energies

MDPI

Review

Reinforcement Learning for Efficient Power Systems Planning: A Review of Operational and Expansion Strategies

Gabriel Pesántez [1,2], Wilian Guamán [1,2,*], José Córdova [1], Miguel Torres [1] and Pablo Benalcazar [3,*]

[1] Faculty of Electrical and Computer Engineering, Escuela Superior Politécnica del Litoral, ESPOL, Campus Gustavo Galindo, Km. 30.5 vía Perimetral, Guayaquil 090902, Ecuador; ganapesa@espol.edu.ec (G.P.); jecordov@espol.edu.ec (J.C.); mitorres@espol.edu.ec (M.T.)
[2] Electrical Engineering Program, Faculty of Engineering and Applied Sciences, Universidad Técnica de Cotopaxi, Campus La Matriz, Latacunga 050108, Ecuador
[3] Division of Energy Economics, Mineral and Energy Economy Research Institute of the Polish Academy of Sciences, ul. J. Wybickiego 7A, 31-261 Kraków, Poland
* Correspondence: wpguaman@espol.edu.ec (W.G.); benalcazar@min-pan.krakow.pl (P.B.)

Abstract: The efficient planning of electric power systems is essential to meet both the current and future energy demands. In this context, reinforcement learning (RL) has emerged as a promising tool for control problems modeled as Markov decision processes (MDPs). Recently, its application has been extended to the planning and operation of power systems. This study provides a systematic review of advances in the application of RL and deep reinforcement learning (DRL) in this field. The problems are classified into two main categories: Operation planning including optimal power flow (OPF), economic dispatch (ED), and unit commitment (UC) and expansion planning, focusing on transmission network expansion planning (TNEP) and distribution network expansion planning (DNEP). The theoretical foundations of RL and DRL are explored, followed by a detailed analysis of their implementation in each planning area. This includes the identification of learning algorithms, function approximators, action policies, agent types, performance metrics, reward functions, and pertinent case studies. Our review reveals that RL and DRL algorithms outperform conventional methods, especially in terms of efficiency in computational time. These results highlight the transformative potential of RL and DRL in addressing complex challenges within power systems.

Keywords: reinforcement learning; optimal power flow; economic dispatch; expansion planning

Citation: Pesántez, G.; Guamán, W.; Córdova, J.; Torres, M.; Benalcazar, P. Reinforcement Learning for Efficient Power Systems Planning: A Review of Operational and Expansion Strategies. *Energies* **2024**, *17*, 2167. https://doi.org/10.3390/en17092167

Received: 24 March 2024
Revised: 21 April 2024
Accepted: 28 April 2024
Published: 1 May 2024

1. Introduction

Power system planning entails the evaluation of short-, medium-, and long-term perspectives, each encompassing distinct characteristics and facing specific challenges. In long-term planning, strategic decisions are often based on general system models aimed at addressing problems such as the expansion of the transmission network and investment in new generation units. In contrast, short- and medium-term planning relies on more detailed models with smaller search spaces [1,2]. Moreover, to minimize an objective function such as the total investment costs or energy production costs, a comprehensive planning approach must incorporate technical, economic, environmental, and social considerations [3]. Thus, the planning of a power grid is frequently structured as a complex optimization problem characterized by a wide range of constraints and variables [4].

The complexity of power system planning increases when considering uncertainty in demand or costs, the integration of flexible alternating current transmission systems (FACTS), and the inclusion of renewable energy sources along with energy storage systems (ESS) [5]. The stochastic nature of electricity production from renewable energy sources and the new challenges that arise in energy systems require advanced planning strategies to ensure reliability and efficiency. These factors significantly amplify the computational effort needed to solve such optimization problems [6]. Moreover, the widespread adoption

of smart meters has transformed electric grids into sources of large volumes of correlated data [7]. According to recent studies, these data could be leveraged using machine learning (ML) to improve the planning and operation of electric power systems. The application of ML, for example, could allow for the extraction of valuable insights for planning or system control from the data analysis [8].

Within the field of ML, reinforcement learning (RL) and deep reinforcement learning (DRL) have shown exceptional performance in solving various control problems in electric power systems, often modeled as Markov decision processes (MDPs) [7,9]. Recent studies such as [10] provide a comprehensive overview of the use of RL to solve decision and control problems in electric grids considering energy storage or electric vehicles. Similarly, [11] gathered some of the most important applications of multi-agent RL algorithms for energy management in microgrids connected to the grid and offers guidelines for the use of transfer learning to improve RL outcomes in complex energy system environments. On the other hand, ref.[12] classified the applications of RL in electric power systems into building energy management systems (BEMS), dispatch, vehicle energy systems, energy devices, grid, and energy markets. The analysis revealed that Q-learning predominates in research addressing energy dispatch problems.

Previous research has examined the diverse applications of DRL and RL in tackling issues within electric power systems. For example, ref. [13] offered an overview of the challenges and opportunities associated with employing DRL approaches in electric power distribution systems. The reviewed applications included active network management, energy management systems, retail electricity market, and demand response. The review by [6] presented models, algorithms, and DRL techniques utilized across various applications in electric power systems, categorizing them into four groups: energy management, demand response, electricity market, and operational control. This categorization also identified the learning algorithm, type of agent, and Q-function estimator for each application. Moreover, ref. [14] identified deep learning methodologies for supervised, unsupervised, and semi-supervised applications in power systems.

Unlike traditional optimization techniques, which often face challenges with the nonlinear and nonconvex nature of power systems [15], RL and DRL offer significant advantages. These methods excel at dynamically adapting to changing scenarios and optimizing decisions over a long-term horizon. This adaptability is particularly crucial in power system planning, where decisions must anticipate future uncertainties and align with long-term sustainability goals. For instance, in transmission network expansion planning (TNEP) and generation expansion planning (GEP), DRL algorithms can evaluate numerous expansion alternatives under various future scenarios, optimizing both the system cost and reliability.

The papers discussed above reveal that RL and DRL have primarily been utilized for load forecasting and controlling variables within power systems. However, the scope of their application has expanded significantly to address a broader spectrum of planning issues. These include innovative solutions for OPF, ED, UC, GEP, TNEP, and DNEP. Unlike other reviews, this work uniquely focused on systematically identifying and analyzing the specific characteristics of these application areas. This focused approach allowed us to not only characterize the current state-of-the-art, but to also highlight gaps and suggest future research directions specific to power system planning. In this context, this review paper offers a comprehensive overview of the principal applications of RL and DRL in the planning and operation of electric power systems, making two significant contributions to the existing literature:

- Presents a detailed analysis of the most relevant publications on the use of RL and DRL in power system operation and expansion planning. The analysis is conducted using a systematic literature review methodology.
- Identifies learning algorithms, function approximators, and reward functions used in the application of RL and DRL in power system operation and expansion planning. It

also highlights relevant case studies to provide a comprehensive perspective on how these technologies are reshaping power system planning and operation.

The remainder of this paper is organized as follows. Section 2 reviews the theoretical foundations of RL and DRL as well as the metrics that evaluate the performance of these algorithms. Section 3 describes the research methodology used for the systematic literature review. Section 4 presents the applications of RL and DRL in the operation and expansion planning of electric power systems. Section 5 offers a detailed discussion of the implications of these applications. Finally, Section 6 presents the conclusions and outlines directions for future research.

2. Reinforcement Learning (RL) Theoretical Background

RL is a subclassification of ML concerned with establishing how an agent takes sequential actions in an uncertain environment to maximize the cumulative reward [16]. Four main sub-elements can be identified in RL: policy, reward, value function, and environment model. The following subsections highlight the fundamentals of RL to illustrate the relationship between this algorithm and power system planning.

2.1. Reinforcement Learning and Markov Decision Process

The global RL process can be described as a Markov decision process (MDP), which is represented by an ordered sequence of elements called a tuple, $M = (S, A, P, R)$. The first element (S) consists of all possible states, the second element (A) includes the possible actions that the agent can take, the third element (P) defines the probability of moving from the current state to a new state, and the last element (R) constitutes the reward, which the agent seeks to maximize [17]. As shown in Figure 1, the agent engages in decision-making within the environment, dynamically interacting with it to execute various actions based on the environmental context and receiving the corresponding rewards. The policy consists of the agent's strategy to determine the action based on the current state [18].

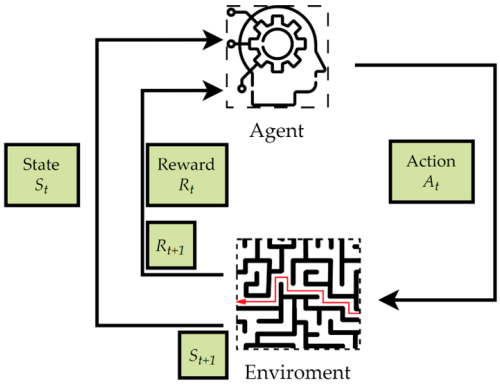

Figure 1. Interaction between the agent and the environment. Based on [17].

The interaction described occurs discretely: at each step, action $A_t \in \mathcal{A}$ receives a representation of the environment, known as the state $S_t \in \mathcal{S}$, and the received reward is determined by $R_t \in \mathcal{R} \subset \mathbb{R}$. For a finite space, $\mathcal{S}, \mathcal{A}$, and $\mathcal{R}$ constitute the sets of states, actions, and rewards, respectively [18]. Equation (1) shows the probability function $p(.|s', a) : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$, which defines the dynamics of the MDP since $p = \mathcal{S} \times \mathcal{R} \times \mathcal{S} \times \mathcal{A} \to [0, 1]$ [16].

$$p(s', r|s, a) \doteq P_r\{S_t = s',\ R_t = R|S_{t-1} = s,\ A_{t-1} = a\} \quad (1)$$

The subscript t denotes the time at which the state is found. On the other hand, p defines a probability distribution P_r, and both s and a are random variables—this probability fully characterizes the dynamics of the environment. Then, every possible value of the state S_t and reward R_t depends on the immediate previous action. During the decision-making

process, in each episode k, the agent will carry out an action and change state after obtaining a reward. Consequently, the accumulated reward is calculated using Equation (2). It is important to note that $\gamma \in [0, 1]$ is the discount factor, reducing the weight of uncertain rewards received in the future. The objective of reinforcement learning is to maximize the total reward [17].

$$R = R_t + \gamma R_{t+1} + \gamma^2 R_{t+2} + \cdots = \sum_{k=0}^{\infty} \gamma^k R_{t+k} \tag{2}$$

When the MDP is solved, a state policy, $\pi(s) \rightarrow A$, is obtained from the actions performed. This policy is considered optimal when the cumulative discounted reward is maximized [7]. This can be achieved by determining the expected return of a state or state action and using it to establish a policy. There are four functions to consider from this point of view: value function (Equation (3)), action-value function (Equation (4)), optimal value function (Equation (5)), and optimal action-value function (Equation (6)). The value function provides the expected return when it starts in a state S_t and a policy π is followed [18]. In Equations (3) and (4), the starting value is the expected reward in the state S_t, plus the value of the new state multiplied by a discount factor γ [19,20].

$$V^{\pi}(S_t) = E[R_t \mid S_t = S] = E\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid S_t = S\right] \tag{3}$$

$$Q^{\pi}(S_t, a_t) = E[R_t \mid S_t = S, A_t = A] = E\left[\sum_{k=0}^{\infty} \gamma^k r_{t+k+1} \mid S_t = S, A_t = A\right] \tag{4}$$

$$V^{*}(A) = \max_{\pi} V^{\pi}(S_t) \tag{5}$$

$$Q^{*}(S, A) = \max_{\pi} Q^{\pi}(S_t, A_t) \tag{6}$$

2.2. Classification of Reinforcement Learning Algorithms

RL algorithms can be classified based on the existence or absence of a model into two broad groups: model-free and model-based. Model-based algorithms can be further subdivided into two categories: learn the model, and given the model [21]. Meanwhile, model-free algorithms can be value-based and policy-based, as shown in Figure 2. Additionally, there are two ways to represent and train agents in model-free algorithms: policy optimization and Q-learning [18]. Methods based on Q-learning use a $Q(s, a)$ approximator to find the optimal function $Q^{*}(s, a)$. The objective function is based on Bellman equations, analogous to Equations (3) and (4), used to determine an estimate of the best policy [7].

