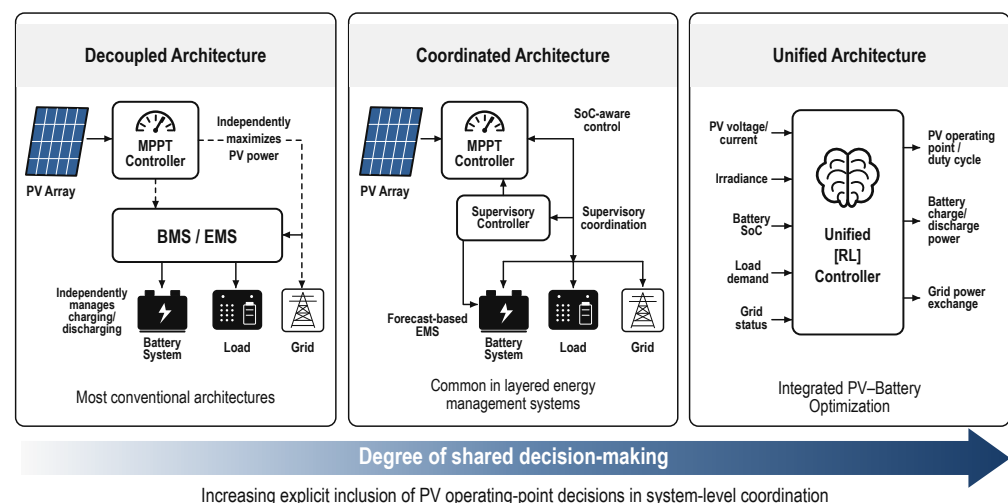


Figure 2. Different distributions of control authority and decision variables of PV–battery control architectures, from decoupled MPPT and battery management to integrated optimization. The leftmost panel represents conventional architectures in which MPPT and battery/EMS functions are treated independently. The middle panel illustrates coordinated or layered approaches where a supervisory controller influences both PV and battery operation while preserving separate control loops. The rightmost panel represents integrated formulations in which a unified controller, such as a reinforcement learning agent, jointly determines PV operating point, battery charging/discharging power, and grid interaction based on system-wide objectives and future-state considerations. The horizontal arrow indicates an increasing degree of shared decision making and does not necessarily imply a progression in technical maturity, stability, or performance.

2.3. Reinforcement Learning for Joint Optimization

Reinforcement learning provides one possible framework for representing PV–battery operation as a sequential decision-making problem. By modeling the system as an MDP, RL enables direct optimization of long-term objectives that incorporate PV power extraction, battery-SoC evolution, and operational constraints.

Several studies demonstrate the effectiveness of RL in energy management in PV–battery systems. Härtel and Bocklisch [29] applied a PPO-based controller for energy cost minimization, achieving performance comparable to model predictive control while outperforming rule-based strategies. Similarly, Xu et al. [30] employed model-based RL to optimize energy scheduling in residential PV–battery systems, improving self-consumption and operational efficiency.

Nevertheless, many works treat PV generation as an exogenous input, assuming that MPPT operates independently of storage management. This separation limits the ability of RL to fully exploit the coupling between PV operating-point selection and battery dynamics.

2.4. Continuous-Action Reinforcement Learning for MPPT

Continuous-action RL formulates MPPT as a sequential control problem in which the agent directly optimizes the converter duty cycle to maximize cumulative performance over time. The distinction between discrete- and continuous-action RL concerns the decision variable produced by the policy, rather than the binary switching state of the semiconductor devices. For example, a continuous-action policy may generate a real-valued duty-ratio command, whereas a discrete-action policy selects that command from a predefined finite set. In both cases, the commanded duty ratio is physically implemented through discrete high-frequency PWM switching. Accordingly, the discrete nature of PWM does not by itself determine the appropriate RL action-space formulation. Discrete actions may be suitable for switching modes or predefined feasible commands, whereas continuous actions may be useful when fine adjustment of a duty-ratio or electrical reference is required. In fact, recent advancements increasingly favor parameterized or hybrid action spaces that combine the deterministic stability of discrete switching with the precision of continuous tuning. Therefore, rather than declaring one approach universally superior, a system-specific evaluation of stability and execution constraints is essential.

Beyond that, and unlike supervised ANN- or deep learning-based approaches that learn a static mapping from measurements to operating points, RL learns a policy through interaction with the PV–converter environment, typically modeled as an MDP. State representations commonly include PV voltage, current, power, and duty-cycle history, while actions are expressed as continuous duty-cycle values or increments. Reward functions generally favor high power extraction while penalizing oscillations, excessive control effort, or other undesirable dynamics [31].

Among continuous-action methods, DDPG has been widely adopted because its actor–critic structure enables direct continuous control. Artetxe et al. [14] reported faster convergence and reduced steady-state oscillations than P&O in both simulation and real-time experiments. However, vanilla DDPG is often sensitive to hyperparameters and may exhibit unstable learning dynamics [32]. To address these limitations, more robust actor–critic variants have been explored, including TD3, which mitigates value overestimation through twin critics and delayed updates; PPO, which improves policy update stability through clipped objectives; and SAC, which enhances exploration robustness through entropy regularization. These structural modifications are intended to reduce particular training pathologies, and reported results indicate reduced settling time and steady-state efficiency exceeding 99% under dynamic irradiance conditions [17]. Rostamipour et al. [33] further proposed an SAC-based MPPT approach augmented with transfer learning, enabling faster adaptation across irradiance and shading scenarios while reducing ripple compared with DDPG and P&O. Building on these continuous-action formulations, Figure 3 presents a generalized RL-based MPPT control architecture in which measured PV variables are processed into a state representation, the agent outputs a continuous duty-cycle increment, and the resulting duty cycle is applied to the converter through the PWM stage.

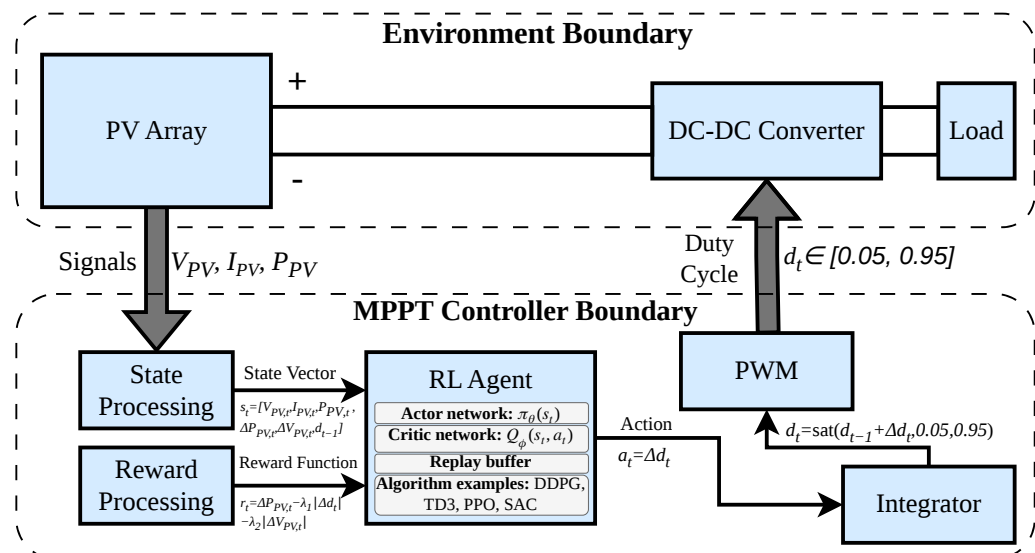


Figure 3. Continuous-action reinforcement learning architecture for the MPPT control of a PV-fed DC–DC converter. The environment consists of the PV array, the DC–DC converter, and the load, while the MPPT controller receives the measured PV variables V_{PV} , I_{PV} , and P_{PV} . These measurements are processed into a state vector s_t , which is supplied to a continuous-action RL agent. The agent outputs a duty-cycle increment $a_t = \Delta d_t$, which is integrated and saturated to generate the converter duty cycle $d_t \in [0.05, 0.95]$. The reward function r_t promotes increased PV power extraction while penalizing excessive duty-cycle variation and oscillatory behavior. The RL agent may be implemented using actor–critic algorithms such as DDPG, TD3, PPO, and SAC. The continuous duty-ratio representation shown at the policy output should be distinguished from the discrete switching events generated by the PWM implementation.

2.4.1. Reward Design and Stability Considerations

Reward design critically shapes continuous-action RL behavior in MPPT. Typical formulations maximize extracted PV power while penalizing excessive current, abrupt duty-cycle variation, or in storage-integrated settings, constraint violations related to battery operation [34]. Because actions influence future system states, the reward implicitly encodes long-term objectives rather than instantaneous optimization alone.

However, actor–critic methods generally provide limited formal stability guarantees. Prior studies report sensitivity to hyperparameters and reward scaling, particularly in DDPG-based implementations [32], and stability assessment remains largely empirical [35]. Robust convergence through principled reward shaping, regularization, and constraint-aware policy design therefore remains an open challenge.

2.4.2. Simulation-to-Real Transfer and Hardware Validation

Most RL-based MPPT studies rely on MATLAB/Simulink environments with simplified PV and converter models. Although such environments are suitable for controlled evaluation, they often neglect sensor noise, switching losses, parameter drift, and other non-idealities, contributing to a persistent simulation-to-reality gap [36]. Hardware validation is typically limited to laboratory prototypes or hardware-in-the-loop platforms [37], and systematic sim-to-real methodologies remain underdeveloped.

For example, Karaki et al. [38] reported improved disturbance resilience using a deep-RL-based MPPT strategy in a fault-tolerant control framework, but validation remained primarily simulation-based. Bridging this sim-to-real gap remains a major barrier to practical deployment in PV systems.

2.4.3. Critical Assessment

Continuous-action RL might provide a suitable formulation when the selected decision variable is represented as a real-valued voltage, current, power, or duty-ratio command. In such cases, actor–critic methods avoid the need to define a finite action grid and can represent fine variations in the commanded reference. Several reviewed MPPT studies report favorable tracking performance, reduced steady-state oscillations, or improved adaptation under nonlinear and time-varying operating conditions. These reported benefits are relevant to applications in which control resolution is important, but they remain dependent on the selected baseline, reward design, training procedure, and evaluation conditions.

This consideration is also relevant in PV–battery systems, where decision variables such as PV voltage or duty ratio, battery charging/discharging power, and current references may be represented as continuous quantities. Compared with finite discretizations of these variables, continuous-action methods can avoid explicit command-level quantization and may scale more conveniently when several real-valued actions are optimized jointly [39]. For example, Jin et al. [40] reported lower operating cost and faster convergence for SAC than for the DQN in the microgrid energy management problem considered in their study. This result supports the potential value of continuous actions in that particular architecture and comparison setting, rather than establishing a general advantage over all discrete-action methods.

Discrete actions nevertheless remain appropriate for operating-mode selection, switching decisions, and explicitly enumerated feasible commands, while mixed-action spaces may be preferable when categorical decisions must be combined with real-valued references. Continuous policies also introduce important challenges, including training instability, sample inefficiency, sensitivity to reward and state design, constraint enforcement, limited formal guarantees, computational requirements, and hardware validation.

Accordingly, continuous-action RL may offer a useful representation for PV–battery control problems involving multiple real-valued decision variables, but it should not be regarded as a universally superior solution or as a replacement for established conventional controllers. Its engineering value must be established through controlled comparisons with discrete-action, mixed-action, and conventional methods under common architectures, constraints, operating conditions, and performance metrics.

2.5. Continuous-Action RL in Integrated-Control Studies

Continuous-action RL is capable of being particularly well suited for integrated MPPT and battery management, as key control variables, such as converter duty cycle and battery charging power, are inherently continuous. Actor–critic methods such as DDPG, PPO, and SAC enable direct optimization of these variables without requiring discretization.

Zhou et al. [41] demonstrated that DDPG-based continuous control improves coordination between PV generation, battery dispatch, and load demand, achieving better supply–demand balance compared with rule-based approaches. Similarly, SAC-based methods have been shown to enhance robustness and convergence in stochastic environments, while supporting constraint-aware optimization [42,43].

Compared with finite discrete-action formulations, continuous RL can offer finer control resolution, reduce command-level quantization effects, and support smoother coordination across coupled subsystems under the evaluated conditions.

Recent non-RL work also illustrates a stronger form of cross-layer coordination than is typically observed in the RL literature. Rizki et al. [44] proposed a standalone PV–battery DC microgrid framework that combines hybrid TCN–LSTM forecasting, predictive energy management, hierarchical load shedding, and converter-level double-integral sliding-mode control within a single coordinated architecture. In that formulation, forecasts of irradiance,

temperature, and load are used to generate PV and battery reference trajectories proactively, while the boost and bidirectional converters track those references through robust low-level control. This type of design is important for the present review because it occupies an intermediate position between fully decoupled EMS formulations and truly unified single-controller PV–battery co-control: PV is not treated as merely exogenous, yet converter-level execution remains layered beneath a supervisory predictive EMS. As such, it provides a useful comparator for assessing how far existing integrated control architectures have progressed, even outside RL.

Additional non-RL studies further reinforce this point, although more selectively. Dagal et al. [45] addressed a PV–battery system under partial shading using a novel gray-wolf-optimization-based MPPT strategy that explicitly incorporates battery state-of-charge considerations into the control objective, reporting high tracking efficiency and reduced ripple in simulation. Likewise, Tavoosi et al. [46] considered a hybrid PV/battery switching system in which separate control inputs regulate the PV-side boost converter and the battery-side converter, using an evolutionary-algorithm-tuned sliding-mode controller to improve transient and steady-state behavior. These studies are relevant because they show that non-RL intelligent or nonlinear control methods can coordinate PV-side and battery-side actuation more tightly than most supervisory EMS formulations; however, they remain primarily converter-control papers rather than long-horizon sequential optimization frameworks.

2.6. Challenges and Research Opportunities

Despite its advantages, integrated RL-based control for PV–battery systems faces several challenges:

- Constraint handling: Battery-SoC limits, power bounds, and safety constraints must be explicitly enforced.
- Training complexity: Learning in high-dimensional, coupled environments increases sample requirements.
- Sim-to-real gap: Most studies rely on simulation, with limited hardware validation.
- State representation: Accurate estimation of battery states (e.g., SoC and SoH) is critical to reliable policy behavior.

These challenges indicate that while continuous-action RL provides a potentially promising framework for unified MPPT and energy management, further research is required to ensure robustness, scalability, and practical deployability. All the same, the reviewed evidence does not establish that an RL controller should necessarily replace mature cascaded, sliding-mode, MPC, droop-based, or other nonlinear control methods.

A practically defensible extension of continuous-action RL-based MPPT is to incorporate battery-aware objectives through coordinated supervisory decision making or reference generation. Under uncertain, multi-objective, and temporally coupled operating conditions, an RL policy may determine PV and battery references, while established voltage, current, DC-bus, and protection loops retain responsibility for fast and safety-critical regulation. In this way, continuous-action RL can be investigated as a cross-layer coordination mechanism rather than as a direct replacement for mature converter controllers. Hybrid implementations may further combine the learned policy with model-based constraints, safety filters, or fallback control laws.

This extension remains to be validated through controlled comparisons that determine whether and under which conditions the inclusion of battery-aware, continuous decision variables provides measurable benefits over separate RL formulations, discrete- or mixed-action methods, and conventional hierarchical control. Such comparisons should use common system architectures, operating scenarios, constraints, time scales, base-

lines, and performance metrics, followed where possible by hardware-in-the-loop or experimental validation.

Such a shift may reinforce an eventual premise of continuous-action RL as a central framework for real-time, system-level optimization in renewable energy systems.

3. Methodology

This section presents the systematic literature review conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [47]. No review protocol was registered. The objective of this review is to ensure a transparent, reproducible, and comprehensive identification of relevant research related to: (i) photovoltaic (PV) and PV–battery energy systems; (ii) maximum power point tracking (MPPT), energy management, and power-electronic control strategies; and (iii) intelligent and learning-based control methods.

With an initial search strategy capturing a broad range of AI-based MPPT and optimization approaches to position the state of the art, the final synthesis prioritizes RL-based formulations, particularly those addressing sequential decision making, continuous-control actions, and battery-integrated PV systems.

3.1. Systematic Review Questions (PRQs)

To guide the systematic review and ensure alignment with the objectives of this work, the following PRISMA research questions (PRQs) were defined:

- PRQ1: In PV–battery systems with variable loads, what limitations are reported for decoupled control architectures where MPPT and battery/EMS control are designed independently?
- PRQ2: What control variables and actuation levels are used in learning-based approaches for PV–battery systems (e.g., PV duty cycle/operating point, battery current/power, and DC-bus voltage), and how often is MPPT treated as a controllable decision rather than an exogenous process?
- PRQ3: How are reinforcement learning formulations constructed for joint PV–battery optimization (state, action space, reward design, and constraint handling), and what evidence comparing discrete-action and continuous-action control in this context exists?
- PRQ4: What practical implementation considerations are addressed in the literature for RL-based PV–battery control (training environments, constraint enforcement mechanisms, observability/SoC estimation, robustness, and validation beyond simulation)?

The PRQs define the coding structure and analytical lens of the systematic review. The evidence synthesized in response to PRQ1–PRQ4 is used to map the reported control architectures, identify comparatively underrepresented combinations, and define requirements for future comparative and experimental evaluation. These findings also inform broader research questions and may motivate the future formulation and assessment of continuous-action RL frameworks for integrated PV–battery control.

3.2. Search Strategy

The literature search was conducted across three major scientific databases: IEEE Xplore, ScienceDirect, and SpringerLink. Google Scholar was used as an indexing database to ensure broad coverage and cross-validation of records retrieved from publisher-specific sources. Google Scholar was additionally used for backward and forward citation tracking to identify relevant studies not captured in the initial database queries.

Four groups of keywords were defined to represent the main conceptual domains of this review:

1. PV, battery, and hybrid energy systems: “photovoltaic”, “PV system”, “solar energy”, “solar power”, “PV–battery”. “PV–BESS”, “energy storage”, “battery energy storage system”, “BESS”, “hybrid renewable system”, “battery management”, “state of charge”, “SoC”.
2. MPPT, energy management, and power electronic control: “MPPT”, “maximum power point tracking”, “energy management”, “power management”, “energy management system”, “EMS”, “DC–DC converter”, “buck converter”, “boost converter”, “bidirectional converter”.
3. Machine learning and deep learning methods: “machine learning”, “deep learning”, “neural network”, “artificial neural network”, “ANN”, “CNN”, “RNN”, “LSTM”, “autoencoder”, “transformer”, “data-driven model”, “fuzzy logic”, “genetic algorithm”, “GA”, “particle swarm optimization”, “PSO”.
4. Reinforcement learning and policy-based control: “reinforcement learning”, “deep reinforcement learning”, “actor–critic”, “policy gradient”, “deep Q-network”, “DQN”, “DDPG”, “TD3”, “SAC”, “continuous action”.

Multiple search strings were formed using partial and complete combinations of the above keyword groups to avoid over-restrictive filtering. The following search string illustrates the primary query used in one of the databases (adapted as required for the other databases):

```
((“photovoltaic” OR “PV system” OR “solar energy”) AND (“battery” OR “energy storage” OR “BESS”) AND (“MPPT” OR “energy management” OR “EMS”) AND (“machine learning” OR “deep learning” OR “neural network” OR “fuzzy logic” OR “genetic algorithm” OR “reinforcement learning” OR “deep reinforcement learning” OR “DQN” OR “DDPG” OR “TD3” OR “SAC”))
```

Table 1 summarizes the database-specific search strings and limits applied in this review. As noted above, the search strategy combined photovoltaic-domain terms with control, energy management, and artificial intelligence keywords, with minor adaptations to match the syntax and filtering options of IEEE Xplore, ScienceDirect, and SpringerLink.

Table 1. PRISMA search keywords by database.

Database	Search Set	Search String/Keywords Used
IEEE Xplore	Broad intelligent control	(“photovoltaic” OR “PV system” OR “solar energy”) AND (“MPPT” OR “maximum power point tracking” OR “energy management” OR “power management”) AND (“machine learning” OR “deep learning” OR “neural network” OR “reinforcement learning” OR “intelligent control”)
IEEE Xplore	PV + battery systems	(“photovoltaic” OR “PV”) AND (“battery” OR “energy storage” OR “BESS”) AND (“energy management” OR “EMS” OR “power control”) AND (“machine learning” OR “deep learning” OR “reinforcement learning” OR “GA” OR “genetic algorithm” OR “particle swarm optimization” OR “PSO” OR “LSTM” OR “RNN” OR “CNN” OR “ANN” OR “artificial neural network” OR “fuzzy logic”)
IEEE Xplore	Converter-level control	(“DC–DC converter” OR “buck converter” OR “boost converter” OR “bidirectional converter” OR “data-driven model”) AND (“photovoltaic” OR “PV”) AND (“machine learning” OR “deep learning” OR “reinforcement learning”)
ScienceDirect (Elsevier)	AI-based Energy Management for PV	TITLE(“photovoltaic” OR “PV”) AND (“energy management” OR “MPPT” OR “control strategy”) AND ABSTRACT(“machine learning” OR “deep learning” OR “reinforcement learning” OR “neural network”)

Table 1. Cont.

Database	Search Set	Search String/Keywords Used
ScienceDirect (Elsevier)	PV-Battery Intelligent Energy Management	("PV battery" OR "hybrid PV" OR "PV-BESS" OR "photovoltaic battery") AND TITLE("power management" OR "energy management" OR "battery management" OR "EMS" OR "SoC" OR "state of charge") AND ABSTRACT("machine learning" OR "deep learning" OR "reinforcement learning")
ScienceDirect (Elsevier)	Converter-Level Intelligent Control	TITLE("DC-DC converter" OR "power converter" OR "bidirectional converter") AND TITLE("photovoltaic" OR "PV") AND ABSTRACT("machine learning" OR "neural network" OR "reinforcement learning")
ScienceDirect (Elsevier)	Intelligent/Optimization-Based Control for PV Systems	TITLE("photovoltaic" OR "PV") AND ("energy management" OR "optimization" OR "intelligent control") AND ABSTRACT("neural network" OR "deep learning" OR "reinforcement learning")
ScienceDirect (Elsevier)	MPPT-Focused AI Control	TITLE("MPPT" OR "maximum power point tracking") AND ("photovoltaic" OR "PV") AND ABSTRACT("machine learning" OR "deep learning" OR "reinforcement learning" OR "neural network")
SpringerLink	Learning-based PV control	("photovoltaic" OR "solar") AND ("control" OR "energy management" OR "MPPT") AND ("machine learning" OR "deep learning" OR "reinforcement learning" OR "actor-critic" OR "policy gradient" OR "deep Q-network" OR "DQN" OR "DDPG" OR "TD3" OR "SAC" OR "continuous action")
SpringerLink	Power electronics focus	("DC-DC converter" OR "power electronic") AND ("photovoltaic") AND ("intelligent control" OR "learning-based")
All databases	PV-battery separate/supervisory control	("PV battery" OR "PV-battery" OR "PV-BESS") AND ("MPPT" OR "maximum power point tracking") AND ("BMS" OR "Battery management" OR "BESS") AND ("hierarchical" OR "decentralized" OR "separate" OR "supervisory")
All databases	PV-battery unified/integrated control	("PV battery" OR "PV-battery" OR "PV-BESS") AND ("MPPT" OR "maximum power point tracking") AND ("BMS" OR "Battery management" OR "BESS") AND ("unified control" OR "integrated control" OR "coordinated control" OR "joint control")

The search was restricted to peer-reviewed journal articles and conference papers published in English between 2015 and 2026. This time window reflects the increased adoption of modern data-driven, deep learning, and RL techniques in power electronics and energy system control during the last decade.

The initial search strategy was deliberately broad and included AI/ML, deep learning, and RL terms in order to capture the full methodological landscape of intelligent control in PV and PV-battery systems and to reduce the risk of prematurely excluding relevant studies. During screening and full-text assessment, it became clear that RL was the most relevant class of methods for addressing the review questions related to sequential decision making, integrated control, and long-horizon optimization. Accordingly, the final analytical synthesis prioritized RL-based studies, while non-RL studies were retained primarily for background framing and contextual comparison; this comparator set includes established hierarchical, cascaded, predictive, nonlinear, and optimization-based control approaches that provide relevant engineering context. Standalone PV MPPT studies were captured during the broader search and screening process because they were useful for establishing the methodological differentiations from conventional and supervised approaches toward RL-based control. However, during full-text eligibility assessment and final coding for synthesis, studies focused exclusively on standalone PV MPPT without storage interaction or without relevance to the sequential PV-battery coordination problem were excluded from the core analytical synthesis. Some such studies were retained only as contextual background where they helped position RL within the broader MPPT literature.

