International Journal of Innovative Research and Scientific Studies, 8(3) 2025, pages: 2438-2453

IJIRSS
International Journal of Innovative Research & Scientific Studies

ISSN: 2617-6548

URL: www.ijirss.com

Comparison of PPO-DRL and A2C-DRL Algorithms for MPPT in Photovoltaic Systems via Buck-Boost Converter

Wiwat Jeungthanasirigool[1], Thanyaphob Sirimaskasem[2], Terapong Boonraksa[3], Promphak Boonraksa[4*]

[1]*Department of Electronic Engineering Technology, College of Industrial Technology, King Mongkut's University of Technology North Bangkok, 1518 Pracharat Road 1, Wong Sawang, Bang Sue, Bangkok, Thailand.*
[2]*Department of Electrical Engineering Technology, Faculty of Industrial Technology, Phranakhon Rajabhat University, 9 Changwattana Road, Bang Khen, Bangkok, Thailand.*
[3]*School of Electrical Engineering, Rajamangala University of Technology Rattanakosin, Nakhon Pathom, Thailand.*
[4]*Department of Mechatronics Engineering, Faculty of Engineering and Architecture, Rajamangala University of Technology Suvarnabhumi, 217 Nonthaburi Road, Suan Yai Subdistrict, Mueang Nonthaburi District,Nonthaburi, Thailand.*

Corresponding author: Promphak Boonraksa (*Email: promphak.b@rmutsb.ac.th*)

Abstract

This research investigates the effectiveness of two deep reinforcement learning algorithms, Proximal Policy Optimization (PPO) and Advantage Actor-Critic (A2C), in achieving the MPPT for PV systems implemented via a Buck-Boost converter. The algorithms were trained and evaluated under varying environmental conditions, including different levels of irradiance and temperature. The results are presented through duty cycle heatmaps, power output heatmaps, and performance curves for power, voltage, and current. The PPO algorithm demonstrated stable and consistent control across all scenarios, maintaining a nearly constant duty cycle and achieving high power output. In contrast, A2C exhibited more adaptive control behavior, adjusting the duty cycle based on environmental changes, but showed lower power output under weak irradiance. Overall, PPO outperformed A2C in terms of stability, accuracy, and ability to reach the optimal operating point, making it a more suitable choice for MPPT applications in PV systems under dynamic conditions.

Keywords: A2C-DRL Algorithms, Deep reinforcement learning, MPPT, DC/DC Buck-boost converter, PPO-DRL algorithms, PV systems.

DOI: 10.53894/ijirss.v8i3.7022
Funding: This research was funded by College of Industrial Technology, King Mongkut's University of Technology North Bangkok (Grant No. Res-CIT360/2024).
History: Received: 24 March 2025 / **Revised:** 28 April 2025 / **Accepted:** 30 April 2025 / **Published:** 14 May 2025

Competing Interests: The authors declare that they have no competing interests.
Authors' Contributions: All authors contributed equally to the conception, methodology, validation, review and editing, formal analysis, and design of the study. All authors have read and agreed to the published version of the manuscript.
Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.
Acknowledgment: The authors thank the King Mongkut's University of Technology North Bangkok for financial support. Special thanks to Phranakhon Rajabhat University and the Faculty of Industrial Technology, as well as Rajamangala University of Technology Suvarnabhumi, for providing facilities and equipment to support the research.
Publisher: Innovative Research Publishing

1. Introduction

The global trend toward renewable energy adoption has been increasing steadily each year, as the international community becomes more attentive to reducing reliance on coal-based energy and mitigating carbon dioxide emissions, which have detrimental impacts on the environment [1-3]. Thailand, in line with these global efforts, has also recognized the importance of such initiatives. Consequently, the Ministry of Energy has formulated policies to promote the use of renewable energy in various sectors [4], including the implementation of feed-in tariffs for electricity generated from solar photovoltaic (PV) systems by both households and private enterprises [5, 6]. At present, the electricity buy-back rate offered by the utility for solar energy is 2.1679 THB per kilowatt-hour (approximately 0.060 USD per unit; note that the exchange rate may fluctuate) [7, 8]. This favorable rate, combined with a short payback period of merely 3-4 years, has further incentivized investment. Additionally, the process for obtaining grid connection approval has been streamlined by the relevant authorities [9]. Governmental and institutional support measures have significantly accelerated the deployment of solar energy systems, positioning them as one of Thailand's most prominent renewable energy sources.

Solar energy is regarded as a clean energy source; as sunlight is freely and directly available to the Earth, it is often considered a cost-free resource [10]. When solar panels are exposed to sunlight, the energy causes electrons to move, thereby generating electrical power [11-13]. The primary parameters affecting the electrical characteristics of photovoltaic cells are temperature and irradiance. An increase in temperature generally leads to a reduction in output power, as illustrated in Table 1.

Table 1.
Summary of Parameter Effect on PV Output Power.

Parameter Effect on PV Output Power	Reference
• Temperature Decrease by ~0.4%–0.5% per °C above 25°C • Irradiance Decrease linearly with irradiance reduction (e.g., 20% drop ~20% drop in power.	Fraunhofer Institute for Solar Energy Systems ISE [14]; National Renewable Energy Laboratory (NREL) [15]; Solar Energy International [16] and Boonraksa, et al. [17]

Conversely, when the irradiance increases, the output power of the system also increases. However, in practical applications, these factors are difficult to control. Consequently, numerous studies have investigated MPPT techniques to ensure that photovoltaic systems operate at their maximum power output at all times, despite changing environmental conditions. Several works have compared the efficiency of various converter circuits used in MPPT, including the Buck converter, Boost converter, Buck-Boost converter, Cuk converter, Sepic converter, Zeta converter, Synchronous Sepic converter, and Synchronous Zeta converter [18-24]. The results indicate that the Synchronous Zeta converter achieves the highest efficiency. Nevertheless, due to its complex structure and large number of components, it is less favored compared to the Buck-Boost converter [25-27].

The MPPT algorithm is another critical component, as it determines how effectively the converter circuit can track the maximum power point. Conventional algorithms include Perturb and Observe (P&O) [28] and Incremental Conductance (IC) [29], as well as advanced approaches based on artificial intelligence (AI) [30] such as Particle Swarm Optimization (PSO), Artificial Neural Networks (ANN) [31], Adaptive Neuro-Fuzzy Inference System (ANFIS) [32], Deep Learning (DL) [33], and Deep Reinforcement Learning (DRL). DRL, a subfield of deep learning, stands out for its adaptability to nonlinear and time-varying conditions. DRL has demonstrated the ability to manage the nonlinear characteristics and rapidly changing environments typical of PV systems, such as fluctuating irradiance and temperature, where conventional algorithms often encounter issues with accuracy and tracking speed. [34, 35].

Moreover, DRL does not require a precise mathematical model of the system, unlike traditional model-based methods. It can learn optimal control policies through real-world interaction and experimentation, bypassing the need to explicitly formulate the dynamics of the PV array and power converter [35, 36]. DRL-based MPPT controllers have also exhibited faster and more accurate tracking performance compared to classical methods like P&O and IC, especially under partial shading or rapidly changing weather conditions [34, 37].

Additionally, DRL demonstrates robustness to measurement noise and system uncertainties; its learning ability enables the system to remain stable and efficient even when subjected to such disturbances [36]. After sufficient training, DRL agents can generalize their learned control strategies to new and previously unseen conditions, making them well-suited for deployment across diverse PV system configurations [34, 37].

Deep Reinforcement Learning (DRL) algorithms can generally be categorized into three main groups based on their methodological architecture: Value-Based Methods, which estimate the value function for each state or state-action pair and select actions that maximize expected rewards. Examples include Deep Q-Network (DQN), Double DQN, Dueling DQN, and Rainbow DQN. Policy-Based Methods, which directly learn the optimal policy without explicitly estimating the value function, are particularly effective for problems with continuous action spaces. Examples include
REINFORCE (Monte Carlo Policy Gradient), Proximal Policy Optimization (PPO), and Trust Region Policy Optimization (TRPO). Actor-Critic Methods combine the strengths of both value-based and policy-based approaches by employing an actor to determine the policy and a critic to evaluate the value function. Notable examples are Advantage Actor-Critic (A2C), Asynchronous Advantage Actor-Critic (A3C), Deep Deterministic Policy Gradient (DDPG), Twin Delayed DDPG (TD3), and Soft Actor-Critic (SAC) [38, 39]. In this research, the focus is narrowed to two state-of-the-art Deep Reinforcement Learning algorithms, namely Proximal Policy Optimization (PPO) and Advantage Actor-Critic (A2C), for the task of maximum power point tracking in PV systems.

Therefore, this research presents a comparative study between PPO-DRL and A2C-DRL algorithms for maximum power point tracking in photovoltaic systems using a buck-boost converter. The comparison considers tracking speed, accuracy, and efficiency under varying irradiance and temperature conditions, with the aim of providing guidance for selecting the most appropriate Deep Reinforcement Learning algorithm for MPPT in PV systems. The application infrastructure of the DRL algorithm for MPPT in PV systems is shown in Figure 1.