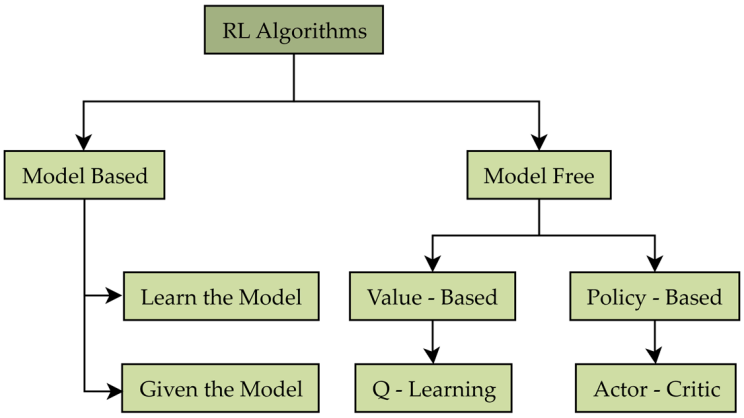

Figure 2. Classification of reinforcement learning algorithms. Based on [7].

2.3. Deep Reinforcement Learning

In RL algorithms, state spaces and approximate value functions are generally represented by tables or matrices. However, in high-dimensional problems such as power system planning, the spaces become too large to be represented in this way. Instead, they are represented using a parametrized functional form with a weight vector. The general strategy of DRL combines the perception function of deep learning with the decision-making capability of reinforcement learning. The main goal of DRL is to train an agent capable of learning an optimal policy, π^*, which maximizes the expected reward return by continuously interacting with the environment. A well-trained DRL agent does not need to rely on complete system models to make control decisions. It can respond to a variety of conditions, making it suitable for many real-time applications. As the agent begins to accumulate information about the environment, it must navigate between learning more about the environment (exploration) or following the most promising strategy with the gained experience (exploitation) [22].

3. Research Methodology

This systematic literature review analyzed existing studies to identify established connections between power system planning and the implementation of RL and DRL techniques. In this context, works that focused on the operation and planning of power grid expansion were reviewed. This review was performed by applying the methodology presented by Kitchenham et al. [23], considering three phases, as shown in Table 1.

Table 1. Systematic literature review phases.

Phases	Steps
A. Planning	Research questions Data sources Search strings Inclusion criteria Quality criteria for study selection
B. Conducting	Primary study selection Data extraction Data synthesis
C. Reporting	Documenting the extracted results

This methodology has been employed in various systematic reviews [24–26] to increase the rigor and transparency of the research process. By adopting this structured approach, this review ensured the comprehensive coverage of relevant literature and an objective evaluation of the findings.

For the development of the three phases, metrics were proposed to determine the selected works, avoiding biases generated by the authors of these papers. In addition, three study subgroups were considered for operation planning: "Optimal Power Flow", "Economic Dispatch", and "Unit Commitment", and two subgroups for expansion planning: "Transmission Network Expansion Planning" and "Distribution Network Expansion Planning".

A-Phase 1: Planning the Review

(Step 1) Research Questions

- The main interest of this review is the various works that implement RL and DRL to solve power system planning problems. In this study, the following research questions were addressed:
    - RQ1: According to the literature, what are the applications of RL and DRL in solving the OPF problem?
    - RQ2: According to the literature, what are the applications of RL and DRL in solving the ED and UC problems?

- RQ3: According to the literature, what are the applications of RL and DRL in TNEP and DNEP?

(Step 2) Data Sources

- In this study, data were collected using a search string optimized to find the most relevant literature. A total of four digital repositories were selected based on the analysis presented in [27]. The selected digital repositories were:
  - IEEE Xplore;
  - ScienceDirect;
  - Springer Link;
  - Wiley Online Library;
  - MDPI.

(Step 3) Search String

- A search string was generated based on the study questions to retrieve the relevant literature from the selected digital sources, called primary studies. The following string was used to search the digital repositories: ((“Power Systems Planning” OR “OPF” OR “Economic Dispatch” OR “Generation Expansion Planning” OR “Transmission Expansion Planning” OR “Distribution Expansion Planning” OR “Grid Planning”) AND (“Reinforcement Learning” OR “Deep Reinforcement Learning”)).

(Step 4) Inclusion Criteria

- For data inclusion, we adopted the following guidelines:
  a. Papers related to the field of Power and Energy Systems;
  b. English language;
  c. Journal papers and conference papers;
  d. Articles published between 2016 and 2024;
  e. Full text available online;
  f. Available in one or more of the selected databases;
  g. Focus on Power Systems Planning: OPF, ED, UC, TNEP, DNEP;
  h. RL or DRL mentioned in the abstract;
  i. Relationship between RL or DRL and power systems planning.

(Step 5) Quality Criteria for Study Selection

- The quality assessment in this study focused on determining the usefulness of the primary studies selected to answer the research questions posed. Simultaneously, data extraction and quality assessment of the selected publications were conducted. To ensure an objective assessment, a checklist was developed (provided in Table 2). This checklist included five quality criteria (Q1–Q5) designed to examine each primary study comprehensively.
  - Articles that fully met each criterion on the checklist were given a score of 1.00;
  - Articles that partially met a criterion received a score of 0.50;
  - Those not addressing a specific criterion on the list received a score of 0.00.

Table 2. Questions designed to establish the study quality criteria.

Questions	Checklist Questions
Q1	Does this paper address issues related to OPF, ED, or GP, whose solution is found by applying RL or DRL technique implementation methodologies?
Q2	Are the learning algorithm, function approximator, agent type, metrics to evaluate algorithm performance, and reward function clearly identified?
Q3	Are the contributions of the document to power system planning clearly stated?
Q4	Is a case study used to validate the methodology presented?
Q5	Are the limitations of the study mentioned?

B-Phase 2: Conducting the Review
(Step 1) Primary Study Selection

- The results shown in Table 3 were obtained based on the considered inclusion criteria. Two additional specific search filters were applied: the first filter considered the search string for the abstract, and the second filter exclusively for the abstract and title, thus obtaining 55 papers.

Table 3. Primary study selection results.

Data Sources	Filter 1	Filter 2
IEEE Xplore	49	23
ScienceDirect	37	15
SpringerLink	62	3
Wiley Online Library	24	7
MDPI	19	7

C-Phase 3: Reporting the Review
(Step 1) Quality Assurance of Primary Selected Studies

- Appendix A presents the complete list of articles reviewed and the score assigned to each, from which 45 papers were selected and classified into three groups: optimal power flow, economic dispatch and unit commitment, and power systems expansion planning, which includes TNEP and DNEP.

4. RL and DRL Applications in Power Systems Operation and Expansion Planning

The evolution of conventional power grids into smart grids poses new challenges for the planning and operation of energy systems. The previous sections provided an overview of the most prevalent RL and DRL algorithms in the electrical sector, followed by a general description of the methodology employed. As a result of this process, specific papers were selected for analysis. This section examines the applications of RL and DRL in solving three planning problems: OPF, economic dispatch, and network expansion planning. the relationship between these three aspects is shown in Figure 3, noting that the OPF can be applied regardless of the planning horizon, from the hourly time resolution of dispatched energy (operation planning) to long-term (expansion planning).

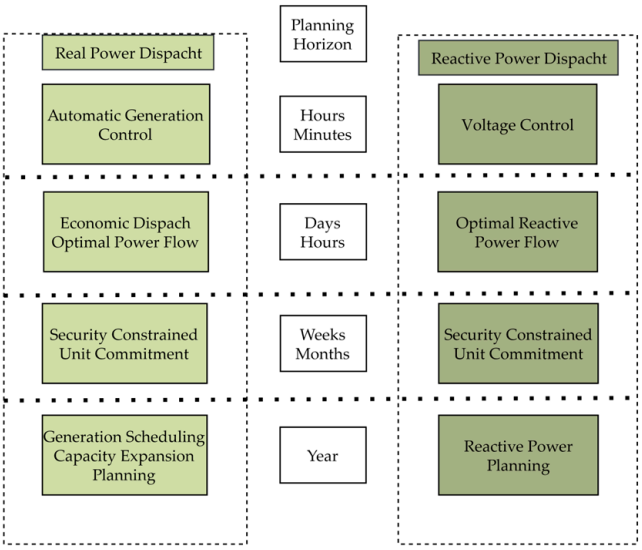

Figure 3. Planning horizon of power system operation. Based on [28].

4.1. Optimal Power Flow

Solving the alternating current (AC) optimal power flow (OPF) with operational constraints remains a significant, yet challenging optimization problem for the safe and economic operation of the electrical grid. The electrical system is modeled as a set of $\mathcal{N}$ buses connected by a set of $\mathcal{L}$ branches, with a subset $G \subseteq \mathcal{N}$ of generators at certain system buses. The cost of utilizing each generator is a function, typically quadratic, that depends on the actual power generated: $C_i\left(P_i^g\right)$. The objective function of the OPF is to minimize the total generation cost; however, the optimization could also target transmission line losses, voltage deviations, total energy transfer capacity, voltage stability, system security, etc. [29,30]. Equation (7) defines the objective function of the problem. Equations (8) and (9) express the active and reactive power balance constraints. Meanwhile, Equations (10)–(13) represent the constraints for the limits of active and reactive power generated, voltage limits, and maximum angular difference, respectively.

$$min \sum_{i \in G} \left(c_{2i} P_i^{g\,2} + c_{1i} P_i^g + c_{0i} \right) \qquad \forall_i \in \mathcal{N} \tag{7}$$

Subject to:

$$P_i(V,\theta) = P_i^{\mathrm{g}} - P_i^{\mathrm{d}} \qquad \forall_i \in \mathcal{N} \tag{8}$$

$$Q_i(V,\theta) = Q_i^{\mathrm{g}} - Q_i^{\mathrm{d}} \qquad \forall_i \in \mathcal{N} \tag{9}$$

$$P_i^{g,min} \le P_i^{\mathrm{g}} \le P_i^{g,max} \qquad \forall_i \in \mathrm{G} \tag{10}$$

$$Q_i^{g,min} \le Q_i^{\mathrm{g}} \le Q_i^{g,max} \qquad \forall_i \in \mathrm{G} \tag{11}$$

$$V_i^{min} \le |V_i| \le V_i^{max} \qquad \forall_i \in \mathcal{N} \tag{12}$$

$$\theta_i^{min} \le \theta_i \le \theta_i^{max} \qquad \forall_i \in \mathcal{N} \tag{13}$$

Carpentier [31] introduced the OPF as an extension of the economic dispatch (ED) problem. His contribution lies in combining the objective function presented in Equation (7) with the power flow Equations (14) and (15) to form the optimization problem.

$$P_i(V,\theta) = |V_i| \sum_{j=1}^{n} |V_j| \left(g_{ij} cos\left(\theta_i - \theta_j\right) + b_{ij} sin\left(\theta_i - \theta_j\right) \right) \quad \forall_{\boldsymbol{i}} \in \mathcal{N} \tag{14}$$

$$Q_i(V,\theta) = |V_i| \sum_{j=1}^{n} |V_j| \left(g_{ij} sin\left(\theta_i - \theta_j\right) - b_{ij} cos\left(\theta_i - \theta_j\right) \right) \quad \forall_i \in \mathcal{N} \tag{15}$$

The vast amount of data available from electric power networks has sparked significant interest in the application of ML algorithms for control, planning, and operation applications in power systems. While there are various mathematical and heuristic approaches to solving the OPF, the use of machine learning to obtain feasible solutions is still in its early stages [32].