3.3. Inclusion and Exclusion Criteria

Studies were included in the review if they satisfied the following criteria:

- Application of ML, DL, RL, DRL, or hybrid data-driven and optimization-based approaches to PV or PV-battery energy systems.

- Focus on MPPT, PV energy optimization, battery management, or power-electronic control.
- Presentation of simulation and/or experimental results with quantitative performance metrics.
- Consideration of continuous or discrete control actions applied to power converters (e.g., buck, boost, or bidirectional DC–DC converters).
- Relevance to PV–battery systems or hybrid renewable energy systems with storage interaction.

Studies were excluded if they met any of the following conditions:

- Purely theoretical discussions of AI methods without application to energy systems.
- Generic ANN, fuzzy logic, or optimization algorithms without relevance to PV, energy storage, or power-electronic control problems.
- PV modeling studies without any optimization, control, or decision-making component.
- Large-scale grid or market-level energy management studies unrelated to standalone or hybrid PV systems.
- Insufficient technical detail, absence of quantitative results, lack of relevance to the defined scope, or lack of peer-reviewed publication status.
- Studies that focus exclusively on standalone PV MPPT without consideration of storage interaction or sequential decision-making effects were excluded from the core analytical synthesis.
- Supervised learning approaches (e.g., ANN, CNN, and LSTM) used solely for static MPP estimation or forecasting, without closed-loop control or policy-based decision making, were excluded from the final reinforcement learning-focused analysis.

Borderline studies were classified according to their primary methodological and application relevance to the review questions. Where a study addressed PV, storage, or intelligent control but did not fully satisfy the core synthesis criteria, it was assessed based on whether it contributed directly to at least one of the following dimensions: (i) sequential decision making, (ii) reinforcement learning-based control, (iii) discrete-versus continuous-action formulation, or (iv) explicit PV–battery coupling. Studies not meeting these criteria were excluded from the core synthesis, although some were retained for contextual background.

3.4. Study Selection Process and PRISMA Flow Diagram

The study selection followed the standard PRISMA workflow consisting of four sequential phases:

1. Identification: The initial database search returned 5120 records. After the removal of duplicate entries, 3667 unique studies remained for screening.
2. Screening: Titles and abstracts of the 3667 records were screened against the predefined inclusion and exclusion criteria. This step resulted in 2147 studies selected for full-text assessment, after 370 not retrieved.
3. Eligibility: Full-text evaluation was conducted to assess methodological relevance, technical depth, and applicability to PV systems, energy storage, and intelligent control. Following this phase, 101 studies were retained for further analysis, 44 of which were retained as contextual non-RL comparators.
4. Included: The final set of 57 publications meeting all eligibility requirements was included in the systematic literature review and used for the thematic and comparative analysis presented in the subsequent literature review sections specifically addressing RL-based control of PV or PV–battery systems, identified for deeper analytical synthe-

sis. This hierarchical classification enabled differentiation between general AI-based MPPT approaches and sequential decision-making frameworks.

Duplicate records were identified using Zotero and removed manually by the author after import from the searched databases. Title/abstract screening was conducted by a single reviewer (the author), with auxiliary organizational support tools. Full-text eligibility assessment was likewise conducted by the author alone. Because screening and full-text assessment were performed by a single reviewer, no inter-reviewer disagreements arose, and no formal arbitration procedure was required.

The screening and eligibility assessment were conducted by the author in accordance with the defined criteria. The complete study selection process is summarized in the PRISMA 2020 flow diagram shown in Figure 4, which illustrates the identification, screening, eligibility, and inclusion stages of the systematic review. To minimize selection bias, eligibility decisions were based strictly on the predefined inclusion and exclusion criteria.

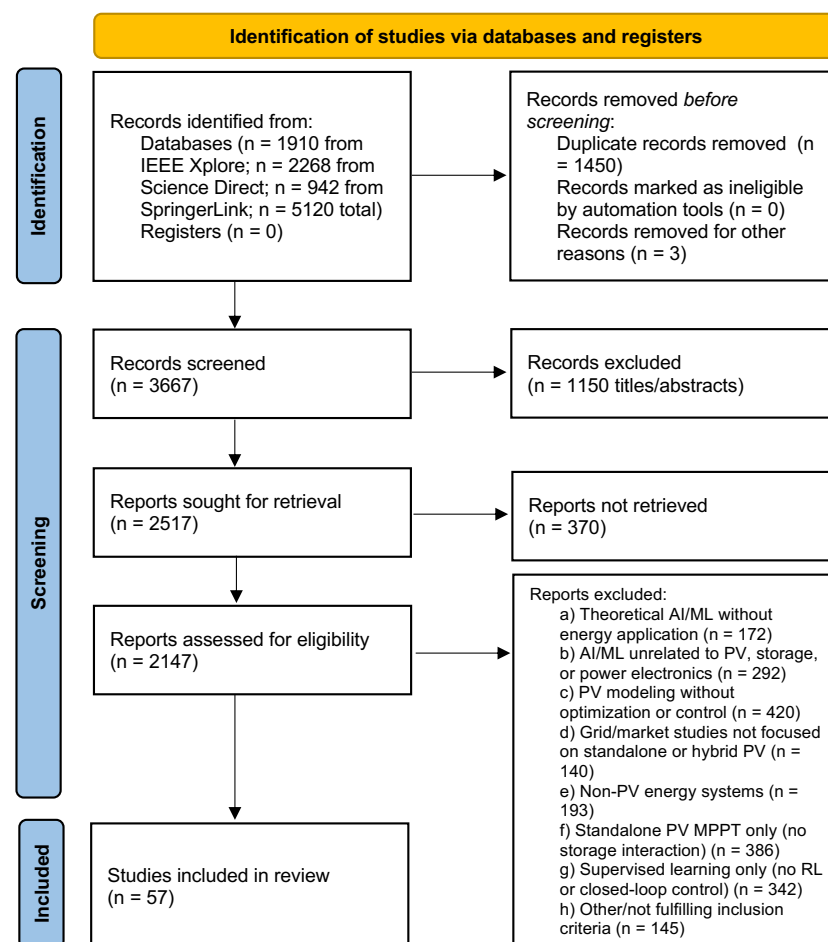


Figure 4. PRISMA flow diagram for the systematic literature review.

4. Results of Systematic Review

4.1. Overview of Included Studies

The PRISMA-guided screening process identified a final corpus of studies for detailed review, with RL papers forming the core analytical set used to address PRQ1–PRQ4. In addition, a secondary set of non-RL studies was retained as a structured comparative reference set. Accordingly, the results are presented in two layers: (i) a descriptive comparison between RL and non-RL studies across architecture, MPPT treatment, actuation, method, and control layer; and (ii) a PRQ-driven synthesis focused on the RL corpus.

In total, $n = 57$ RL studies were included in the core synthesis, while $n = 44$ non-RL studies were retained for contextual comparison. This comparative coding does not broaden the analytical scope of the review beyond RL; rather, it is used to clarify where RL-based work is positioned within the wider PV–battery control literature and to make visible the structural differences between RL and non-RL approaches.

The comparative analysis shows that the literature is not divided simply by algorithmic family but also by control layer, actuation level, and treatment of MPPT. In particular, RL studies are concentrated at the EMS/supervisory layer and often treat PV generation as exogenous, whereas non-RL studies more frequently engage converter-level PV control or explicitly coordinated MPPT structures. These patterns are summarized in Figures 5–14 and are subsequently used to contextualize the RL-focused PRQ synthesis.

Figure 5 shows the distribution of studies by publication year for the RL corpus and the non-RL comparator set. The temporal pattern indicates that RL-based work is concentrated in more recent years, with a marked increase from 2023 onward, whereas the non-RL literature is distributed more broadly across earlier years. This pattern is consistent with the recent rise of RL in PV–battery energy management and supports the interpretation of RL as a comparatively newer methodological strand within the broader intelligent control literature.

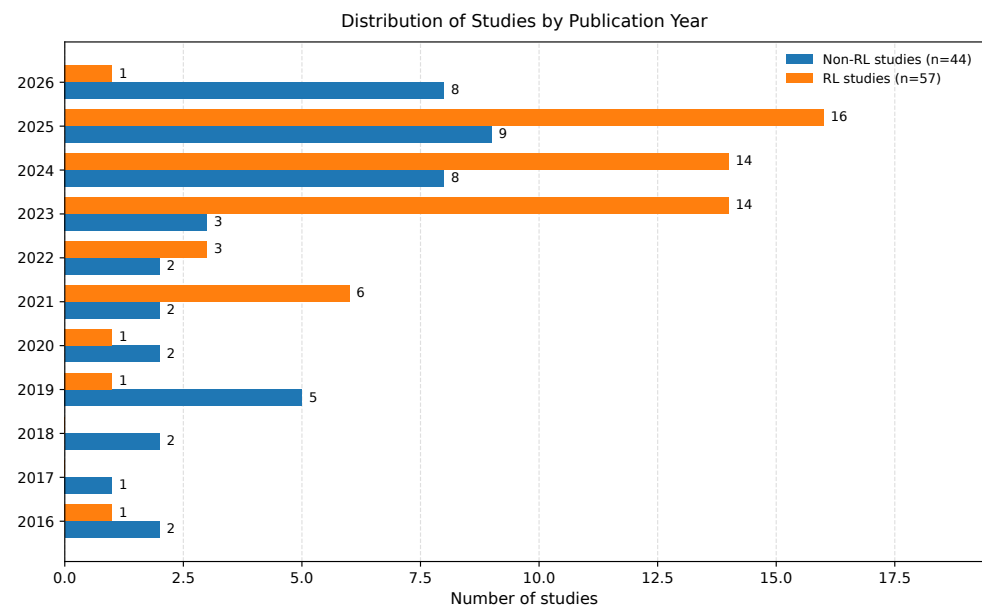


Figure 5. Publication-year distribution of the reviewed literature, comparing RL studies with the non-RL comparator set. The RL corpus is concentrated in recent years, indicating the comparatively recent emergence of reinforcement learning within PV–battery control research.

Figure 6 shows a clear difference in publication format between the two bodies of literature. The non-RL comparator set is dominated by journal articles (34 studies), with a smaller number of conference papers (10), whereas the RL corpus contains a larger share of conference papers (32), alongside 24 journal articles and 1 master’s thesis. This pattern suggests that reinforcement learning-based work in PV–battery control is still evolving as a comparatively newer and faster-moving research strand, with many contributions appearing first in conference venues, while the non-RL literature appears more established and consolidated in archival journal publications. In the context of this review, the figure complements the broader temporal trend already observed in the manuscript by showing that the recent growth of RL research is accompanied not only by increased volume but also by a publication profile that is more conference-oriented than the non-RL comparator set.

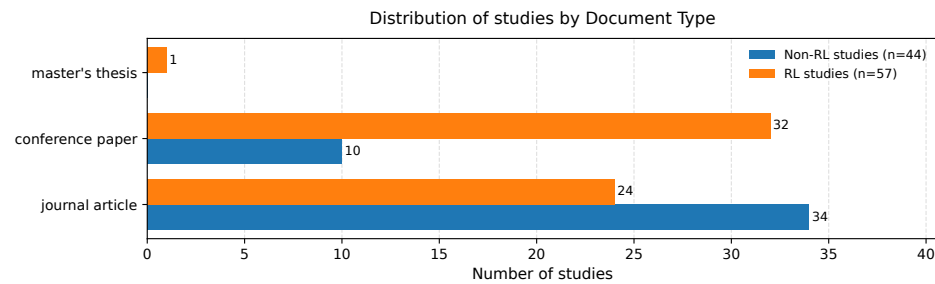


Figure 6. Distribution of studies by document type, comparing the RL corpus and the non-RL comparator set. The RL literature is concentrated in conference papers and journal articles, whereas the non-RL literature is dominated by journal articles, with only a small number of thesis represented.

Figure 7 presents the top publication venues for the RL corpus. The studies are distributed across a diverse set of journals and conference proceedings, with IEEE Access and the IEEE Internet of Things Journal contributing the largest numbers of papers, while most other venues contain only one or two studies. This pattern suggests that RL-based PV–battery research is dispersed across multiple adjacent engineering and energy domains rather than concentrated in a single dedicated outlet.

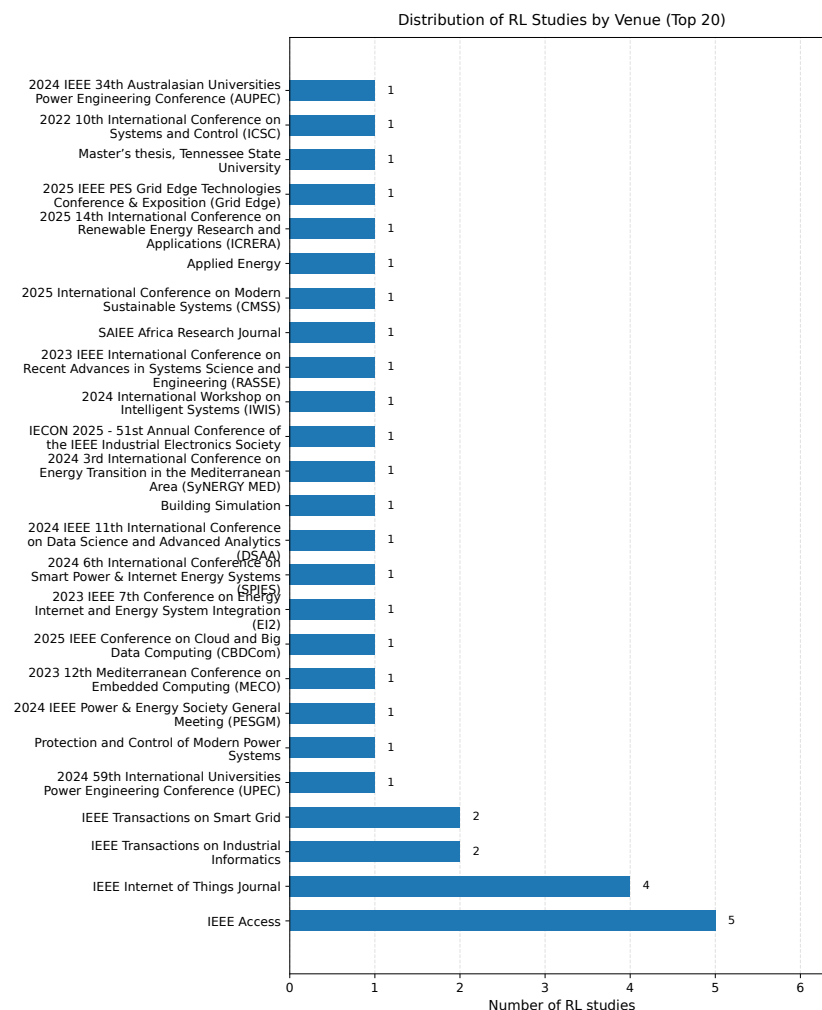


Figure 7. Distribution of RL studies by publication venue (top 20 venues shown). The RL literature is spread across a broad range of journals and conference proceedings, indicating a fragmented but multidisciplinary publication landscape.

With that, Figure 8 illustrates a VOSviewer keyword co-occurrence map of the studies retained after the PRISMA screening process. The visualization reveals several strongly interconnected thematic clusters, showing that the literature is organized around a few central concepts. Terms such as algorithm, performance, model, energy management, system, and microgrid occupy the most central positions, indicating that the reviewed studies are largely focused on optimization and control of PV–battery and microgrid energy systems. The map also highlights the prominence of reinforcement learning-related terms, including deep RL, agent, state, PPO, TD3, and deep deterministic policy gradient, which confirms the growing role of RL-based approaches in this research area. At the same time, the dense links between these terms and words such as battery, load, grid, energy cost, and HEMS suggest that most of the literature emphasizes supervisory energy management and battery dispatch rather than fully unified converter-level MPPT and battery co-control. Overall, the figure supports the interpretation that the field is evolving toward more intelligent and integrated decision-making frameworks while remaining structurally centered on system-level energy management.

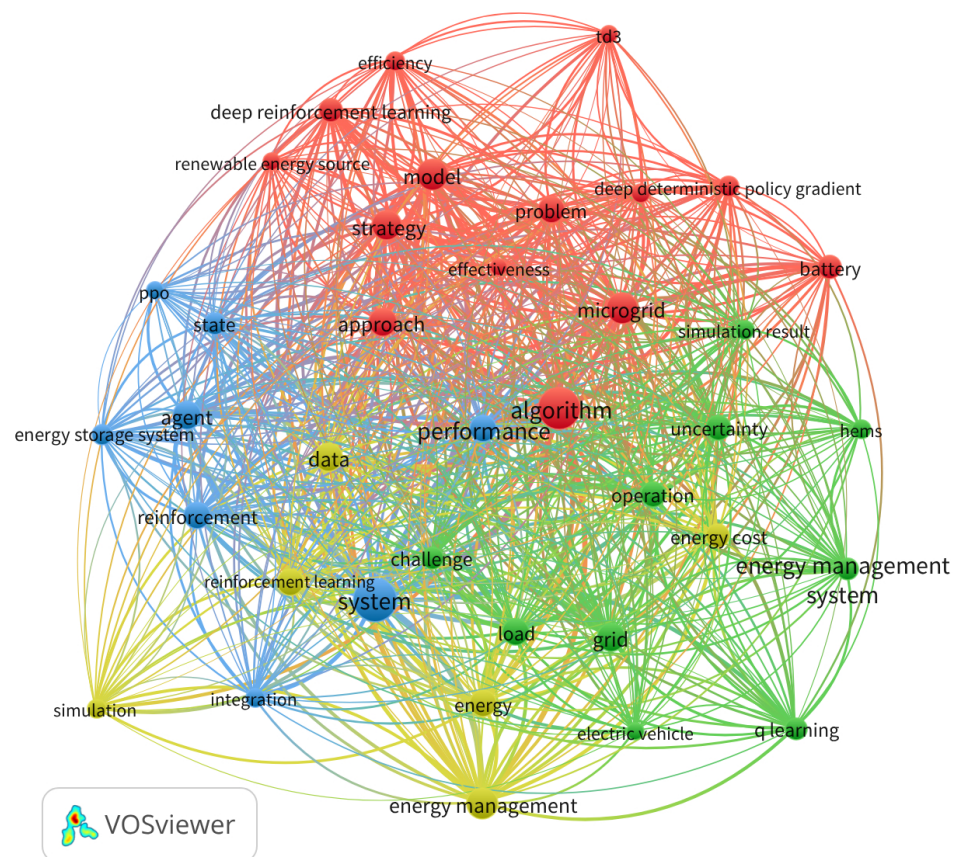


Figure 8. Visualization of co-occurrence patterns among keywords in the studies retained after the PRISMA screening process, produced using VOSviewer 1.6.20.

4.2. Comparative Study Characteristics: RL and Non-RL Literature

To enable consistent cross-study comparison, each included article was coded using a structured classification framework. The framework captures the degree of integration of PV control, MPPT, battery control, and supervisory energy management functions, as well as the methodological approach, control layer, and relevance to the primary review questions. Each study was assigned the category that best reflects its dominant control structure and contribution.

The control-layer classification was used as an analytical coding dimension rather than as an eligibility criterion. Studies were not excluded merely because their control authority was located at the EMS, reference-generation, or converter-control layer. Moreover, these layers were not treated as mutually exclusive: coordinated and hierarchical architectures spanning multiple layers were explicitly identified and coded. The resulting distributions should therefore be interpreted as a descriptive mapping of the reviewed corpus rather than as proof that the broader literature is divided into isolated or competing control domains.

Table 2 summarizes the coding framework used to classify the selected studies.

Table 2. Classification criteria used for coding the reviewed articles.

Dimension	Code	Definition
Architecture	A1	Separated architecture: PV MPPT and battery/EMS are largely treated as distinct subsystems with little direct coordination.
	A2	Coordinated/layered integrated architecture: multiple controllers or loops are present but explicitly coordinated through supervision, hierarchy, or shared system structure.
	A3	Strongly unified architecture: one controller or one decision framework directly co-determines both PV and battery behavior.
MPPT treatment	M0	MPPT not mentioned explicitly.
	M1	MPPT assumed implicitly, or PV treated as exogenous, fixed, or uncontrollable.
	M2	Conventional external MPPT: A separate standard MPPT loop exists and is not subordinated to system-level objectives.
	M3	Supervisory-coordinated MPPT: A conventional MPPT loop exists but can be overridden, constrained, or coordinated by battery or system objectives.
	M4	MPPT is itself a reinforcement learning decision.
	M5	MPPT and battery control are unified within the same converter/control scheme or decision framework.
PV actuation	B0	No meaningful PV actuation.
	B1	PV treated as exogenous or influenced only indirectly.
	B2	PV actively controlled, but in a separate or loosely coupled manner.
	B3	PV control is strongly integrated into the central decision or control structure.
Battery actuation	C0	No battery included.
	C1	Battery present but not meaningfully controlled.
	C2	Battery actively controlled.
Method	D0	No AI or advanced optimization beyond standard control.
	D1	Rule-based, heuristic, or deterministic scheduling/control.
	D2	Classical control, PI, MPC, or optimization without learning-based control.
	D3	Intelligent non-RL method, such as ANN, ANFIS, fuzzy logic, PSO, or GWO.
	D4	Reinforcement learning.
Layer	E1	Planning, forecasting, day-ahead, or economic scheduling layer.
	E2	EMS, supervisory control, or dispatch layer.
	E3	Converter-level, fast control, or directly electrical control layer.

4.2.1. Architecture of PV–Battery Coordination

Figure 9 compares the architectural structure of RL and non-RL studies. Three broad classes were coded: separated architectures (A1), coordinated/layered integrated architectures (A2), and strongly unified architectures (A3). A clear contrast emerges. The non-RL comparator set is dominated by coordinated/layered architectures, whereas the RL set

contains a larger proportion of separated architectures. Strongly unified architectures remain rare in both groups.

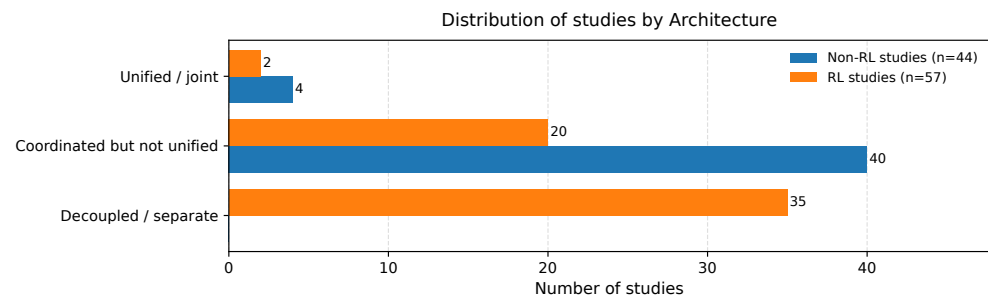


Figure 9. Comparison of architectural integration between RL studies ($n = 57$) and non-RL studies ($n = 44$). A1 denotes separated architectures, A2 coordinated/layered architectures, and A3 strongly unified architectures.

This distribution is important because it indicates that even when RL is used, most studies do not yet formulate PV and battery behavior within a single tightly unified decision framework. Instead, RL is commonly inserted into an already layered control structure, often at the supervisory or scheduling level. Conversely, non-RL studies more often reflect coordinated multi-loop architectures in which PV, battery, and converter behavior are linked through explicit engineering structure, even if not optimized through a unified learning framework.

This pattern is also visible in specific representative studies. Jha et al. [28] provide a clear example of a strongly unified non-RL converter-level architecture in which MPPT and battery-current regulation are carried out through a shared bidirectional converter. By contrast, Rizki et al. [44] represent a coordinated but layered architecture in which forecasting, a predictive EMS, and converter-level control are tightly linked but remain structurally differentiated across supervisory and execution layers. Dagal et al. [45] and Tavooosi et al. [46] further illustrate converter-oriented PV–battery coordination using metaheuristic and nonlinear control, respectively, but without extending the formulation to a broader sequential decision-making EMS framework. Together, these studies help clarify that stronger PV–battery integration is already present in parts of the non-RL literature, particularly at the converter layer, whereas RL-based work remains concentrated at supervisory levels with PV often treated as exogenous.

4.2.2. Treatment of MPPT

The treatment of MPPT is one of the most revealing dimensions in the dataset. As shown in Figure 10, the RL literature is heavily concentrated in categories where MPPT is either not explicitly addressed, assumed implicitly, or left external to the learning problem. Only a very small number of RL studies formulate MPPT itself as an RL decision or embed MPPT and battery control in the same unified decision framework.