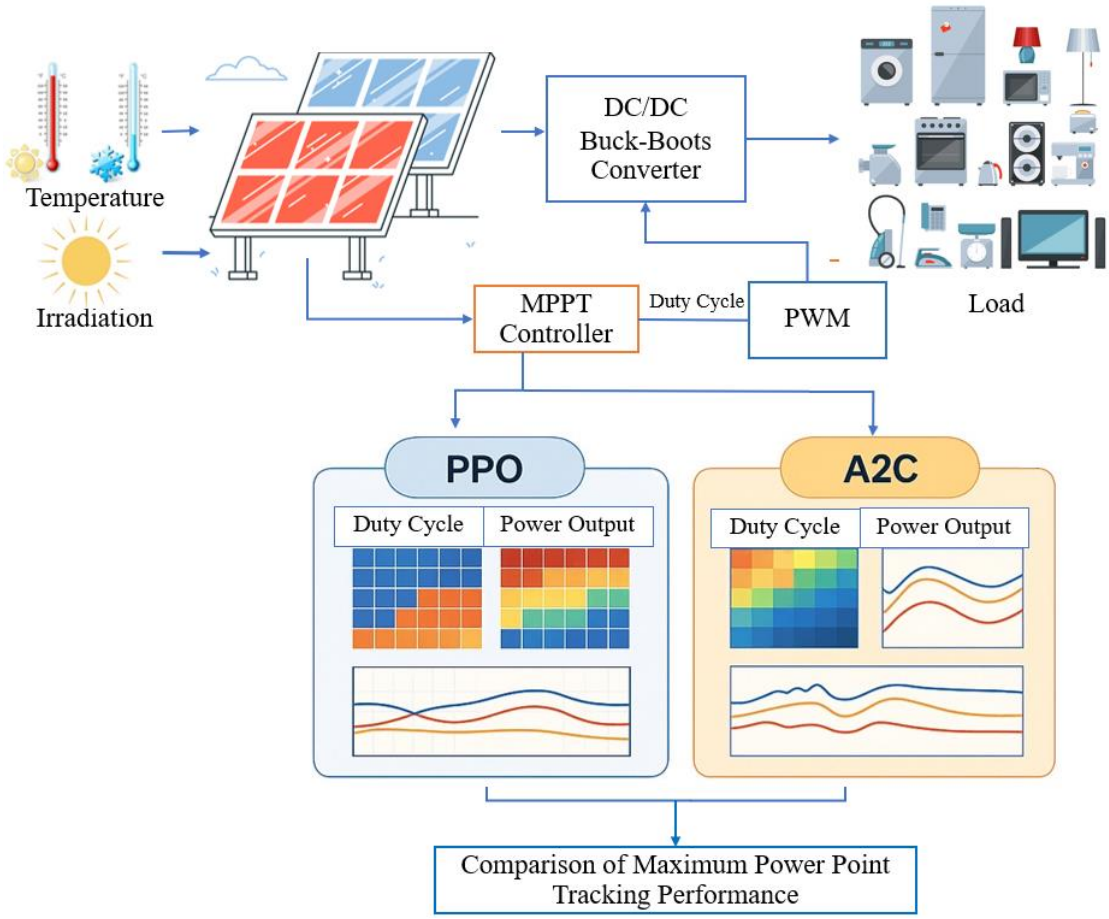

Figure 1.
Application infrastructure of DRL algorithm for MPPT in PV systems.

2. Theory and Related Works

This section provides a comprehensive summary of the fundamental theories essential for the formulation and execution of the research. It encompasses key topics including PV System Fundamentals, MPPT, DC-DC Converters for MPPT applications, and the application of DRL techniques for MPPT optimization. In addition, a review of related research studies is presented to contextualize the current study within the existing body of knowledge. The theoretical background and prior research discussed herein serve as the foundation for the development of the proposed research methodology.

2.1. The PV System Fundamentals

The PV system converts solar energy directly into electrical energy using semiconductor materials. The PV module equivalent circuit is shown in Figure 2. In this model, R_s represents the series resistance within the cell, while R_p denotes the parallel (shunt) resistance. The relationship between the current and voltage of the solar panel can be expressed mathematically by the following Equations 1-2.

$$I_{PV} = I_{out} - I_D\left[\exp\left(\frac{V_{PV}}{V_T}\right) - 1\right] \quad (1)$$

$$V_{PV} = V_T I_n\left[\frac{I_{out} - I_{PV}}{I_D} + 1\right] \quad (2)$$

The following parameters are defined:

I_{PV} is the current generated by the solar panel (A)
I_D is the reverse saturation current (A)
I_{out} is the output current (A)
V_{PV} is the voltage generated when the solar panel is exposed to sunlight (V)
V_T is the temperature-dependent voltage

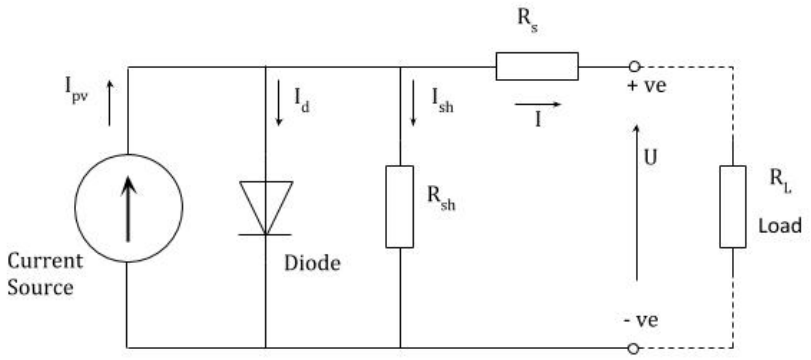

Figure 2.
PV module equivalent circuit [40].

The performance of PV cells depends significantly on environmental factors such as irradiance and temperature. The output voltage and current characteristics are nonlinear, and the point at which maximum power can be extracted (Maximum Power Point, MPP) varies with changing conditions [41, 42]. The current-voltage (I-V) characteristics of a conventional silicon-based solar cell are depicted in Figure 3. The electrical power generated by the solar cell is given by the product of current and voltage, expressed as $P = I \times V$. The curve identifies the MPP, which corresponds to the operating conditions at which both the current (I_{max}) and voltage (V_{max}) yield the highest power output.

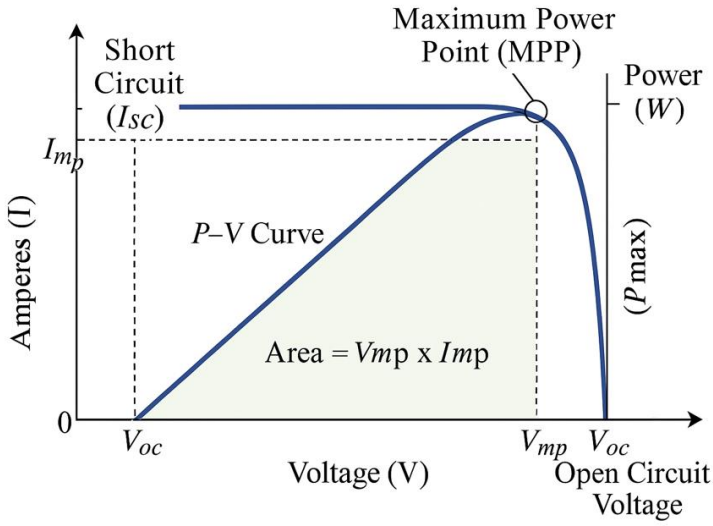

Figure 3.
The current–voltage (I–V) characteristics of a conventional silicon-based PV panel [43].

2.2. Maximum Power Point Tracking

MPPT techniques are used to continuously track and operate PV modules at their maximum power point (MPP), maximizing energy harvest. These techniques ensure that the PV system operates at its maximum power point under varying environmental conditions. MPPT methods can generally be classified into three main categories:

1. OFF-line MPPT Techniques

Off-line MPPT techniques are based on predetermined data or characteristics of the PV system. These methods do not require continuous real-time monitoring and adjustment during operation. Instead, they use mathematical models, lookup tables, or fixed parameters to estimate the maximum power point. These techniques are generally simpler and easier to implement but may not always provide the highest accuracy under rapidly changing conditions.

2. ON-line MPPT Techniques

Online MPPT techniques involve real-time measurement and adjustment to continuously track the maximum power point as environmental conditions change. They are dynamic and adaptive, making them more effective in practical applications where irradiance and temperature can fluctuate.

3. Intelligent MPPT Techniques

Intelligent MPPT techniques utilize advanced computational algorithms and artificial intelligence to enhance tracking performance, especially in complex or rapidly changing environments.

Popular algorithms include Perturb and Observe (P&O), Incremental Conductance (INC), and modern artificial intelligence-based approaches. The effectiveness of MPPT algorithms is critical for ensuring optimal performance of PV systems under dynamic weather conditions [43-45]. The MPPT methods are shown in Figure 4.