A notable work proposing an OPF solution with RL is presented in [33], where the authors formulate a probabilistic OPF method to manage the risk of the electricity market price. In electricity markets based on locational marginal prices (LMP), the OPF allows for the calculation of the LMPs at each bus or zone. In their work, an adaptive importance sampling (AIS) method was developed to improve the efficiency of the simulation calculation while maintaining the accuracy of the estimation. The result of the conventional Monte Carlo simulation estimation was utilized as a reference. Moreover, a case study using the IEEE 39-Bus system was conducted to compare the proposed method with the point estimate method (PEM), demonstrating the feasibility and efficiency of the method.

To assist power system operators in making decisions that ensure system security, ref. [22] presented a method to obtain fast OPF solutions with constraints using a DRL

algorithm. The proposed method employed imitation learning to generate initial weights for the neural network (NN), and a proximal policy optimization (PPO) algorithm to train and test stable and robust RL and DRL agents. The training and testing processes were carried out on the IEEE 14-bus and 200-bus Illinois systems. The results showed that the optimal costs of the proposed method were nearly identical to those computed using the interior-point solver; however, the execution time was reduced by at least seven times compared to a conventional solver.

Table 4 presents a summary of the primary studies concerning the solution of the OPF using RL and DRL. It identifies the learning algorithm, the approximation function, the algorithm's performance metrics, the reward function, and the case study.

Table 4. Summary of the literature review on RL/DRL and OPF.

Ref.	Application	Learning Algorithm	Function Approximator	Reward Function	Metrics	Test System
[34]	OPF of distribution networks	PPO algorithm with clipped surrogate loss	Value function and policy function with DNN. The actor and critic networks have three hidden layers with 200, 100, and 100 neurons, respectively.	Penalties associated with voltage restrictions, power capacity, and storage limits.	Proportion of satisfied constraints (PSC).	Modified IEEE 33-bus system trained by a 5500 dataset.
[22] [35]	AC OPF	PPO algorithm with clipped surrogate loss	Value function and policy function with DNN. The actor-critic structure: Three hidden layers with (380, 195, 100) neurons are applied in the actor NN, and three hidden layers with (380, 44, 5) are applied in the critic NN in PPO.	Negative reward (-5000) if the OPF does not converge. Also, penalties are associated with the total number of violations of active power, voltage, and line loading constraints.	Cost comparison in percentage as an MAE, feasibility rate, and running time.	IEEE 14-bus: 55,000 training dataset, 17,364 testing dataset I, 2000 testing dataset II. Illinois 200-bus systems: 60,000 training dataset, 17,364 testing dataset I, 2000 testing dataset II.
[36]	AC OPF	Modified DDPG with Lagrangian-based gradient	At the offline stage, a policy model optimizes the augmented cost and iteratively updates the parameters of a deep neural network (DNN) agent using the deep deterministic policy gradient.	Penalties are in the form of coefficients that correspond to equality and in-equality constraints.	Generation power average as MAE, generation cost, operating cost comparison of different OPF methods.	IEEE 118-bus system
[37]	OPF in a multi-objective optimization	Combination of Monte Carlo tree search and reinforcement learning MCTS-RL	Q-value: The tree state is randomly built up, and the accumulated experience in each state is updated by random sampling during the optimization and exploration policy process.	γ is a discount factor that indicates the effect of the current decision on the long-term reward.	Power transfer distribution factor (PTDF).	IEEE 33-bus test system.
[20]	Real-time OPF solution	Deep deterministic policy gradient (DDPG)	DQN: The actor is updated by following the applying the chain rule to the expected return from the start distribution concerning the actor parameters	Considers the network losses, penalty factor σ, and the quadratic number of violations, the reward is determined.	Network losses, batch average critic training cost.	IEEE 9-bus system.
[38]	Distributed optimal power flow	Inverse reinforcement learning (IRL)	The value function $Q_{(i,j)}$ represents the experience value of the agent acting is the learning rate	A general indicator is defined based on the self-fitting error to evaluate the model's accuracy.	A general indicator is defined based on the self-fitting error, which is obtained from the lower-level optimization and denoted as an optimization error.	IEEE 57-bus power system is utilized in the model, and OPF considers N $-$ 1 static security constraints.

Table 4. *Cont.*

Ref.	Application	Learning Algorithm	Function Approximator	Reward Function	Metrics	Test System
[39]	OPF	Multi-agent reinforcement learning (MARL)	The Q-value of the player is defined as a function of all players' actions	A reward function of the agent after bidding at the demand level (payoff of each generator after clearing the market).	Learning rate, the cost function.	IEEE-30-bus power system.
[40]	Distribution network planning	Deep Q-network (DQN)	Neural network trained by Q-values	Minimum network loss and voltage deviation are taken as the reward function.	Network loss distribution.	IEEE-37 bus distribution network
[41]	Distribution network	Traditional and accelerated Q learning	Deep neural network	Node voltage	Convergence time.	IEEE 33-bus system.
[42]	Optimal power flow	Twin delayed deep deterministic policy gradient (TD3)	Deep neural networks	The value is determined by calculating the following factors: (1) line current exceeding the limit, (2) consumption of renewable energy units, (3) balanced unit power exceeding the limit, (4) unit operating costs, and (5) reactive power output exceeding the limit.	Renewable energy consumption under different weights.	IEEE-30 bus networks.
[43]	Operation of distribution networks	Double deep Q network	Deep neural network	It is determined by running a power flow with input state information and selected actions.	Output power.	IEEE 33-bus networks.
[44]	Optimal power flow	Partially observable Markov game (POMG)	Q-value	The penalty function is analogous to the reward function and employs active power load and active power loss.	Daily routing and scheduling decisions.	6-bus and 33-bus power networks.

4.2. Economic Dispatch and Unit Commitment

Economic dispatch (ED) models aim to find a generation schedule that minimizes the generation costs while satisfying the power constraints of the generation units. The main difference between OPF and ED is that the former solves ED and the system power flows simultaneously, while ED ignores the system and the consequences that flows have on lines and buses [30]. In general, the ED problem is first solved without constraints on the maximum and minimum production of the generator and without transmission losses. It is then extended to include inequality constraints on the production of generation units and to account for the impact of transmission losses. In the case of thermal and hydraulic plants, this concept is extended to hydrothermal dispatch, where the time horizon is determined based on the capacity of the reservoirs. On the other hand, the unit commitment (UC) problem can be extended to higher time granularity such as hours within a day. The solution procedures often incorporate the economic dispatch problem as a subproblem. That is, for each of the subsets of the total number of units connected to the load, the subset must be operated in an economically optimal manner [4].

Optimization-based methods such as heuristic, dynamic programming, and mixed-integer quadratic programming (MIQP) typically yield effective solutions for the UC problem. However, the computational time of optimization-based methods increases exponentially with the number of generating units, which poses a significant bottleneck in practice. To address this issue, [45] proposed a reinforcement learning (RL) algorithm that approximates the value-action function with neural networks to determine the feasible action space. Numerical studies conducted on a five-generator test case demonstrated that the proposed algorithm achieved a similar level of performance to MIQP-based optimization in terms of optimality.

In order to apply an RL algorithm, the state space is defined by $S_t = (P_t^g, P_t^d)$, $S_t \epsilon S$, while the actions that define the scheduling of the generation units are considered as $A_t = (P_{i,t}^G)$, $A_t \, \epsilon \, A$. The reward R_t is given by the environment as an indicator to guide the direction of the policy updates. In the example shown in Equation (16), the reward should guide the agent to minimize the operating costs while satisfying the generation power limit constraints and the power balance presented in Section 4.1. Note that σ_1 and σ_2 are used to control the trade-off between cost minimization and the penalty incurred in the case of power imbalance.

$$R_t(S_t, A_t) = -\sigma_1 \left[\sum_{i \in \mathcal{G}} C_{i,t}^G \right] - \sigma_2 \Delta P_t \qquad \forall_{\boldsymbol{i}} \in \mathcal{N} \tag{16}$$

In Equation (17), ΔP_t represents the power imbalance at the time t product of the difference between power generation and load.

$$\Delta P_t = \left| \sum_{i \in \mathcal{G}} P_{i,t}^G - \sum_{k \in \mathcal{L}} P_{k,t}^L \right| \qquad \forall_{\boldsymbol{i}} \in \mathcal{N} \tag{17}$$

Table 5 provides a summary of the primary research works concerning the application of RL and DRL in addressing economic dispatch and unit commitment problems.

Table 5. Summary of the literature review on RL/DRL and ED/UC.

Ref.	Application	Learning Algorithm	Function Approximator	Reward Function	Metrics	Study Case
[46]	Unit commitment	Q-learning-based	Adjust power output with ε-greedy.	Reflects the negative of the operation cost.	Generation cost.	New England 10-unit system.
[47]	Unit commitment and dispatch with multistage stochastic programming	Q-learning-based	DNN with state action value function to minimize operation.	Penalty ratios associated with violations of voltage and current limits, respectively.	Energy cost, Network losses cost, curtailment penalty, total cost, and CPU time.	Modified IEEE 39-bus two-region system.
[48]	Optimal dispatch	Nash-Q learning	Q-value function incorporating a Nash equilibrium.	Reward obtained by performing action a from state s to state s'.	Mean value of the objective function, variance, standard deviation, and relative standard deviation.	IEEE 39-bus two-region system.
[45]	Optimal dispatch	Multi-step deep Q-learning	DQN using stochastic gradient descent.	Penalties associated with generation operating costs.	MAE, mean-squared temporal difference error.	5-unit UC test case.