By contrast, the non-RL comparator set contains more studies in which MPPT is explicitly coordinated or implemented through a separate conventional loop that remains visible in the system formulation. This difference supports one of the central arguments of the review: much of the RL literature in PV–battery systems focuses on battery dispatch, energy management, or scheduling while leaving PV operating-point selection outside the learning loop. In other words, the key gap is not merely the absence of learning, but the persistent separation between energy harvesting and energy management.

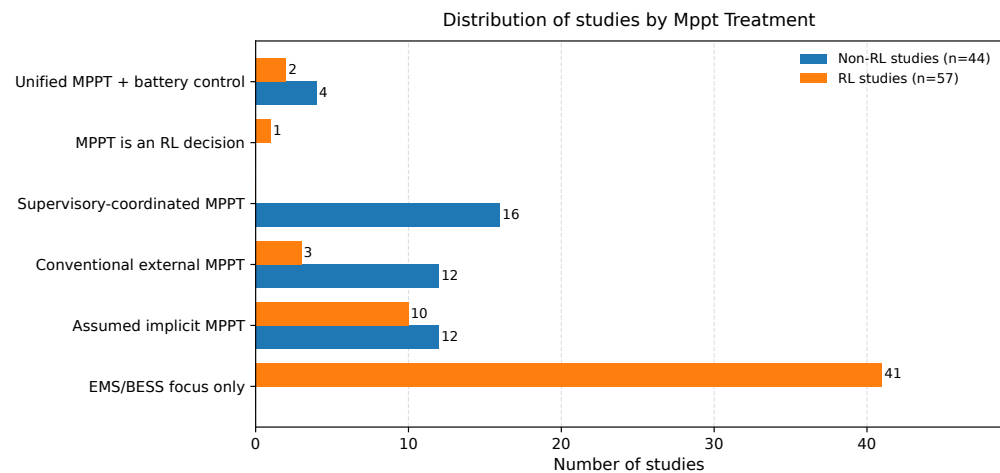


Figure 10. Comparison of MPPT treatment between RL studies ($n = 57$) and non-RL studies ($n = 44$). The coding distinguishes whether MPPT is absent, implicit, externally conventional, supervisory-coordinated, explicitly an RL decision, or unified with battery control.

4.2.3. PV Actuation and Battery Actuation

Figures 11 and 12 compare how directly RL and non-RL studies control PV-side and battery-side variables. The RL literature is strongly concentrated in categories where PV is treated as exogenous or only indirectly represented, whereas the non-RL comparator set contains many more studies with active PV-side control at the reference or converter level. This reinforces the view that RL-based PV–battery work has largely evolved as EMS-oriented research rather than as converter-level PV-control research.

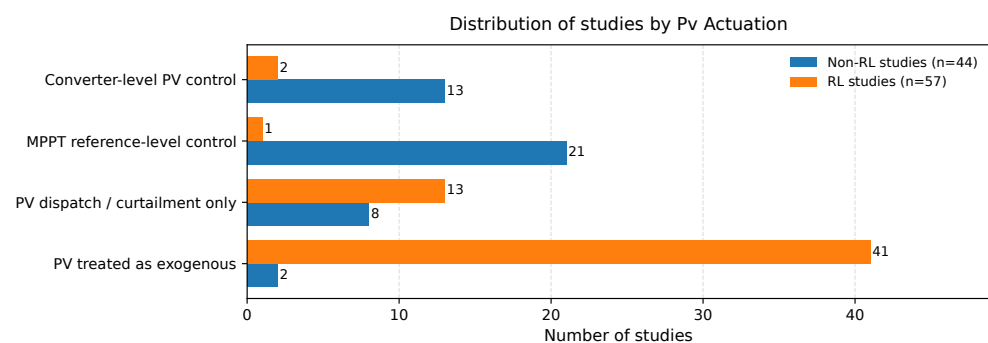


Figure 11. Comparison of PV actuation between RL studies ($n = 57$) and non-RL studies ($n = 44$). RL studies are concentrated in categories where PV is treated as exogenous or only indirectly controlled, while non-RL studies more often include active PV-side control.

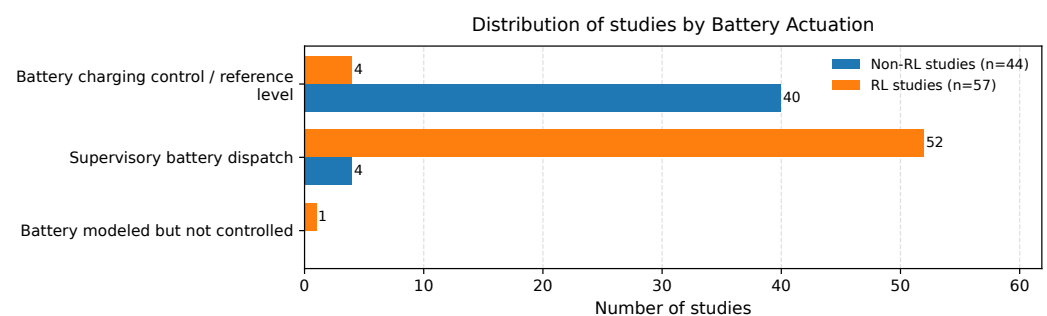


Figure 12. Comparison of battery actuation between RL studies ($n = 57$) and non-RL studies ($n = 44$). RL studies are strongly concentrated in active battery-control categories, highlighting the battery-centric nature of most RL formulations.

The opposite pattern is observed on the battery side. RL studies overwhelmingly involve active battery dispatch or charging/discharging decisions, confirming that storage is the main actuation target in RL-based formulations. Non-RL studies, in contrast, more frequently operate closer to charging-control or converter-reference implementations. Taken together, these two figures show that the dominant RL paradigm in this domain is battery-centric supervisory control with exogenous PV, rather than genuinely integrated PV–battery co-control.

4.2.4. Methodological Distribution

Figure 13 summarizes the methodological distribution across the two sets. As expected, the RL corpus is dominated by reinforcement learning-based and hybrid formulations, while the non-RL comparator set is concentrated in classical control, optimization-based approaches, and intelligent non-RL methods such as ANN, fuzzy, or metaheuristic-based control.

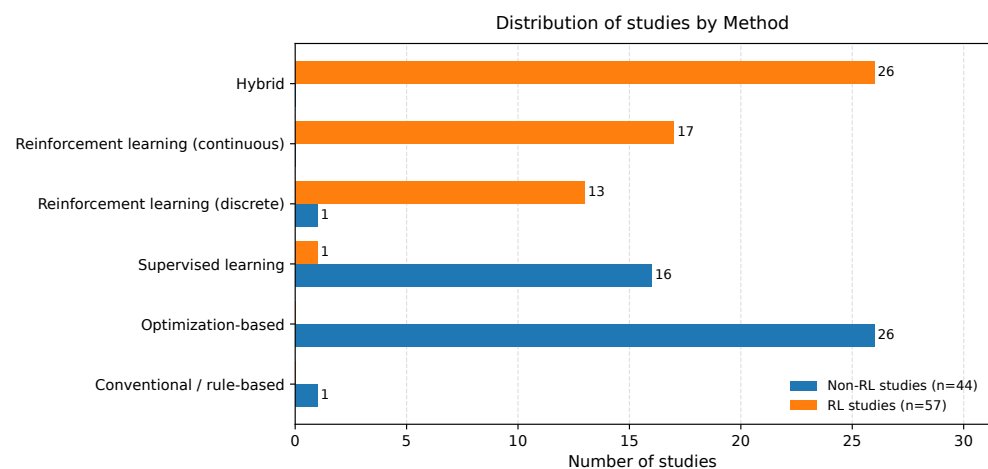


Figure 13. Methodological distribution across RL studies ($n = 57$) and non-RL studies ($n = 44$). The non-RL set includes conventional, rule-based, optimization-based, and intelligent non-RL methods, while the RL set includes reinforcement learning and hybrid approaches.

This figure is less important as evidence of the research gap by itself, since the classification partly reflects the sampling logic of the review. However, it remains useful because it shows that the contrast between the two groups is not simply “RL versus not RL” in name only; it also corresponds to broader methodological tendencies. The non-RL set provides a meaningful baseline for comparison because it includes multiple established alternatives that have historically addressed PV control, MPPT coordination, converter regulation, and battery management without RL.

4.2.5. Control Layer

The clearest structural distinction appears in the control-layer comparison shown in Figure 14. RL studies are concentrated at the EMS/supervisory and scheduling layers, whereas non-RL studies have substantially stronger representation at the converter and directly electrical control layer. This pattern is consistent with the findings above on MPPT treatment and PV actuation. It indicates that most RL-based PV–battery studies do not formulate the problem at the electrical actuation layer where converter duty cycle, voltage references, or MPPT-related operating points are directly optimized.

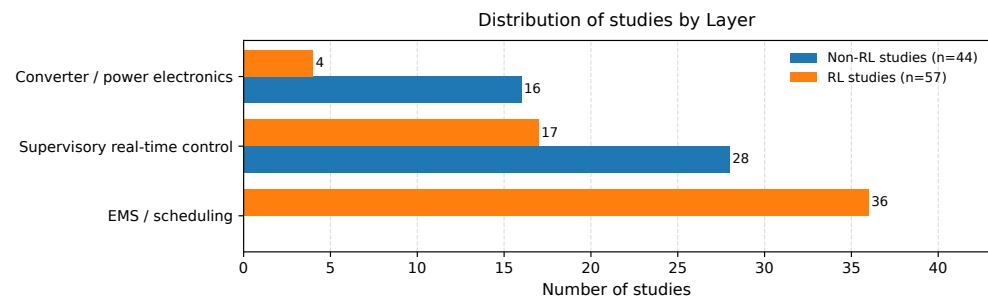


Figure 14. Comparison of control-layer emphasis between RL studies ($n = 57$) and non-RL studies ($n = 44$). RL studies are concentrated at the EMS/supervisory and scheduling layers, while non-RL studies are more strongly represented at the converter-level control layer.

This layer separation helps explain why integrated MPPT and battery control remains uncommon in the RL literature. It is not only that MPPT is often omitted; rather, the majority of RL work is posed at a control layer where PV generation enters as an input, forecast, or upper bound, not as a directly optimized electrical decision variable. Consequently, the converter-level and EMS-level strands of the literature remain only weakly connected.

4.3. Implications of the Comparative Coding

Taken together, the comparative coding reveals a consistent structural pattern. The RL literature is primarily organized around storage dispatch, supervisory control, and scheduling, with PV generation often treated as fixed, forecasted, or implicitly MPPT-operated. The non-RL literature, by contrast, more frequently engages the converter layer and explicit PV control, including externally implemented or supervisory-coordinated MPPT structures. Strongly unified PV–battery control remains uncommon in both sets.

These observations are important because they sharpen the interpretation of the RL corpus reviewed in this article. The research gap is not simply that RL has not yet been used widely enough; rather, the gap is that RL has been used predominantly at higher supervisory abstraction levels for fully integrated PV–battery optimization. The central observation is not that EMS-level RL is incomplete or should be replaced by converter-level learning. The reviewed EMS studies make substantial contributions within battery dispatch, scheduling, economic, grid-interaction, and uncertainty-management objectives. A different and comparatively underrepresented combination is the explicit connection of converter-level PV operating-point decisions with battery-aware long-horizon objectives within the same RL formulation or coordinated benchmark.

4.4. Summary of Findings

The reviewed literature demonstrates significant progress in the application of artificial intelligence and RL to PV–battery systems, particularly at the energy management and scheduling layer. RL has been successfully applied to battery dispatch, grid exchange optimization, EV coordination, multi-agent energy trading, and degradation-aware scheduling under uncertainty. Continuous actor–critic methods, hybrid RL–optimization architectures, and structured constraint-handling mechanisms are increasingly common.

However, this methodological maturity is concentrated at the supervisory EMS level. PV generation is treated in the overwhelming majority of studies as an exogenous input or as an upper-bounded dispatchable source with implicit MPPT, rather than as a controllable operating-point decision. A small number of recent studies do move closer to integrated control by embedding RL-based MPPT and battery charging/discharging within a single controller; however, these efforts remain small-scale and primarily converter-centric, rather than establishing a broadly adopted unified framework for joint PV operating-point op-

timization and storage-aware EMS across realistic microgrid settings [48,49]. As a result, most existing formulations optimize energy allocation and curtailment rather than energy harvesting itself, even though the converter-control literature now clearly shows that DRL can act directly on duty cycle, switching, and MPPT objectives when the problem is posed at the power-electronics layer [50].

This persistent separation between PV operating-point control and battery dispatch highlights a structural gap in the literature: sequential decision-making frameworks rarely integrate converter-level PV actuation with storage-aware energy optimization. The synthesis in the following section therefore examines this gap systematically through PRQ1–PRQ4 and establishes the foundation for a unified continuous-control framework.

4.5. Synthesis of Evidence by PRQ

This subsection synthesizes the evidence to address PRQ1–PRQ4. Across the reviewed studies, RL is predominantly applied at the energy management and scheduling layer, including grid exchange, battery dispatch, and load or EV coordination, as listed in Table A1. In contrast, inverter-level PV operating-point control (i.e., MPPT treated as an explicit decision variable) is typically assumed, abstracted, or implemented via conventional control schemes external to the learning problem. At the same time, several non-RL studies, as listed in Table A2, show that this separation is not inevitable. Jha et al. [28] demonstrate that MPPT and battery-current regulation can be unified at the converter layer through a shared bidirectional converter, while Rizki et al. [44] show that PV references, battery references, and converter control can be coordinated through a predictive architecture spanning forecasting, an EMS, and low-level execution. Dagal et al. [45] and Tavoosi et al. [46] provide additional evidence that non-RL metaheuristic and nonlinear-control approaches can directly actuate both PV-side and battery-side dynamics. These studies are therefore useful because they indicate that the main limitation in much of the RL literature is not the presence or absence of coordination in principle, nor whether supervisory or converter-level control is preferable. It is whether the learning formulation directly includes PV operating-point decisions, battery-side objectives, or both and how coordination between the associated control layers is evaluated.

4.5.1. PRQ1: Limitations of Decoupled Architectures (MPPT vs. Storage/EMS)

Across the reviewed literature, PV generation is modeled in almost all cases as an exogenous input to the EMS, implicitly assuming independent MPPT or a fixed PV yield model. Recent converter-oriented DRL studies confirm that RL can, in principle, control MPPT and converter duty-cycle decisions directly, but this strand of work remains methodologically separate from most PV–battery EMS formulations [43,50,51]. RL actions regulate battery and grid power, while PV enters the state as measured or forecast output without a controllable operating-point variable [29,42,52–57]. When PV “control” is included, it is typically expressed as curtailment or scaling at an aggregated DER level rather than converter duty-cycle or V_{PV}/I_{PV} regulation [58]. The separation between energy harvesting and energy allocation is therefore preserved. A representative example is the PSCAD-based PV–battery hybrid system by Go and Choi [59], in which the PV converter performs P&O-based MPPT, the BESS regulates the DC-link through a bidirectional converter, and the inverter controls system output power. Although this constitutes a genuinely integrated PV–battery control architecture at the hardware and converter level, the control objectives remain partitioned across separate subsystems rather than being learned or optimized jointly within a unified policy.

A limited exception is provided by recent small-scale PPO-based studies that jointly control PV MPPT, BESS charging/discharging, and load-side voltage regulation within a

centralized RL controller, using duty-cycle and modulation actions rather than dispatch-only decisions. These studies are important because they demonstrate feasibility of integrated learning-based control, yet they remain focused on compact DER test systems rather than the broader uncertainty-aware, degradation-aware, or market-coupled EMS formulations that dominate the literature [48,49].

This abstraction spans diverse system types, including wind–PV–storage scheduling [60], residential PV–battery EMSs [61,62], EV charging stations [63,64], multi-asset microgrids [65,66], and actor–critic-based energy trading [67,68]. Even in degradation-aware, forecast-integrated, hybrid RL–optimization, Lyapunov-stable, or robust H_∞ -inspired formulations, PV power remains a disturbance or upper-bounded source rather than an electrically optimized variable. The MPPT–EMS separation is thus primarily architectural rather than algorithmic: converter-level RL for MPPT and duty-cycle control clearly exists, but it is rarely co-developed with storage-aware supervisory optimization in a unified formulation [48,50].

Practical implications of layered design are illustrated by Qiu et al. [69], who show that legacy charge-control logic can become mismatched under heterogeneous storage chemistries, leading to inefficiencies and premature limiting. Although focused on batteries, this result highlights a broader limitation: embedded assumptions in decoupled layers may fail when dynamics are tightly coupled.

Recent continuous and uncertainty-aware EMS approaches maintain this separation. Multi-agent DDPG scheduling [70], multi-timescale SAC corrective control [71], and various continuous-action EMS formulations [41,72–77] all treat PV as forecasted or measured input while policies act on storage and grid exchange. Even when a PV utilization variable is introduced, it reflects dispatch or curtailment rather than converter operating-point control.

A smaller subset incorporates converter-adjacent variables while retaining MPPT as an external conventional loop. TD3 has been used to augment DC-bus-voltage regulation on top of classical PI control, with MPPT implemented via fixed P&O [78]. Similarly, RL-based supervisory controllers for PV feed-in and voltage quality retain conventional voltage/reactive-power control and treat PV generation as exogenous [79]. Even near the converter layer, MPPT and storage dispatch are usually not co-optimized, although a small number of recent centralized RL controllers do jointly regulate PV MPPT, bidirectional battery charging, and load voltage in compact DER systems [48]. The system presented therein illustrates a centralized, integrated converter-level RL architecture in which a single agent observes coupled electrical and energy-state variables, including changes in PV power and voltage, BESS power, state of charge, load power, and output voltage, and issues PWM-related duty-cycle commands to both the PV-side and battery-side converters. This differs fundamentally from mainstream PV–battery EMS formulations, where PV generation is treated as an exogenous disturbance and control authority is limited to supervisory dispatch. By placing MPPT-related PV control and battery charging/discharging within a single learned control loop, the architecture represents a rare counterexample to the dominant separation between energy harvesting and energy management in the reviewed literature.

These studies represent rare counterexamples to the dominant separation between energy harvesting and energy management, because PV-side control actions and battery-side actions are placed within the same learned control loop. Abstracting from these emerging integrated formulations, Figure 15 summarizes the target co-control structure implied by the architectural gap identified in this review. The controller observes PV electrical variables, battery state information, and load/grid conditions, and jointly produces PV-side, battery-side, and grid/load-side control actions. The key distinction from mainstream EMS-oriented RL is that the PV operating point is no longer treated only as an exogenous

input or forecasted resource but becomes an explicit decision variable through commands such as a PV duty cycle or PV voltage reference.

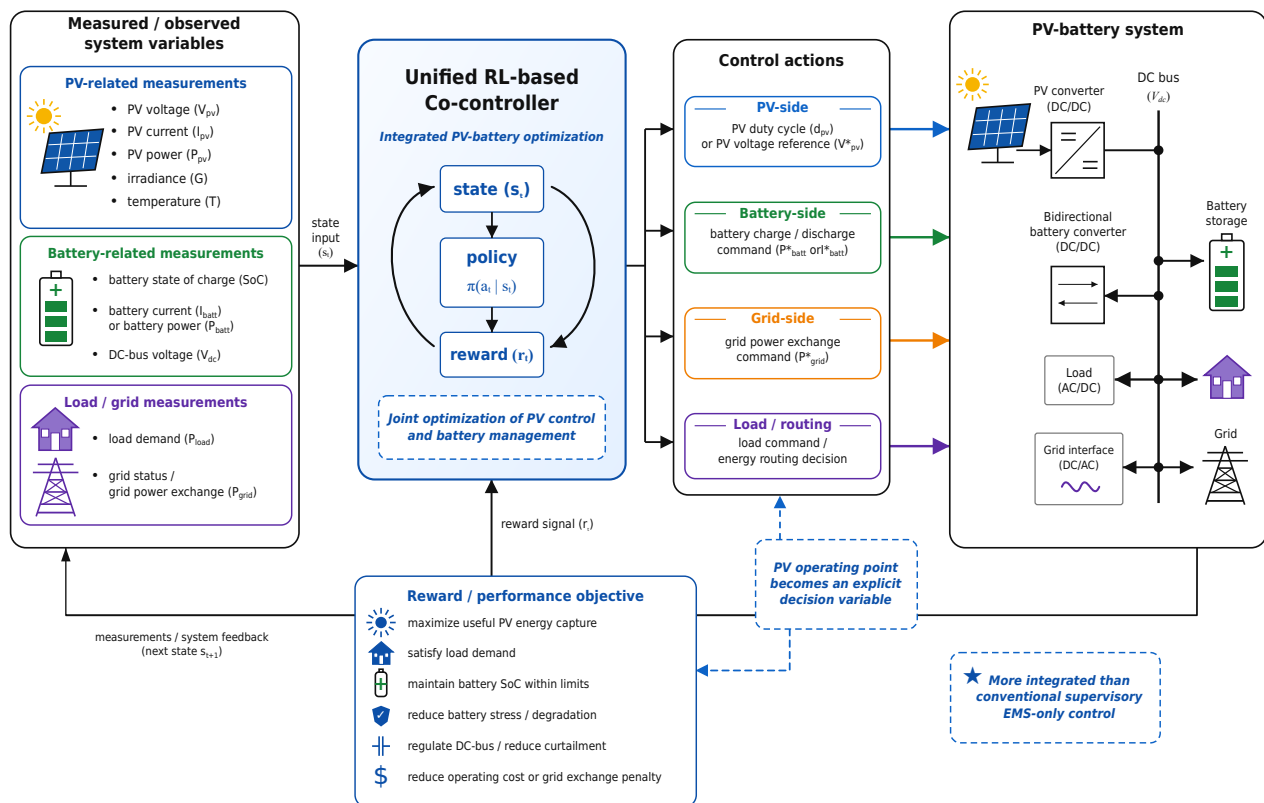


Figure 15. Conceptual synthesis of a unified PV–battery RL co-control architecture. The RL controller receives PV measurements, battery state information, and load/grid variables and jointly determines PV-side, battery-side, and grid/load-side control actions. Unlike purely supervisory EMS formulations, the PV operating point is treated as an explicit decision variable through commands such as the PV duty cycle or PV voltage reference, enabling tighter coordination of MPPT, battery management, and system-level objectives.

With that, decoupling simplifies modeling and training but prevents unified reasoning over PV extraction, DC-bus dynamics, inverter stress, and battery constraints. MPPT trade-offs therefore remain outside the learning loop in the vast majority of reviewed studies, underscoring the structural nature of the architectural gap addressed in this review, while also indicating that recent proof-of-concept integrated controllers have begun to narrow that gap at the small scale [43,48,51].

4.5.2. PRQ2: Control Variables and Actuation Levels in Learning-Based PV–Battery Approaches

Across the reviewed literature, control variables are overwhelmingly defined at the EMS supervisory layer, even though a parallel converter-control literature shows that DRL can also act directly on PWM duty cycle, modulation variables, phase shift, and MPPT objectives [43,50,51]. Some recent works formulate RL control directly at the MPPT layer, using a continuous-action SAC agent to adjust the DC–DC converter duty cycle for global maximum power point tracking under irradiance variation, temperature changes, and partial shading, showing a clearly converter-level rather than supervisory EMS-level control formulation [33]. Others have also modeled PV MPPT under partial shading as an MDP and compare SAC and DDPG using converter duty-cycle perturbation as the control action, showing a clearly converter-level continuous-control formulation [80]. Typical actions include (i) battery charge/discharge and grid exchange power [29,42,55,57,81]; (ii) allocation

ratios among PV, load, storage, and grid [56,82]; and (iii) EV charging and tariff-driven decisions in EV-integrated systems [52–54,81,83]. In all cases, actuation occurs at the energy scheduling level. This distinction matters because “RL for PV–battery systems” now spans at least two different actuation regimes: dominant EMS-level scheduling formulations that regulate battery, grid, and flexible-load power and a separate converter-level strand in which DRL manipulates duty cycle, switching, or MPPT-related variables. A representative example is the HEMS formulation by Si et al. [84], which explicitly defines different action sets for different device classes: binary on/off actions for interruptible and non-interruptible constant-power appliances, continuous power actions for thermostatically controlled loads, and bidirectional continuous charge/discharge actions for energy storage devices. Their state design likewise combines supervisory variables such as electricity price, PV output, and non-schedulable load with device-specific states such as task progress, indoor temperature, and battery energy state. This is useful because it shows that recent DRL-based PV–battery/home-energy studies can employ rich mixed discrete–continuous-action designs while still locating control authority at the EMS layer; PV appears as forecasted available generation rather than as a directly controlled MPPT duty-cycle or operating-voltage decision.