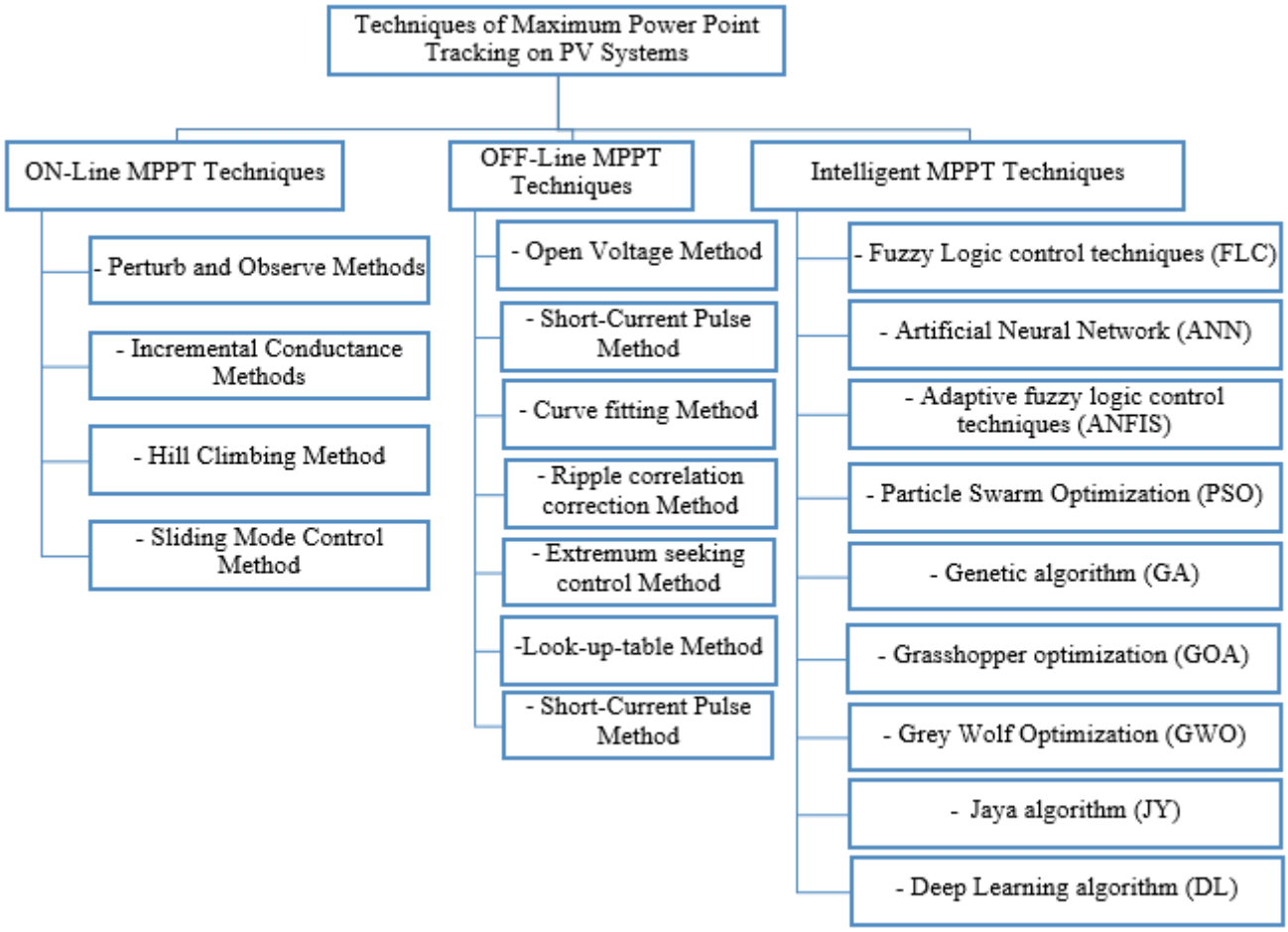

Figure 4.
Algorithm of Maximum Power Point Tracking [43].

2.3. DC-DC Converter for MPPT

A DC-DC converter (such as Buck, Boost, or Buck-Boost) is typically used as an interface between the PV module and the load or battery. The converter's duty cycle is regulated by the MPPT controller, which adjusts the PV operating point in real time to maximize output power [45]. This integration enables the system to follow the MPP despite changes in solar irradiance or temperature. In this research, the researcher chose to use the Buck-Boost converter circuit. Therefore, the theoretical background specific to this circuit is presented in detail as follows.

1. Buck-Boost Converter Theory

A Buck-Boost Converter is a DC-DC converter capable of both stepping up and stepping down the input voltage to achieve a desired output level. By combining the operating principles of buck and boost converters, it can deliver output voltages that are either higher or lower than the input. Additionally, this converter can invert the polarity of the output voltage, making it suitable for applications requiring a flexible voltage range and polarity control. The Buck-Boost converter typically consists of a switch (such as a MOSFET), a diode, an inductor, and a capacitor. Its operation can be divided into two main modes: When the switch is ON, the inductor stores energy from the input source while the diode prevents current from flowing to the output. When the switch is OFF, the energy stored in the inductor is released to the output through the diode, causing the output voltage to either rise above or drop below the input, depending on the duty cycle of the switch. Buck-boost converter circuit diagram as shown in Figures 5-6.

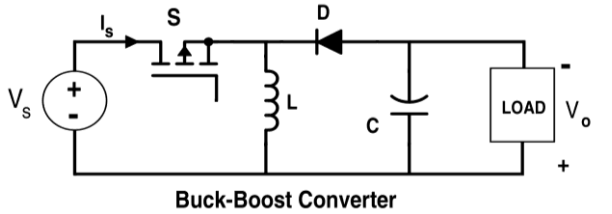

Figure 5.
Buck-boost converter circuit diagram [46].

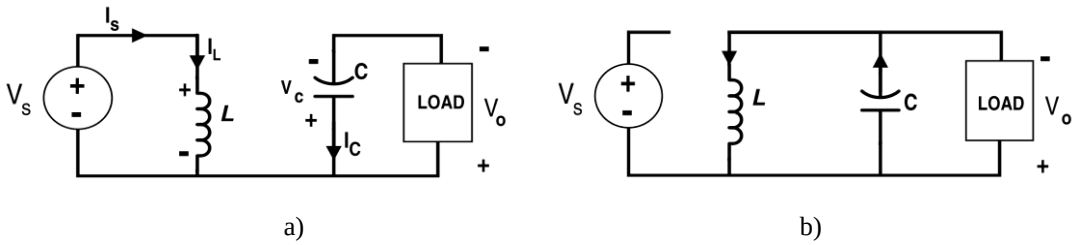

Figure 6.
Buck-boost converter circuit diagram: a) Switch S is opened and b) Switch S is closed) [46].

2.4. Deep Reinforcement Learning for MPPT

Recent research has applied the DRL algorithms, such as PPO and A2C, to MPPT problems in PV systems. DRL methods can learn optimal policies for complex, nonlinear systems without explicit modeling, making them suitable for real-time power optimization. Several studies have demonstrated that DRL-based MPPT can outperform traditional algorithms in terms of tracking accuracy and speed, especially under rapidly changing environmental conditions [47-49].

2.4.1. Proximal Policy Optimization algorithm

PPO is a reinforcement learning algorithm designed to update policies in a stable and efficient manner. It improves training by limiting how much the policy can change at each update step, using a clipped objective function to prevent overly large shifts. By balancing exploration and exploitation, PPO maintains a steady learning process, making it suitable for a wide range of complex environments [50]. Figure 7 shows the Proximal Policy Optimization algorithm pseudocode.

Algorithm 2 Proximal Policy Optimization (PPO)

1: Initialize actor $\mu: S \rightarrow R^{m+1}$ and $\sigma: S \rightarrow diag(\sigma_1, \sigma_2, \ldots, \sigma_{m+1})$
2: **for** i = 1 to M **do**
Run policy $\pi\theta \sim N(\mu(s), \sigma(s))$ for T timesteps and collect (s_t, a_t, r_t)
Estimate advantages $\hat{A}_t = \sum_{t'>t} \gamma^{t'-t} r_{t'} - v(s_t)$
Update old policy $\pi_{old} \leftarrow \pi_0$
3: **for** j = 1 to N **do**
Update actor policy by policy gradient:

$$\sum_i \nabla_\theta L_i^{CLIP}(\theta)$$

Update critic by:

$$\nabla L(\phi) = -\sum_{t=1}^{T} \nabla \hat{A}_t{}^2$$

4: **end for**
5: **end for**

Figure 7.
Proximal Policy Optimization algorithm pseudocode [50].

2.4.2. Advantage Actor-Critic algorithm

A2C is a synchronous reinforcement learning method that combines value-based and policy-based approaches. It uses two neural networks: one (the actor) to select actions and another (the critic) to evaluate the actions by estimating the advantage function. By updating both networks simultaneously, A2C aims to accelerate learning and improve decision-making stability in dynamic environments [51]. Pseudocode for the A2C algorithm as shown in Figure 8.

Algorithm 1 Advantage Actor-Critic (A2C)

1: //Assume global shared $\theta, \theta^{-}, and\ counter\ T = 0$.
2: Initialize thread step counter $t \leftarrow 0, \theta^{-} \leftarrow \theta, d\theta \leftarrow 0$.
Get initial state s.
3: **repeat**
Take action a with ϵ-greedy policy base on $Q(s, a;\ \theta)$
4: Receive new state s' and reward r

$$y = \begin{cases} r & for\ terminal\ s' \\ r + \gamma max_{a'} Q(s', a'; \theta^{-}) & for\ non-terminal\ s' \end{cases}$$

$s = s'$
$T \leftarrow T + 1\ and\ t \leftarrow t + 1$
5: **If** $T\ mod\ I_{target} == 0$ **then**
Update the target network $\theta^{-} \leftarrow \theta$
end if
if t $mod\ I_{AsyncUpdate} == 0$ or s is terminal **then**
Perform asynchronous update of θ using $d\theta$.
Clear gradients $d\theta \leftarrow 0$.
end if
until $T > T_{max}$

Figure 8.
Advantage Actor-Critic algorithm pseudocode [51].

2.5. Literature Review Summary

This research explores the application of DRL algorithms, specifically PPO and A2C, for MPPT control in photovoltaic systems. Previous studies show that traditional MPPT methods have limitations under dynamic conditions, while DRL offers a more adaptive and intelligent control approach. This section briefly reviews related works on MPPT techniques and the use of PPO and A2C in energy systems. Summary of relevant research as shown in Table 2.

Table 2.
Summary of relevant research.