Table 5. *Cont.*

Ref.	Application	Learning Algorithm	Function Approximator	Reward Function	Metrics	Study Case
[49]	Economic dispatch model	Least square policy iteration (LSPI)	Radial basis functions (RBFs).	Two terms for each PV agent: the first reduces the amount of PV active power constrained, and the second penalizes actions that cause a voltage magnitude violation.	Total power curtailed PV, total reward, and voltage magnitude.	25-node unbalanced distribution system test.
[50]	Economic dispatch	NSGA-RL, an enhancement of the non-dominated sorting genetic algorithm II (NSGA-II)	Q-value function using NSGA.	The NSGA-RL uses an implicit reward function, rewarding efficient parameter values during its evolutionary process.	Generational distance for convergence, extent of spread achieved among the obtained solutions.	IEEE 30-bus system model.
[51]	Economic dispatch	Adam algorithm	The DQN (deep Q-network) algorithm computes the action-state value function.	It is defined by the scale constraint, upper and lower limit constraints of the generators, and the power balancing compensation, which is added together to obtain the reward.	The mean square error is used to define the error function in DQN training.	The IEEE-14 and IEEE-162 node systems are analyzed.
[52]	Economic dispatch	Multi-level backtracking prioritized experience replay-twin delayed deep deterministic policy gradient (MBEPR-TD3)	An actor neural network which maps the environment states of combined heat and power-virtual power plant.	The reward function is composed of the operation cost of virtual power plants and the penalty cost.	The metrics evaluated in the study include the increase in profits and reduction in carbon emissions due to the incorporation of power-to-gas in CHP-VPP.	Proposed 4-bus CHP-VPP system considering carbon capture and P2G technologies.
[53]	Power grid operational planning	Intelligent reschedule algorithm Q-learning based	DQN, which approximates the value function of the rescheduled action through the Q network.	It includes three aspects of rescheduling: the average node voltage fluctuation, the system fragile line load safety margin, and the generation cost.	Voltage fluctuation, the variance between the line load and the base value power generation cost index.	9-bus radial distribution feeder. 34-bus radial distribution feeders.
[54]	Economic dispatch	Novel graph-based deep reinforcement learning	GraphSAGE network.	Correlation between power loss and operating costs.	Correlation between power loss and operating costs.	IEEE 118-bus system
[55]	Economic dispatch	Proximal policy optimization (PPO)	Neural network.	Renewable energy consumption, line overload, unit operating cost penalties, penalties for power imbalances, penalties for exceeding the unit power limit, and penalties for exceeding the thermal unit power limit.	Renewable energy output.	The grid has 126 nodes, 35 thermal power units, 18 renewable energy units, 1 balancing unit, 91 loads, and 185 load lines.

Table 5. *Cont.*

Ref.	Application	Learning Algorithm	Function Approximator	Reward Function	Metrics	Study Case
[56]	Economic dispatch	A soft actor-critic	Neural network.	Minimization of carbon emissions costs and carbon dioxide emissions during dispatch operations.	Electric load curtailment.	Community-integrated energy system with electricity–gas–cooling coupling.
[57]	CHP economic dispatch	Q-learning	Q-value.	Linear sum of profit, unserved heat, maximum inlet supply temperature, minimum inlet supply temperature, minimum inlet return temperature, and maximum mass flow.	Profit.	System constructed with data obtained online.
[58]	Economic dispatch	Twin-delayed deep deterministic policy gradient (TD3)	Q-value or neural network.	Total market profit. Defined as the sum of the profits of all attacker generators. It can be employed in the reward function as an incentive for the agent.	The summed market profits, the attacker market share, and constraint violations are categorized by undervoltage, overvoltage, and branch overload.	97-bus rural MV Simbench system.
[59]	Economic dispatch	Deep deterministic policy gradient	Deep neural networks.	Negative equivalent of the microgrid operational cost.	Fuel cost and power limits of generators in microgrid.	Cimei Island power system.
[60]	Economic dispatch	Bacteria foraging reinforcement learning	Neural network.	Fuel cost.	Calculation time.	IEEE RTS-79 system.
[61]	Hydro-thermal economic dispatch	DQN and A2C	Neural network.	An aggregate level of volume water stored in the reservoir in the system.	MAPE and Pearson's correlation coefficient.	Hydro-thermal economic dispatch case study.
[62]	Economic Dispatch	Based crisscross optimization (CSO)	Neural network.	The reward function includes the cost of all units while considering the balance constraints.	Discount factor.	48 units, 96 units as well as 192 units
[63]	Economic dispatch	Deep deterministic policy gradient (DDPG)	Q-network.	Consists of two components: look-ahead economic dispatch model and total generation cost of generators.	Power generation costs.	IEEE30-bus and SG126-bus systems.
[64]	Economic dispatch	Distributed proximal policy optimization (DPPO)	Neural network.	The reward function is divided into two aspects: objective function and power deviation reward.	Total training time (s).	Real data from a region in the Liaoning Province of China to build a test system.

Table 5. *Cont.*

Ref.	Application	Learning Algorithm	Function Approximator	Reward Function	Metrics	Study Case
[65]	Economic dispatch	Distributed proximal policy optimization (DPPO)	Neural network.	The reward consists of 3 sub-targets: total operating costs, power mismatch, and storage tank status.	Economic performance.	Two different systems with four decision variables (gas turbine (GT), gas boiler (GB), power grid, and thermal storage tank (TST)) and four random variables (wind turbine, energy price, heat load, and electricity load), which was adopted to test whether our method could cope with variable operating states without recalculation.

4.3. RL and DRL Applications in Power Systems Expansion Planning

Planning models must consider the expansion of generation and transmission infrastructure, taking into account the projection of energy demand. Generally, power system planning is treated as an optimization problem, where the objective function is to minimize the future infrastructure's operating and investment costs while satisfying the model's constraints. For instance, if the model is AC, the constraints shown in Equations (8)–(13) should be satisfied. In this scenario, because of the nonlinearity in the constraints, there are two potential approaches to solving the problem: use a commercial solver that employs an interior point method (IPM) or linearize the constraints. The latter solution simplifies the problem and reduces the computational complexity by turning network planning into a linear programming problem. However, these simplifications move the solution away from the global optimum. Moreover, power systems have high-reliability requirements, and physical constraints must be handled carefully when building an RL and DRL model [7].

A notable application of RL in power system network planning was presented in [66], where an algorithm was proposed for the sizing and location of capacitors in 9-bus radial distribution feeders as well as with a 34-bus radial distribution feeder. In the study, the Q-learning algorithm was adopted as the "agent", the dimensionality of the state vectors corresponded to the number of buses available for capacitor installation, and the action vectors were the discrete values of the possible capacitors. The algorithm works as follows: The agent observes the power flow solution as the system's initial state (S) and chooses an action (A) from the predefined action vector. The process is repeated so that the agent observes the resulting state and returns a reward that expresses the degree of satisfaction of the agent with the operating limits of the restricted variables (voltages). Then, a new action is selected that leads to a new power flow solution and a new reward. The selection of new control actions is repeated until the voltage limit constraints at the radial network buses are met. The goal of the agent is to learn the optimal Q-function by mapping states to actions in such a way that the long-term reward is maximized. Thus, the agent finds the set of actions that results in the optimal policy.

Modern network expansion planning models consider the integration of renewable energies, energy storage systems, flexible AC transmission systems (FACTS), and uncertainty conditions. This makes the planning process suitable for different system configurations and scenarios, but also more complex [67,68]. Furthermore, network planning can be formulated as a single- or multi-objective problem (e.g., minimization of losses, costs, CO_2 emissions, etc.), where numerous decision variables are involved such as the real and reactive power injected by the generation units, the voltage at the generation buses, the size and location of the generators, and the investment costs, among others [69]. Multi-objective

optimization can be performed using DRL, given its capacity for high-dimensional data perception. For instance, the authors of [53] proposed a DQN-based algorithm for reprogramming the operational planning of the electrical grid by using the state feature vector (network data) as input to the Q-network and outputting the value of the reprogramming action. The trained algorithm provided the optimal discrete action strategy to achieve the planning goal and was tested on the IEEE-39 bus system, resulting in good convergence and relatively short computation times.

Table 6 provides a comprehensive summary of the primary research works concerning the application of RL and DRL on capacity expansion and transmission network expansion problems.

Table 6. Summary of the literature review on RL/DRL and expansion planning.

Ref.	Application	Learning Algorithm	Function Approximator	Reward Function	Metrics	Study Case
[70]	Plan for the deployment of shunts for power system resilience enhancement	Multi-agent based hybrid soft actor critic (HSAC) algorithm	Policy Q-function with Monte Carlo estimator	Penalties are associated with bus voltage magnitude deviation, energy not supplied, and transmission cost during contingencies.	Amount of rewards of training episodes.	IEEE 57-bus and IEEE 300-bus systems.
[9]	Power grid planning and operation under uncertainties	SAC algorithm with automated temperature coefficient calculation is adopted for training effective SAC agent	Q-function with the batch normalization technique is applied.	Contingency reward and base case reward consider the power flow through the line and the line capacity impact ratio.	Average reward and training step curves.	SGCC Zhejiang Electric Power Company study cases.
[71]	Transmission network expansion planning	Double deep-Q network with deep ResNet	The deep learning has two main branches: the deep convolutional networks and the deep confidence networks.	The reward is based on expected energy not supplied, electrical interconnection, and global cost.	Total cost, EENS, increase in load, and generator capacity.	IEEE New England 39-bus test system. IEEE RTS 24-bus test system.
[72]	Distribution network planning	Monte Carlo tree search-based reinforcement learning	Policy network function with DNN.	The reward is a function of the total investment cost and device installation investment.	Investment cost, load curtailment, and PV curtailment.	IEEE 33-bus test system. The nodes 14, 22, and 33 are equipped with ESS, gas generator, and CB.
[73]	Transmission network expansion planning	Deep Q-network (DQN)	The action's Q-value can be calculated based on the feedback of the action.	The final benchmark cost is appropriately increased on this basis, and the N-1 security constraints are considered so that the reward.	Comparison of network loss after cutting different lines.	IEEE 24-bus reliability test system is selected for calculation and analysis.
[74]	Transmission network expansion planning	Multi-agent double deep Q network (DDQN) based on deep reinforcement learning.	The value function can be calculated iteratively through dynamic programming.	The reward is considered based on meeting the upper and lower bounds of the constraints of the TNEP optimization model.	Accumulation and change rate as indicators to measure the data uncertainty.	Modified IEEE 24-bus system and New England 39-bus system.
[75]	Transmission network expansion planning	Q-learning-based with a preprocessing step	Random forest based algorithm using synthetic dataset.	A storage expansion planning framework using reinforcement learning and simulation-based optimization.	Monetary savings. The number of episodes required for convergence.	The microgrid is in Westhampton, NY.

Table 6. *Cont.*

Ref.	Application	Learning Algorithm	Function Approximator	Reward Function	Metrics	Study Case
[76]	Power grid planning and operation	Deep Q-network (MDQN)	Neural network for action value function Q.	Minimization of overall operational expenses.	Cumulative Unbalance (kW).	Virtual power plant consisting of photovoltaic (PV), energy storage, and three micro gas turbines as distributed energy resources.
[77]	Power planning for distribution network	Q-learning	Q-table.	Rewards include: construction, operation costs, and constraint function.	Voltage node.	IEEE-18 system.
[78]	Power planning for distribution network	Q-learning	Convolutional neural network (CNN).	Active power loss.	Accuracy, security, and dependability.	IEEE 33 bus radial distribution networks.
[79]	Power planning for distribution network	Dynamic distribution network reconfiguration (DDNR)	Q-table.	Active energy losses, price of the switching, penalty value,	Losses reduction	IEEE 33-bus radial system.

5. Discussion

The evaluation of an ML algorithm's performance represents a fundamental aspect of its development. In this context, metrics are commonly applied to assess the specific outcomes of interest. General metrics such as mean absolute error (MAE) and root mean square error (RMSE) are widely utilized and applicable across various problems [80,81]. Additionally, there are metrics specifically designed to evaluate RL algorithms in electric power systems. For example, energy cost, network loss cost, and curtailment penalty are standard metrics in transmission expansion planning and ED and UC problems. On the other hand, the proportion of satisfied constraints, generation power average, generation cost, and operating cost comparison are often considered in OPF problems. Furthermore, the total cost and CPU time are applicable to nearly all problems related to power system planning, primarily for comparing solutions obtained through reinforcement learning with those from conventional optimization methods. Another common metric in RL and DRL algorithms is the average reward, since the higher the reward, the better the performance of the learning algorithm.