Discrete formulations typically use quantized battery power levels or enumerated operating modes [61–63], sometimes restricting actions to precomputed feasible power-balance combinations [64,66]. Deep RL approaches replace tabular representations with neural approximators while retaining EMS-level control [65]. Continuous actor–critic methods directly parameterize battery or grid power as real-valued variables [67,68] and are now widely adopted using DDPG, TD3, or SAC across residential, community, and grid-scale applications [41,70–77]. Multi-agent decompositions distribute scheduling across time-indexed agents or interacting prosumers [70,74]. Some works treat residential PV–battery energy management as a discrete–continuous hybrid control problem rather than a purely discrete or purely continuous one. Huang et al. [85], for instance, couple discrete appliance scheduling with continuous HVAC and battery actions, highlighting that many practical RL-based formulations remain hybrid supervisory EMS problems rather than converter-level MPPT control problems.

Parallel to EMS-level developments, a growing body of literature investigates RL-based control at the power-electronics layer, where actions correspond to converter duty cycles, voltage references, or switching behavior rather than energy dispatch variables [50].

Despite this progression from discrete to continuous neural policies, the dominant actuation target in integrated PV–battery RL remains supervisory dispatch rather than converter-level PV control. In most studies, “continuous control” refers to battery, grid, or load power references, whereas direct optimization of converter duty cycle or PV operating point is treated separately or assumed implicitly through external MPPT loops.

MPPT is therefore rarely modeled as a decision variable in EMS-focused PV–battery studies, with only limited recent exceptions in which PPO-based controllers explicitly command PV MPPT duty cycle alongside BESSs and load-control actions [48,49]. PV generation is typically treated as irradiance-derived or forecasted available power, or as an upper-bounded source with implicit MPPT. Even when PV flexibility is included, it is expressed as curtailment or allocation of already-produced energy rather than DC–DC operating-point control [52,56,58]. Recent continuous-action EMS formulations introduce PV utilization or feed-in variables [79,86], but these remain dispatch-level quantities.

Thus, learning-based PV–battery approaches show increasing sophistication in EMS-level scheduling, yet control authority consistently remains above the converter layer in almost all studies, leaving PV operating-point optimization outside the RL loop except for a small number of recent converter-centric integrated controllers [43,48,51].

A limited number of studies extend RL action spaces to include both converter-level and supervisory variables. For example, centralized DRL controllers have been demonstrated for simultaneous MPPT, battery charging, and voltage regulation using direct duty-cycle control [48]. However, such approaches are typically confined to single-system prototypes, without incorporation of multi-timescale scheduling, market interaction, or distributed coordination. Consequently, despite isolated examples of low-level actuation, the dominant paradigm in the literature continues to define control variables at the EMS layer, leaving converter-level decisions structurally separated from system-level optimization.

4.5.3. PRQ3: RL Formulation Design for Joint PV–Battery Optimization and Evidence on Discrete vs. Continuous Control

MDP/POMDP Construction

Most works adopt a standard MDP in which the state combines battery SoC with exogenous drivers such as PV generation, load, prices, and time features. This pattern is consistent across residential HEMSs [53,54,87], community dispatch [42,57], and microgrid scheduling [56]. States reflect supervisory energy variables rather than electrical-layer dynamics. Härtel and Bocklisch [29] used PPO with an LSTM-based structure for PV–battery energy management, reporting performance comparable to MPC and superior to rule-based strategies. Xu et al. [30] formulated a residential PV–battery system as an MDP and applied model-based RL to improve training efficiency while increasing self-consumption and reducing operating cost.

Extensions beyond fully observable MDPs are limited. Härtel and Bocklisch [29] formulated a POMDP and employed an LSTM encoder within PPO to infer latent states from short histories. Multi-agent formulations address decentralized storage coordination with continuous actions and explicit coupling [88]. Despite structural variations, PV generation remains modeled as an exogenous input at the EMS layer. By contrast, converter-level DRL formulations use electrical measurements such as voltage, current, temperature, and duty cycle within the state, and define actions directly over the PWM duty cycle, switching signals, or other modulation variables. This confirms that the omission of PV operating-point control in most PV–battery EMS studies is not due to an intrinsic limitation of RL formulations but to the control layer at which the problem is posed [50].

Offline-to-Online Policy Approximation

Several works derive near-optimal dispatch policies offline and approximate them for real-time execution. Keerthisinghe et al. [89] generated optimal state–action trajectories using dynamic programming and trained a policy function approximator, enabling near-optimal real-time decisions without repeated online optimization. Mbuwir et al. [90] similarly combined fitted Q-iteration with clustering-based transfer learning, approaching MILP performance with reduced convergence time. These methods often treat PV output as an external input rather than a controllable variable, limiting MPPT–storage co-optimization.

Action Spaces and Algorithm Choices

A central design axis concerns discrete- versus continuous-action spaces. Discrete formulations, often paired with value-based methods, remain common in residential HEMSs and small microgrids, where battery power is quantized or operating modes enumerated [53,54,69,83,87]. Tabular Q-learning and Watkins Q(λ) are still used in degradation-aware scheduling [61,62], while SARSA has been applied to wind–PV–storage coordination [60]. Hybrid variants embed deterministic solvers within discrete RL loops (e.g., MINLP-guided Q-learning) [66]. Deep value-based extensions such as DDQN expand state dimensionality but retain discretized actions [65,91]. Discrete-action formulations can

improve energy capture but restrict fine-grained modulation [83]. Continuous actor–critic methods address this limitation by parameterizing actions as real-valued variables. For example, Zhou et al. [41] used DDPG for continuous coordination of PV–battery dispatch and EV demand, improving supply–demand balance and reducing operating cost relative to rule-based control. However, the term “continuous action” is used across different physical actuation layers in the literature. As summarized conceptually in Figure 16, continuous actions may correspond to supervisory dispatch variables, intermediate setpoint references, or direct converter-level commands. Figure 16 also distinguishes control actions according to their abstraction level, physical actuation variables, and typical operating time scale. This classification does not imply a hierarchy of technical maturity or importance. Supervisory EMS controllers address long-horizon and system-wide objectives, including power balance, battery utilization, economic operation, and grid interaction, whereas converter-level controllers provide fast and accurate electrical regulation. The intermediate coordination layer communicates system-level objectives and constraints to the converter controllers through appropriate references. Consequently, effective PV–battery control generally depends on coordinated operation across these layers, often through bidirectional coordination between them, rather than replacement of one layer by another.

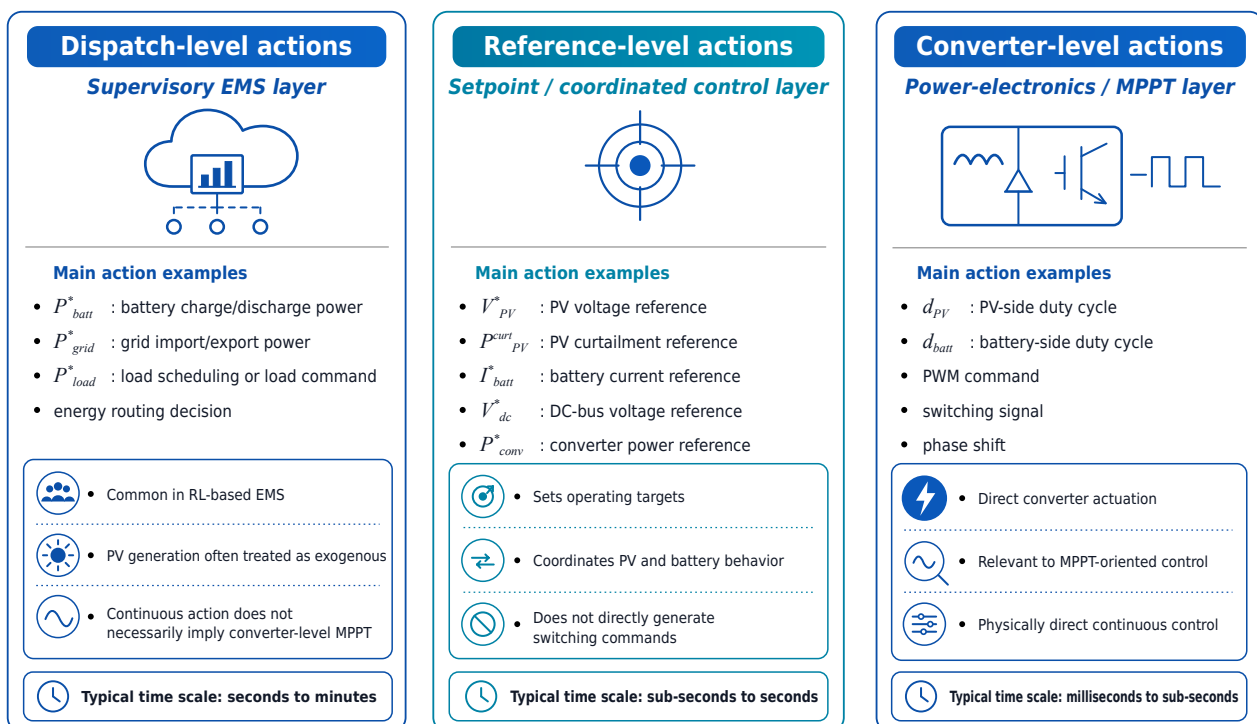


Figure 16. Complementary control layers and associated action variables in RL-based PV–battery systems. Supervisory EMS actions address system-level objectives such as battery power scheduling, grid exchange, load management, and energy routing. The coordination layer translates these objectives and operational constraints into electrical references, including PV-voltage, battery-current, DC-bus-voltage, and converter-power references. Converter-level controllers implement these references through duty-cycle, modulation, switching, or phase-shift actions. The layers operate at different time scales and abstraction levels but are complementary rather than hierarchical in technical value; practical high-performance systems generally rely on their coordinated operation. Converter-level actions directly manipulate power-electronic variables and are therefore specifically relevant when MPPT is formulated as a fast electrical-control problem.

Continuous-action actor–critic methods directly parameterize battery or grid power as real-valued variables, avoiding discretization at the EMS layer. In the converter-control literature, continuous actions are instead naturally associated with duty-cycle adjustment,

phase-shift control, and MPPT-related modulation, again underscoring that “continuous control” in the broader RL literature spans materially different actuation layers [50]. Examples include PPO-based scheduling [82], SAC under uncertainty [42], TD3 for microgrid and charging dispatch [55,81,88], and integral Q-learning or neuro-evolution-enhanced actor–critic schemes [67,68]. Continuous control is now widely adopted for multi-device coordination across storage, EVs, flexible loads, and multi-energy systems [41,71–73,75–77]. Multi-agent actor–critic architectures further scale coordination temporally or across prosumers [70,74].

Recent trends include safety-aware continuous RL with differentiable projection layers for hard constraint enforcement [86] and the structured benchmarking of PPO, SAC, TD3, and A2C against MILP or rule-based baselines [29,30,39,92,93].

Discrete vs. Continuous Evidence

Empirical evidence suggests advantages of continuous control when decisions correspond to power setpoints rather than mode switches. PPO has been shown to outperform DDQN/DDPG variants in convergence and profitability [82]. SAC demonstrates improved cost performance and constraint behavior under uncertainty relative to PPO and DDPG [42]. Projection-based continuous PPO reduces optimality gaps versus discrete DQNs while guaranteeing feasibility [56].

Conversely, discrete HEMS studies reveal quantization limitations. MILP outperforms DQNs under deterministic conditions [53], and MPC surpasses DQNs under accurate forecasts, with DQN performance being sensitive to reward shaping and action design [54]. Tabular methods remain competitive in small-scale problems but face scalability and resolution limits [60,62], while deep value-based discrete methods mitigate state complexity but retain action quantization [65]. Continuous actor–critic approaches typically yield smoother trajectories and improved convergence, particularly when degradation or stability-aware objectives are included [67,68].

However, in the majority of the reviewed RL studies, “continuous control” refers to supervisory dispatch decisions operating at hourly or multi-minute resolution rather than converter-level electrical control. Only a limited number of works extend continuous decision variables to converter duty-cycle control or direct PV operating-voltage optimization within integrated PV–battery formulations, although converter-level DRL studies and a small number of recent PPO-based integrated DER controllers demonstrate that such actions are technically feasible [48,50]. Consequently, the discrete-versus-continuous distinction in much of the reviewed literature primarily concerns scheduling and dispatch granularity rather than true cross-layer PV–battery control integration, with MPPT still being treated as external to the learning loop in most EMS-oriented formulations.

4.5.4. PRQ4: Practical Implementation Considerations (Training Environments, Constraints, Observability, Robustness, and Validation)

Training Environments and Data

Training and evaluation are predominantly simulation-based, typically using real-world datasets for PV generation, load demand, and electricity prices (e.g., Pecan Street or market datasets) [42,53,54,56,57]. Some works employ richer pipelines, such as SAM-generated PV data or simulator-based forecasts [52,58]. Forecast integration is increasingly common, including RNN-based 24-h PV/load prediction [65] and tariff forecasting in EV-integrated EMSs [63]. Pre-training on historical data followed by limited online adaptation has also been explored [60].

Uncertainty handling includes scenario-based learning [70], multi-timescale formulations under forecast errors [71], and offline-to-online approaches combining imitation learning from MIQP experts with SAC refinement [75]. However, robustness to model

mismatch, sensor noise, and forecast bias is rarely quantified explicitly. Some works illustrate that recent EMS-focused RL studies increasingly address practical issues such as network-constraint enforcement, uncertainty adaptation, and scalability but still do so at the supervisory dispatch layer rather than through integrated MPPT/converter-level control [94]. Validation remains largely simulation-bound. Early hardware-in-the-loop (HIL) demonstrations exist for continuous PV–battery EMS architectures [95], but PV remains exogenous and converter-level MPPT or duty-cycle control is not integrated. More broadly, the recent review literature also emphasizes that hardware validation remains limited and benchmarking practices remain non-standardized in standalone and converter-oriented microgrid control research. Esfahani et al. [96] employed GAN-based scenario generation to expose controllers to realistic variability patterns, improving resilience under high-variance operating conditions. Although isolated integrated RL demonstrations now exist for small DER systems with MPPT, bidirectional charging, and load-voltage control, the literature still lacks a mature validation pipeline for unified PV operating-point and battery co-optimization under realistic converter dynamics and deployment constraints [48,50,97].

Constraint Enforcement Mechanisms

Practical deployment requires explicit consideration of physical constraints such as SoC bounds, charge/discharge limits, voltage stability, and converter operating regions [57,94]. Constraint handling ranges from heuristic penalty shaping to optimization-based feasibility guarantees. Most approaches rely on reward penalties for SoC and power violations [53–55,87] or action clipping [52], which are simple but lack guarantees. More structured methods include pre-filtered feasible action spaces and hybrid RL–optimization schemes [64,66], Lagrangian relaxation [57], Lyapunov-based updates [67], robust H_∞ -inspired reward design [68], and externally coordinated multi-agent balancing [88]. More structured approaches explicitly enforce admissibility during policy execution. Huang and Wang [82] introduced a safety layer to project actions onto feasible regions for continuous battery dispatch via PPO. Fan and Wang [42] applied SAC with entropy regularization for robust community battery scheduling under uncertainty. Eyimaya [43] incorporated a safe action projection layer for an off-grid PV–battery nanogrid, ensuring physical admissibility during learning and execution. The general constraint-aware control pattern emerging from these studies is summarized conceptually in Figure 17.

Projection-based safety layers provide stronger feasibility guarantees. This trend is reinforced by recent SAC-based nanogrid work, where safe action projection is used to enforce SoC and charge/discharge limits before actions reach the plant. However, these guarantees remain focused on supervisory energy variables rather than converter switching constraints or PV operating-point feasibility [43]. Early feasibility checks [82] have evolved into differentiable optimization-based projections enabling gradient backpropagation through convex solvers [56,86]. Despite these advances, constraint handling remains focused on supervisory scheduling variables [41,70,73,76,77]. Converter-level protection, DC-bus dynamics, and subsecond constraints are generally not addressed in EMS-oriented studies, even though recent perspective and review papers explicitly identify the continued separation between RL-based EMS decisions and certified low-level protection/control layers as a major deployment constraint [50,98].

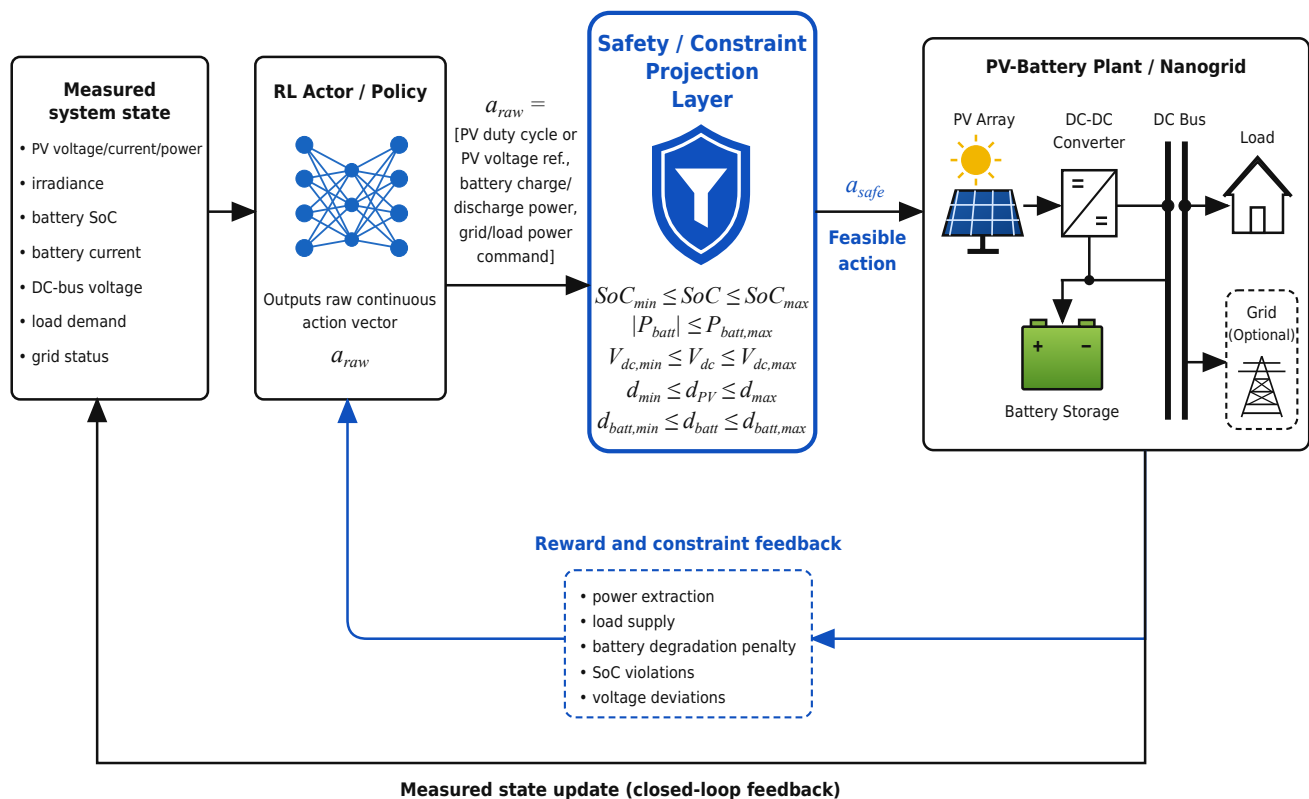


Figure 17. Constraint-aware reinforcement learning loop for integrated PV–battery control. The RL policy generates a raw continuous-action vector, including PV-side and battery-side control commands. A safety or projection layer then enforces operational limits related to battery SoC, battery power, DC-bus voltage, and converter duty cycle before applying the feasible action to the PV–battery plant. Feedback from the plant updates the measured state and provides reward and constraint information for policy learning.

Observability and SoC Estimation

Most studies assume direct, noise-free SoC availability [55,58], with little discussion of estimation uncertainty. Reliable state estimation is equally critical. Talha et al. [99] proposed neural network-based SoC estimation using voltage and current measurements, highlighting the importance of accurate observability for learning-based control. Partial observability is rarely modeled explicitly. Notable exceptions formulate the problem as a POMDP and use LSTM-based encoders within PPO to reconstruct latent states [29]. However, systematic analysis of estimation errors, sensor noise, and their impact on closed-loop stability remains limited. Recent perspective work likewise notes that standardized state representations and uncertainty-modeling frameworks remain underdeveloped, which further limits reproducibility and robustness assessment in RL-driven DC microgrid applications [98].

Validation and Physical Realism

Validation is predominantly simulation-based. Explainability tools have been introduced to interpret policy dependence on price and demand signals [87]. Some works verify RL schedules against physical microgrid models to ensure PCC voltage compliance [100], though this remains offline validation.

Physical degradation modeling is increasingly incorporated. Inverter (IGBT) wear has been embedded via mission-profile-based modeling [101] and battery aging via semi-empirical degradation laws integrated through structured reward shaping [102]. While these improve physical relevance, validation remains at supervisory time scales. Converter-control reviews similarly identify sim-to-real transfer, safe exploration, hardware deploy-

ment constraints, and benchmarking gaps as unresolved issues, indicating that the challenge is not merely the missing algorithms but the absence of validated cross-layer integration procedures [50].

Synthesis and Implication

The literature demonstrates methodological maturity of RL at the EMS layer, including actor–critic methods, hybrid RL–optimization architectures, multi-agent coordination, degradation-aware objectives, and projection-based safety mechanisms. At the same time, recent converter-control reviews show that DRL is also capable of MPPT, duty-cycle, and inverter-level control when the problem is formulated at the power-electronics layer. Yet these two strands remain largely disconnected. In the dominant PV–battery EMS literature, PV generation is still treated as exogenous or implicitly MPPT-operated, and converter-level operating-point control is almost never included as a learning variable. A few recent proof-of-concept studies have begun to unify PV MPPT, bidirectional battery charging, and load-voltage regulation within a single RL controller, but they remain limited in scale and do not yet establish a general framework for safety-aware, storage-aware, and converter-feasible joint optimization. This persistent disconnect motivates unified continuous-control formulations that jointly optimize PV operating-point control and battery dispatch while enforcing converter-level feasibility, protection constraints, and dynamic coupling.

5. Discussion and Research Gaps

This article reviewed the progression of MPPT techniques from conventional perturbation-based algorithms to intelligent approaches based on ANNs, deep learning, and RL. The trajectory is clear: increasing PV system complexity and operating non-stationarity have driven methods toward higher adaptability and richer representations, particularly under fast irradiance variations and partial shading.

Conventional MPPT methods such as P&O and IncCond remain attractive due to low complexity, yet comparative studies consistently report the same structural weaknesses: steady-state oscillations, sensitivity to irradiance transients, and vulnerability to local maxima under partial shading. Improved variants and optimization-assisted approaches mitigate specific symptoms but remain rooted in instantaneous, local-search logic.

Supervised learning methods (ANN and deep learning) reduce oscillations by replacing perturbation with nonlinear regression from measurements to an estimated optimum. While these approaches often achieve high tracking efficiency under trained conditions, their performance remains tightly coupled to offline dataset coverage, and they generally lack reliable online adaptation when conditions deviate from the training distribution.

Reinforcement learning introduces a different paradigm by optimizing policies through interaction and cumulative reward. Discrete-action RL validates the sequential decision formulation but remains constrained by action quantization. Continuous-action actor–critic methods better match power-electronic control variables and frequently report smoother duty-cycle trajectories and improved transient performance. However, much of the RL MPPT literature still optimizes instantaneous PV power and relies predominantly on simulation-based validation.

A central finding of this review is that most RL-based PV–battery studies are formulated at the EMS or supervisory layer, where battery dispatch, grid exchange, or load coordination are optimized, while PV operating-point selection is assumed, abstracted, or delegated to a separate MPPT loop. Thus, the main limitation in the literature is not simply that RL has been used too little but also that it has often remained concentrated primarily at supervisory control layers rather than extending to converter-level PV operating-point optimization and unified MPPT–battery coordination.

A consistent insight across PV–battery studies is that strict instantaneous MPPT is not necessarily system-optimal once storage is present. PV operating-point selection influences battery charging current, SoC trajectory, conversion losses, and potentially degradation. Consequently, treating MPPT and battery energy management as separate layers can yield conflicting actions, avoidable curtailment, and inefficient cycling, particularly under variable loads and constrained SoC headroom.

These observations support reframing MPPT in PV–battery systems as part of a coupled, constrained, long-horizon optimization problem rather than an isolated instantaneous maximization task.

More specifically, the literature reveals a persistent separation between converter-level PV control and system-level energy management. This layer mismatch helps explain why unified formulations that jointly optimize duty-cycle-level PV actuation and battery-aware long-horizon objectives remain relatively uncommon.

The literature on PV–battery systems reveals several consistent insights:

- Strict instantaneous MPPT can become sub-optimal when storage, converter, or grid constraints are present.
- Decoupled MPPT and EMS architectures may insufficiently represent temporal interdependence.
- Reinforcement learning provides one possible framework for unified, long-horizon optimization.
- Continuous-action formulations are particularly well aligned with converter-level control.

