Analysis of Deep Reinforcement Learning and Conventional MPPT Control under Fast-Changing Irradiance

Ameze Big-Alabo[1,2]

[1]E. ON Energy Research Center, RWTH Aachen University, Germany
[2]Department of Electrical/Electronic Engineering, Faculty of Engineering, University of Port Harcourt, PMB 5323, Port Harcourt, Rivers Nigeria

*Corresponding Author: [1,2]Ameze Big-Alabo (Email: ameze.big-alabo@eonerc.rwth-aachen.de)

(Received: 16-October-2025; accepted: 12-January-2026; published: 31-March-2026)
http://dx.doi.org/10.55579/jaec.2026101.522

Abstract. *This paper investigates a deep reinforcement learning (DRL) maximum power point tracking (MPPT) strategy for a photovoltaic (PV) boost converter system using the Deep Deterministic Policy Gradient (DDPG) algorithm. The proposed controller is designed to overcome limitations of conventional perturb-and-observe (P&O) methods combined with PID control, particularly under non-uniform irradiance and load impedance variations. The DDPG agent is trained offline and learns the nonlinear mapping between PV electrical states and duty-cycle control, while explicitly accounting for DC-link voltage regulation. A comparative performance evaluation is conducted against conventional P&O–PID and a PSO–NN MPPT scheme under fast irradiance transients, and varying load conditions. Simulation results show that while the P&O–PID and PSO–NN controller achieves marginally higher instantaneous PV power under rapid irradiance changes, the proposed DRL controller provides superior DC-bus voltage regulation and sustained stability within a simulated irradiance levels ranging from 1000 W/m^2 to 400 W/m^2. This reflects a trade-off between energy extraction and system-level stability but overall, the DRL MPPT approach demonstrates improved robustness under fast environmental transients and realistic operating conditions, highlighting its suitability for standalone DC microgrid and advanced PV power conversion applications.*

Keywords: *Deep Reinforcement Learning, Maximum Power Point Tracking, Perturb and Observe, Particle swarm optimization – Neural Network.*

1. Introduction

Solar photovoltaic, PV systems is a type of renewable energy source that is clean, noiseless and eco-friendly and free of GHG emissions during operation. It is one of the cheapest sources of electricity generation due to its reduced average levelized cost of electricity from the year 2010 to 2022 [1]. However, a major drawback of PV system is its low energy conversion efficiency and thus, the need for improvements in PV technology that will increase PV output [2]. It has been suggested that advanced PV technologies can help to increase energy output, making solar PV more efficient and scalable [3]. The PV system exhibits a nonlinear current–voltage characteristic that is dependent on temperature and solar irradiance. The operating point of the PV array directly determines the amount of power extracted.

It is significantly affected by environmental conditions as well as the nature of the connected load. This can cause the PV system to operate away from its maximum power point, MPP that can result to substantial energy losses. Therefore, MPPT is essential to ensure that PV systems consistently deliver optimal power under dynamic environmental scenarios [4].

A typical PV system comprises of solar cells connected in series and parallel to yield a desired power output. The solar cell can be represented by a single diode model and the governing equation is stated in Equation 1. This describes the current - voltage I-V characteristics of a photovoltaic, PV cell [5].

$$I = I_l - I_S \left(\exp\left(\frac{V + IR_s}{mn_{se}V_T} \right) - 1 \right) - \frac{V + IR_s}{R_{sh}} \tag{1}$$

Where I is the current and V is the voltage, I_l is the light induced current, I_S is the diode dark saturation current, m is the diode quality factor, R_s is the series resistance, R_{sh} is the shunt resistance, n_{se} is the number of series-connected cells in the PV module, V_T is the thermal voltage. The induced current in Equation 1 is affected by changes in the irradiance and temperature [6]. The dynamics of solar panels that captures the interdependencies between current, voltage, power, solar irradiation, and temperature has been considered extensively [7]. The efficiency of the solar PV system depends upon its power output and this output is dependent on the panel's output voltage and current. Thus, the maximum power point MPP is the operating point at which the PV current and voltage is maximum. Hence, maintaining the MPP via MPPT strategies is essential for maximizing the power delivered by PV systems [8, 9].

Various MPPT control strategies have been developed to optimize the performance of PV systems. MPPT control strategies such as the P&O [10] and incremental conductance [11] offers low-cost solutions but may exhibit steady-state oscillations and slow convergence under rapidly changing environmental and partial shading conditions. To overcome these drawbacks, artificial intelligence such as fuzzy logic–based methods [12,13], can help to improve adaptability. However, it often relies on heuristic tuning that is parameter sensitive. With the evolution of biologically inspired algorithms, neural networks and metaheuristic optimizers, such as particle swarm optimization [14, 15], Genetic Algorithm [16, 17], and Differential Evolution [18], have been introduced due to its global search capabilities. However, these approaches are inherently iterative optimizers rather than learning based controllers. When irradiation conditions change, the P–V curve shifts, requiring the algorithm to reinitialize its population and run the search again. This makes them relatively slow to track dynamic variations in the Global MPP, GMPP as they cannot be pre-trained over a range of operating conditions to directly generalize new scenarios.