Historically, power system planning has relied on traditional optimization methods grounded in mathematical programming and economic theories. Such methods, encompassing linear programming for economic dispatch and mixed-integer programming for generation [82] and transmission expansion planning [83], have been the foundation of the industry for decades. These strategies have provided a robust framework for addressing large-scale and long-term planning challenges under conditions of relative stability. However, the advent of renewable energy sources and the rise in distributed generation have introduced a degree of variability and uncertainty that strains the capacity of these traditional models. While RL and DRL offer notable advances in managing stochastic inputs and enhancing adaptability to real-time data, it is imperative to acknowledge the enduring relevance of core planning principles such as economics, reliability, and cost-efficiency. Thus, integrating a hybrid model that combines the predictive strengths of traditional methods with the dynamic adaptability of contemporary machine learning techniques constitutes a holistic approach to navigating the complex modern landscape of power system planning.

Regarding the OPF, most authors agree that new research should be directed toward algorithms capable of solving the multi-period AC OPF problem, considering the security constraints and high renewable energy penetration scenarios. Moreover, research works that have used ML suggest a great potential for RL to obtain solutions to the OPF. For

instance, ref. [84] proposed a spatial network decomposition for fast and scalable AC-OPF learning. Another application with enormous potential is the application of reinforcement learning approaches to solve real-time OPF, as proposed by [36,85]. Looking to the future, network expansion planning derived from OPF would benefit from a multi-agent approach. This approach would allow for the simultaneous training of different types of agents, thus improving the convergence speed in the search for feasible solutions. Integrating a wider range of devices into the optimization problem such as feeders, charging installations, and switches is also a promising avenue. Moreover, the collaborative dynamics of multi agent systems (MASs) offer unique advantages in managing the complexities of modern electric grids. By enabling multiple agents to work together in a coordinated manner, systems can dynamically adjust to operational demands in real-time, enhancing both their efficiency and resilience against failures or unexpected changes.

As the energy landscape evolves, microgeneration at the household level, often paired with energy storage systems, is progressively challenging the traditional reliance on centralized distribution networks. RL approaches in distribution networks must now address the integration of photovoltaic (PV) systems and the dynamic interactions within increasingly distributed energy resources. The rise in household microgeneration necessitates advanced communication and control technologies to coordinate a multitude of small-scale energy producers effectively. This trend underscores the necessity for MAS [11], where individual energy-producing agents operate cooperatively to maintain grid stability and reliability. These advances are particularly effective in the context of microgrids, which can operate independently and are often powered by various energy sources [59]. This capability not only ensures a continuous power supply during grid failures, but also highlights the significance of microgrids in enhancing the resilience of energy systems. The ability of microgrids to operate in island mode necessitates robust planning frameworks that can integrate these decentralized sources effectively, thereby ensuring operational flexibility and enhanced strategic planning across the energy network.

Despite the rapid expansion of renewable energy technologies, combined heat and power (CHP) production continues to play a foundational role in many national energy systems. Notably, CHP systems including combined cycle gas turbines and extensive district heating networks are integral in countries with established infrastructures [86]. These systems not only provide reliable energy output, but also help in managing the variability and intermittency associated with renewable sources. Modern computational techniques such as RL and DRL can significantly enhance the operational planning and efficiency of these systems. For example, RL techniques can optimize the operational dynamics of CHP plants by predicting and adjusting to demand fluctuations in real-time, thus reducing unnecessary energy wastage and enhancing system reliability. Furthermore, DRL can be employed to automate and improve decision-making processes regarding the dispatch of both electrical and thermal energy outputs based on the current grid conditions and forecasted demand.

6. Conclusions

This study provides a comprehensive review of the application of RL and DRL techniques in power system planning and offers an analysis of the most relevant publications concerning the use of RL and DRL in power system operation and expansion planning. Additionally, it identifies learning algorithms, function approximators, and reward functions used in the application of RL and DRL in power system operation and expansion planning. Furthermore, it highlights key case studies to provide a comprehensive perspective of how these technologies are reshaping the planning and operation of electric power systems. Considering these insights, the following conclusions can be drawn from this review:

- The use of RL and DRL in power system operation and planning is a relatively recent development. In this study, RL and DRL algorithms applied to problems such as OPF, ED, UC, and expansion planning have been examined in detail. In all of these areas,

the results indicate that RL and DRL algorithms outperform conventional methods, especially in terms of efficiency in computational time.

- The metrics used to evaluate the performance of RL and DRL algorithms in the context of electrical power systems are not uniform. Many of the studies reviewed in this paper resorted to the mean absolute error (MAE) to compare their results with solutions obtained from traditional optimization methods. In addition, the use of average reward was common, reflecting the intrinsic nature of RL problems, which seek to maximize the reward.
- The strategies and approximation functions used in DRL and RL for planning electrical systems converge on a common goal: minimizing the costs associated with generation, network operation, and the construction of new infrastructure. However, there is significant potential to extend their application to additional objectives such as minimizing CO_2 emissions and maximizing network reliability.

While the works reviewed in this paper described the learning algorithms used to solve the planning problems of operation and expansion of electrical systems, the methodology for defining the architectures of the neural networks used for the approximation function was not explained with the same level of detail. Therefore, future work should study the criteria used to define, for example, the number of hidden layers and the type of neural network, among other network architecture components. Beyond neural network configurations, there is a significant opportunity to explore how different RL and DRL approaches can be tailored to more specific power system applications. This includes enhancing the adaptability of these algorithms to real-time operations such as dynamic pricing in markets and real-time grid stability management. Furthermore, integrating advanced simulation models to predict and simulate the impact of RL and DRL in large-scale deployments will be crucial. These models can help understand the scalability of RL techniques in managing distributed energy resources and their interactions within smart grid environments. Moreover, investigating the role of RL in facilitating the transition to renewable energy sources by optimizing the placement of these resources within the grid could provide critical insights into sustainable power system planning.

Author Contributions: Conceptualization, G.P.; Methodology, G.P. and W.G; Formal analysis, G.P. and W.G.; Investigation, G.P., W.G, J.C., and P.B; Resources, J.C., M.T., and P.B.; Data curation, J.C. and P.B.; Writing—original draft preparation, G.P. and W.G.; Writing—review and editing, G.P., W.G., and P.B.; Visualization, J.C., M.T., and P.B.; Supervision, J.C. and M.T.; Project administration, P.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

Nomenclature

Abbreviations	
RL	Reinforcement learning
MDP	Markov decision process
DRL	Deep reinforcement learning
ML	Machine learning
SAC	Soft actor-critic
OPF	Optimal power flow
UC	Unit commitment
ED	Economic dispatch
TNEP	Transmission network expansion planning
PEM	Point estimate method
NN	Neural network
PPO	Proximal policy optimization
Sets, indices, and dimensions	
$\mathcal{N}$	Set of Busbar system
$\mathcal{L}$	Set of branch

M	Tuple containing the state space, agent, policy, and reward.
G	Set of generators
$\mathcal{A}$	Action Space
A_t	Action t
$\mathcal{S}$	State space
S_t	State t
$\mathcal{R}$	Rewards Space
R_t	Reward t
i, j	Subindex of bus i, and bus j
Parameters	
b_{ij}	Susceptance ij(p.u.)
g_{ij}	Conductance ij(p.u.)
Y_{ij}	Admittance ij(p.u.)
λ, ς	Lagrange Vector
c_{ki}	$k-th$ Cost coefficient for generator$i$(\$)
$C^G_{i,t}$	Operational cos ts of generator i during period t(\$)
Variables	
V_i	Voltage magnitude at bus i(p.u.)
θ_i	Voltage phase angle at bus i (radians)
V_j	Voltage magnitude at bus j(p.u.)
θ_i^{min}	Minimum voltage phase angle at bus i (radians)
θ_i^{max}	Maximum voltage phase angle at bus i (radians)
P_i	Active power at bus i(p.u.)
Q_i	Reactive power at bus i(p.u.)
P_i^g	Active power generation in the bus i(p.u.)
P_i^d	Active power demand in the bus i(p.u.)
Q_i^g	Reactive power generation in the bus i(p.u.)
Q_i^d	Reactive power demand in the bus i(p.u.)
$P_i^{g,min}$	Minimum active power generation in the bus i(p.u.)
$P_i^{g,max}$	Maximum active power generation in the bus i(p.u.)
$Q_i^{g,min}$	Minimum reactive power generation in the bus i(p.u.)
$Q_i^{g,max}$	Maximum reactive power generation in the bus i(p.u.)
V_i^{min}	Minimum voltage at the bus i(p.u.)
V_i^{max}	Maximum voltage at the bus i(p.u.)
$F_i(P_i)$	Cost function of the thermal units (\$)

Appendix A

Item	Ref.	Paper Title	Year	Data Sources	Q1	Q2	Q3	Q4	Q5	Total
1	[9]	Reinforcement learning-based solution to power grid planning and operation under uncertainties	2020	IEEE Xplore	1.0	1.0	1.0	1.0	1.0	5.0
2	[34]	Deep reinforcement learning based approach for optimal power flow of distribution networks embedded with renewable energy and storage devices	2021	IEEE Xplore	1.0	1.0	0.5	1.0	0.5	4.0
3	[22]	A data-driven method for fast AC optimal power flow solutions via deep reinforcement learning	2020	IEEE Xplore	1.0	1.0	1.0	1.0	0.0	4.0
4	[35]	Deep reinforcement learning based real-time AC optimal power flow considering uncertainties	2022	IEEE Xplore	1.0	1.0	1.0	1.0	1.0	5.0
5	[36]	Real-time optimal power flow: A Lagrangian-based deep reinforcement learning	2020	IEEE Xplore	1.0	1.0	1.0	1.0	1.0	5.0
6	[37]	Distributed optimal power flow for electric power systems with high penetration of distributed energy resources	2019	IEEE Xplore	1.0	1.0	0.5	1.0	0.5	4.0
7	[20]	A general real-time OPF algorithm using DDPG with multiple simulation platforms	2019	Wiley Online Library	1.0	0.5	1.0	1.0	0.5	4.0
8	[38]	Two-level area-load modeling for OPF of power system using reinforcement learning	2019	Wiley Online Library	1.0	1.0	1.0	1.0	0.5	4.5
9	[39]	Markov game approach for multi-agent competitive bidding strategies in the electricity market	2016	IEEE Xplore	1.0	1.0	1.0	0.5	0.0	3.5