Ref.	Authors & Year	DRL Algorithm	Converter Type	Key Contributions
[52]	Saha et al., 2023	PPO	Boost	Demonstrated that PPO-based control outperforms traditional methods in terms of settling time and stability in DC-DC boost converters.
[53]	Cui et al., 2020	DQN	Buck	Proposed a DRL-based intelligent control strategy for buck converters, enhancing voltage stability under varying loads.
[54]	Liu et al., 2020	DQN & DDPG	Buck-Boost	Developed DRL-based MPPT algorithms (DQN and DDPG) for PV systems under partial shading, outperforming traditional methods.
[55]	Wongsathan, 2024	ANN	Buck-Boost	Integrated a neural network-based MPPT with ant colony optimization-tuned PI controller, achieving improved energy efficiency.

3. Methodology

This research, titled Comparison of PPO-DRL and A2C-DRL Algorithms for MPPT in Photovoltaic Systems via Buck-Boost Converter is structured into three main stages of methodology, detailed as follows: Preparation of Solar Panel Parameters, Calculation of Buck-Boost Converter Parameters and Simulation of MPPT Algorithms.

3.1. Preparation of Solar Panel Parameters

In this step, the key electrical parameters of the PV panel are identified and prepared. These parameters include open-circuit voltage, short-circuit current, maximum power point voltage, and current under standard test conditions. Accurate parameterization is essential for developing a reliable PV system model, which serves as the foundation for subsequent simulations and algorithm evaluations. Table 3 shows parameters of the PV module in this case study.

Table 3.
PV module Parameters.

Parameters (at STC)	**Values**
Maximum Power (P_{max})	340 W
Maximum Voltage at P_{max} (V_{mp})	38.5 V
Maximum Current at P_{max} (I_{mp})	8.33 A
Open-Circuit Voltage (V_{oc})	47.2 V
Short-Circuit Current (I_{sc})	9.40 A
Temp. Coefficiency of V_{oc} (%/deg.^{o}C or ^{o}K)	-0.38
Temp. Coefficiency of I_{sc} (%/deg.oC or ^{o}K)	0.065

3.2. Calculation of Buck-Boost Converter Parameters

This stage involves determining the electrical characteristics of the Buck-Boost converter, which functions as the power electronic interface between the PV panel and the load. Important parameters such as duty cycle range, inductor and capacitor values, and switching frequency are calculated based on the operational requirements of the PV system. Proper design of the converter ensures efficient energy conversion across varying environmental conditions. The various circuit parameters were determined according to Equations 3 through 6, which describe the fundamental relationships governing the operation of the system [56]. Parameters of the Buck-Boost converter are shown in Table 4.

3.2.1. Inductor

Use the buck-boost inductor sizing formula (worst-case at max duty cycle):

$$L=\frac{V_{in}\cdot D}{\Delta L\cdot f_s} \quad (3)$$

2. Output Capacitor

$$C=\frac{I_{at}\cdot D}{\Delta V_{at}\cdot f_s} \quad (4)$$

3. Load Resistance (Max)

$$R=\frac{V^2}{P} \quad (5)$$

4. Duty Cycle Range

$$D=\frac{V_{at}}{V_{at}+V_{in}} \quad (6)$$

Table 4.
Parameters of the Buck-Boots converter.

Parameters	**Values**
Capacitor C_1	1000 µF
Inductor L_1,	330 µH
Forward Voltage of Diode (V_f)	0.5-0.7 V(Schottky)
Load Resistance maximum	4.62 Ω
Duty cycle	0.1-0.8

3.3. Simulation of MPPT Algorithms

In the final stage, the Proximal Policy Optimization Deep Reinforcement Learning (PPO-DRL) and Advantage Actor-Critic Deep Reinforcement Learning (A2C-DRL) algorithms are implemented and simulated. The objective is to track the maximum power point (MPP) of the PV system under dynamic irradiance and temperature conditions. The performance of each algorithm is assessed based on criteria such as tracking speed, stability, and overall energy harvesting efficiency.

The fundamental equations representing the operation of the Advantage Actor-Critic (A2C) algorithm for Maximum Power Point Tracking (MPPT) in photovoltaic (PV) systems are presented in Equations 7-9. These equations illustrate the policy gradient update, advantage estimation, and value function loss used by the actor and critic networks during the training process [57-59].

Policy Gradient Update:

$$\nabla_\theta J(\theta) = \mathrm{E}_t\left[\nabla_\theta \log \pi_\theta\left(a_t|s_t\right)\cdot A_t\right] \tag{7}$$

Advantage Estimation:

$$A_t = R_t - V(s_t) \tag{8}$$

Critic Loss:

$$L_{critic} = \left(R_t - V(s_t)\right)^2 \tag{9}$$

When π_θ is Policy with parameters θ
R_t is Return at timestep t
A_t is the Advantage function

In addition, the working principles of the Proximal Policy Optimization (PPO) algorithm applied to PV MPPT are formulated in Equations 10-13. These equations define the clipped surrogate objective function, the policy probability ratio, the value function loss, and the entropy bonus, which collectively enhance training stability and ensure robust policy updates under varying irradiance and temperature conditions [57-59].

Clipped Surrogate Objective:

$$L^{CLIP}(\theta) = \mathrm{E}_t\left[\min\left(r_t(\theta)A_t, \mathrm{clip}\left(r_t(\theta), 1-\epsilon, 1+\epsilon\right)A_t\right)\right] \tag{10}$$

Probability Ratio:

$$r_t(\theta) = \frac{\pi_\theta\left(a_t|s_t\right)}{\pi_{\theta \mathrm{old}}\left(a_t|s_t\right)} \tag{11}$$

Value Function Loss:

$$L^{VF} = \left(V(s_t) - R_t\right)^2 \tag{12}$$

Entropy Bonus(optional)

$$L^{entropy} = -\beta \cdot \mathrm{E}_t\left[\mathrm{Entropy}\left(\pi_\theta\left(\cdot|s_t\right)\right)\right] \tag{13}$$

When ϵ is Clipping threshold (typically 0.1-0.3)
β is the Entropy coefficient to encourage exploration

A flowchart of the A2C algorithm MPPT Tracking Process is shown in Figure 9, and the Flowchart PPO algorithm Tracking Process is shown in Figure 10.

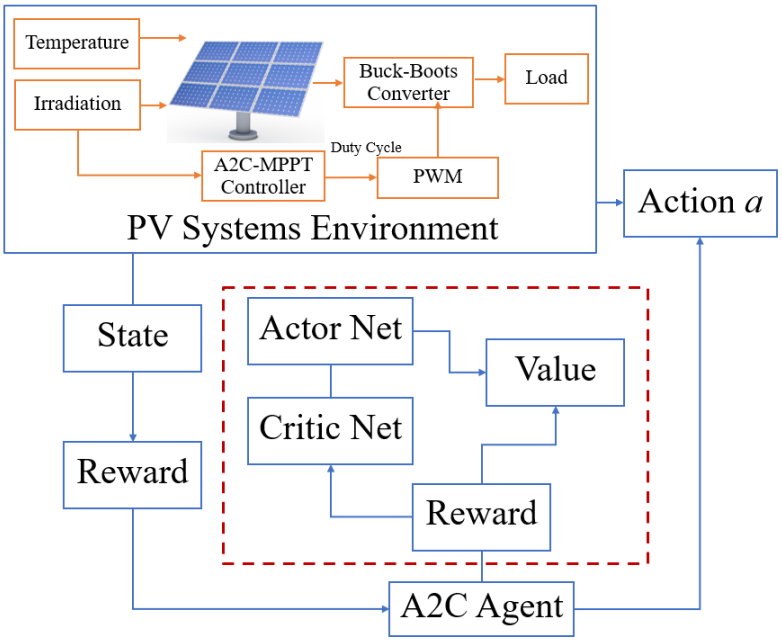

Figure 9.
Flowchart of the A2C algorithm MPPT Tracking Process.

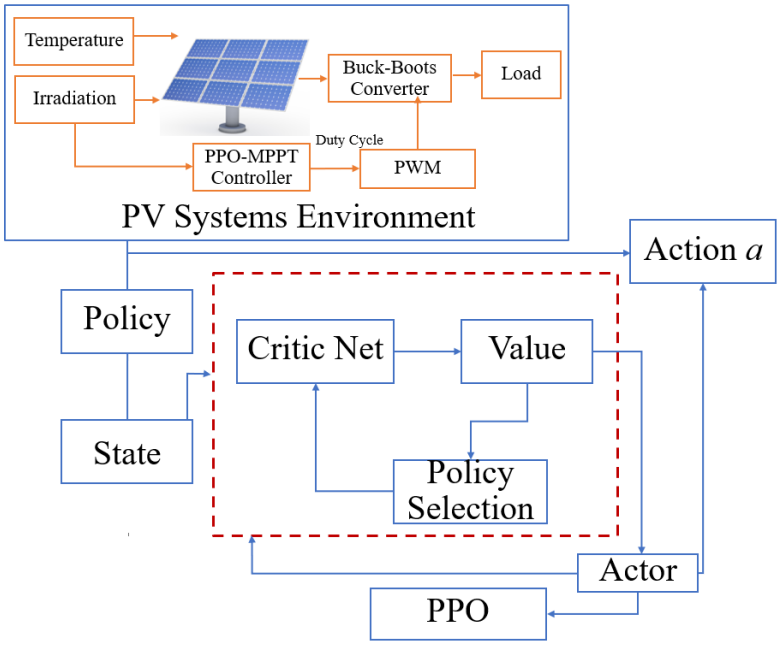

Figure 10.
Flowchart the PPO algorithm Tracking Process.