Nevertheless, fully integrated continuous-action frameworks that simultaneously optimize PV operating-point selection and battery dispatch remain comparatively underexplored. Many existing works either assume fixed MPPT operation or treat PV generation as exogenous.

This gap motivates the investigation of unified formulations in which duty-cycle control and storage scheduling are optimized jointly within a single sequential decision-making framework.

5.1. Identified Research Gaps

Based on the reviewed literature, several gaps remain:

- Objective formulation gap: Many approaches, including RL-based MPPT, continue to optimize instantaneous PV power even when long-term metrics (battery stress, degradation-aware operation, and efficiency over time) are dominant.
- Coordination gap: MPPT and battery management are often treated as separable problems, leaving the PV–converter–battery coupling underexploited.
- Control-layer gap: Much of the RL literature is formulated at the supervisory or EMS layer, where PV generation is treated as exogenous or implicitly MPPT-operated, rather than at the converter layer where PV operating-point selection can be directly optimized.
- State observability gap: Many integrated control formulations assume reliable access to battery internal states such as SoC and, sometimes, degradation-related indicators, even though these variables must be estimated online and may be uncertain, noisy, or model-dependent in practice.
- Adaptation gap: Supervised MPPT methods remain largely offline-trained and can degrade under operating regimes not represented during training (e.g., evolving parameters, atypical shading patterns, and aging).
- Control fidelity and constraint gap: Discrete abstractions, reward-only constraint penalties, and action clipping can be poorly aligned with continuous converter control

and strict safety constraints (SoC, current, and voltage), motivating more principled constrained/safe RL formulations.

- Training-efficiency gap: Coupled PV–battery environments are high-dimensional and temporally dependent, which increases sample requirements and computational burden for RL training and limits reproducibility and practical online adaptation.
- Validation gap: Integrated PV–battery learning-based controllers are still predominantly evaluated in simulation; systematic hardware validation and sim-to-real methodologies remain limited.
- Benchmarking gap: Cross-study comparison remains difficult because the literature uses heterogeneous PV–battery architectures, irradiance and load profiles, reward definitions, constraint treatments, and evaluation metrics, making it hard to isolate the true benefit of integrated continuous-action control.

5.2. Time-Scale Separation and Cross-Layer Coordination

Classical time-scale separation remains justified in PV–battery systems because fast current, voltage, and DC-link regulation operate at different bandwidths and impose distinct stability, protection, communication, and computational requirements from supervisory energy management [103,104]. The principal limitation is therefore not the layered architecture itself but insufficient coordination among the objectives assigned to its layers. For example, an MPPT controller may maximize instantaneous PV power without considering battery-SoC headroom, charging limits, degradation, or future demand. Cross-layer coordination can address these conflicts by sharing states, constraints, and references while retaining established fast inner loops. Thus, the decision context may be integrated without combining every physical control time scale.

The reviewed evidence illustrates both the potential and the limitations of greater integration:

- Multi-time-scale and hierarchical coordination: Watari et al. [105] reported an electricity-cost reduction of up to 47.5% using a coordinated multi-time-scale formulation relative to the two evaluated single-time-scale formulations. Hu et al. [71] combined hourly optimization with real-time DRL corrective control, noting that complete reliance on a single learning agent may reduce reliability and interpretability. Liu et al. [83] directly compared a hierarchical formulation with a packaged agent that jointly selected power-purchasing and charging decisions. The hierarchical structure achieved a higher reward and less oscillatory convergence, indicating that an enlarged joint action space can impair learning.
- Centralized converter control: Wiencek and Ghosh [48] used one PPO agent to control PV MPPT, battery charging/discharging, and inverter voltage regulation. Voltage THD decreased from approximately 1.70% to 0.23% relative to the evaluated conventional controllers. However, the agent occasionally extracted less PV power than P&O, and validation was limited to a simulated 200 W system without hardware testing or a formal closed-loop stability proof.
- Computational trade-offs: Härtel and Bocklisch [29] trained 216 PPO–LSTM policy configurations, with full-run performance assessed after 48 h, but obtained energy costs only similar to or slightly better than MPC. Conversely, Cai et al. [93] reported a 37.7% improvement in their evaluated performance measure relative to TD3 using MPC–RL. Although offline training was longer, each online MPC calculation required only 0.08 s within a 15 min supervisory interval.

These results do not prove that physical time-scale separation should be abandoned. Rather, they support selective integration in which shared supervisory objectives coordinate PV and storage decisions while conventional inner control, protection, and fallback

functions remain active. Fully joint control should therefore be regarded as a testable architectural alternative, not as an inherently superior replacement for hierarchical control.

A definitive assessment requires matched comparisons under identical models, operating profiles, constraints, control intervals, and computing platforms. Such comparisons should report energy and life-cycle performance together with training effort, online computation, constraint violations, closed-loop robustness, and hardware performance.

5.3. Future Directions

The literature suggests potential for unified continuous-action control frameworks that jointly optimize PV operating-point selection and battery dispatch under realistic constraints. Continuous-action RL may be a potential candidate due to its ability to optimize long-horizon objectives and operate directly over real-valued control actions. Progress toward deployable solutions will likely require: (i) constraint-aware policy optimization (e.g., safe layers or CMDP formulations); (ii) reward structures that reflect battery health, curtailment trade-offs, and system-level efficiency over time; (iii) improved state estimation and observability for variables such as SoC and degradation-related indicators; (iv) robustness to model mismatch and non-stationarity via uncertainty-aware or domain-randomized training; (v) improved sample efficiency through hybrid or model-informed learning architectures; and (vi) broader experimental validation and benchmarking beyond simulation.

This work motivates these directions by positioning PV–battery MPPT as a coupled sequential decision-making problem where control fidelity, safety, and long-term optimality must be addressed simultaneously.

In this sense, the key open challenge is not only to improve MPPT accuracy but also to bridge converter-level PV actuation and battery-aware energy management within a single practical learning framework.

6. Conclusions

This article presented a structured literature review on RL for integrated MPPT and battery management in PV systems. This review showed that the development from conventional perturbation-based MPPT methods toward intelligent and learning-based control has been driven by the increasing complexity, variability, and coupling of modern PV–battery energy systems. While classical MPPT methods remain attractive because of their simplicity, their limitations under fast irradiance changes, partial shading, and storage-coupled operation motivate the investigation of more adaptive control strategies.

The reviewed literature indicates that RL provides a useful framework for formulating control as a sequential decision-making problem and for considering long-term objectives beyond instantaneous power extraction. Continuous-action actor–critic methods are relevant to converter-based energy systems because they can operate directly over real-valued control variables and may provide smoother control actions than discretized approaches. However, their performance depends strongly on assumptions regarding state observability, reward formulation, model fidelity, constraint enforcement, and the training environment. Moreover, most existing RL-based studies remain concentrated at the supervisory or energy management layer, where PV generation is often treated as exogenous and MPPT is implemented outside the learning loop.

A key conclusion of this review is therefore that the main research gap is not simply the limited use of RL but also the limited coordination between converter-level PV operating-point control and battery-aware long-horizon optimization. In PV–battery systems, strict instantaneous MPPT is not always system-optimal, since PV operating-point decisions affect battery state of charge, charging stress, efficiency, curtailment, and long-term performance. This motivates explicit coordination between MPPT and battery man-

agement, whether implemented through a hierarchical layered architecture or a unified decision framework.

Finally, under specific assumptions and constraints, continuous-action RL may provide new ideas for integrated control, but its challenges in constraint handling, stability, and hardware deployment still need to be overcome. Future work should therefore investigate safe and constraint-aware learning formulations, stability assessment, improved state estimation, hybrid model-informed approaches, and systematic sim-to-real and experimental validation. Addressing these issues is necessary before such methods can be considered practical and deployable for PV–battery energy systems.

Author Contributions: Conceptualization, F.F. and R.B.; Methodology, F.F. and R.B.; Formal Analysis, F.F.; Investigation, F.F.; Writing—Original Draft Preparation, F.F.; Writing—Review and Editing, F.F. and R.B.; Visualization, F.F.; Supervision, R.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT GPT-5.5 for the purposes of generating preliminary versions of selected figures and graphical illustrations. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

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

Abbreviations

The following abbreviations are used in this manuscript:

ANN	Artificial Neural Network
BESS	Battery Energy Storage System
CNN	Convolutional Neural Network
DDPG	Deep Deterministic Policy Gradient
DRL	Deep Reinforcement Learning
EMS	Energy Management System
GA	Genetic Algorithm
GMPP	Global Maximum Power Point
INC	Incremental Conductance
LMPP	Local Maximum Power Point
LSTM	Long Short-Term Memory
MDP	Markov Decision Process
MPP	Maximum Power Point
MPPT	Maximum Power Point Tracking
OCV	Open-Circuit Voltage
P&O	Perturb and Observe
PSO	Particle Swarm Optimization
PV	Photovoltaic
RL	Reinforcement Learning
RNN	Recurrent Neural Network
SAC	Soft Actor–Critic
SoC	State of Charge
TD3	Twin Delayed Deep Deterministic Policy Gradient

Appendix A. Summary of Included Studies

Table A1. Assessment of PRISMA-reviewed articles against the PRQs, with justification and control-layer notes.

Paper	Architecture	Battery Actuation	Method	Layer	RL State	RL Action	RL Reward	RL Constraints	PRQ Coverage	Objective
[95]	A2	C1	D5	E2	PV forecast; load forecast; SoC; price; hour	Battery charge/discharge; grid import/export	Cost ↓; battery health ↑; grid benefit ↑	SoC bounds; over-/deep-discharge prevention; ToU grid scheduling; DoD penalty	PRQ2–PRQ4	Forecast-informed residential EMS; cost- and battery-aware continuous dispatch; HIL/embedded validation and PV exogenous/forecasted; implicit MPPT; no PV actuation
[64]	A1	C1	D4	E1	Time; EVCS load; PV forecast; tariff forecast; battery SoE	Discrete battery power; grid power	Energy cost ↓; degradation cost ↓; feasible action reward	DC-bus balance; SoC/SoE bounds; grid-power limits; feasible action set	PRQ2–PRQ4	Discrete EMS scheduling for PV-EVCS battery system; improved Q-learning; cost/degradation minimization and PV exogenous/forecasted; implicit MPPT; no PV operating-point control
[61]	A1	C1	D4	E1	Net demand; SoC	Charge; discharge; grid import; grid export	Grid cost ↓; export revenue ↑; battery use ↑; violations ↓	Power balance; grid limit; SoC/energy bounds; charge/discharge efficiency; episode violation check	PRQ2, PRQ3	Residential PV–battery EMS; discrete Q-learning; household cost minimization and battery-use policy, and PV exogenous; self-consumption/export only; implicit MPPT
[58]	A2	C1	D5	E2	Time/date features; weather; SoC; PV/load/heat-pump nowcast and forecast; grid setpoint	Battery power; PV output scaling; heating/cooling power	Grid-setpoint error ↓; grid smoothing; self-sufficiency ↑	Battery power/energy bounds; comfort limits; PV scaling 0–1; benchmark PV-surplus charging rule	PRQ2–PRQ4	Community DER coordination with DDPG; battery, PV curtailment, and climate-load control; MILP benchmark and PV curtailment/reference control; external MPPT overridden; no duty-cycle control
[87]	A1	C1	D4	E1	Net demand; grid price; SoC	Discrete battery charge/discharge levels; idle	Grid cost ↓; battery contribution ↑; state-weighted reward	Discrete 10% action steps; SoC evolution; net-demand balance; residual grid demand	PRQ2–PRQ4	Building EMS with PV and battery; DQN/PPO comparison; explainable battery scheduling and PV exogenous via net demand; implicit MPPT; no PV actuation
[93]	A2	C1	D5	E2	Thermal states; tank temperatures; battery energy; PV/load/weather forecasts	Battery charge/discharge; grid buy/sell; heat-pump power; valve opening	Energy cost ↓; comfort penalty ↓	Tank temperature bounds; heat-pump/valve limits; battery bounds; no simultaneous charge/discharge; grid limits; no simultaneous buy/sell; power balance	PRQ2–PRQ4	HEMS multi-energy coordination; MPC-based actor–critic vs TD3; cost, comfort, uncertainty, and safety focus, and PV exogenous/forecasted; implicit MPPT; no PV operating-point control

Table A1. Cont.

Paper	Architecture	Battery Actuation	Method	Layer	RL State	RL Action	RL Reward	RL Constraints	PRQ Coverage	Objective
[91]	A1	C1	D5	E1	Time; PV generation; SoC; load; real-time price	Charge; discharge; idle	Transaction cost ↓; degradation cost ↓; violations ↓	Battery power limits; SoC bounds; energy balance; storage-operation penalties	PRQ2–PRQ4	Real-time microgrid ESS dispatch with DQN; price-aware energy trading; prioritized replay and PV exogenous; implicit MPPT; no PV actuation
[106]	A2	C1	D6	E2	TES SoC; indoor/setpoint temperature; price forecast; time/occupancy/weather; BESS SoC; PV production; building load	Cooling mode; cooling fraction	Electricity cost ↓; comfort error ↓; peak shaving	TES limits; mode feasibility; BESS rule-based limits; comfort band	PRQ2–PRQ4	Heterogeneous transfer learning for building DRL; cooling/TES control with PV-BESS target system and PV/BESS rule-based external control; no PV actuation; no unified PV–battery RL
[56]	A2	C1	D6	E2	Month; hour; workday; load forecast; PV forecast; market price forecast; SoC	Renewable allocation; battery–grid exchange; PV/battery/grid routing	Grid transaction cost ↓; battery cost ↓; profit ↑	Power balance; PV limit; SoC bounds; charge/discharge limits; no simultaneous C/D; grid consistency; projection layer	PRQ2–PRQ4	Decentralized PV–battery microgrid EMS; continuous MicroPPO; differentiable safety projection and PV power routing/curtailment only; implicit MPPT; no operating-point control
[101]	A2	C1	D6	E2	SoC; PV power; load; inverter power; curtailed PV; ambient temperature; 4-step horizon	Battery power; diesel-generator power	Diesel cost ↓; curtailment cost ↓; IGBT degradation ↓; SoC violations ↓	DC/AC balance; SoC bounds; battery limits; inverter/DG ratings; PV curtailment bounds; 4-step dispatch	PRQ1–PRQ4	Reliability-aware isolated microgrid EMS; SAC dispatch with inverter electrothermal degradation and PV curtailment computed; MPPT external/implicit; no converter operating-point RL
[43]	A2	C1	D6	E2	SoC; recent load/PV history; short-horizon forecasts; time encodings	Battery power setpoint; generator on/off intent	Unmet load ↓; fuel use ↓; battery stress ↓; starts ↓	SoC bounds; charge/discharge limits; ramp limits; generator up-/downtimes; startup delay; safe action projection	PRQ1–PRQ4	Off-grid nanogrid EMS with SAC; forecast-informed safe resource dispatch and PV exogenous/available; de facto MPPT; implicit curtailment; no PV operating-point control
[42]	A1	C1	D6	E1	Battery energy; buy/sell prices; residential load; PV generation	Battery charge/discharge; grid-interaction fraction	System cost ↓; degradation cost ↓; grid violations ↓	Battery power limits; no simultaneous C/D; energy bounds; power balance; import/export logic; grid capacity limits	PRQ1–PRQ4	Community battery scheduling with SAC; uncertainty-aware shared-storage coordination; NoisyNet exploration and PV exogenous; implicit MPPT; storage/grid scheduling only
[102]	A1	C1	D6	E1	Load; SoC; PV generation; electricity price	Continuous battery charge/discharge power	Grid cost ↓; sell revenue ↑; degradation ↓; SoC preservation ↑	Power balance; SoC dynamics; battery capacity degradation; inverter efficiency; reward-shaped safe operation	PRQ2–PRQ4	Residential PPO battery EMS; degradation-aware reward shaping; year-long simulation and PV exogenous; no curtailment action; implicit/absent MPPT

Table A1. Cont.

Paper	Architecture	Battery Actuation	Method	Layer	RL State	RL Action	RL Reward	RL Constraints	PRQ Coverage	Objective
[107]	A2	C1	D5	E2	Weather; time; PV generation; non-HVAC load; TESS state	HVAC thermal-storage charge/discharge	PV-load mismatch ↓; consumption matching ↑	TESS bounds; ESS/EV constraints; controllable-load limits; power balance; iterative coupling consistency	PRQ1–PRQ4	Hybrid connected-building EMS; DDPG for HVAC/TESS plus optimization for ESS, EVs, and loads and PV exogenous; EMS-level PV matching; no MPPT/PV actuation
[72]	A1	C1	D6	E1	Prices; time; load; PV generation; SoC; feasible action bounds	Continuous BESS charge/discharge	Social welfare ↑; battery cost ↓; constraint violations ↓	BESS power/action bounds; PV availability; SoC dynamics; energy balance; import/export logic; clipping and penalties	PRQ1–PRQ4	Renewable energy community BESS scheduling; SAC with MILP-assisted training; real-time welfare optimization and PV exogenous; no curtailment/MPPT action; storage-only dispatch
[65]	A1	C1	D4	E1	SoC; fuel state; PV power; load; tariff; time; 24-h grid-power forecast	Discrete battery command; microturbine command	Operating cost ↓; SoC/fuel/grid violations ↓	Battery limits; SoC/end-day SoC bounds; microturbine limits; fuel bound; feeder limit; power balance	PRQ1–PRQ4	Forecast-informed microgrid EMS; DDQN battery and microturbine scheduling; DQN/PPO benchmark and PV forecast/exogenous; no curtailment action; implicit/absent MPPT
[108]	A2	C2	D4	E2	BES reference; measured BES power; tracking error/integral; power-limit errors/integrals	Per-battery charge/discharge power setpoints	BES tracking error ↓; power-limit violations ↓; action penalty ↓	SoC safety bounds; SOC-based power limits; grid-power tracking; charge/discharge mode selection; conventional DC-link/dq control	PRQ1–PRQ4	TD3 supervisory storage allocation for PV-BES quasi-Z-source CHB inverter; module-level battery coordination and PV under conventional MPPT; RL dispatches BES only; no unified MPPT-battery control
[71]	A2	C1	D6	E2	Net load; battery SoC; supercapacitor SoC	Grid-power correction; battery-power correction; supercapacitor power	Grid deviation ↓; battery deviation ↓; SC SoC error ↓; SC degradation ↓	Power balance; grid limits; battery/SC power limits; SoC bounds; storage dynamics	PRQ1–PRQ4	Two-stage hybrid-storage EMS; hourly optimization plus SAC real-time correction and PV/wind exogenous; no PV actuation; MPPT absent
[82]	A2	C1	D6	E2	SoC feasible space; LMP; regulation signal; regulation price; available PV power	PV-charging ratio; arbitrage ratio; regulation–capacity ratio	Net profit ↑; degradation cost ↓	SoC bounds; battery power capacity; minimum regulation bid; PV-charging headroom; serial safety control	PRQ1–PRQ4	Market-facing PV–battery EMS; PPO service-allocation policy with safety filter and PV available power allocated; no operating-point control; MPPT absent
[29]	A1	C1	D6	E1	SoC; past load energy; past PV energy; LSTM history encoding	Continuous grid-power setpoint	Energy cost ↓; feed-in revenue ↑	Inverter power limit; battery capacity; feed-in cap; curtailment accounting; efficiency/SoC dynamics	PRQ2–PRQ4	Recurrent PPO PV–battery EMS; cost minimization; rule-based and MPC comparison and PV exogenous; battery via grid setpoint; implicit/external MPPT

Table A1. Cont.

Paper	Architecture	Battery Actuation	Method	Layer	RL State	RL Action	RL Reward	RL Constraints	PRQ Coverage	Objective
[52]	A1	C1	D6	E1	Time; irradiance; load; EV state/input; buy/sell prices	EV charge/discharge; V2H/V2G; grid sharing	Cost error ↓; monthly cost ↓	Grid balance; EV time window; EV SoC bounds; EV power limits; PV from irradiance model	PRQ2–PRQ4	Residential PV-EV HEMSs; PPO V2H/V2G scheduling; cost-minimization benchmark and PV available/exogenous; EV battery dispatch; MPPT absent
[86]	A1	C1	D6	E1	Irradiance; inflexible load; price; ESS SoC; temperatures; peak demand; appliance/EV states; time	Grid import; ESS power; PV utilization; heat-pump power; dishwasher; EV charging	Operating cost ↓; O&M cost ↓; peak cost ↓	ESS SoC; PV upper bound; comfort bound; EV feasibility; dishwasher completion; power balance; projection layer	PRQ2–PRQ4	Safe building EMS with DDPG; differentiable projection for feasible storage/load scheduling and PV utilization variable only; no PV operating-point control; MPPT absent
[70]	A1	C1	D6	E1	Transformer loading; microgrid exchange; BESS power; SoC; ToU/FIT/carbon rates; hour	Continuous BESS charge/discharge signal	Energy cost ↓; degradation ↓; emissions ↓; constraint penalties ↓	AC power flow; voltage/current limits; BESS limits; transformer thermal limit; SoC dynamics	PRQ2–PRQ4	Predictive microgrid EMS; 24-agent DDPG for day-ahead BESS scheduling and PV exogenous; no PV action; MPPT absent
[38]	A1	C0	D4	E3	PV voltage; PV current; duty cycle; duty-cycle change	Continuous duty-cycle perturbation	PV power ↑; MPP tracking ↑; invalid duty cycle ↓	Duty-cycle bounds; perturbation bounds; fault/shading scenarios; DDPG continuous control	PRQ1–PRQ4	Converter-level DRL MPPT; fault-tolerant boost-converter control; comparison with INC and direct PV duty-cycle actuation; RL-MPPT; no battery control
[89]	A1	C1	D3	E1	SoC; demand; PV output; tariff; next-step PV/load predictions	Grid-power policy; reconstructed battery charge/discharge	Electricity cost ↓; DP/optimization objective	Energy balance; inverter/battery efficiency; self-discharge; SoC bounds; charge/discharge limits; peak-period flow filters	PRQ2, PRQ4	Policy-approximation PV-battery EMS; offline optimization-to-policy learning; low-computation deployment and PV exogenous; battery dispatch only; MPPT absent
[39]	A1	C1	D5	E1	Battery size; battery charge; household load; controlled load; PV output; residual demands	Solar-to-battery charge; tariff-to-battery charge; battery discharge	Electricity cost ↓; solar service ↑	Battery capacity/SoC limit; solar-first charging; tariff charging; residual grid purchase; fixed tariff structure	PRQ2–PRQ4	Domestic DDPG PV-battery EMS; continuous battery scheduling; battery-size and hyperparameter analysis, and PV available/exogenous; no PV-side action; MPPT absent
[109]	A1	C1	D6	E1	Load; PV output; SoC; history window for POMDP	Diesel-generator dispatch	Diesel cost ↓; power imbalance ↓; unserved/wasted energy ↓	DG output limits; battery power limit; energy bounds; efficiency losses; imbalance penalties	PRQ2–PRQ4	Isolated microgrid DRL dispatch; FH-DDPG/FH-RDPG under uncertainty and partial observability, and PV stochastic/exogenous; battery balances surplus/deficit; MPPT absent

Table A1. Cont.