Item	Ref.	Paper Title	Year	Data Sources	Q1	Q2	Q3	Q4	Q5	Total
10	[46]	Distributed Q-learning-based online optimization algorithm for unit commitment and dispatch in smart grid	2020	IEEE Xplore	1.0	1.0	0.50	1.0	1.0	4.5
11	[47]	Day-ahead optimal dispatch strategy for active distribution network based on improved deep reinforcement learning	2022	Science Direct	1.0	1.0	1.0	1.0	1.0	5
12	[48]	Nash-Q learning-based collaborative dispatch strategy for interconnected power systems	2020	IEEE Xplore	1.0	0.5	1.0	1.0	0.5	4.0
13	[45]	Solving unit commitment problems with multi-step deep reinforcement learning	2021	Science Direct	1.0	1.0	1.0	1.0	0.0	4.0
14	[49]	Optimal dispatch of PV inverters in unbalanced distribution systems using reinforcement learning	2022	IEEE Xplore	1.0	1.0	1.0	1.0	1.0	5.0
15	[87]	Evaluation of look-ahead economic dispatch using reinforcement learning	2022	Science Direct	1.0	0.5	0.5	1.0	0.5	3.5
16	[50]	Multi-objective optimization of the environmental-economic dispatch with reinforcement learning based on a non-dominated sorting genetic algorithm	2019	IEEE Xplore	1.0	1.0	0.5	1.0	0.5	4.0
17	[88]	Deep reinforcement learning for scenario-based robust economic dispatch strategy in Internet of energy	2021	IEEE Xplore	1.0	1.0	1.0	0.0	0.5	3.5
18	[89]	Deep reinforcement learning for economic dispatch of virtual power plant in Internet of energy	2020	Wiley Online Library	1.0	1.0	1.0	0.5	0.0	3.5
19	[51]	The distributed economic dispatch of smart grid based on deep reinforcement learning	2021	Wiley Online Library	1.0	0.5	1.0	1.0	1.0	4.5
20	[52]	Low-carbon economic dispatch of the combined heat and power-virtual power plants: An improved deep reinforcement learning-based approach	2023	Wiley Online Library	1.0	1.0	0.5	1.0	1.0	4.5
21	[90]	Hierarchical learning optimization method for the coordination dispatch of the inter-regional power grid considering the quality-of-service index	2020	Wiley Online Library	1.0	1.0	1.0	0.5	0.0	3.5
22	[39]	Markov game approach for multi-agent competitive bidding strategies in the electricity market	2016	Wiley Online Library	1.0	1.0	1.0	0.5	0.5	4.0
23	[70]	A deep reinforcement learning-based multi-agent framework to enhance power system resilience using shunt resources	2021	IEEE Xplore	1.0	1.0	1.0	1.0	0.50	4.5
24	[53]	Deep-Q-network-based intelligent reschedule for power system operational planning	2020	IEEE Xplore	1.0	1.0	0.50	1.0	0.50	4.0
25	[71]	Transmission network dynamic planning based on a double deep-Q network with deep ResNet	2021	Science Direct	1.0	1.0	1.0	1.0	0.50	4.5
26	[72]	Reinforcement learning for active distribution network planning based on Monte Carlo tree search.	2022	MDPI	1.0	1.0	1.0	1.0	0.50	4.5
27	[73]	Flexible transmission network expansion planning based on DQN algorithm	2021	MDPI	1.0	1.0	1.0	1.0	1.0	5.0
28	[74]	Transmission network expansion planning considering wind power and load uncertainties based on multi-agent DDQN	2021	IEEE Xplore	1.0	1.0	1.0	1.0	1.0	5.0
29	[75]	A storage expansion planning framework using reinforcement learning and simulation-based optimization	2021	Science Direct	1.0	1.0	1.0	0.50	0.50	4.0
30	[91]	Machine learning approaches to the unit commitment problem: Current trends, emerging challenges, and new strategies	2021	IEEE Xplore	1.0	0.5	0.5	0.5	0.0	2.5
31	[49]	Optimal dispatch of PV inverters in unbalanced distribution systems using reinforcement learning	2022	IEEE Xplore	1.0	1.0	1.0	0.50	0.0	3.5
32	[40]	Reactive power optimization of distribution network based on deep reinforcement learning and multi-agent system	2021	IEEE Xplore	1.0	0.5	1.0	1.0	1.0	4.5
33	[54]	A graph-based deep reinforcement learning framework for autonomous power dispatch on power systems with changing topologies	2022	IEEE Xplore	1.0	1.0	0.5	1.0	0.5	4
34	[92]	A new power system dispatching optimization method based on reinforcement learning	2023	IEEE Xplore	1.0	0.5	0.5	0.0	1.0	3
35	[41]	Reinforcement learning-based optimal power flow of distribution networks with high permeation of distributed PVs	2023	IEEE Xplore	1.0	0.5	1.0	1.0	1.0	4.5

Item	Ref.	Paper Title	Year	Data Sources	Q1	Q2	Q3	Q4	Q5	Total
36	[76]	Application of improved reinforcement learning technology for real time operation and scheduling optimization of virtual power plant	2023	Springer Link	1.0	0.5	1.0	1.0	1.0	4.5
37	[77]	Planning for network expansion based on prim algorithm and reinforcement learning	2023	Springer Link	1.0	1.0	1.0	1.0	0.5	4.5
38	[78]	Integrating distributed generation and advanced deep learning for efficient distribution system management and fault detection	2024	MDPI	1.0	1.0	1.0	1.0	0.5	4.5
39	[79]	Solving dynamic distribution network reconfiguration using deep reinforcement learning	2021	MDPI	1.0	1.0	1.0	1.0	0.5	4.5
40	[60]	Bacteria foraging reinforcement learning for risk-based economic dispatch via knowledge transfer	2017	MDPI	1.0	1.0	0.5	1.0	0.5	4
41	[55]	Research on data-driven optimal scheduling of power system	2023	MDPI	1.0	0.5	1.0	0.5	1.0	4
42	[56]	Deep-reinforcement-learning-based low-carbon economic dispatch for community-integrated energy system under multiple uncertainties	2023	Springer Link	1.0	1.0	1.0	0.5	0.5	4
43	[57]	Unlocking the flexibility of district heating pipeline energy storage with reinforcement learning	2022	MDPI	1.0	1.0	1.0	0.0	1.0	4
44	[58]	Towards reinforcement learning for vulnerability analysis in power-economic systems	2021	Science Direct	1.0	0.5	1.0	1.0	1.0	4.5
45	[59]	A deep reinforcement learning method for economic power dispatch of microgrid in OPAL-RT environment	2023	Science Direct	1.0	1.0	1.0	0.5	1.0	4.5
46	[61]	Deep reinforcement learning approaches for the hydro-thermal economic dispatch problem considering the uncertainties of the context	2023	Science Direct	1.0	1.0	1.0	1.0	0.5	4.5
47	[62]	Solving large-scale combined heat and power economic dispatch problems by using deep reinforcement learning-based crisscross optimization algorithm	2024	Science Direct	1.0	1.0	1.0	1.0	1.0	5
48	[63]	Adaptive look-ahead economic dispatch based on deep reinforcement learning	2024	Science Direct	1.0	1.0	1.0	1.0	0.5	4.5
49	[64]	Economic dispatch of industrial park considering uncertainty of renewable energy based on a deep reinforcement learning approach	2023	Science Direct	1.0	0.5	1.0	1.0	0.5	4
50	[65]	Combined heat and power system intelligent economic dispatch: A deep reinforcement learning approach	2020	Science Direct	1.0	0.5	1.0	1.0	0.5	4
51	[42]	Multi-objective solution of optimal power flow based on TD3 deep reinforcement learning algorithm	2023	Science Direct	1.0	0.5	1.0	0.5	1.0	4
52	[43]	Real-time operation of distribution network: A deep reinforcement learning-based reconfiguration approach	2022	Science Direct	1.0	1.0	1.0	1.0	0.5	4.5
53	[44]	Multi-agent deep reinforcement learning for resilience-driven routing and scheduling of mobile energy storage systems	2022	Science Direct	1.0	0.5	1.0	1.0	0.5	4
54	[93]	A scalable graph reinforcement learning algorithm based stochastic dynamic dispatch of power system under high penetration of renewable energy Junbin	2023	Science Direct	1.0	0.5	1.0	1.0	0.0	3.5
55	[94]	Emergency fault affected wide-area automatic generation control via large-scale deep reinforcement learning	2021	Science Direct	0.0	1.0	1.0	1.0	0.5	3.5

References

1. Wood, A.; Wollemberg, B.; Sheblé, G. *Power Generation, Operation and Control*, 3rd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2015; ISBN 9780471790556.
2. Glover, J.D.; Overbye, T.J.; Sarma, M.S. *Power System Analysis and Design*, 6th ed.; Cengage Learning: Boston, MA, USA, 2017; ISBN 9781305632134.
3. Natividad, L.E.; Benalcazar, P. Hybrid Renewable Energy Systems for Sustainable Rural Development: Perspectives and Challenges in Energy Systems Modeling. *Energies* **2023**, *16*, 1328. [CrossRef]
4. Conejo, A.J.; Baringo Morales, L.; Kazempour, S.J.; Siddiqui, A.S. *Investment in Electricity Generation and Transmission*; Springer: Berlin/Heidelberg, Germany, 2016; ISBN 9783319294995.
5. Cordova-Garcia, J.; Wang, X. Robust Power Line Outage Detection with Unreliable Phasor Measurements. In Proceedings of the 2017 IEEE 33rd International Conference on Data Engineering (ICDE), San Diego, CA, USA, 19–22 April 2017; pp. 1309–1319. [CrossRef]
6. Zhang, Z.; Zhang, D.; Qiu, R.C. Deep Reinforcement Learning for Power System: An Overview. *CSEE J. Power Energy Syst.* **2019**, *6*, 213–225. [CrossRef]