4. Results and Discussion

4.1. Comparison of the Mean Reward Per Episode Between the PPO and A2C Algorithms

To evaluate the effectiveness of DRL algorithms for MPPT in PV systems, this study implemented and trained two state-of-the-art DRL algorithms: PPO and A2C. Both algorithms were applied to control a Buck-Boost converter to maximize the power output of the PV system under dynamic environmental conditions. The performance of each algorithm was assessed based on the mean reward obtained during training episodes, which reflects the agent's ability to accurately and efficiently track the maximum power point. The experimental results are illustrated in Figure 11, which presents the smoothed mean reward per episode for both PPO and A2C. The following section presents a comparative analysis of the training results.

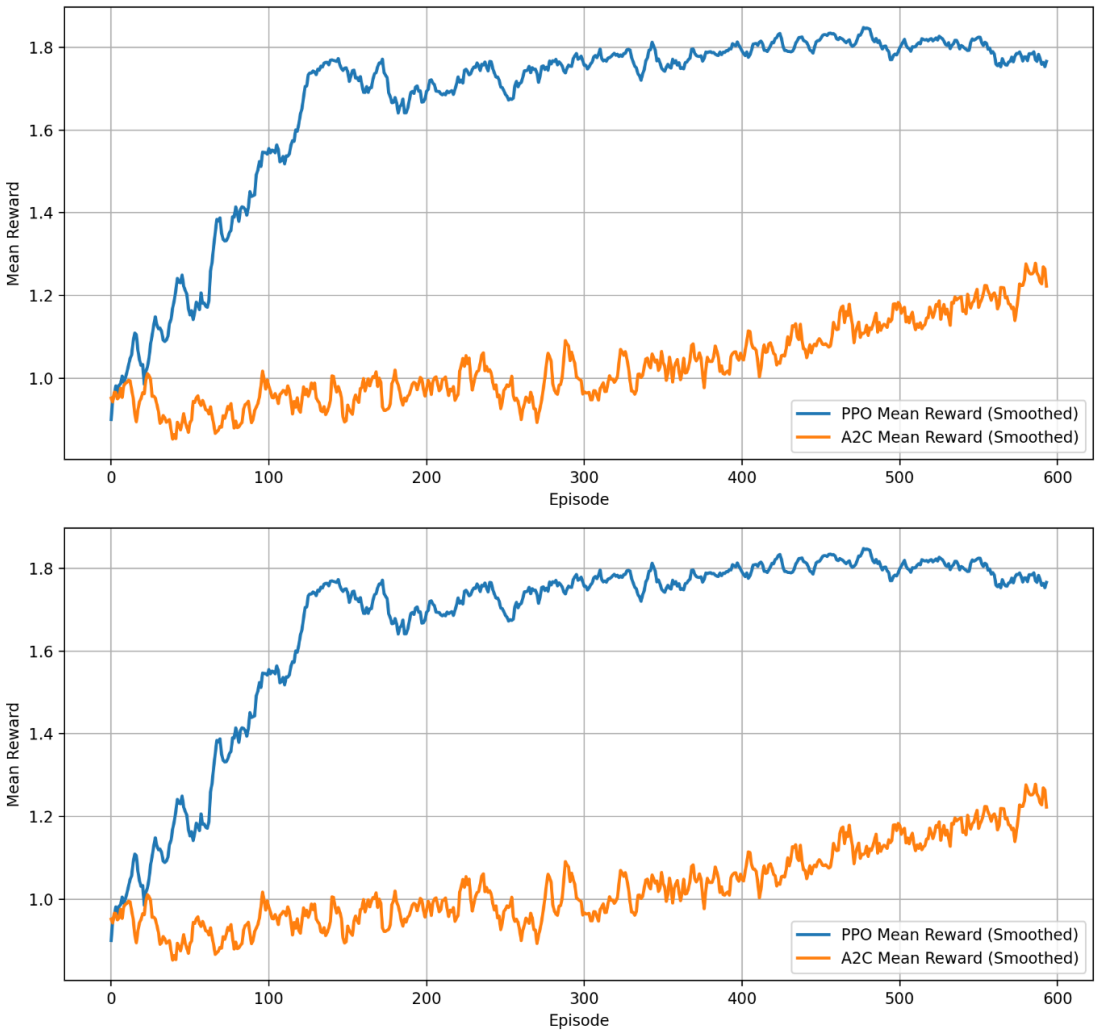

Figure 11.
Mean Reward per Episode of the PPO and A2C.

Figure 11 presents the comparison of the mean reward per episode between the PPO and A2C algorithms, applied to the MPPT task in a photovoltaic system using a Buck-Boost converter. The x-axis represents the number of training episodes, while the y-axis shows the smoothed mean reward, which reflects the performance of the agent in each episode. The blue curve corresponds to the PPO algorithm, and the orange curve represents the A2C algorithm. From the graph, it is evident that the PPO algorithm achieves faster convergence and higher performance compared to A2C. Specifically, PPO's reward increases rapidly within the first 100 episodes and stabilizes at a higher value of around 1.8. In contrast, A2C demonstrates slower learning progress and reaches a lower average reward of around 1.2. These results indicate that PPO performs better in identifying and tracking the maximum power point, making it more suitable for MPPT control in photovoltaic systems under the given simulation conditions.

4.2. Performance Comparison of PPO and A2C Algorithms under Varying Irradiance and Temperature Conditions

To further evaluate the performance and control behavior of the PPO and A2C algorithms for MPPT in PV systems, heatmaps of duty cycle outputs were generated under varying temperature and irradiance conditions. The goal was to observe how each algorithm adjusts the duty cycle of the buck-boost converter to maintain optimal power extraction in response to environmental changes. The figures below illustrate the variation in duty cycle values as predicted by each algorithm across different irradiance levels (200-1,000 W/m²) and PV cell temperatures (25-75°C).

The heatmap of the PPO algorithm demonstrates consistent behavior, where the duty cycle remains fixed at approximately 0.80 across all temperature and irradiance conditions. This indicates that PPO has learned a stable policy that generalizes well to different environmental states but may lack adaptability in fine-tuning the response under varying inputs. In contrast, the heatmap of the A2C algorithm exhibits a more dynamic adjustment of the duty cycle. At lower irradiance and temperature values, the duty cycle varies significantly, increasing gradually as irradiance and temperature rise. The duty cycle values reach 0.80 under high irradiance and temperature conditions, showing A2C's sensitivity and responsiveness to environmental variations. Figure 12 shows a heatmap of the duty cycle generated by the A2C algorithm under varying irradiance and temperature conditions, and Figure 13 shows a heatmap of the duty cycle generated by the PPO algorithm under varying irradiance and temperature conditions.

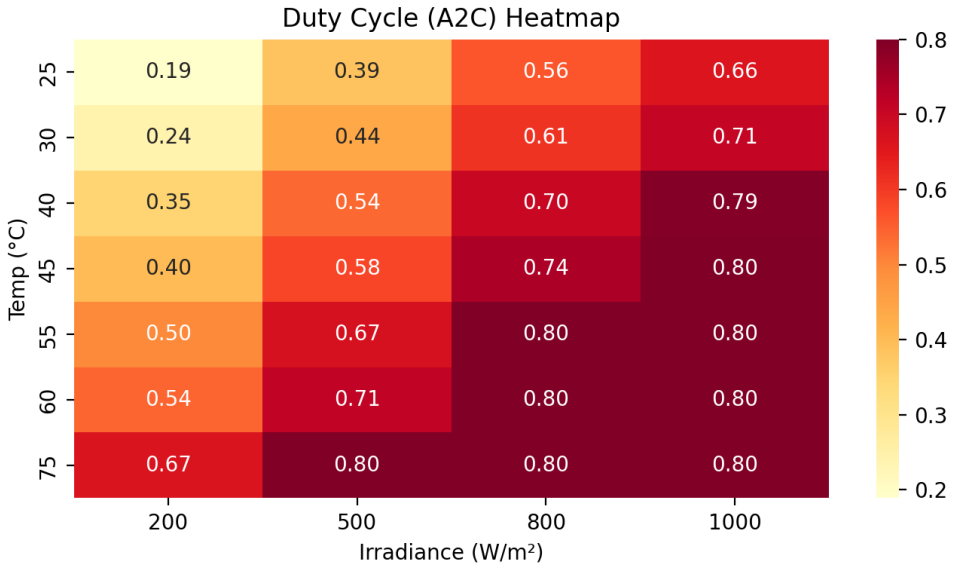

Figure 12.
Heatmap of Duty Cycle Generated by A2C algorithm under Varying Irradiance and Temperature Conditions.