Paper	Architecture	Battery Actuation	Method	Layer	RL State	RL Action	RL Reward	RL Constraints	PRQ Coverage	Objective
[54]	A1	C1	D4	E1	Load; EV presence; EV SoC; BSS SoC; prices; PV production; time/date	Discrete EV/BSS charge–discharge-grid combinations	Grid cost ↓; EV demand satisfaction ↑; low-SoC penalty ↓	House balance; EV/BSS power limits; SoC bounds; EV availability; grid limits; no simultaneous C/D	PRQ1–PRQ4	Comparative HEMS study; RL vs rule-based, optimization, and MPC; PV-BSS-EV coordination and PV exogenous; energy-use management only; MPPT absent
[83]	A2	C1	D6	E2	EV set/SoC; PV/grid prices; PV forecast; ESS energy; tier-2 EV SoC	PV/grid/ESS procurement; ESS arbitrage; EV charging rates	Cost ↓; PV self-consumption ↑; ESS utility ↑; CVaR risk ↓	PV production/demand limits; ESS energy/power limits; EV power/SoC bounds; discrete trading/charging levels	PRQ1–PRQ4	Hierarchical PV-CS-EV RL; two-tier power trading and EV charging; risk-aware multi-objective EMS and PV forecast/procurement only; no operating-point control; MPPT absent
[77]	A1	C1	D6	E1	Wind/PV output; load; prior thermal outputs; ESS SoC; real-time price	Thermal-generator dispatch; ESS charge/discharge	Emissions ↓; grid cost ↓; degradation ↓; imbalance ↓; correct ESS charging ↑	Demand balance; generator/ramp limits; grid limits; ESS power/SoC bounds	PRQ2–PRQ4	GAN-assisted DRL microgrid dispatch; enhanced DDPG with ESS reward shaping and PV/wind exogenous; ESS dispatch only; MPPT absent
[67]	A1	C1	D6	E1	Grid-power state; PV generation; load; EV demand; battery energy; price	Battery charge/discharge policy	Trading cost ↓; battery performance ↑	PV model; EV bounds; battery power/energy bounds; hourly power balance	PRQ2–PRQ4	Integral Q-learning microgrid battery management; energy trading and battery-life focus and PV available/exogenous; battery-centered dispatch; MPPT absent
[62]	A1	C1	D4	E1	Time; PV output; load; battery energy; tariff	Discrete battery charge/discharge levels; idle	Availability ↑; grid cost ↓; degradation ↓; overflow penalty ↓	Battery energy/SoC limits; grid import/export limits; DoD degradation cost; 1-h step	PRQ2–PRQ4	Residential PV–battery MDP; Watkins Q(λ); degradation-aware tariff scheduling and PV available/exogenous; battery scheduling only; MPPT absent
[63]	A1	C1	D4	E1	Time; load; available PV; tariff; EV battery SoE	Discrete EV-battery SoE target; charge/discharge level	Grid cost ↓; degradation cost ↓	Power balance; SoE bounds; battery power limits; no grid export; 1-h step	PRQ2–PRQ4	PV-EV charging-station EMS; Q-learning vs PSO; LSTM tariff forecasting and PV available/exogenous; EV battery scheduling; MPPT absent
[60]	A1	C1	D4	E1	ES energy; wind forecast; PV forecast	ES charge/discharge; wind output increment; PV output increment	Tracking deviation ↓; ES operating cost ↓	Wind/PV bounds and ramps; ES power/energy limits; no simultaneous C/D; equal start/end energy	PRQ2–PRQ4	Centralized renewable-storage EMS; SARSA plan tracking; online learning vs MPC/DQN and EMS-level PV output adjustment; no converter actuation; MPPT absent

Table A1. Cont.

Paper	Architecture	Battery Actuation	Method	Layer	RL State	RL Action	RL Reward	RL Constraints	PRQ Coverage	Objective
[78]	A2	C2	D6	E3	DC-link voltage; DC-link voltage reference	PI correction for DC–DC duty cycle	DC-link error ↓; action variation ↓	DC-link reference; cascaded battery PI loops; external P&O MPPT; converter duty-cycle regulation	PRQ1–PRQ4	Hybrid PI+TD3 DC-link control for PV–battery system; converter-level stabilization benchmark and external P&O MPPT; battery PI control; coordinated but not unified RL-MPPT
[74]	A2	C1	D6	E1	Purchased electricity; central ESS SoC/capacity; household demands; hour	Household ESS request intensity	Purchase cost ↓; selfish request penalty ↓; ESS degradation ↓	ESS availability; greedy allocation; SoC transition; capacity degradation	PRQ2–PRQ4	MARL community PV sharing; central ESS allocation among households; MATD3 comparison and PV collected as energy; ESS allocation only; MPPT absent
[110]	A1	C1	D6	E1	PV generation bin; load bin; battery energy bin	Charge; discharge; idle	Grid cost ↓; PV/battery utilization ↑	Battery energy update; DoD lower bound; max energy; mutually exclusive actions; action eligibility	PRQ2–PRQ4	Microgrid Q-learning EMS; battery scheduling for self-consumption and grid-cost reduction and PV available/forecasted; battery-only control; MPPT absent
[69]	A1	C1	D4	E1	VRB SoC range; LAB SoC range; battery terminal power range	VRB on/off; LAB on/off	Battery loss ↓; unavailable selection penalty ↓	Source–load power balance; SoC bounds; power ratings; master–slave converter architecture	PRQ1–PRQ4	Heterogeneous ESS Q-learning; high-level storage selection; low-level converter/MPPT separation and PV MPPT handled externally; RL coordinates storage only
[111]	A1	C1	D6	E1	Time; ESS SoC; appliance waiting times; PV/load forecasts	ESS charge/idle/discharge; appliance on/off	HEMS profit ↑; electricity cost ↓	ESS SoC update/bounds; fixed charge power; appliance waiting limits; device-load constraints	PRQ2–PRQ4	Seq2Seq+Q-learning smart-home EMS; ESS and appliance scheduling under PV forecasts and PV forecast/exogenous; no PV actuation; MPPT absent
[79]	A2	C1	D6	E2	Three-phase voltage; available PV; SoC; recognized-load outputs	PV feed-in share; battery share; phase-distribution rates	Voltage deviation ↓; voltage unbalance ↓	SoC 20–80%; PV/battery policy restrictions; power-flow feasibility; Q(U) law; phase-rate limits	PRQ1–PRQ4	Grid-supportive PV–battery node control; voltage-based load recognition; active-power balancing and PV active-power regulation; no MPPT duty-cycle control; grid-support focus
[100]	A1	C1	D4	E1	Net demand; BSS energy; electricity cost	BSS power setpoint; grid-power setpoint	Daily electricity cost ↓	Power balance; BSS energy bounds; charge/discharge limits; PCC voltage check	PRQ2–PRQ4	Community microgrid Q-learning/SARSA EMS; battery–grid scheduling with electrical-model validation and PV available/exogenous; battery–grid scheduling; MPPT absent

Table A1. Cont.

Paper	Architecture	Battery Actuation	Method	Layer	RL State	RL Action	RL Reward	RL Constraints	PRQ Coverage	Objective
[68]	A1	C1	D6	E1	Grid power; battery energy deviation; price/load/solar profiles; disturbance	Battery charge/discharge/idle law	Grid cost ↓; battery deviation ↓; cycling ↓; disturbance rejection ↑	Battery energy bounds; no simultaneous C/D; load balance; solar-to-load priority; no export	PRQ2–PRQ4	Smart-house solar–battery EMS; Q-learning with H_{∞} /NEAT control design and solar available/exogenous; battery scheduling only; MPPT absent
[55]	A2	C1	D5	E2	PV power; battery energy; load power	Battery power; grid interaction; flexible charging power	PV consumption ↑; curtailment ↓; reserve loss ↓; charging adjustment reward ↑	Power balance; battery capacity; grid-power limit; charge/discharge limits	PRQ2–PRQ4	PEDF microgrid EMS; improved TD3 for PV–storage–grid-flexible charging coordination and PV available generation; flexible load/storage dispatch; MPPT absent
[76]	A1	C1	D5	E1	Wind/PV generation; electric/heat loads; energy prices; electrical/thermal storage states	CHP/boiler/heat-pump output; electrical/thermal storage; grid trading	Operating cost ↓; balance/capacity violations ↓	Electric/thermal balance; storage limits; generator limits; grid limits; no simultaneous C/D	PRQ2–PRQ4	Multi-energy community EMS; improved TD3 with CNN-attention scheduling and one PV exogenous source; storage scheduling; MPPT absent
[49]	A3	C2	D5	E3	Microgrid observations; PV/BESS/WECS/generator/grid/load states	PV MPPT control; BESS charge/discharge; DER PI tuning	Voltage stability ↑; power quality ↑; DER power optimization ↑	Voltage bounds; irradiance/wind/load variation; grid-tied/islanded modes; fault scenarios	PRQ2, PRQ3	Centralized RL microgrid controller; PV MPPT, BESS dispatch, and DER controller tuning, and direct RL-MPPT and BESS control; unified converter-level coordination
[48]	A3	C2	D5	E3	PV power/voltage changes; BESS power; SoC; load power; voltage error	PV PWM duty cycle; BESS converter duty cycle; inverter modulation index	PV power ↑; SoC change reward; voltage error ↓	Physical converter dynamics; reward-shaped feasibility; no formal safety projection	PRQ1–PRQ4	Centralized PPO converter-level controller; joint MPPT, BESS charge/discharge, and voltage regulation, and direct PV duty-cycle MPPT; direct BESS converter control; unified PV–battery–load control
[73]	A1	C1	D6	E1	Hour/day; PV generation; price; ESS SoC	ESS charge/discharge ratio; schedulable-load ratios; appliance on/off	Electricity cost ↓; SoC/load violations ↓	SoC bounds; no simultaneous C/D; TSL/PSL satisfaction; home power balance	PRQ2–PRQ4	Residential HEMS; multi-step TD3 with PER; ESS and appliance scheduling and PV available/exogenous; no PV actuation; MPPT absent
[51]	A2	C1	D5	E2	Battery SoC; UC SoC; EPDM output; PV generation; load; price; stored-energy price	Battery power reference; grid-converter power reference	Power balance ↑; unsafe SoC alarms ↓; trading profit ↑	Battery/UC SoC bounds; converter tracking; UC buffering; soft unsafe-region penalties	PRQ2–PRQ4	PV–battery–grid EMS with ultracapacitor buffer; DDPG real-time power references; EPDM pricing and PV under external MPPT; battery/grid dispatch only; no RL-MPPT

Table A1. Cont.

Paper	Architecture	Battery Actuation	Method	Layer	RL State	RL Action	RL Reward	RL Constraints	PRQ Coverage	Objective
[88]	A1	C1	D6	E2	Local SoC; local PV; total demand; neighbor SoCs; average-SoC estimate	ESU output power	SoC variance ↓; degradation cost ↓	Global power balance; ESU power bounds; SoC bounds; communication limits; action adjustment	PRQ2–PRQ4	Decentralized MARL for DC microgrid DESS; SoC balancing and demand support, and PV exogenous; distributed battery allocation; MPPT absent
[30]	A1	C1	D5	E1	PV power; load; SoC; real-time price	Battery charge/discharge power	Grid transaction cost ↓; safe battery action ↑; SoC violations ↓	SoC 20–95%; battery power limits; price interaction; soft unsafe-action penalties	PRQ2–PRQ4	ZEH PV–battery EMS; Q-learning/DQN/DDPG comparison; self-consumption and cost focus, and PV available/exogenous; battery scheduling only; MPPT absent
[75]	A1	C1	D6	E1	PV generation; load; previous DG output; ESS SoC; price	DG output; ESS charge/discharge	Operating cost ↓; grid cost ↓; imbalance ↓	Power balance; DG limits/commitment/ramping; ESS limits; grid limits	PRQ2–PRQ4	Offline-to-online microgrid EMS; GAIL initialization plus SAC refinement and PV available/exogenous; DG-ESS scheduling; MPPT absent
[66]	A1	C1	D4	E1	SoC; PV array outputs; load; real-time price	Battery charge/discharge; diesel-generator output	Operation cost ↓; degradation/fuel/grid/emission cost ↓	SoC bounds; battery/DG/grid limits; AC power flow; line/voltage limits; power balance	PRQ2–PRQ4	MINLP-guided Q-learning microgrid EMS; battery–diesel–grid dispatch with network constraints and PV available/exogenous; dispatch only; MPPT absent
[57]	A1	C1	D5	E1	SoC; PV forecast; load forecast; time encodings; price context	Continuous battery charge/discharge power	Self-sufficiency ↑; cost ↓; degradation ↓; violations/bad logic ↓	SoC bounds; battery power bounds; no simultaneous C/D; FEC degradation; grid logic	PRQ2–PRQ4	REC battery dispatch; SAC vs Lagrange-SAC; constraint-aware self-consumption management and PV community generation; battery dispatch only; MPPT absent
[92]	A1	C1	D5	E1	Time features; load/PV forecasts; tariffs; previous load/PV/tariffs; battery energy	Continuous battery charge/discharge power	Cost ↓; PV-surplus charging ↑; deficit discharging ↑; capacity violations ↓	Energy balance; SoC bounds; battery power limits; no simultaneous C/D	PRQ2–PRQ4	Microgrid EMS; PPO/A2C/SAC/TD3 vs MILP; battery trading schedule and PV forecast/exogenous; battery scheduling only; MPPT absent
[81]	A2	C1	D5	E2	ToU tariff; hazard warning; PV output; base load; EV load; ESS SoC	ESS charge/discharge; dynamic EV service fee	Operator benefit ↑; emergency resilience ↑	Normal/emergency balance; PV limit; ESS/EV SoC and power limits; travel expectation; grid capacity; supply-duration constraint	PRQ1–PRQ4	PV–storage–charging EMS; TD3 normal/emergency operation; warning-based pre-storage and PV available/exogenous; resilience-oriented ESS dispatch; MPPT absent

Table A1. Cont.

Paper	Architecture	Battery Actuation	Method	Layer	RL State	RL Action	RL Reward	RL Constraints	PRQ Coverage	Objective
[41]	A2	C1	D5	E1	PV generation; non-dispatchable load; SoC; grid price; EV charging price	Battery power; grid exchange; dispatchable EV charging load	Operating benefit ↑; balance violations ↓; SoC violations ↓	Power/energy balance; PV limits; storage power limits; charging-load and pile limits	PRQ2–PRQ4	PV–storage–charging station EMS; DDPG battery–grid–load scheduling; peak–valley reduction and PV available/exogenous; battery and EV-load dispatch; MPPT absent
[112]	A2	C1	D5	E1	Market price; load; PV generation; SoC; seasonal context	Dispatch adjustment; DoD modulation; capacity expansion; objective-weight tuning	Cost ↓; peak power ↓; curtailment ↓; degradation ↓	SoC bounds; charge/discharge limits; energy balance; hard checks and soft penalties	PRQ1–PRQ3	Industrial PV-ESS planning and EMS; adaptive multi-objective RL; sizing plus operation and PV available/exogenous; planning/dispatch focus; MPPT absent

Table A2. Assessment of non-RL articles, with objective and actuation notes.

Paper	Architecture	MPPT Treatment	PV Actuation	Battery Actuation	Method	Layer	Objective	PV Actuation Details
[113]	A2	M3	B2	C2	D3	E3	Integrated standalone PV–battery converter control; PV boost converters use SSA-PSO MPPT; PV switches to mode-based MPPT/DC-bus coordination; layered PV and battery loops	DC-bus regulation when battery unavailable; active but mode–logic-coordinated
[114]	A2	M1	B2	C2	D2	E2	Islanded PV-BESS microgrid supervision; SOC/power-limit protection; fuzzy PV curtailment and auxiliary support	PV normally under local MPPT; supervisory frequency signaling curtails PV indirectly; no direct unified MPPT-battery control
[115]	A2	M2	B2	C2	D2	E3	Residential LVDC PV–battery integration; grid/islanded power management; laboratory converter validation	Dedicated P&O MPPT boost loop; separate bidirectional battery PI control; active PV control but not unified with battery law
[116]	A2	M3	B3	C1	D2	E3	Single-switch PV–battery–load converter; simultaneous MPPT and voltage regulation; converter-level partial unification	PV operating point controlled by switching frequency; duty cycle regulates output voltage; conventional P&O MPPT within integrated converter
[117]	A2	M1	B2	C2	D2	E2	Off-grid PV-BESS-EV EMS; DC-bus regulation; rule-based battery and EV coordination	PV uses separate P&O MPPT boost converter; EMS manages battery/EV demand; MPPT not jointly optimized
[118]	A2	M3	B2	C2	D3	E2	Industrial PV–battery predictive EMS; GEP-ANFIS forecasting, MPPT adjustment, and battery cycling	GEP-ANFIS MPPT block with proactive forecast-based adjustment; separate EMS for battery dispatch; active but layered PV control
[119]	A2	M3	B2	C2	D2	E3	Dual-PV battery converter; single-stage DC/AC supply; DC-link-led MPPT and battery support	Separate P&O MPPT per PV input; DC-link master loop subordinates MPPT; battery supports DC link under PV deficit
[120]	A2	M1	B1	C2	D2	E2	DC microgrid EMS; PV–wind–battery–EV coordination; droop-based storage and EV control	PV uses local Incremental-Conductance MPPT; separate from battery/EV control; no supervisory MPPT override

Table A2. Cont.

Paper	Architecture	MPPT Treatment	PV Actuation	Battery Actuation	Method	Layer	Objective	PV Actuation Details
[121]	A2	M2	B2	C2	D3	E2	Centralized DC microgrid EMS; PV/BESS/grid/load-shedding coordination under uncertainty	PV normally P&O MPPT; CEMS can switch to OFF-MPPT voltage-control mode; MPPT coordinated but not unified
[122]	A2	M1	B1	C1	D3	E2	DC microgrid strategy; MTSO MPPT plus RMPC EMS; coordinated but decoupled PV extraction and storage management	PV boost converter has dedicated MTSO MPPT layer; battery/grid/load managed by separate RMPC EMS; no joint PV–battery action
[123]	A2	M1	B1	C2	D3	E2	Grid-integrated LVDC PV-BESS EMS; SOC-aware charging, export, and islanding; battery-health-oriented coordination	PV buck converter uses local MPPT; EMS uses estimated PV power and SOC for battery/grid decisions; MPPT remains local
[124]	A2	M2	B2	C2	D3	E3	PV/BESS DC microgrid robust converter control; ADRC-based BESS current and DC-bus regulation	PV boost converter tracks P&O MPPT reference with cascaded loops; separate from BESS/DC-bus ADRC control
[45]	A3	M5	B3	C2	D3	E3	Converter-level PV–battery MPPT; NGWO duty-cycle optimization with SOC-aware update; partial-shading validation	NGWO directly updates PV boost duty cycle; battery SOC included in MPPT update; PV operating point not external
[125]	A2	M1	B1	C2	D2	E2	SPV-BESS microgrid monitoring and control; PMU-enabled fault-aware supervisory coordination	SPV uses dedicated P&O MPPT; battery/grid handled by separate converter/VSI control; MPPT local and decoupled
[126]	A2	M3	B2	C2	D2	E2	Hybrid wind–PV–battery EMS; MPPT/APPT and battery CC/CV mode coordination; layered backstepping converter control	PV boost converter supports MPPT, APPT, CC, and CV modes; EMS switches modes using frequency, SoC, and power-balance conditions
[127]	A2	M3	B2	C2	D3	E2	Hybrid DC microgrid EMS; PV–battery–hydrogen coordination; MPPT/APPT with energy-flow optimization	PV boost converter uses backstepping MPPT; APPT curtails surplus when load, battery, and electrolyzer needs are satisfied
[128]	A2	M2	B2	C2	D2	E2	PV–battery energy management architecture; sliding-mode MPPT plus battery/load current regulation	PV buck converter uses dedicated sliding-mode MPPT; battery/load interface controlled by separate buck–boost and supervisory logic
[129]	A2	M1	B1	C2	D3	E2	PV-HESS microgrid EMS; optimized PI storage control; DC-bus regulation and battery-stress reduction	PV boost converter uses local P&O MPPT; HESS and DC-bus handled by separate EMS/storage-control layer
[130]	A2	M3	B2	C2	D2	E2	Distributed PV-BESS hierarchical EMS; centralized weighted-sum optimization; constraint-aware MPP relaxation	Local PV modules use P&O MPPT; upper optimizer may force off-MPP operation under grid, converter, or SoC constraints
[131]	A3	M3	B3	C1	D1	E2	Unified PV MPPT–droop control; DC-bus regulation; load sharing under unequal/shaded conditions	Single PV controller updates duty cycle for both MPPT and droop; PV shifts between MPP extraction and bus-voltage/power-sharing support
[132]	A2	M1	B1	C2	D2	E2	Off-grid PV–battery EMS; MPPT, battery management, and DC-bus stability coordination	PV uses dedicated P&O/INC MPPT; battery BMS and PV MPPT remain separate functions under supervisory logic
[133]	A2	M2	B2	C2	D3	E3	Hybrid PV/BESS DC microgrid control; coordinated multi-loop ADRC framework; CHIL validation	PV boost loop tracks MPPT voltage reference; battery/DC-bus regulated by separate cascaded loops; active but not unified

Table A2. Cont.

Paper	Architecture	MPPT Treatment	PV Actuation	Battery Actuation	Method	Layer	Objective	PV Actuation Details
[134]	A2	M2	B2	C2	D2	E3	Islanded AC microgrid predictive control; PV/battery DG coordination; DC-link and SoC balancing	PV boost converter uses MPC-MPPT and operates at MPP; battery absorbs/supplies surplus via separate DC-link control
[28]	A3	M5	B3	C2	D2	E3	Unified DG-PV-BES converter control; shared MPPT and battery-current regulation; islanded power-quality support	Single-stage PV connection; bidirectional buck–boost converter performs both PV MPPT and battery-current regulation through shared DC-link control
[135]	A2	M2	B2	C2	D3	E3	Hybrid PV–battery microgrid control; GWO MPPT plus ANFIS-PI battery regulation; simulation-based coordination	PV boost converter uses GWO MPPT duty-cycle control; battery converter uses separate ANFIS-PI loop; active but not unified
[136]	A2	M3	B3	C2	D2	E2	Hybrid DC source management; battery-SOC-aware PV power limiting; decentralized droop coordination	PV power reference adjusted by DC-bus voltage; PV shifts between MPPT and power-limit mode based on bus/SOC conditions
[137]	A2	M1	B0	C2	D2	E2	Distributed feeder voltage control; BESS coordination under high PV penetration; event-triggered communication	PV runs local P&O MPPT; coordinated layer uses only BESS charge/discharge; PV operating point not controlled
[138]	A2	M3	B3	C2	D2	E3	Integrated three-port PV–battery converter; duty-ratio-based MPPT, DC-bus, and battery management	Single duty-ratio variable coordinates PV operating point, DC bus, and battery charge/discharge; MPPT partly integrated with storage control
[139]	A2	M3	B3	C2	D2	E3	PV–battery SRM-drive coordination; MPPT/CV switching; MPC load-side control	PV converter switches between P&O MPPT and constant-voltage mode using irradiance, SoC, DC-link, and supervisory conditions
[140]	A2	M2	B2	C2	D3	E2	Hybrid PV–wind–FC–battery EMS; GGWO-TSK MPPT, PV duty cycle set by dedicated GGWO-TSK MPPT from EMS, and inverter control stages	PV duty cycle set by dedicated GGWO-TSK MPPT from PV voltage/current; separate from battery dispatch stage
[141]	A2	M2	B2	C2	D3	E2	PV-BESS microgrid LEM; improved MPPT estimator; droop-based supervisory coordination	PV converters use autonomous MPPT and provide references to primary control; battery/LEM coordination remains separate
[142]	A2	M1	B0	C2	D2	E2	PV-HESS DC microgrid EMS; battery-supercapacitor power sharing; DC-bus regulation and battery-stress reduction	PV boost converter extracts MPP locally; main control acts through HESS power sharing; MPPT not co-optimized
[143]	A2	M2	B2	C2	D2	E2	Dual-battery PV system supervision; drift-avoidance MPPT plus centralized power handling; DC-link regulation	PV TSIBC duty cycle set by dedicated drift-avoidance MPPT; battery protection handled by separate power unit
[44]	A2	M5	B3	C2	D3	E2	Predictive PV–battery–load EMS; forecast-driven PV/battery references; optimized load shedding and DISMC tracking	EMS computes PV MPP voltage/current/power references and battery references together; separate DISMC loops track them
[144]	A2	M3	B2	C2	D3	E2	PV-HESS microgrid control; RBFNN MPPT plus NARX storage EMS; DC-bus regulation and battery protection	PV quadratic boost converter uses RBFNN MPPT; storage EMS may reduce PV generation, but PV and HESS loops remain separate

Table A2. Cont.