7. Cao, D.; Hu, W.; Zhao, J.; Zhang, G.; Zhang, B.; Liu, Z.; Chen, Z.; Blaabjerg, F. Reinforcement Learning and Its Applications in Modern Power and Energy Systems: A Review. *J. Mod. Power Syst. Clean Energy* **2020**, *8*, 1029–1042. [CrossRef]
8. Nazari-Heris, M.; Asadi, S.; Abdar, B.M.-I.M.; Jebelli, H.; Sadat-Mohammadi, M. *Application of Machine Learning and Deep Learning Methods to Power System Problems*; Springer: Berlin/Heidelberg, Germany, 2021; ISBN 9783030776954.
9. Shang, X.; Ye, L.; Zhang, J.; Yang, J.; Xu, J.; Lyu, Q.; Diao, R. Reinforcement Learning-Based Solution to Power Grid Planning and Operation Under Uncertainties. In Proceedings of the 2020 IEEE/ACM Workshop on Machine Learning in High Performance Computing Environments (MLHPC) and Workshop on Artificial Intelligence and Machine Learning for Scientific Applications (AI4S), Atlanta, GA, USA, 12 November 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 72–79.
10. Glavic, M.; Fonteneau, R.; Ernst, D. Reinforcement Learning for Electric Power System Decision and Control: Past Considerations and Perspectives. *IFAC-PapersOnLine* **2017**, *50*, 6918–6927. [CrossRef]
11. Arwa, E.O.; Folly, K.A. Reinforcement Learning Techniques for Optimal Power Control in Grid-Connected Microgrids: A Comprehensive Review. *IEEE Access* **2020**, *8*, 208992–209007. [CrossRef]
12. Perera, A.T.D.; Kamalaruban, P. Applications of Reinforcement Learning in Energy Systems. *Renew. Sustain. Energy Rev.* **2021**, *137*, 110618. [CrossRef]
13. Gao, Y.; Yu, N. Deep Reinforcement Learning in Power Distribution Systems: Overview, Challenges, and Opportunities. In Proceedings of the IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 16–18 February 2021; pp. 1–5.
14. Khodayar, M.; Liu, G.; Wang, J.; Khodayar, M.E. Deep Learning in Power Systems Research: A Review. *CSEE J. Power Energy Syst.* **2021**, *7*, 209–220. [CrossRef]
15. Frank, S.; Rebennack, S. An Introduction to Optimal Power Flow: Theory, Formulation, and Examples. *IIE Trans.* **2016**, *48*, 1172–1197. [CrossRef]
16. Chen, X.; Qu, G.; Tang, Y.; Low, S.; Li, N. Reinforcement Learning for Selective Key Applications in Power Systems: Recent Advances and Future Challenges. *arXiv* **2022**, arXiv:2102.01168. [CrossRef]
17. Wang, Y.; Chai, B.; Lu, W.; Zheng, X. A Review of Deep Reinforcement Learning Applications in Power System Parameter Estimation. In Proceedings of the 2021 International Conference on Power System Technology (POWERCON), Haikou, China, 8–9 December 2021. [CrossRef]
18. Sutton, R.; Barto, A. *Reinforcement Learning: An Introduction*, 2nd ed.; Bach, F., Ed.; The MIT Press: Cambridge, MA, USA, 2020; ISBN 9780262039246.
19. Coronado, C.A.; Figueroa, M.R.; Roa-Sepulveda, C.A. A Reinforcement Learning Solution for the Unit Commitment Problem. In Proceedings of the 2012 47th International Universities Power Engineering Conference (UPEC), Uxbridge, UK, 4–7 September 2012; pp. 2–7. [CrossRef]
20. Nie, H.; Chen, Y.; Song, Y.; Huang, S. A General Real-Time OPF Algorithm Using DDPG with Multiple Simulation Platforms. In Proceedings of the 2019 IEEE Innovative Smart Grid Technologies—Asia (ISGT Asia), Chengdu, China, 21–24 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 3713–3718.
21. Sanghi, N. *Deep Reinforcement Learning with Python*; Apress: New York, NY, USA, 2021; ISBN 9781484268087.
22. Zhou, Y.; Zhang, B.; Xu, C.; Lan, T.; Diao, R.; Shi, D.; Wang, Z.; Lee, W.-J. A Data-Driven Method for Fast AC Optimal Power Flow Solutions via Deep Reinforcement Learning. *J. Mod. Power Syst. Clean Energy* **2020**, *8*, 1128–1139. [CrossRef]
23. Kitchenham, B.; Charters, S. *Guidelines for Performing Systematic Literature Reviews in Software Engineering*; EBSE Technical Report; Version 2.3; Elsevier: Amsterdam, The Netherlands, 2007; Volume 1, pp. 1–54.
24. Al Naqbi, A.; Alyieliely, S.S.; Talib, M.A.; Nasir, Q.; Bettayeb, M.; Ghenai, C. Energy Reduction in Building Energy Management Systems Using the Internet of Things: Systematic Literature Review. In Proceedings of the2021 International Symposium on Networks, Computers and Communications (ISNCC), Dubai, United Arab Emirates, 31 October–2 November 2021; pp. 1–7. [CrossRef]
25. Chiu, P.C.; Selamat, A.; Krejcar, O.; Kuok, K.K.; Bujang, S.D.A.; Fujita, H. Missing Value Imputation Designs and Methods of Nature-Inspired Metaheuristic Techniques: A Systematic Review. *IEEE Access* **2022**, *10*, 61544–61566. [CrossRef]
26. Mendoza-Pitti, L.; Calderon-Gomez, H.; Vargas-Lombardo, M.; Gomez-Pulido, J.M.; Castillo-Sequera, J.L. Towards a Service-Oriented Architecture for the Energy Efficiency of Buildings: A Systematic Review. *IEEE Access* **2021**, *9*, 26119–26137. [CrossRef]
27. Khan, R.A.; Khan, S.U.; Khan, H.U.; Ilyas, M. Systematic Literature Review on Security Risks and Its Practices in Secure Software Development. *IEEE Access* **2022**, *10*, 5456–5481. [CrossRef]
28. Kim, J.Y.; Kim, K.S. Integrated Model of Economic Generation System Expansion Plan for the Stable Operation of a Power Plant and the Response of Future Electricity Power Demand. *Sustainability* **2018**, *10*, 2417. [CrossRef]
29. Ebeed, M.; Kamel, S.; Jurado, F. *Optimal Power Flow Using Recent Optimization Techniques*; Elsevier Inc.: Amsterdam, The Netherlands, 2018; ISBN 9780128124420.
30. Guamán, W.P.; Pesántez, G.N.; Torres R., M.A.; Falcones, S.; Urquizo, J. Optimal Dynamic Reactive Power Compensation in Power Systems: Case Study of Ecuador-Perú Interconnection. *Electr. Power Syst. Res.* **2023**, *218*, 109191. [CrossRef]
31. Carpentier, J. Contribution a.'l'etude Du Dispatching Economique. *Bull. Soc. Fr. Electr.* **1962**, *3*, 431–447.
32. Hasan, F.; Kargarian, A.; Mohammadi, A. A Survey on Applications of Machine Learning for Optimal Power Flow. In Proceedings of the 2020 IEEE Texas Power and Energy Conference (TPEC), College Station, TX, USA, 6–7 February 2020; pp. 1–6. [CrossRef]

33. Huang, J.; Xue, Y.; Dong, Z.Y.; Wong, K.P. An Adaptive Importance Sampling Method for Probabilistic Optimal Power Flow. In Proceedings of the 2011 IEEE Power and Energy Society General Meeting, Detroit, MI, USA, 24–28 July 2011; pp. 1–6. [CrossRef]
34. Cao, D.; Hu, W.; Xu, X.; Wu, Q.; Huang, Q.; Chen, Z.; Blaabjerg, F. Deep Reinforcement Learning Based Approach for Optimal Power Flow of Distribution Networks Embedded with Renewable Energy and Storage Devices. *J. Mod. Power Syst. Clean Energy* **2021**, *9*, 1101–1110. [CrossRef]
35. Zhou, Y.; Lee, W.; Diao, R.; Shi, D. Deep Reinforcement Learning Based Real-Time AC Optimal Power Flow Considering Uncertainties. *J. Mod. Power Syst. Clean Energy* **2022**, *10*, 1098–1109. [CrossRef]
36. Yan, Z.; Xu, Y. Real-Time Optimal Power Flow: A Lagrangian Based Deep Reinforcement Learning Approach. *IEEE Trans. Power Syst.* **2020**, *35*, 3270–3273. [CrossRef]
37. Al-Saffar, M.; Musilek, P. Distributed Optimal Power Flow for Electric Power Systems with High Penetration of Distributed Energy Resources. In Proceedings of the 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), Edmonton, AB, Canada, 5–8 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–5.
38. Jiang, C.; Li, Z.; Zheng, J.H.; Wu, Q.H.; Shang, X. Two-level Area-load Modelling for OPF of Power System Using Reinforcement Learning. *IET Gener. Transm. Distrib.* **2019**, *13*, 4141–4149. [CrossRef]
39. Rashedi, N.; Tajeddini, M.A.; Kebriaei, H. Markov Game Approach for Multi-agent Competitive Bidding Strategies in Electricity Market. *IET Gener. Transm. Distrib.* **2016**, *10*, 3756–3763. [CrossRef]
40. Gao, Z.; Zheng, Z.; Wu, J.; Qi, L.; Li, W.; Yang, Y. Reactive Power Optimization of Distribution Network Based on Deep Reinforcement Learning and Multi Agent System. In Proceedings of the 2021 IEEE 5th Conference on Energy Internet and Energy System Integration (EI2), Taiyuan, China, 22–24 October 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1052–1057.
41. Yao, Z.; Chen, W.; Sun, L.; Wu, X. Reinforcement Learning-Based Optimal Power Flow of Distribution Networks with High Permeation of Distributed PVs. In Proceedings of the 2023 IEEE 6th International Electrical and Energy Conference (CIEEC), Hefei, China, 12–14 May 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 3421–3426.
42. Sun, B.; Song, M.; Li, A.; Zou, N.; Pan, P.; Lu, X.; Yang, Q.; Zhang, H.; Kong, X. Multi-Objective Solution of Optimal Power Flow Based on TD3 Deep Reinforcement Learning Algorithm. *Sustain. Energy Grids Netw.* **2023**, *34*, 101054. [CrossRef]
43. Bui, V.-H.; Su, W. Real-Time Operation of Distribution Network: A Deep Reinforcement Learning-Based Reconfiguration Approach. *Sustain. Energy Technol. Assess.* **2022**, *50*, 101841. [CrossRef]
44. Wang, Y.; Qiu, D.; Strbac, G. Multi-agent deep reinforcement learning for resilience-driven routing and scheduling of mobile energy storage systems. *Appl. Energy* **2022**, *310*, 118575. [CrossRef]
45. Qin, J.; Yu, N.; Gao, Y. Solving Unit Commitment Problems with Multi-Step Deep Reinforcement Learning. In Proceedings of the 2021 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm), Aachen, Germany, 25–28 October 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 140–145.
46. Li, F.; Qin, J.; Zheng, W.X. Distributed Q-Learning-Based Online Optimization Algorithm for Unit Commitment and Dispatch in Smart Grid. *IEEE Trans. Cybern.* **2020**, *50*, 4146–4156. [CrossRef] [PubMed]
47. Li, X.; Han, X.; Yang, M. Day-Ahead Optimal Dispatch Strategy for Active Distribution Network Based on Improved Deep Reinforcement Learning. *IEEE Access* **2022**, *10*, 9357–9370. [CrossRef]
48. Li, R.; Han, Y.; Ma, T.; Liu, H. Nash-Q Learning-Based Collaborative Dispatch Strategy for Interconnected Power Systems. *Glob. Energy Interconnect.* **2020**, *3*, 227–236. [CrossRef]
49. Vergara, P.P.; Salazar, M.; Giraldo, J.S.; Palensky, P. Optimal Dispatch of PV Inverters in Unbalanced Distribution Systems Using Reinforcement Learning. *Int. J. Electr. Power Energy Syst.* **2022**, *136*, 107628. [CrossRef]
50. Bora, T.C.; Mariani, V.C.; dos Santos Coelho, L. Multi-Objective Optimization of the Environmental-Economic Dispatch with Reinforcement Learning Based on Non-Dominated Sorting Genetic Algorithm. *Appl. Therm. Eng.* **2019**, *146*, 688–700. [CrossRef]
51. Fu, Y.; Guo, X.; Mi, Y.; Yuan, M.; Ge, X.; Su, X.; Li, Z. The Distributed Economic Dispatch of Smart Grid Based on Deep Reinforcement Learning. *IET Gener. Transm. Distrib.* **2021**, *15*, 2645–2658. [CrossRef]
52. Tan, Y.; Shen, Y.; Yu, X.; Lu, X. Low-carbon Economic Dispatch of the Combined Heat and Power-virtual Power Plants: A Improved Deep Reinforcement Learning-based Approach. *IET Renew. Power Gener.* **2023**, *17*, 982–1007. [CrossRef]
53. Liu, J.; Liu, Y.; Qiu, G.; Gu, Y.; Li, H.; Liu, J. Deep-Q-Network-Based Intelligent Reschedule for Power System Operational Planning. In Proceedings of the 2020 12th IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC), Nanjing, China, 20–23 September 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
54. Zhao, Y.; Liu, J.; Liu, X.; Yuan, K.; Ren, K.; Yang, M. A Graph-Based Deep Reinforcement Learning Framework for Autonomous Power Dispatch on Power Systems with Changing Topologies. In Proceedings of the 2022 IEEE Sustainable Power and Energy Conference (iSPEC), Perth, Australia, 4–7 December 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–5.
55. Luo, J.; Zhang, W.; Wang, H.; Wei, W.; He, J. Research on Data-Driven Optimal Scheduling of Power System. *Energies* **2023**, *16*, 2926. [CrossRef]
56. Mo, M.; Xiong, X.; Wu, Y.; Yu, Z. Deep-Reinforcement-Learning-Based Low-Carbon Economic Dispatch for Community-Integrated Energy System under Multiple Uncertainties. *Energies* **2023**, *16*, 7669. [CrossRef]
57. Stepanovic, K.; Wu, J.; Everhardt, R.; de Weerdt, M. Unlocking the Flexibility of District Heating Pipeline Energy Storage with Reinforcement Learning. *Energies* **2022**, *15*, 3290. [CrossRef]
58. Wolgast, T.; Veith, E.M.; Nieße, A. Towards Reinforcement Learning for Vulnerability Analysis in Power-Economic Systems. *Energy Inform.* **2021**, *4*, 21. [CrossRef]