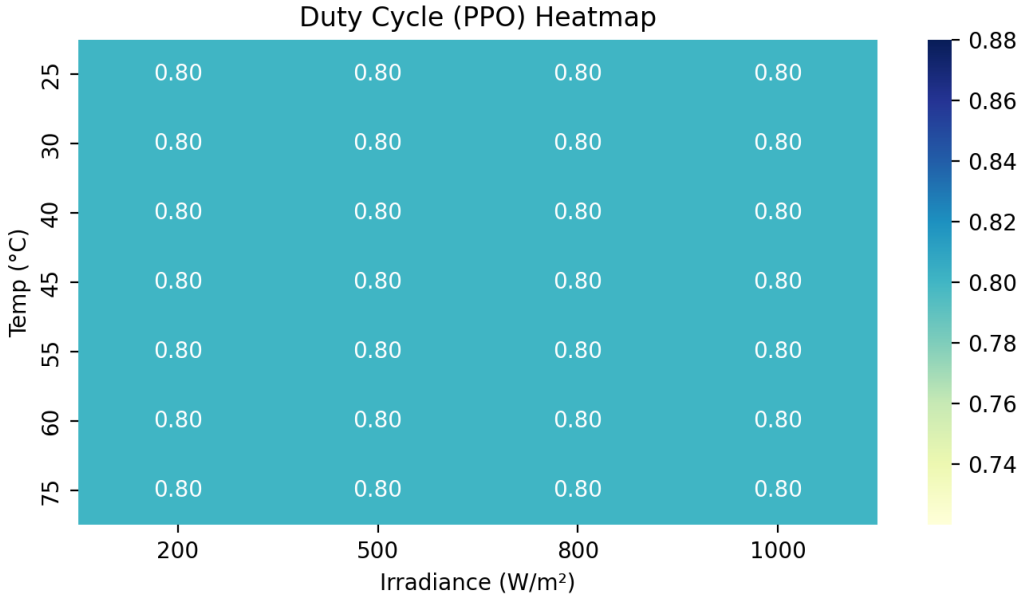

Figure 13.
Heatmap of Duty Cycle Generated by PPO algorithm under Varying Irradiance and Temperature Conditions.

These results suggest that while PPO provides a more stable and possibly conservative control policy, A2C demonstrates better adaptability in response to fluctuating environmental conditions, though it may be more complex to stabilize. Therefore, the choice between the two algorithms should consider the trade-off between stability and adaptability depending on the application requirements.

To comprehensively assess the control behavior and energy harvesting performance of PPO-DRL and A2C-DRL algorithms for MPPT in photovoltaic systems via a Buck-Boost converter, this study analyzes the variation of duty cycles and corresponding power output under diverse environmental conditions. Heatmaps were generated to visualize the output of both algorithms with respect to changes in irradiance (200–1,000 W/m²) and temperature (25-75°C). The objective is to evaluate how effectively each algorithm responds to external factors and maximizes power output under dynamic PV system conditions.

The heatmaps for duty cycle reveal distinct differences in control strategy. PPO outputs a constant duty cycle of approximately 0.80 across all irradiance and temperature levels, indicating that it has learned a fixed policy that generalizes well but lacks responsiveness to environmental variation. Conversely, A2C adjusts the duty cycle dynamically, with values ranging from 0.19 to 0.80 depending on irradiance and temperature. This shows that A2C is more reactive and adapts its control to optimize for changing conditions.

When comparing power output, the PPO algorithm consistently produces high power, especially under high irradiance levels, achieving values up to 340 W. Its output decreases steadily with rising temperature and lower irradiance, yet it remains

relatively stable and reliable. In contrast, the A2C algorithm exhibits weaker performance at low irradiance and low temperature (e.g., only 0.74 W at 200 W/m² and 25°C), but it gradually improves and matches PPO performance under higher irradiance. At full sunlight (1,000 W/m²), both algorithms reach the maximum rated output of 340 W. The heatmap of power output generated by the A2C and PPO algorithms under varying irradiance and temperature conditions is shown in Figures 14 and Figure 15.

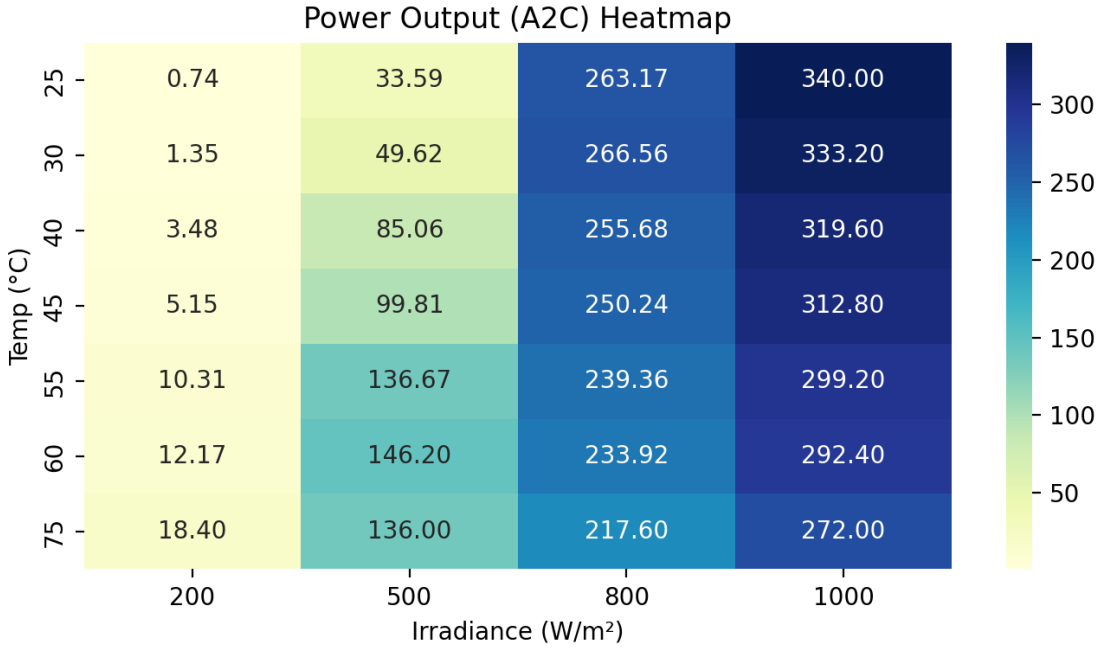

Figure 14.
Heatmap of Power Output Generated by A2C algorithm under Varying Irradiance and Temperature Conditions.

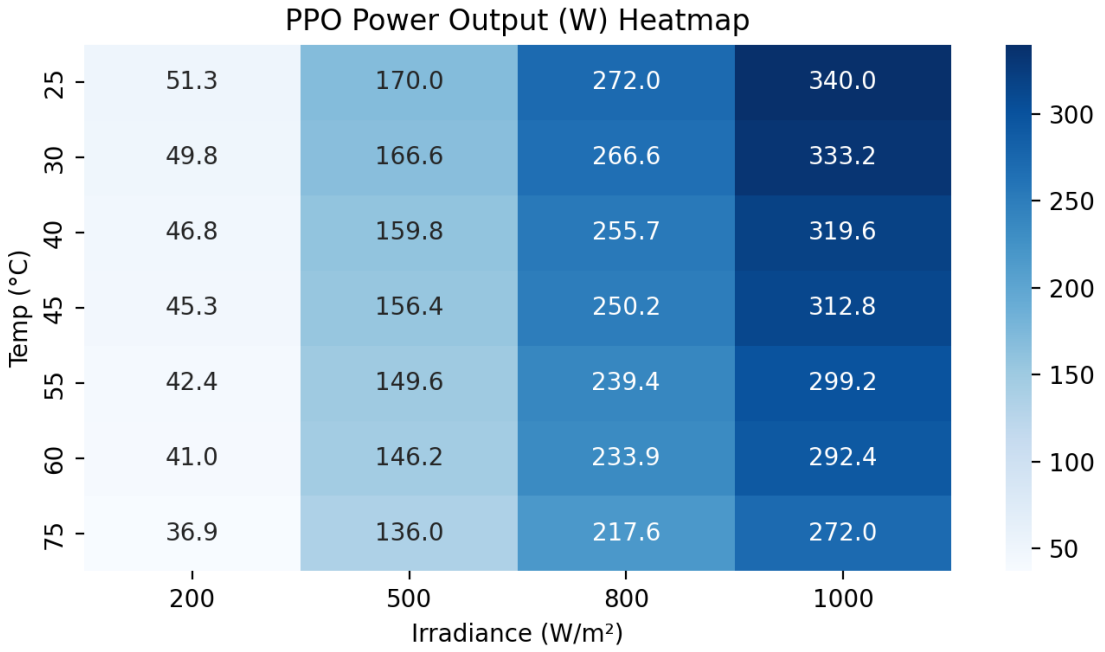

Figure 15.
Heatmap of Power Output Generated by PPO algorithm under Varying Irradiance and Temperature Conditions.

In summary, PPO shows superior stability and maintains high power output across a wide range of conditions with a fixed control approach. A2C, while initially less efficient under weak conditions, demonstrates a stronger adaptive capacity. Therefore, PPO may be preferred for robust, consistent environments, whereas A2C offers greater flexibility for rapidly changing operating conditions.

4.3. Comparison of Power, Voltage, and Current Output between PPO and A2C Algorithms under Test Conditions

This experiment compares the performance of the PPO and A2C algorithms in MPPT by analyzing power, voltage, and current outputs across 28 test cases. The black line represents the ideal output, while the red (PPO) and blue (A2C) lines show the actual outputs.

From the power output graph (Figure 16a), it is evident that the PPO algorithm closely follows the reference (best) values in most test cases, particularly in mid to high irradiance scenarios. A2C, while improving gradually, tends to lag behind PPO in the early test cases but converges better as irradiance increases.

In the voltage output graph (Figure 16b), PPO maintains a relatively consistent voltage profile close to the reference, whereas A2C shows significant underestimation in lower test cases and slightly overshoots in later ones. This indicates that PPO maintains more stable control over voltage adjustment through the buck-boost converter.

Regarding the current output (Figure 16c), both algorithms produce similar trends, with PPO slightly outperforming A2C in certain regions. However, the current output remains more stable and accurate under PPO, aligning better with the reference current at higher irradiance.