Paper	Architecture	MPPT Treatment	PV Actuation	Battery Actuation	Method	Layer	Objective	PV Actuation Details
[145]	A2	M3	B3	C2	D2	E2	PV–battery–diesel power management; SOC-triggered off-MPPT–droop operation; standalone AC microgrid co-ordination	PV boost converter switches from MPPT to droop/off-MPPT when battery full and excess power cannot be absorbed
[146]	A2	M2	B2	C2	D4	E3	PV–battery microgrid physical control; PSO-ANFIS MPPT with inverter P-Q/droop support	PV boost converter uses PSO-ANFIS MPPT duty ratio; battery and inverter controls remain separate classical loops
[46]	A3	M5	B3	C2	D3	E3	Unified PV–battery sliding-mode control; dual converter duty ratios; evolutionary coefficient tuning	One SMC framework generates PV and battery duty ratios; PV MPP condition embedded in sliding surface with battery/load regulation
[147]	A2	M2	B2	C2	D2	E2	Islanded LVDC microgrid-power-flow control; MPPT/droop PV switching; SOC-aware battery management	PV boost converter normally uses INC MPPT; switches to droop mode under high bus voltage and high battery SoC
[148]	A2	M1	B1	C2	D2	E2	Decentralized PV-BESS droop optimization; transmission-loss minimization; BESS-supported inverter dispatch	PV treated as non-dispatchable under MPPT; control acts through BESS inverter droop, not PV operating point
[149]	A2	M3	B3	C2	D2	E3	Standalone PV–battery supervisory control; virtual battery emulation; battery-discharge reduction and cycle-life extension	PV finite-state control switches among charging, MPPT, and VBE; VBE drives PV off-MPP to match load and protect battery
[150]	A2	M1	B1	C2	D2	E2	PV–battery–hydrogen DC microgrid EMS; TAB multiport storage coordination; rule-based energy management	PV uses separate boost-converter MPPT on common DC bus; battery/hydrogen coordination occurs in separate EMS layer
[151]	A2	M3	B3	C2	D2	E2	Standalone PV–battery efficiency optimization; bus-voltage reference tuning; SOC-aware coordinated control	PV uses conventional MPPT, but battery-side hierarchical control shifts DC-bus/PV operating point for efficiency and SOC objectives
[152]	A2	M3	B3	C1	D2	E3	Unified PV MPC; integrated MPPT and DC-bus regulation; multi-PV source coordination	Modified MPC directly sets PV boost duty cycle and current reference; switches between MPPT and voltage-regulation objectives

References

1. Mirbagheri, S.Z.; Aldeen, M.; Saha, S. A comparative study of MPPT algorithms for standalone PV systems under RCIC. In *Proceedings of the 2015 IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Brisbane, QLD, Australia, 15–18 November 2015*; IEEE: Piscataway, NJ, USA, 2015; pp. 1–5. [\[CrossRef\]](#)
2. Zakki, M.I.M.; Hussain, M.N.M.; Seroji, M.N. A hybrid fast distributed maximum power point tracking technique in rapid mismatching condition. In *Proceedings of the 2017 IEEE 2nd International Conference on Automatic Control and Intelligent Systems (I2CACIS), Kota Kinabalu, Malaysia, 21 October 2017*; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6. [\[CrossRef\]](#)
3. Rimouche, W.; Talha, A.; Boumaaraf, H. Comparative Analysis of Hardware Resource Utilization for MPPT Techniques: P&O, Incremental Conductance, and ANN Using Processor-in-the-Loop Testing on STM32F407. In *Proceedings of the 2024 12th International Conference on Systems and Control (ICSC), Batna, Algeria, 3–5 November 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 53–59. [\[CrossRef\]](#)
4. Mumtaz, W.; Sheina, A.; Hasan, H.; Geumpana, T.A.; Roslidar, R. Incremental Conductance MPPT Optimization System Modeling using PID Control with Fuzzy Gain Scheduling. In *Proceedings of the 2023 IEEE Fifth International Conference on DC Microgrids (ICDCM), Auckland, New Zealand, 15–17 November 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6. [\[CrossRef\]](#)
5. Chaudhary, S.; Singh, A. Analysis of AI Techniques for Maximum Power Point Tracking Under Partial Shading Conditions. In *Proceedings of the 2020 IEEE 17th India Council International Conference (INDICON), New Delhi, India, 10–13 December 2020*; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6. [\[CrossRef\]](#)
6. Mwine, I.N.; Han, H.; Liu, H.; Wang, H.; Zheng, X.; Fei, L. A Novel Neural Network Based MPPT Method under Partial Shading Conditions. In *Proceedings of the 2023 Panda Forum on Power and Energy (PandaFPE), Chengdu, China, 27–30 April 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 1057–1062. [\[CrossRef\]](#)
7. Sefriti, B.; Boumhidi, I. Neural network Incremental conductance MPPT algorithm for photovoltaic water pumping system. In *Proceedings of the 2015 10th International Conference on Intelligent Systems: Theories and Applications (SITA), Rabat, Morocco, 20–21 October 2015*; IEEE: Piscataway, NJ, USA, 2015; pp. 1–6. [\[CrossRef\]](#)
8. Kofinas, P.; Doltsinis, S.; Dounis, A.I.; Vouras, G.A. A reinforcement learning approach for MPPT control method of photovoltaic sources. *Renew. Energy* **2017**, *108*, 461–473. [\[CrossRef\]](#)
9. Lin, D.; Li, X.; Ding, S. Maximum power point tracking based on reinforcement learning in photovoltaic system. In *Proceedings of the IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES), Jaipur, India, 16–19 December 2020*; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6. [\[CrossRef\]](#)
10. Chou, K.Y.; Yang, S.T.; Yang, C.S.; Chen, Y.P. Maximum Power Point Tracking of Photovoltaic System Based on Reinforcement Learning. In *Proceedings of the 2019 IEEE International Conference on Consumer Electronics-Taiwan (ICCE-TW), Yilan, Taiwan, 20–22 May 2019*; IEEE: Piscataway, NJ, USA, 2019; pp. 1–2. [\[CrossRef\]](#)
11. Phan, B.C.; Lai, Y.; Lin, C.E. A Deep Reinforcement Learning-Based MPPT Control for PV Systems under Partial Shading Condition. *Sensors* **2020**, *20*, 3039. [\[CrossRef\]](#) [\[PubMed\]](#)
12. Giraldo, L.F.; Gaviria, J.F.; Torres, M.I.; Alonso, C.; Bressan, M. Deep reinforcement learning using deep-Q-network for Global Maximum Power Point tracking: Design and experiments in real photovoltaic systems. *Heliyon* **2024**, *10*, e37974. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Chen, Y.; Lin, D.; Fei, X.; Li, X.; Wang, W.; Ding, S. Research on Q-Table Design for Maximum Power Point Tracking-Based Reinforcement Learning in PV Systems. *Energies* **2023**, *16*, 7286. [\[CrossRef\]](#)
14. Artetxe, E.; Uralde, J.; Barambones, O.; Calvo, I.; Martin, I. Maximum Power Point Tracker Controller for Solar Photovoltaic Based on Reinforcement Learning Agent with a Digital Twin. *Mathematics* **2023**, *11*, 2166. [\[CrossRef\]](#)
15. Enderica, C.A.S.; Llata, J.; Torre-Ferrero, C. Guided Reinforcement Learning with Twin Delayed Deep Deterministic Policy Gradient for a Rotary Flexible-Link System. *Robotics* **2025**, *14*, 76. [\[CrossRef\]](#)
16. Chevtchenko, S.F.; Barbosa, E.J.; Cavalcanti, M.C.; Azevedo, G.M.; Ludermit, T.B. Combining PPO and incremental conductance for MPPT under dynamic shading and temperature. *Appl. Soft Comput.* **2022**, *131*, 109748. [\[CrossRef\]](#)
17. Nwachukwu, S.E.; Folly, K.A.; Awodele, K.O. Soft Actor-Critic-Based MPPT Control of Solar PV Systems Under Partial Shading Conditions. *IEEE Open Access J. Power Energy* **2025**, *12*, 194–208. [\[CrossRef\]](#)
18. Yaqobi, M.A.; Matayoshi, H.; Danish, M.S.S.; Urasaki, N.; Howlader, A.M.; Senjyu, T. Control and Energy Management Strategy of Standalone DC Microgrid Cluster using PV and Battery Storage for Rural Application. *Int. J. Power Energy Res.* **2018**, *2*, 53–68. [\[CrossRef\]](#)
19. Hong, J.; Yin, J.; Liu, Y.; Peng, J.; Jiang, H. Energy Management and Control Strategy of Photovoltaic/Battery Hybrid Distributed Power Generation Systems With an Integrated Three-Port Power Converter. *IEEE Access* **2019**, *7*, 82838–82847. [\[CrossRef\]](#)
20. Wu, X.; Tang, Z.; Stroe, D.I.; Kerekes, T. Overview and Comparative Study of Energy Management Strategies for Residential PV Systems with Battery Storage. *Batteries* **2022**, *8*, 279. [\[CrossRef\]](#)
21. Li, X.; Ma, T.; Zhang, Z.; Zhang, D.; Xü, Y.; Wang, K. A Black Start Recovery Strategy for a PV-Based Energy Storage Microgrid, Considering the State of Charge of Energy Storage. *Electronics* **2025**, *14*, 1696. [\[CrossRef\]](#)

22. Naseri, N.; Hani, S.E.; Machmoum, M.; Elbouchikhi, E.; Daghour, A. Energy Management Strategy for a Net Zero Emission Islanded Photovoltaic Microgrid-Based Green Hydrogen System. *Energies* **2024**, *17*, 2111. [\[CrossRef\]](#)
23. Gillot, V.; Khalil, J.; Kreiss, J.; Mardaneh, M.; Jamshidpour, E. Energy management and control method for a three port non-isolated DC-DC converter for PV applications. In *Proceeding of the 2024 International Conference on Materials and Energy: Energy in Electrical Engineering (ICOME-EE), Bangkok, Thailand, 30 October–1 November 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–4. [\[CrossRef\]](#)
24. Voigt, N.; Wawer, T.; Albert, T. Optimizing prosumers' battery energy storage management using machine learning. In *Proceedings of the 2019 16th International Conference on the European Energy Market (EEM), Ljubljana, Slovenia, 18–20 September 2019*; IEEE: Piscataway, NJ, USA, 2019; pp. 1–6. [\[CrossRef\]](#)
25. Mahamat, C.; Ilunga, G.; Bechet, J.; Zermani, S.; Linguet, L. Artificial Neural Network in Photovoltaic-Battery Microgrid System for Controlling the Battery Current in French Guiana: Battery Life Improvement. In *Proceedings of the 2022 13th International Renewable Energy Congress (IREC), Hammamet, Tunisia, 13–15 December 2022*; IEEE: Piscataway, NJ, USA, 2022; pp. 1–6. [\[CrossRef\]](#)
26. Bresnav, K.V.; Singaravelu, S. ANN-Based MPPT Control of a Standalone PV System with Bi-Directional Battery Management and NPC Inverter for Dynamic 3-Phase AC Loads. *Res. Sq.* **2025**. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Rashid, A. A grid-integrated PV-powered EV charging station featuring artificial neural network-based MPPT and comprehensive power flow control. *E3S Web Conf.* **2025**, *650*, 01015. [\[CrossRef\]](#)
28. Jha, S.; Singh, B.; Mishra, S. Performance Enhancement of PV–DG–BS Distributed Generation System in Islanded Mode. *J. Inst. Eng. India Ser. B* **2021**, *102*, 311–321. [\[CrossRef\]](#)
29. Härtel, F.; Bocklisch, T. Minimizing Energy Cost in PV Battery Storage Systems Using Reinforcement Learning. *IEEE Access* **2023**, *11*, 39855–39865. [\[CrossRef\]](#)
30. Xu, W.; Li, Y.; He, G.; Xu, Y.; Gao, W. Performance Assessment and Comparative Analysis of Photovoltaic-Battery System Scheduling in an Existing Zero-Energy House Based on Reinforcement Learning Control. *Energies* **2023**, *16*, 4844. [\[CrossRef\]](#)
31. Bavarinos, K.; Dounis, A.; Kofinas, P. Maximum Power Point Tracking Based on Reinforcement Learning Using Evolutionary Optimization Algorithms. *Energies* **2021**, *14*, 335. [\[CrossRef\]](#)
32. De Silva, W.M.P.; Krishnan, T.N.H.; Ho, P.W.C.; Sarimuthu, C.R. A Novel Meta-Learning-Based Reinforcement Controller for Voltage Regulation of an Interleaved Boost Converter in Solar Photovoltaic Systems. *IET Renew. Power Gener.* **2025**, *19*, e70156. [\[CrossRef\]](#)
33. Rostamipour, A.; Ghayeni, A.H.; Abdi, M. Optimizing Maximum Power Point Tracking with Soft Actor-Critic Reinforcement Learning and Transfer Learning Techniques. *Arab J. Sci. Eng.* **2026**, *51*, 11083–11103. [\[CrossRef\]](#)
34. Sousa, J.T.; Barbosa, R.S. Comparison of Classical and Artificial Intelligence Algorithms to the Optimization of Photovoltaic Panels Using MPPT. *Algorithms* **2025**, *18*, 493. [\[CrossRef\]](#)
35. Kwon, R.; Kwon, G. Safety Constraint-Guided Reinforcement Learning with Linear Temporal Logic. *Systems* **2023**, *11*, 535. [\[CrossRef\]](#)
36. Muratore, F.; Gienger, M.; Peters, J. Assessing Transferability From Simulation to Reality for Reinforcement Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *43*, 1172–1183. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Chen, Y.; Zeng, C.; Wang, Z.; Lu, P.; Yang, C. Zero-shot sim-to-real transfer of reinforcement learning framework for robotics manipulation with demonstration and force feedback. *Robotica* **2022**, *41*, 1015–1024. [\[CrossRef\]](#)
38. Karaki, T.; Saied, M.; Shraim, H. Deep Reinforcement Learning-Based Approach for Fault-Tolerant Control of PV Systems in Smart Grids. In *Proceedings of the 2022 10th International Conference on Systems and Control (ICSC), Marseille, France, 23–25 November 2022*; IEEE: Piscataway, NJ, USA, 2022; pp. 283–288. [\[CrossRef\]](#)
39. Kell, A.; McGough, A.S.; Forshaw, M. Optimizing a domestic battery and solar photovoltaic system with deep reinforcement learning. *arXiv* **2021**, arXiv:2109.05024. [\[CrossRef\]](#)
40. Jin, P.; Qin, J.; Ma, C.; Sha, G.; Cong, X. Comparing SAC and DQN in Microgrid EMS: Baseline Results under Fixed Electricity Price. In *Proceedings of the 2025 10th International Conference on Power and Renewable Energy (ICPRE), Hangzhou, China, 19–22 September 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 930–933. [\[CrossRef\]](#)
41. Zhou, H.; Zhang, M.; Ju, S.; Shi, T. Research on Energy Management of Photovoltaic-Storage-Charging Integrated System Based on Deep Reinforcement Learning. In *Proceedings of the 2025 37th Chinese Control and Decision Conference (CCDC), Xiamen, China, 16–19 May 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 2022–2028. [\[CrossRef\]](#)
42. Fan, J.; Wang, H. Deep Reinforcement Learning for Community Battery Scheduling Under Uncertainties of Load, PV Generation, and Energy Prices. In *Proceedings of the 2023 IEEE 7th Conference on Energy Internet and Energy System Integration (EI2), Hangzhou, China, 15–18 December 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 4871–4876. [\[CrossRef\]](#)
43. Eyimaya, S.E. A Soft Actor-Critic—Based Deep Reinforcement Learning for Nanogrid Energy Management. In *Proceedings of the 2025 14th International Conference on Renewable Energy Research and Applications (ICRERA), Vienna, Austria, 27–30 October 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 1978–1983. [\[CrossRef\]](#)

44. Rizki, H.; Aiboudi, J.; Boufounas, E.M.; Amrani, A.E.; Bejjit, L. Predictive energy management for standalone PV–battery DC microgrids using hybrid TCN–LSTM forecasting and double integral sliding mode control. *Energy Convers. Manag.* **2026**, *351*, 121063. [\[CrossRef\]](#)
45. Dagal, I.; Ibrahim, A.W.; Harrison, A. Leveraging a novel grey wolf algorithm for optimization of photovoltaic-battery energy storage system under partial shading conditions. *Comput. Electr. Eng.* **2025**, *122*, 109991. [\[CrossRef\]](#)
46. Tavoosi, J.; Danyali, S.; Shirkhani, M. Evolutionary algorithm- based sliding mode control of a photovoltaic/battery system. *Electr. Power Syst. Res.* **2025**, *246*, 111717. [\[CrossRef\]](#)
47. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ* **2021**, *372*, n71. [\[CrossRef\]](#) [\[PubMed\]](#)
48. Wiencek, R.; Ghosh, S. Centralized controller for photovoltaic mppt, bidirectional charging, and dynamic load control using deep reinforcement learning. In *Proceedings of the 2025 IEEE PES Grid Edge Technologies Conference & Exposition (Grid Edge)*; IEEE: New York, NY, USA, 2025; pp. 1–5.
49. Wiencek, R. Microgrid Control Using Reinforcement Learning Operating in 1877 Grid Tied and Islanded Modes. Master's Thesis, Tennessee State University, Nashville, TN, USA, 2024.
50. Rajamallaiyah, A.; Naresh, S.; Raghuvamsi, Y.; Manmadharao, S.; Bingi, K.; R, A.; Guerrero, J.M. Deep Reinforcement Learning for Power Converter Control: A Comprehensive Review of Applications and Challenges. *IEEE Open J. Power Electron.* **2025**, *6*, 1769–1802. [\[CrossRef\]](#)
51. Wu, J.; Li, S.; Yu, Y.; Guan, Y.; Wei, B.; Chen, Z. Deep reinforcement learning-based multi-objective energy management system for microgrids under flexible energy market. *Appl. Energy* **2026**, *406*, 127223. [\[CrossRef\]](#)
52. Irfan, M.; Tahir, T.; Deilami, S.; Huang, S.; Veettil, B.P. A Novel Data-Driven Optimization for Cost-Effective Home Energy Management with PV-EV Integration. In *Proceedings of the 2024 IEEE 34th Australasian Universities Power Engineering Conference (AUPEC), Sydney, Australia, 20–22 November 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6. [\[CrossRef\]](#)
53. Kappos, I.; Martinez-Hermida, S.; Barja-Martinez, S.; Menos-Aikateriniadis, C.; Aragüés-Peñalba, M. Balancing Cost and Self-Consumption in Home Energy Management: A Deep Reinforcement Learning and Model-Based Perspective with EV and Storage Systems. In *Proceedings of the 2025 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT Europe), Valletta, Malta, 20–23 October 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–5. [\[CrossRef\]](#)
54. Leva, S.; Giacomelli, G.; Magri, C.; Piccoli, G.; Pierini, M.; Lazzaroni, P.; Cuccovillo, A.; Tosarello, M.; Rosso, M.; Lupo, V. Control Strategies for Home Energy Management Systems. In *Proceedings of the 2024 IEEE International Conference on Artificial Intelligence & Green Energy (ICAIGE), Yasmine Hammamet, Tunisia, 10–12 October 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6. [\[CrossRef\]](#)
55. Wang, D.; Li, L.; Ma, X.; Dai, M.; Kong, S. Energy Management Method for Building Photovoltaic-Energy Storage-Direct Current-Flexible Charging Microgrid Based on Improved TD3. In *Proceedings of the 2025 4th International Conference on Power System and Energy Technology (ICPSET), Chengdu, China, 18–20 July 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 137–141. [\[CrossRef\]](#)
56. Ebi, D.; Fouché, E.; Heyden, M.; Böhm, K. MicroPPO: Safe Power Flow Management in Decentralized Micro-Grids with Proximal Policy Optimization. In *Proceedings of the 2024 IEEE 11th International Conference on Data Science and Advanced Analytics (DSAA), San Diego, CA, USA, 6–10 October 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–10. [\[CrossRef\]](#)
57. Zendejdel, D.; De Santis, E.; Capillo, A.; Odonkor, P.; Rizzi, A. Multi-Objective Battery Dispatching using an Enhanced SAC Algorithm. In *Proceedings of the 2025 International Joint Conference on Neural Networks (IJCNN), Rome, Italy, 30 June–5 July 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–8. [\[CrossRef\]](#)
58. Bauer, P.; Kapfinger, J.; Konrad, M.; Leopold, T.; Wilker, S.; Sauter, T. DER Control: Harnessing DDPG and MILP for Enhanced Performance in Active Energy Management. In *Proceedings of the 2024 International Workshop on Intelligent Systems (IWIS), Ulsan, Republic of Korea, 28–30 August 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6. [\[CrossRef\]](#)
59. Go, S.I.; Choi, J.H. Design and Dynamic Modelling of PV-Battery Hybrid Systems for Custom Electromagnetic Transient Simulation. *Electronics* **2020**, *9*, 1651. [\[CrossRef\]](#)
60. Meng, Q.; Hussain, S.; Luo, F.; Wang, Z.; Jin, X. An Online Reinforcement Learning-Based Energy Management Strategy for Microgrids With Centralized Control. *IEEE Trans. Ind. Appl.* **2024**, *61*, 1501–1510. [\[CrossRef\]](#)
61. Baberwal, K.; Yadav, A.K.; Saini, V.K.; Lamba, R.; Kumar, R. Data Driven Energy Management of Residential PV-Battery System Using Q-Learning. In *Proceedings of the 2023 IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE), Kerala, India, 8–11 November 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6. [\[CrossRef\]](#)
62. Alpha, M.M.; Kumar, H.; P S, L.P. Optimal Scheduling Reinforcement Algorithm for Residential Grid Tied PV Microgrid. In *Proceedings of the 2025 International Conference on Power, Instrumentation, Control, and Computing (PICC), Thrissur, India, 9–11 October 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–6. [\[CrossRef\]](#)
63. Matore, T.N.; Folly, K.A. Energy Management System of an Electric Vehicle Charging Station Using Q-Learning and Artificial Intelligence. In *Proceedings of the 2023 IEEE PES GTD International Conference and Exposition (GTD), Istanbul, Turkiye, 22–25 May 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 246–250. [\[CrossRef\]](#)

64. Arwa, E.O.; Folly, K.A. Improved Q-learning for Energy Management in a Grid-tied PV Microgrid. *SAIEE Afr. Res. J.* **2021**, *112*, 77–88. [\[CrossRef\]](#)
65. Hossain, M.S.; Mahmood, H. Intelligent Energy Management of a Microgrid Using Reinforcement Learning. In *Proceedings of the 2024 IEEE Power & Energy Society General Meeting (PESGM), Seattle, WA, USA, 21–25 July 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–5. [\[CrossRef\]](#)
66. Yoldas, Y.; Goren, S.; Onen, A. Optimal Control of Microgrids with Multi-stage Mixed-integer Nonlinear Programming Guided Q-learning Algorithm. *J. Mod. Power Syst. Clean Energy* **2020**, *8*, 1151–1159. [\[CrossRef\]](#)
67. Lv, Y.; Wu, Z.; Zhao, X. Data-Based Optimal Microgrid Management for Energy Trading with Integral Q-Learning Scheme. *IEEE Internet Things J.* **2023**, *10*, 16183–16193. [\[CrossRef\]](#)
68. Vy, H.T.; Anh, H.P.H. Optimal Renewable Energy Management System Using Hybrid Evolutionary Neural Q-Learning Technique. *J. Mod. Power Syst. Clean Energy* **2021**, *36*, 956–969. [\[CrossRef\]](#)
69. Qiu, X.; Nguyen, T.A.; Crow, M.L. Heterogeneous Energy Storage Optimization for Microgrids. *IEEE Trans. Smart Grid* **2016**, *7*, 1453–1461. [\[CrossRef\]](#)
70. Kaewdornhan, N.; Chatthaworn, R. Predictive Energy Management for Microgrid Using Multi-Agent Deep Deterministic Policy Gradient With Random Sampling. *IEEE Access* **2024**, *12*, 95071–95090. [\[CrossRef\]](#)
71. Hu, C.; Cai, Z.; Zhang, Y.; Yan, R.; Cai, Y.; Cen, B. A soft actor-critic deep reinforcement learning method for multi-timescale coordinated operation of microgrids. *Prot. Control Mod. Power Syst.* **2022**, *7*, 29. [\[CrossRef\]](#)
72. Guiducci, L.; Palma, G.; Stentati, M.; Rizzo, A.; Paoletti, S. A Reinforcement Learning approach to the management of Renewable Energy Communities. In *Proceedings of the 2023 12th Mediterranean Conference on Embedded Computing (MECO), Budva, Montenegro, 6–10 June 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–8. [\[CrossRef\]](#)
73. Wu, X.; Zhang, Y.; Chen, L.; Luo, X.; Jiang, H.; Ou, J. Multi-step Prioritized Deep Reinforcement Learning Method for Home Energy Management. In *Proceedings of the 2024 IEEE International Conference on Smart City (SmartCity), Wuhan, China, 13–15 December 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 9–14. [\[CrossRef\]](#)
74. Park, J.; Kwon, T.; Kim, B.; Hwang, Y.; Sim, M.K. Multi-Agent Reinforcement Learning Based Optimal PV-ESS Control In Grid. In *Proceedings of the 2023 IEEE IAS Global Conference on Renewable Energy and Hydrogen Technologies (GlobConHT), Male, Maldives, 11–12 March 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6. [\[CrossRef\]](#)
75. Yang, A.; Lin, Z.; Lin, K.; Li, L. Optimal Energy Scheduling of a Microgrid based on Offline-to-Online Deep Reinforcement Learning. In *Proceedings of the 2024 6th International Conference on Energy Systems and Electrical Power (ICESEP), Wuhan, China, 21–23 June 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 1088–1092. [\[CrossRef\]](#)
76. Wen, M.; Wang, Y.; Chen, J.; Zhu, R. Integrated Multi-Energy Scheduling Strategy for Smart Community based on Twin Delayed Deep Deterministic Policy Gradient. In *Proceedings of the 2023 International Conference on Smart Electrical Grid and Renewable Energy (SEGRE), Changsha, China, 16–19 June 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 311–317. [\[CrossRef\]](#)
77. Luo, R.; Gao, W.; Zhao, X.; Zhang, K.; Chen, X.; Guan, Y.; Liu, S.; Liu, J. A Two-Stage Green Energy Dispatch Scheme for Microgrid Using Deep Reinforcement Learning. *IEEE Trans. Green Commun. Netw.* **2025**, *9*, 2279–2291. [\[CrossRef\]](#)
78. Nicola, M.; Nicola, C.I. Improved Control of DC-DC Converter for PV Array and Battery Hybrid System Based on Sine Cosine Algorithm and RL-TD3 Agent. In *Proceedings of the 2023 3rd International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), Tenerife, Canary Islands, Spain, 19–21 July 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6. [\[CrossRef\]](#)
79. Schlachter, H.; GeißEndörfer, S.; Behrends, H.; Maydell, K.V.; Agert, C. Voltage-Based Load Recognition and Its Integration Into an Application Use Case. *IEEE Access* **2025**, *13*, 134423–134440. [\[CrossRef\]](#)
80. Nwachukwu, S.E.; Folly, K.A.; Awodele, K.O. A Comparative Study Between Soft Actor-Critic (SAC) and Deep Deterministic Policy Gradient (DDPG) Algorithms for Solar PV MPPT Control Under Partial Shading Conditions. *IEEE Access* **2025**, *13*, 71738–71754. [\[CrossRef\]](#)
81. Zhang, J.; Lu, W.; Dou, S.; Chen, L. Emergency Accident Handling Strategy for Photovoltaic-Storage-Charging Systems Based on Deep Reinforcement Learning. In *Proceedings of the 2025 IEEE International Conference on Smart Power and Energy Technologies (SPET), Wuhan, China, 19–20 September 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–8. [\[CrossRef\]](#)
82. Huang, B.; Wang, J. Deep-Reinforcement-Learning-Based Capacity Scheduling for PV-Battery Storage System. *IEEE Trans. Smart Grid* **2021**, *12*, 2272–2283. [\[CrossRef\]](#)
83. Liu, J.; Zhu, J.; Guan, Q.; Luo, Y.; Tang, X. Hierarchical Energy Management and Charging Scheduling in the PV-CS-EV Integrated System. *IEEE Internet Things J.* **2025**, *12*, 24417–24430. [\[CrossRef\]](#)
84. Si, C.; Tao, Y.; Qiu, J.; Lai, S.; Zhao, J. Deep reinforcement learning based home energy management system with devices operational dependencies. *Int. J. Mach. Learn. Cybern.* **2021**, *12*, 1687–1703. [\[CrossRef\]](#)
85. Huang, C.; Zhang, H.; Wang, L.; Luo, X.; Song, Y. Mixed Deep Reinforcement Learning Considering Discrete-continuous Hybrid Action Space for Smart Home Energy Management. *J. Mod. Power Syst. Clean Energy* **2022**, *10*, 743–754. [\[CrossRef\]](#)