59. Lin, F.-J.; Chang, C.-F.; Huang, Y.-C.; Su, T.-M. A Deep Reinforcement Learning Method for Economic Power Dispatch of Microgrid in OPAL-RT Environment. *Technologies* **2023**, *11*, 96. [CrossRef]
60. Han, C.; Yang, B.; Bao, T.; Yu, T.; Zhang, X. Bacteria Foraging Reinforcement Learning for Risk-Based Economic Dispatch via Knowledge Transfer. *Energies* **2017**, *10*, 638. [CrossRef]
61. Arango, A.R.; Aguilar, J.; R-Moreno, M.D. Deep Reinforcement Learning Approaches for the Hydro-Thermal Economic Dispatch Problem Considering the Uncertainties of the Context. *Sustain. Energy Grids Netw.* **2023**, *35*, 101109. [CrossRef]
62. Meng, A.; Rong, J.; Yin, H.; Luo, J.; Tang, Y.; Zhang, H.; Li, C.; Zhu, J.; Yin, Y.; Li, H.; et al. Solving Large-Scale Combined Heat and Power Economic Dispatch Problems by Using Deep Reinforcement Learning Based Crisscross Optimization Algorithm. *Appl. Therm. Eng.* **2024**, *245*, 122781. [CrossRef]
63. Wang, X.; Zhong, H.; Zhang, G.; Ruan, G.; He, Y.; Yu, Z. Adaptive Look-Ahead Economic Dispatch Based on Deep Reinforcement Learning. *Appl. Energy* **2024**, *353*, 122121. [CrossRef]
64. Feng, J.; Wang, H.; Yang, Z.; Chen, Z.; Li, Y.; Yang, J.; Wang, K. Economic Dispatch of Industrial Park Considering Uncertainty of Renewable Energy Based on a Deep Reinforcement Learning Approach. *Sustain. Energy Grids Netw.* **2023**, *34*, 101050. [CrossRef]
65. Zhou, S.; Hu, Z.; Gu, W.; Jiang, M.; Chen, M.; Hong, Q.; Booth, C. Combined Heat and Power System Intelligent Economic Dispatch: A Deep Reinforcement Learning Approach. *Int. J. Electr. Power Energy Syst.* **2020**, *120*, 106016. [CrossRef]
66. Ahrari Nouri, M.; Hesami, A.; Seifi, A. Reactive Power Planning in Distribution Systems Using a Reinforcement Learning Method. In Proceedings of the 2007 International Conference on Intelligent and Advanced Systems, Kuala Lumpur, Malaysia, 25–28 November 2007; IEEE: Piscataway, NJ, USA, 2007; pp. 157–161.
67. MingKui, W.; ShaoRong, C.; Quan, Z.; Xu, Z.; Hong, Z.; YuHong, W. Multi-Objective Transmission Network Expansion Planning Based on Reinforcement Learning. In Proceedings of the 2020 IEEE Sustainable Power and Energy Conference (iSPEC), Chengdu, China, 23–25 November 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 2348–2353.
68. Choi, J.; Lee, K. *Probabilistic Power System Expansion Planning with Renewable Energy Resources and Energy Storage Systems*; IEEE Press Editorial Board, Ed.; Wiley: Hoboken, NJ, USA, 2022; ISBN 9781119684138.
69. Papadimitrakis, M.; Giamarelos, N.; Stogiannos, M.; Zois, E.N.; Livanos, N.A.-I.; Alexandridis, A. Metaheuristic Search in Smart Grid: A Review with Emphasis on Planning, Scheduling and Power Flow Optimization Applications. *Renew. Sustain. Energy Rev.* **2021**, *145*, 111072. [CrossRef]
70. Kamruzzaman, M.; Duan, J.; Shi, D.; Benidris, M. A Deep Reinforcement Learning-Based Multi-Agent Framework to Enhance Power System Resilience Using Shunt Resources. *IEEE Trans. Power Syst.* **2021**, *36*, 5525–5536. [CrossRef]
71. Wang, Y.; Zhou, X.; Zhou, H.; Chen, L.; Zheng, Z.; Zeng, Q.; Cai, S.; Wang, Q. Transmission Network Dynamic Planning Based on a Double Deep-Q Network With Deep ResNet. *IEEE Access* **2021**, *9*, 76921–76937. [CrossRef]
72. Zhang, X.; Hua, W.; Liu, Y.; Duan, J.; Tang, Z.; Liu, J. Reinforcement Learning for Active Distribution Network Planning Based on Monte Carlo Tree Search. *Int. J. Electr. Power Energy Syst.* **2022**, *138*, 107885. [CrossRef]
73. Wang, Y.; Chen, L.; Zhou, H.; Zhou, X.; Zheng, Z.; Zeng, Q.; Jiang, L.; Lu, L. Flexible Transmission Network Expansion Planning Based on DQN Algorithm. *Energies* **2021**, *14*, 1944. [CrossRef]
74. Wang, Y.; Zhou, X.; Shi, Y.; Zheng, Z.; Zeng, Q.; Chen, L.; Xiang, B.; Huang, R. Transmission Network Expansion Planning Considering Wind Power and Load Uncertainties Based on Multi-Agent DDQN. *Energies* **2021**, *14*, 6073. [CrossRef]
75. Tsianikas, S.; Yousefi, N.; Zhou, J.; Rodgers, M.D.; Coit, D. A Storage Expansion Planning Framework Using Reinforcement Learning and Simulation-Based Optimization. *Appl. Energy* **2021**, *290*, 116778. [CrossRef]
76. Chao, F.A.Z.; Ying, S.B.Z.; Yu, T.C.J. Application of Improved Reinforcement Learning Technology for Real Time Operation and Scheduling Optimization of Virtual Power Plant. In Proceedings of the 2023 IEEE Sustainable Power and Energy Conference (iSPEC), Chongqing, China, 28–30 November 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
77. Dong, F.; Li, Z.; Xu, Y.; Zhu, D.; Huang, R.; Zou, H.; Wu, Z.; Wang, X. Planning for Network Expansion Based on Prim Algorithm and Reinforcement Learning. In Proceedings of the 2023 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia), Chongqing, China, 7–9 July 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 252–258.
78. Bhatnagar, M.; Yadav, A.; Swetapadma, A. Integrating Distributed Generation and Advanced Deep Learning for Efficient Distribution System Management and Fault Detection. *Arab. J. Sci. Eng.* **2024**, *49*, 7095–7111. [CrossRef]
79. Kundačina, O.B.; Vidović, P.M.; Petković, M.R. Solving Dynamic Distribution Network Reconfiguration Using Deep Reinforcement Learning. *Electr. Eng.* **2022**, *104*, 1487–1501. [CrossRef]
80. Davis, J.V.; Dhillon, I.S. Structured Metric Learning for High Dimensional Problems. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, Las Vegas, NV, USA, 24–27 August 2008; pp. 195–203. [CrossRef]
81. Hossin, M.; Sulaiman, M. A Review on Evaluation Metrics for Data Classification Evaluations. *Int. J. Data Min. Knowl. Manag. Process* **2015**, *5*, 1–11. [CrossRef]
82. Koltsaklis, N.E.; Dagoumas, A.S. State-of-the-Art Generation Expansion Planning: A Review. *Appl. Energy* **2018**, *230*, 563–589. [CrossRef]
83. Mahdavi, M.; Sabillon Antunez, C.; Ajalli, M.; Romero, R. Transmission Expansion Planning: Literature Review and Classification. *IEEE Syst. J.* **2019**, *13*, 3129–3140. [CrossRef]
84. Chatzos, M.; Mak, T.W.K.; Vanhentenryck, P. Spatial Network Decomposition for Fast and Scalable AC-OPF Learning. *IEEE Trans. Power Syst.* **2021**, *37*, 2601–2612. [CrossRef]

85. Woo, J.H.; Wu, L.; Park, J.B.; Roh, J.H. Real-Time Optimal Power Flow Using Twin Delayed Deep Deterministic Policy Gradient Algorithm. *IEEE Access* **2020**, *8*, 213611–213618. [CrossRef]
86. Benalcazar, P.; Kamiński, J.; Stós, K. An Integrated Approach to Long-Term Fuel Supply Planning in Combined Heat and Power Systems. *Energies* **2022**, *15*, 8339. [CrossRef]
87. Yu, Z.; Ruan, G.; Wang, X.; Zhang, G.; He, Y.; Zhong, H. Evaluation of Look-Ahead Economic Dispatch Using Reinforcement Learning. In Proceedings of the 2022 IEEE 6th Conference on Energy Internet and Energy System Integration (EI2), Chengdu, China, 11–13 November 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1708–1713.
88. Fang, D.; Guan, X.; Hu, B.; Peng, Y.; Chen, M.; Hwang, K. Deep Reinforcement Learning for Scenario-Based Robust Economic Dispatch Strategy in Internet of Energy. *IEEE Internet Things J.* **2021**, *8*, 9654–9663. [CrossRef]
89. Lin, L.; Guan, X.; Peng, Y.; Wang, N.; Maharjan, S.; Ohtsuki, T. Deep Reinforcement Learning for Economic Dispatch of Virtual Power Plant in Internet of Energy. *IEEE Internet Things J.* **2020**, *7*, 6288–6301. [CrossRef]
90. Lv, K.; Tang, H.; Bak-Jensen, B.; Radhakrishna Pillai, J.; Tan, Q.; Zhang, Q. Hierarchical Learning Optimisation Method for the Coordination Dispatch of the Inter-regional Power Grid Considering the Quality of Service Index. *IET Gener. Transm. Distrib.* **2020**, *14*, 3673–3684. [CrossRef]
91. Yang, Y.; Wu, L. Machine Learning Approaches to the Unit Commitment Problem: Current Trends, Emerging Challenges, and New Strategies. *Electr. J.* **2021**, *34*, 106889. [CrossRef]
92. Wang, D. A New Power System Dispatching Optimization Method Based on Reinforcement Learning. In Proceedings of the 2023 2nd Asian Conference on Frontiers of Power and Energy (ACFPE), Chengdu, China, 20–22 October 2023; IEEE: Piscataway, NJ, USA, 2023; Volume 4, pp. 145–149.
93. Chen, J.; Yu, T.; Pan, Z.; Zhang, M.; Deng, B. A Scalable Graph Reinforcement Learning Algorithm Based Stochastic Dynamic Dispatch of Power System under High Penetration of Renewable Energy. *Int. J. Electr. Power Energy Syst.* **2023**, *152*, 109212. [CrossRef]
94. Li, J.; Yu, T.; Zhang, X. Emergency Fault Affected Wide-Area Automatic Generation Control via Large-Scale Deep Reinforcement Learning. *Eng. Appl. Artif. Intell.* **2021**, *106*, 104500. [CrossRef]