In summary, PPO demonstrates more accurate and stable MPPT performance across voltage, current, and power dimensions compared to A2C. The results suggest that PPO is more effective in matching the system's optimal operating point, particularly under dynamic and varied conditions. Figure 16 shows the comparison of power, voltage, and current output between PPO and A2C algorithms under test conditions.

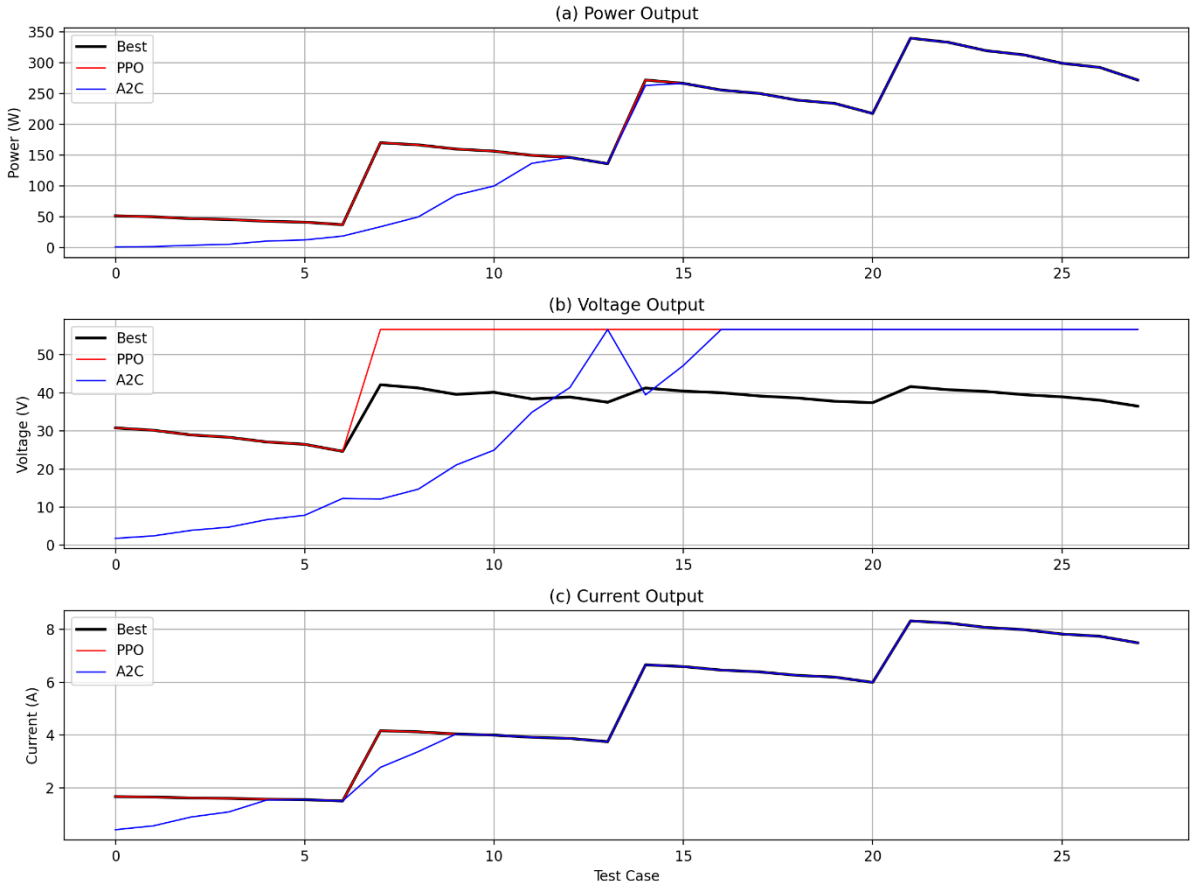

Figure 16.
Comparison of Power, Voltage, and Current Output between PPO and A2C Algorithms under Test Conditions.

5. Conclusion

This study aimed to compare the performance of two prominent DRL algorithms: PPO and A2C in the application of MPPT for PV systems using a Buck-Boost converter. Through simulation-based experiments under various irradiance (200–1,000 W/m²) and temperature (25-75°C) conditions, both algorithms were evaluated based on their control behavior (duty cycle) and energy extraction performance (power output, voltage, and current). The results show that the PPO algorithm consistently achieved a high and stable duty cycle of approximately 0.80, regardless of environmental conditions. This indicates that PPO successfully learned a generalized policy for power maximization, offering robustness and ease of implementation. Its power output remained close to the theoretical maximum (340 W) in most cases, and its voltage and current outputs aligned well with the expected optimal values. This reflects PPO's capability to maintain system stability and efficiency, particularly under changing environmental conditions. In contrast, the A2C algorithm demonstrated more dynamic behavior, adjusting the duty cycle in response to irradiance and temperature variations. While this adaptability is beneficial in principle, A2C struggled to produce high power output under low irradiance conditions. Its voltage output showed significant deviations from the optimal reference, especially in early test cases. However, under higher irradiance levels, A2C's performance gradually improved and matched PPO's in later test scenarios. When comparing real-time output across

28 test cases, PPO outperformed A2C in terms of power accuracy, voltage regulation, and current stability. While A2C displayed promising adaptability, it requires more fine-tuning and training stability to achieve results comparable to PPO. In conclusion, the PPO algorithm is better suited for MPPT control in PV systems that operate under dynamic or uncertain environmental conditions. It provides a strong balance between learning efficiency, stability, and energy extraction performance. A2C remains a viable option for scenarios where adaptive behavior is more critical, but it may require additional training optimization. Future work could explore hybrid approaches or fine-tuned reward structures to enhance A2C's practical performance.

References

[1] International Energy Agency, *Renewables – global energy & CO_2 status report 2019*. Paris: IEA, 2019.

[2] United Nations, "Renewable energy – powering a safer future. UN," Retrieved: https://www.un.org/en/climatechange/raising-ambition/renewable-energy, 2022.

[3] International Renewable Energy Agency (IRENA), *World energy transitions outlook 2023*. Abu Dhabi: IRENA, 2023.

[4] U.S. Department of Commerce, *Energy resource guide - thailand - renewable energy*. Washington, D.C: U.S. Department of Commerce, 2021.

[5] Thailand Ministry of Energy, *Alternative energy development plan: AEDP 2018*. Bangkok: Thailand Ministry of Energy, 2019.

[6] International Renewable Energy Agency (IRENA), *Renewable energy market analysis: Southeast Asia*. Abu Dhabi: International Renewable Energy Agency, 2022.

[7] B. Santos, "Thailand introduces FIT scheme for solar, storage. PV Magazine," Retrieved: https://www.pv-magazine.com/2022/10/31/thailand-introduces-fit-scheme-for-solar-storage/, 2022.

[8] Thailand Board of Investment (BOI), *Thailand: Alternative energy industry*. Bangkok, Thailand: Thailand Board of Investment, 2014.

[9] F. Watson and Williams, *Thailand powers Up: New renewable energy incentives and opportunities in 2024*. Bangkok, Thailand: Watson Farley & William, 2024.

[10] Solar Energy Industries Association, *About solar energy*. Washington, D.C., USA: Solar Energy Industries Association, 2023.

[11] U.S. Energy Information Administration, *olar explained—photovoltaics and electricity. U.S. Department of Energy*. Washington, D.C., USA: Solar explained—photovoltaics and electricity. U.S. Department of Energy, 2023.

[12] American Chemical Society, *How a solar cell works*. Washington, D.C., USA: American Chemical Society, 2014.

[13] Valur, "How solar panels generate electricity. Valur," Retrieved: https://learn.valur.com/solar-panels/. [Accessed 2022.

[14] Fraunhofer Institute for Solar Energy Systems ISE, "Photovoltaics report. Fraunhofer ISE," Retrieved: https://www.ise.fraunhofer.de/en/publications/studies/photovoltaics-report.html, 2024.

[15] National Renewable Energy Laboratory (NREL), *Best practices for PV system performance*. Golden, CO: NREL, 2019.

[16] Solar Energy International, "PV module temperature technical note," Retrieved: https://www.solarenergy.org/wp-content/uploads/2015/02/PV-Module-Temperature.pdf, 2015.

[17] P. Boonraksa, T. Booraksa, and B. Marungsri, "Comparison of the cuk, sepic, and zeta converters circuit efficiency for improving the maximum power point tracking on photovoltaic systems," in *2021 International Conference on Power, Energy and Innovations (ICPEI)*, 2021: IEEE, pp. 150-154.

[18] D. Verma, S. Nema, A. Shandilya, and S. K. Dash, "Maximum power point tracking (MPPT) techniques: Recapitulation in solar photovoltaic systems," *Renewable and Sustainable Energy Reviews*, vol. 54, pp. 1018-1034, 2016. https://doi.org/10.1016/j.rser.2015.10.068

[19] S. R. Bull, "Renewable energy today and tomorrow," *Proceedings of the IEEE*, vol. 89, no. 8, pp. 1216-1226, 2001. https://doi.org/10.1109/5.942924

[20] M. A. G. De Brito, L. Galotto, L. P. Sampaio, G. d. A. e Melo, and C. A. Canesin, "Evaluation of the main MPPT techniques for photovoltaic applications," *IEEE transactions on industrial electronics*, vol. 60, no. 3, pp. 1156-1167, 2012.