86. Jiang, X.; Xu, Q. Safe Deep Reinforcement Learning Based Energy Management for Educational Buildings With Guaranteed Constraints. In *Proceedings of the IECON 2025—51st Annual Conference of the IEEE Industrial Electronics Society, Madrid, Spain, 14–17 October 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–6. [\[CrossRef\]](#)
87. Butt, H.S.; Huang, Q.; Schäfer, B. Explainable Reinforcement Learning for Optimizing Electricity Costs in Building Energy Management. In *Proceedings of the 2024 3rd International Conference on Energy Transition in the Mediterranean Area (SyNERGY MED), Limassol, Cyprus, 21–23 October 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6. [\[CrossRef\]](#)
88. Xiong, Z.; Luo, B.; Wang, B.C.; Xu, X.; Liu, X.; Huang, T. Decentralized Multiagent Reinforcement Learning Based State-of-Charge Balancing Strategy for Distributed Energy Storage System. *IEEE Trans. Ind. Inform.* **2024**, *20*, 12450–12460. [\[CrossRef\]](#)
89. Keerthisinghe, C.; Chapman, A.C.; Verbič, G. Energy Management of PV-Storage Systems: Policy Approximations Using Machine Learning. *IEEE Trans. Ind. Inform.* **2019**, *15*, 257–265. [\[CrossRef\]](#)
90. Mbuwir, B.V.; Paridari, K.; Spiessens, F.; Nordström, L.; Deconinck, G. Transfer learning for operational planning of batteries in commercial buildings. In *Proceedings of the 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), New Delhi, India, 10–13 December 2020*; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6. [\[CrossRef\]](#)
91. Cambambi, C.A.C.; Lautert, R.R.; Milani, I.P.; Rangel, C.A.S.; Canha, L.N.; Wolter, M. Real-Time Power Dispatch in Grid-Connected Microgrids using Deep Q-Learning. In *Proceedings of the 2024 59th International Universities Power Engineering Conference (UPEC), Cardiff, UK, 2–6 September 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–6. [\[CrossRef\]](#)
92. Zhadan, A.; Wu, H.; Kudin, P.S.; Zhang, Y.; Petrosian, O. Microgrid control for renewable energy sources based on deep reinforcement learning and numerical optimization approaches. *Vestn. St. Petersburg Univ. Appl. Math. Comput. Sci. Control Processes* **2023**, *19*, 391–402. [\[CrossRef\]](#)
93. Cai, W.; Sawant, S.; Reinhardt, D.; Rastegarpour, S.; Gros, S. A Learning-Based Model Predictive Control Strategy for Home Energy Management Systems. *IEEE Access* **2023**, *11*, 145264–145280. [\[CrossRef\]](#)
94. Ye, Y.; Wang, H.; Chen, P.; Tang, Y.; Strbac, G. Safe Deep Reinforcement Learning for Microgrid Energy Management in Distribution Networks with Leveraged Spatial–Temporal Perception. *IEEE Trans. Smart Grid* **2023**, *14*, 3759–3775. [\[CrossRef\]](#)
95. Alagusundari, N.; Arthi, D.; Kumaran, E.; Dikesh, K. AI-based Predictive and Optimal Control for Microgrid Energy Management: Design and Implementation. In *Proceedings of the 2025 International Conference on Modern Sustainable Systems (CMSS), Shah Alam, Malaysia, 12–14 August 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 222–228. [\[CrossRef\]](#)
96. Esfahani, F.N.; Suri, N.; Ma, X. Machine Learning Forecasting and GAN-Based Scenario Control for EV Charging and PV Integration. In *Proceedings of the IECON 2025—51st Annual Conference of the IEEE Industrial Electronics Society, Madrid, Spain, 14–17 October 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–6. [\[CrossRef\]](#)
97. Shehu, M.A.; Talapiden, K.; Chau, T.T.; Haruna, A.; Aly, M.; Gali, V.; Do, T.D.; Alhassan, A.B. Control and energy management of standalone microgrids in remote areas: A review of recent advances, challenges, and opportunities for future research. *Eng. Sci. Technol. Int. J.* **2023**, *75*, 102288. [\[CrossRef\]](#)
98. Shi, J.; Xue, W.; Li, K. Reinforcement Learning-Driven Control Strategies for DC Flexible Microgrids: Challenges and Future. *Energies* **2026**, *19*, 648. [\[CrossRef\]](#)
99. Talha, M.; Asghar, F.; Kim, S.H. A Neural Network-Based Robust Online SOC and SOH Estimation for Sealed Lead–Acid Batteries in Renewable Systems. *Arab. J. Sci. Eng.* **2019**, *44*, 1869–1881. [\[CrossRef\]](#)
100. Swibki, T.; Salem, I.B.; Kraiem, Y.; Amraoui, L.E.; Abbes, D. Reinforcement Learning-Based Energy Management Algorithms Effect on Microgrid Physical Properties. In *Proceedings of the 2023 IEEE 11th International Conference on Systems and Control (ICSC), Sousse, Tunisia, 18–20 December 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 366–371. [\[CrossRef\]](#)
101. Elshamy, A.; Osama, A.; Elsir, M.; Muhaidat, S.; Al Hosani, K.; El Moursi, M.; Sumaiti, A.A. Improving Isolated Microgrid Inverter Lifetime with Reinforcement Learning-based Energy Management. In *Proceedings of the 2024 6th International Conference on Smart Power & Internet Energy Systems (SPIES), Abu Dhabi, United Arab Emirates, 4–6 December 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 100–105. [\[CrossRef\]](#)
102. Fatima; Shimizu, R.; Cheah, W.; Banno, R.; Shao, X. Reward-Shaped PPO for Economic and Degradation-Aware Battery Control in Microgrids. In *Proceedings of the 2025 IEEE Conference on Cloud and Big Data Computing (CBDDCom), Hakodate, Japan, 21–24 October 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–8. [\[CrossRef\]](#)
103. Gao, F.; Kang, R.; Cao, J.; Yang, T. Primary and Secondary Control in DC Microgrids: A Review. *J. Mod. Power Syst. Clean Energy* **2019**, *7*, 227–242. [\[CrossRef\]](#)
104. Li, H.; Zhang, Y.; Ju, Y.; Li, K.; Zhang, B.; Vazquez, S. A Multi-Timescale Stability Analysis Framework for Isolated DC Microgrids. *Chin. J. Electr. Eng.* **2025**, *11*, 216–227. [\[CrossRef\]](#)
105. Watari, D.; Taniguchi, I.; Goverde, H.; Manganiello, P.; Shirazi, E.; Catthoor, F.; Onoye, T. Multi-Time Scale Energy Management Framework for Smart PV Systems Mixing Fast and Slow Dynamics. *Appl. Energy* **2021**, *289*, 116671. [\[CrossRef\]](#)
106. Coraci, D.; Brandi, S.; Hong, T.; Capozzoli, A. An innovative heterogeneous transfer learning framework to enhance the scalability of deep reinforcement learning controllers in buildings with integrated energy systems. *Build. Simul.* **2024**, *17*, 739–770. [\[CrossRef\]](#)

107. Gao, Y.; Li, S.; Xiao, Y.; Dong, W.; Fairbank, M.; Lu, B. An Iterative Optimization and Learning-Based IoT System for Energy Management of Connected Buildings. *IEEE Internet Things J.* **2022**, *9*, 21246–21259. [\[CrossRef\]](#)
108. Hosseini, E.; García-Triviño, P.; Sarrias-Mena, R.; Vázquez, C.G.; Fernández-Ramírez, L.M. Reinforcement Learning based Energy Management System for Grid-Connected PV plants and Energy-Stored Quasi-Z-source Cascaded H-Bridge Multilevel Inverter. In *Proceedings of the 2023 IEEE International Conference on Environment and Electrical Engineering and 2023 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), Madrid, Spain, 6–9 June 2023*; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6. [\[CrossRef\]](#)
109. Lei, L.; Tan, Y.; Dahlenburg, G.; Xiang, W.; Zheng, K. Dynamic Energy Dispatch Based on Deep Reinforcement Learning in IoT-Driven Smart Isolated Microgrids. *IEEE Internet Things J.* **2021**, *8*, 7938–7953. [\[CrossRef\]](#)
110. Perera, M.K.; Hemapala, K.T.M.U.; Wijayapala, W.D.A.S. Developing a Reinforcement Learning model for energy management of microgrids in Python. In *Proceedings of the 2021 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), Dubai, United Arab Emirates, 17–18 March 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 68–73. [\[CrossRef\]](#)
111. Razghandi, M.; Zhou, H.; Erol-Kantarci, M.; Turgut, D. Smart Home Energy Management: Sequence-to-Sequence Load Forecasting and Q-Learning. In *Proceedings of the 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 7–11 December 2021*; IEEE: Piscataway, NJ, USA, 2021; pp. 1–6. [\[CrossRef\]](#)
112. Zournatzis, A.C.; Nicolopoulou, E.P.; Kontargyri, V.T.; Christodoulou, C.A. An Adaptive Multi-Objective Decision System With Reinforcement Learning for PV-ESS Sizing and Operation for Grid-Connected Industrial PV-ESS Systems in Distribution Networks. *IEEE Access* **2025**, *13*, 148547–148558. [\[CrossRef\]](#)
113. AL-Wesabi, I.; Zhijian, F.; Farh, H.M.H.; Dagal, I.; A. Al-Shamma'a, A.; Al-Shaalan, A.M.; kai, Y. Hybrid SSA-PSO based intelligent direct sliding-mode control for extracting maximum photovoltaic output power and regulating the DC-bus voltage. *Int. J. Hydrogen Energy* **2024**, *51*, 348–370. [\[CrossRef\]](#)
114. Al Badwawi, R.; Issa, W.R.; Mallick, T.K.; Abusara, M. Supervisory Control for Power Management of an Islanded AC Microgrid Using a Frequency Signalling-Based Fuzzy Logic Controller. *IEEE Trans. Sustain. Energy* **2025**, *10*, 94–104. [\[CrossRef\]](#)
115. Ali, Z.; Raisz, D. Development and laboratory validation of an LVDC back-to-back system with residential inverter, PV, and battery-based power generation. *Energy Convers. Manag. X* **2025**, *26*, 100997. [\[CrossRef\]](#)
116. Alli, S.S.; Jovanović, S.; Poure, P.; Jamshidpour, E. MPPT and output voltage control of Photovoltaic systems using a Single-Switch DC-DC converter. In *Proceedings of the 2016 IEEE International Energy Conference (ENERGYCON), Leuven, Belgium, 4–8 April 2016*; IEEE: Piscataway, NJ, USA, 2016; pp. 1–6. [\[CrossRef\]](#)
117. Awad, M.M.; Bhattacharjee, A.; Hasan, K.N. Design and Energy Management of Solar PV-BESS Integrated Off-Grid Electric Vehicle Charging Station. In *Proceedings of the 2025 12th National Power Electronics Conference (NPEC), Kozhikode, India, 14–16 December 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–6. [\[CrossRef\]](#)
118. Bakare, M.S.; Abdulkarim, A.; Shuaibu, A.N.; Muhamad, M.M. Predictive energy control for grid-connected industrial PV-battery systems using GEP-ANFIS. *e-Prime* **2024**, *9*, 100647. [\[CrossRef\]](#)
119. Barman, S.R.; Ahmed, M.; Ahmad, A.; Sarmah, N. Efficient power conversion for renewable energy: Hybrid multi-output system with solar photovoltaic inputs and battery backup. *Renew. Energy Focus* **2026**, *57*, 100811. [\[CrossRef\]](#)
120. Bhargavi, K.M.; Jayalakshmi, N.S. A New Control Strategy for Plug-in Electric Vehicle of DC Microgrid with PV and Wind Power Integration. *J. Electr. Eng. Technol.* **2019**, *14*, 13–25. [\[CrossRef\]](#)
121. Bhattar, C.L.; Chaudhari, M.A. Centralized Energy Management Scheme for Grid Connected DC Microgrid. *IEEE Syst. J.* **2025**, *17*, 3741–3751. [\[CrossRef\]](#)
122. Buchibabu, P.; Somlal, J. Green energy management in DC microgrids enhanced with robust model predictive control and muddled tuna swarm MPPT. *Electr. Eng.* **2024**, *106*, 2799–2819. [\[CrossRef\]](#)
123. Chaudhary, N.; Verma, K.S.; Pandey, J.P. Energy management scheme in microgrid with LVDC using smart energy estimation technique. *Electr. Eng.* **2025**, *107*, 4143–4154. [\[CrossRef\]](#)
124. Chebabhi, A.; Belkhier, Y.; Ghadbane, H.E.; Barkat, S.; Fellah, M.K.; Rezk, H.; Rekioua, T.; Benbouzid, M. Optimized generalized super twisting observer based active disturbance rejection control method for PV/Battery energy storage system within DC microgrids under uncertainty and disturbances. *Energy Convers. Manag. X* **2025**, *28*, 101362. [\[CrossRef\]](#)
125. Dhinu Lal, M.; Varadarajan, R. Synchrophasor Enabled Energy Monitoring, Management and Control of Solar PV-Battery Interfaced Microgrid. In *Proceedings of the 2024 Second International Conference on Smart Technologies for Power and Renewable Energy (SPECon), Ernakulam, India, 2–4 April 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 1–8. [\[CrossRef\]](#)
126. Aadouli, N.E.; Magri, A.E.; Lajouad, R.; Myasse, I.E.; Mansouri, A.; Mezdi, K.E.; Kumar, P. Adaptive nonlinear control and energy management for grid-connected hybrid wind–PV systems using Vienna converters. *Energy Convers. Manag. X* **2026**, *30*, 101790. [\[CrossRef\]](#)
127. el Mezdi, K.; El Magri, A.; El Myasse, I.; Giri, F.; Kumar, P. Energy flow optimization for a hybrid -microgrid integrating hydrogen production via a electrolyzer and fuel cell backup. *Int. J. Electr. Power Energy Syst.* **2026**, *175*, 111568. [\[CrossRef\]](#)

128. Fekik, A.; Hamida, M.L.; Azar, A.T.; Ghanes, M.; Hakim, A.; Denoun, H.; Hameed, I.A. Robust power control for PV and battery systems: Integrating sliding mode MPPT with dual buck converters. *Front. Energy Res.* **2024**, *12*, 1380387. [\[CrossRef\]](#)
129. Firah, A.; Birane, M.; Degla, A.; Hadroug, N. Energy Management and Improved Metaheuristic Optimization-Based Control of Photovoltaic/Hybrid Energy Storage System-Based Microgrid. *Arab. J. Sci. Eng.* **2025**, *50*, 17617–17635. [\[CrossRef\]](#)
130. Franzese, P.; Ribera, M.; Cervone, A.; Iannuzzi, D. Optimized control strategy for single-phase multilevel cascaded converter in a distributed PV-BESS system. *Electr. Power Syst. Res.* **2023**, *214*, 108818. [\[CrossRef\]](#)
131. Golla, M.; Sudarsan Reddy, D.V.; Thangavel, S. A unified single-controller approach of an MPPT and droop control for solar PVs in DC microgrid under unequal and partial shading conditions. *Renew. Energy* **2026**, *256*, 123940. [\[CrossRef\]](#)
132. Hossain, M.I.; Tanvir Islam Mim, M.; Banik, T.; Jany, M.R.; Mukter Hossain Emon, M.; Shufian, A. Decentralized Solar PV/Battery-based Energy Management for Isolated Areas. In *Proceedings of the 2024 IEEE 3rd International Conference on Robotics, Automation, Artificial-Intelligence and Internet-of-Things (RAAICON)*, Dhaka, Bangladesh, 29–30 November 2024; IEEE: Piscataway, NJ, USA, 2024; pp. 188–193. [\[CrossRef\]](#)
133. Ihammouchen, S.; Chebabhi, A.; Ziane, D.; Ghadbane, H.E.; Rekioua, T.; Rezk, H.; Benkhoris, M.F. An improved active disturbance rejection control for voltage stabilization and performance enhancement in hybrid photovoltaic battery based DC microgrids via red-tailed hawk algorithm. *J. Energy Storage* **2026**, *152*, 120743. [\[CrossRef\]](#)
134. Jayachandran, M.; Ravi, G. Predictive power management strategy for PV/battery hybrid unit based islanded AC microgrid. *Int. J. Electr. Power Energy Syst.* **2019**, *110*, 487–496. [\[CrossRef\]](#)
135. Lachheb, A.; Akoubi, N.; Salem, J.B. Advanced control strategy for energy optimization in microgrids using intelligent algorithms: Gray wolf optimizer and ANFIS. *Int. J. Dyn. Control* **2026**, *14*, 84. [\[CrossRef\]](#)
136. Li, X.; Wu, X. Autonomous energy management strategy for a hybrid power system of more-electric aircraft based on composite droop schemes. *Int. J. Electr. Power Energy Syst.* **2021**, *129*, 106828. [\[CrossRef\]](#)
137. Mahmud, N.; Zahedi, A.; Rahman, M.S. An event-triggered distributed coordinated voltage control strategy for large grid-tied PV system with battery energy storage. In *Proceedings of the 2017 Australasian Universities Power Engineering Conference (AUPEC)*, Melbourne, VIC, Australia, 19–22 November 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 1–6. [\[CrossRef\]](#)
138. Mounica, P.; Sivadnam, N. Battery Management for Integrating Renewable Energy Systems to the DC Micro-grid. In *Proceedings of the 2022 IEEE 7th International Conference for Convergence in Technology (I2CT)*, Mumbai, India, 7–9 April 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–6. [\[CrossRef\]](#)
139. Mukhopadhyay, S.; Mansani, S.; Sekaran, S. Model predictive control-based PV-battery-fed SRM drive with reduced torque ripple. *Electr. Eng.* **2026**, *108*, 9. [\[CrossRef\]](#)
140. Ozcan, O.F.; Kilic, H.; Ozguven, O.F. A unified robust hybrid optimized Takagi–Sugeno fuzzy control for hydrogen fuel cell-integrated microgrids. *Int. J. Hydrogen Energy* **2025**, *143*, 1279–1294. [\[CrossRef\]](#)
141. Priyadarshini, L.; Dash, P.K.; Dhar, S. A new Exponentially Expanded Robust Random Vector Functional Link Network based MPPT model for Local Energy Management of PV-Battery Energy Storage Integrated Microgrid. *Eng. Appl. Artif. Intell.* **2020**, *91*, 103633. [\[CrossRef\]](#)
142. Punna, S.; Manthathi, U.B. Optimum design and analysis of a dynamic energy management scheme for HESS in renewable power generation applications. *SN Appl. Sci.* **2020**, *2*, 495. [\[CrossRef\]](#)
143. Purwar, A.; Mal, R.; Pandit, S.; Ray, S. Supervisory control strategy for dual battery assisted solar water pumping using a 3-stage interleaved boost converter. *J. Energy Storage* **2024**, *101*, 113929. [\[CrossRef\]](#)
144. Sarra, Z.; Bouziane, M.; Bouddou, R.; Benboughenni, H.; Mekhilef, S.; Elbarbary, Z.M.S. Intelligent control of hybrid energy storage system using NARX-RBF neural network techniques for microgrid energy management. *Energy Rep.* **2024**, *12*, 5445–5461. [\[CrossRef\]](#)
145. Sharma, R.K.; Mudaliyar, S.; Mishra, S. Power management and economic load dispatch based control of hybrid PV-battery-diesel standalone AC system. In *Proceedings of the 2018 IEEMA Engineer Infinite Conference (eTechNxT)*, New Delhi, India, 13–14 March 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–6. [\[CrossRef\]](#)
146. Siddaraj, S.; Yarangatti, U.R.; Harischandrapa, N. Coordinated PSO-ANFIS-Based 2 MPPT Control of Microgrid with Solar Photovoltaic and Battery Energy Storage System. *J. Sens. Actuator Netw.* **2023**, *12*, 45. [\[CrossRef\]](#)
147. Varghese, A.; Chandran, L.R.; Rajendran, A. Power flow control of solar PV based islanded low voltage DC microgrid with battery management system. In *Proceedings of the 2016 IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, Delhi, India, 4–6 July 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 1–6. [\[CrossRef\]](#)
148. Vazquez, N.; Yu, S.S.; Chau, T.K.; Fernando, T.; Iu, H.H.C. A Fully Decentralized Adaptive Droop Optimization Strategy for Power Loss Minimization in Microgrids with PV-BESS. *IEEE Trans. Energy Convers.* **2019**, *34*, 385–395. [\[CrossRef\]](#)
149. Venkatramanan, D.; John, V. Supervisory Control Architecture for Standalone Solar Photo- Voltaic Power Generation System. In *Proceedings of the 2018 IEEE Energy Conversion Congress and Exposition (ECCE)*, Portland, OR, USA, 23–27 September 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 6688–6695. [\[CrossRef\]](#)

150. Wang, Q.; Gunawardane, K.; Li, L. Triple-Active-Bridge Power Conversion in Hydrogen-Based DC Microgrids: Topology Design and Coordinated Control. In *Proceedings of the 2025 IEEE International Conference on Energy Technologies for Future Grids (ETFG), Wollongong, NSW, Australia, 7–11 December 2025*; IEEE: Piscataway, NJ, USA, 2025; pp. 1–6. [[CrossRef](#)]
151. Yang, Y.; Qin, Y.; Tan, S.C.; Hui, S.Y.R. Efficient Improvement of Photovoltaic-Battery Systems in Standalone DC Microgrids Using a Local Hierarchical Control for the Battery System. *IEEE Trans. Power Electron.* **2019**, *34*, 10796–10807. [[CrossRef](#)]
152. Yang, P.; Peng, Y.; Xia, Y.; Wei, W.; Yu, M.; Feng, Q. A unified bus voltage regulation and MPPT control for multiple PV sources based on modified MPC in the DC microgrid. *Front. Energy Res.* **2022**, *10*, 10425. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.