[21] H. Patel and V. Agarwal, "MATLAB-based modeling to study the effects of partial shading on PV array characteristics," *IEEE transactions on energy conversion*, vol. 23, no. 1, pp. 302-310, 2008. https://doi.org/10.1109/TEC.2007.909537

[22] T. Nagadurga, R. Devarapalli, and Ł. Knypiński, "Comparison of meta-heuristic optimization algorithms for global maximum power point tracking of partially shaded solar photovoltaic systems," *Algorithms*, vol. 16, no. 8, p. 376, 2023. https://doi.org/10.3390/a16080376

[23] A. Safari and S. Mekhilef, "Simulation and hardware implementation of incremental conductance MPPT with direct control method using cuk converter," *IEEE Transactions on Industrial Electronics*, vol. 58, no. 4, pp. 1154-1161, 2010. https://doi.org/10.1109/TIE.2010.2052282

[24] S. Jain and V. Agarwal, "Comparison of the performance of maximum power point tracking schemes applied to single-stage grid-connected photovoltaic systems," *IET Electric Power Applications*, vol. 1, no. 5, pp. 753-762, 2007. https://doi.org/10.1049/iet-epa:20060077

[25] M. A. Elgendy, B. Zahawi, and D. J. Atkinson, "Assessment of perturb and observe MPPT algorithm implementation techniques for PV pumping applications," *IEEE Transactions on Sustainable Energy*, vol. 3, no. 1, pp. 21-33, 2011. https://doi.org/10.1109/TSTE.2011.2168543

[26] S. K. Kollimalla and M. K. Mishra, "A novel adaptive P&O MPPT algorithm considering sudden changes in the irradiance," *IEEE Transactions on Energy Conversion*, vol. 29, no. 3, pp. 602-610, 2014. https://doi.org/10.1109/TEC.2014.2317110

[27] A. K. Abdelsalam, A. M. Massoud, S. Ahmed, and P. N. Enjeti, "High-performance adaptive perturb and observe MPPT technique for photovoltaic-based microgrids," *IEEE Transactions on Power Electronics*, vol. 26, no. 4, pp. 1010-1021, 2011. https://doi.org/10.1109/TPEL.2010.2048479

[28] N. K. Ali and T. Petrov, "Design and analysis of MPPT for PV system by perturb and observe algorithm," in *E3S Web of Conferences*, 2024, vol. 542: EDP Sciences, p. 01010.

[29] S. Chahar and D. K. Yadav, "Retrospection and investigation of ANN-based MPPT technique in comparison with soft computing-based MPPT techniques for PV solar and wind energy generation system," *International Journal of Mathematical Modelling and Numerical Optimisation*, vol. 14, no. 1-2, pp. 69-83, 2024. https://doi.org/10.1504/ijsi.2023.10055513

[30] K. Ishaque, Z. Salam, M. Amjad, and S. Mekhilef, "An improved particle swarm optimization (PSO)–based MPPT for PV with reduced steady-state oscillation," *IEEE transactions on Power Electronics*, vol. 27, no. 8, pp. 3627-3638, 2012. https://doi.org/10.1109/TPEL.2012.2185713

[31] S. K. Roy, S. Hussain, and M. A. Bazaz, "Implementation of MPPT technique for solar PV system using ANN," presented at the 2017 Recent Developments in Control, Automation & Power Engineering (RDCAPE), Noida, India. https://doi.org/10.1109/RDCAPE.2017.8358286, 2017.

[32] P. Boonraksa, K. Palachai, P. Chotipintu, T. Chaisa-Ard, T. Boonraksa, and B. Marungsri, "Design and simulation of MPPT for PV systems using ANFIS algorithm," in *2023 International Electrical Engineering Congress (iEECON)*, 2023: IEEE, pp. 425-428.

[33] K. Ramu *et al.*, "Smart solar power conversion: Leveraging deep learning MPPT and hybrid cascaded H-bridge multilevel inverters for optimal efficiency," *Biomedical Signal Processing and Control*, vol. 105, p. 107582, 2025. https://doi.org/10.1016/j.bspc.2025.107582

[34] L. Avila, M. De Paula, M. Trimboli, and I. Carlucho, "Deep reinforcement learning approach for MPPT control of partially shaded PV systems in Smart Grids," *Applied Soft Computing*, vol. 97, p. 106711, 2020. https://doi.org/10.1016/j.asoc.2020.106711

[35] E. Artetxe, J. Uralde, O. Barambones, I. Calvo, and I. Martin, "Maximum power point tracker controller for solar photovoltaic based on reinforcement learning agent with a digital twin," *Mathematics*, vol. 11, no. 9, p. 2166, 2023. https://doi.org/10.3390/math11092166

[36] T. Dong, "Maximum power point tracking control for photovoltaic battery systems using deep q network algorithm," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 3, pp. 1288–1296, 2021.

[37] M. Khan, A. S. Khan, S. Sardar, and M. I. Khan, "A photovoltaic system maximum power point tracking techniques comparison under variable atmospheric condition," *Zhongguo Kuangye Daxue Xuebao*, vol. 29, no. 3, pp. 291-303, 2024.

[38] PV Education University of Cambridge, "Effects of irradiance and temperature," Retrieved: https://www.pveducation.org/pvcdrom/solar-cell-operation/effects-of-irradiance-and-temperature, 2023.

[39] Penn State University, "Irradiance and PV performance optimization," Retrieved: https://www.energy.psu.edu/, 2023.

[40] McFadyen, "Photovoltaic (PV) - electrical calculations, myElectrical," Retrieved: https://myelectrical.com/notes/entryid/225/photovoltaic-pv-electrical-calculations. [Accessed Apr. 27, 2025], 2013.

[41] T. Markvart and L. Castañer, *Practical handbook of photovoltaics: Fundamentals and applications*. Netherlands: Elsevier, 2003.

[42] J. A. Nelson, *The physics of solar cells*. World Scientific Publishing Company. https://doi.org/10.1142/p276, 2003.

[43] L. Bhukya, N. R. Kedika, and S. R. Salkuti, "Enhanced maximum power point techniques for solar photovoltaic system under uniform insolation and partial shading conditions: A review," *Algorithms*, vol. 15, no. 10, p. 365, 2022. https://doi.org/10.3390/a15100365

[44] T. Esram and P. L. Chapman, "Comparison of photovoltaic array maximum power point tracking techniques," *IEEE Transactions on Energy Conversion*, vol. 22, no. 2, pp. 439-449, 2007. https://doi.org/10.1109/TEC.2006.874230

[45] B. Subudhi and R. Pradhan, "A comparative study on maximum power point tracking techniques for photovoltaic power systems," *IEEE Transactions on Sustainable Energy*, vol. 4, no. 1, pp. 89-98, 2012.

[46] Peterson, "Analysis of four DC-DC converters in equilibrium, All About Circuits," Retrieved: https://www.allaboutcircuits.com/technical-articles/analysis-of-four-dc-dc-converters-in-equilibrium/, 2021.

[47] M. Glavic, "Deep Reinforcement learning for electric power system control and related problems: A short review and perspectives," *Annual Review of Control*, vol. 48, pp. 22–35, 2019. https://doi.org/10.1016/j.arcontrol.2019.05.005

[48] P. Kofinas, S. Doltsinis, A. I. Dounis, and G. A. Vouros, "A reinforcement learning approach for MPPT control method of photovoltaic sources," *Renewable Energy*, vol. 108, pp. 461–473, 2017. https://doi.org/10.1016/j.renene.2017.02.059

[49] S. van Der Walt, "Understanding actor-critic methods, Medium," Retrieved: https://medium.com/data-science/understanding-actor-critic-methods-931b97b6df3f, 2018.

[50] G. Huang, X. Zhou, and Q. Song, "A deep reinforcement learning framework for dynamic portfolio optimization: Evidence from China's stock market," 2025. https://doi.org/10.21203/rs.3.rs-6415851/v1

[51] M.-Y. Day, C.-Y. Yang, and Y. Ni, "Portfolio dynamic trading strategies using deep reinforcement learning," *Soft Computing*, vol. 28, no. 15, pp. 8715-8730, 2024. https://doi.org/10.1007/s00500-023-08973-5

[52] U. Saha, A. Jawad, S. Shahria, and A. H.-U. Rashid, "Proximal policy optimization-based reinforcement learning approach for DC-DC boost converter control: A comparative evaluation against traditional control techniques," *Heliyon*, vol. 10, no. 18, p. e37823, 2024.

[53] C. Cui, N. Yan, and C. Zhang, "An intelligent control strategy for buck DC-DC converter via deep reinforcement learning," *arXiv preprint arXiv:2008.04542*, 2020. https://doi.org/10.48550/arXiv.2008.04542

[54] B. C. Phan, Y.-C. Lai, and C. E. Lin, "A deep reinforcement learning-based MPPT control for PV systems under partial shading condition," *Sensors*, vol. 20, no. 11, p. 3039, 2020. https://doi.org/10.3390/s20113039

[55] R. Wongsathan, "Integrated neural network-based MPPT and ant colony optimization-tuned PI bidirectional charger-controller for PV-powered motor-pump system," *Engineering and Applied Science Research*, vol. 51, no. 5, pp. 605-617, 2024.

[56] R. Srinivasan and C. Ramalingam Balamurugan, "Deep neural network based MPPT algorithm and PR controller based SMO for grid connected PV system," *International Journal of Electronics*, vol. 109, no. 4, pp. 576-595, 2022. https://doi.org/10.1080/00207217.2021.1914192

[57] V. Mnih *et al.*, "Asynchronous methods for deep reinforcement learning," in *International Conference on Machine Learning*, 2016: PmLR, pp. 1928-1937.

[58] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017. https://doi.org/10.48550/arXiv.1707.06347

[59] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, 2nd ed. Cambridge, MA: MIT Press, 2018.