Also, they do not undergo an offline training stage but execute the optimization process online. This involves repeated population evolution and fitness evaluations that leads to high computational complexity, making them unsuitable for real-time MPPT under rapidly changing irradiance conditions. Recent studies show that Artificial Intelligence – based controllers achieve faster and more reliable GMPP tracking under variable irradiance compared to classical MPPT techniques [19]. Moreover, an ANN-based MPPT scheme was designed to address the high complexity and computational demands of traditional model predictive control, MPC methods [20]. A stacked LSTM controller was applied to a 100 kWp grid-tied system and results showed over 99% tracking efficiency achieved under dynamic irradiance [21]. These findings confirm the growing suitability of intelligent control for PV applications, particularly in dynamic conditions.

Despite these advances, Classical MPPT algorithms with PID controllers remain widely adopted in practice due to their ease of implementation and established industrial acceptance as highlighted in the literature [22]. The PID helps to ensure accurate tracking of the reference voltage generated by the MPPT controller. In contrast, the DRL controller directly maps PV system measurements to the converter duty cycle thereby unifying the roles of both the MPPT and PID. This single learned policy achieves maximum power extraction and stable converter operation without the need for manual tuning and thus simplifies the control architecture while maintaining performance. Most of the reviewed literatures have only considered the comparisons of DRL with Conventional or Heuristic approaches for MPPT control. But scarcely is there any literature that considered the comparison for both MPPT control and voltage tracking simultaneously. This motivates a systematic comparison between conventional MPPT schemes, such as P&O combined with a PID-controlled boost converter and a DRL controller. The DRL directly generates the duty cycle without relying on either an outer-loop MPPT algorithm or an inner-loop PID regulator. Such a comparison allows quantification of the performance gains achievable when the entire MPPT and control process is handled by DRL. Insights into the specific scenarios where DRL offers significant benefits over conventional control methods are presented in this study.

Contributions of study

This study advances the state-of-the-art by reframing DRL MPPT as a system-level power management problem, in which PV power extraction and DC-link voltage stability are addressed simultaneously within a single control framework. Specifically, the proposed DDPG controller incorporates duty-cycle constraints and power–load balance considerations directly into the learning objective, enabling the agent to implicitly regulate the DC-bus voltage without relying on auxiliary voltage controllers. A key contribution of this work is the explicit comparative evaluation of DC-link stability under severe irradiance reduction. Through detailed simulations, it is demonstrated that the proposed DRL controller maintains stable DC – bus voltage operation down to an irradiance level of 400 W/m^2, whereas a PSO–ANN-based MPPT controller exhibits significant DC-link voltage deviation under the same conditions. By demonstrating robust DC-bus regulation under dynamic load and low-irradiance conditions, this study extends DRL MPPT beyond conventional power-tracking performance metrics and establishes its effectiveness as a unified control solution for PV systems operating under realistic and challenging conditions.

2. Theoretical Framework for Conventional and DRL MPPT control

2.1. Conventional MPPT control

The conventional P&O algorithm achieves MPPT via perturbation of the duty cycle of the DC–DC converter and observes the corresponding change in PV output power. The direction of perturbation is maintained with an increase in power and is reversed with a decrease in power. This process converges around the maximum power point, MPP. The classical P&O with fixed step size has inherent limitations: a large step size accelerates convergence but causes oscillations around the MPP, whereas a small step size reduces oscillations at the cost of slower convergence. To mitigate this, several studies have proposed Variable Step-Size, VSS extensions of P&O, in which the perturbation magnitude adapts dynamically according to operating conditions. Although VSS-P&O is not as widely implemented in real-world as classical P&O, it serves as a useful academic benchmark to highlight the trade-offs in tracking speed, accuracy, and adaptability, and therefore provides a fair basis for comparison with DRL approach.

The VSS-P&O approach adapts the perturbation size based on the slope, dP/dV of the Power – Voltage, P –V curve. Large step-size is applied when far from MPP to speed convergence, while step-size decreases as the system tends towards MPP to minimize oscillations. This approach can be expressed mathematically as stated in Equation 2 [23]

$$D_{VSS} = M \times \frac{dP}{dV} \tag{2}$$

where D_{VSS} is the duty cycle increment and M is the scaling factor, $\frac{dP}{dV}$ shows how power changes when voltage is slightly perturbed. This approach is easy to implement and requires low computational demand. It tracks better with improved convergence under moderate irradiance fluctuations compared to the classical fixed P&O [24].

2.2. Deep Reinforcement Learning Approach

Reinforcement learning is a machine learning technique that is typically formulated as a Markov Decision Process (MDP), where the agent learns to make sequential decisions by interacting with an environment. Sequential implies that each action taken by the agent not only yields an immediate reward but also influences the next state of the environment that affects future rewards. In the context of MPPT, this dependency is formalized by the MDP whereby the adjustment of the duty cycle at a given moment affects subsequent PV voltage and current responses such that the controller must learn a policy that accounts for these temporal dependencies rather than treating each decision as isolated. The MDP is described by (s, a, P, r, γ) where 's' and 'a' represent state and action spaces respectively. The function P is the transition probability function of moving from one state to another following an action 'a'. Then 'r' is the non − negative reward function while $\gamma \in [0, 1]$ is the discount factor that modifies the value of future rewards [25]. When the agent takes an action, it receives feedback in the form of rewards or penalties. It therefore learns the optimal policy that guides its decision making for maximum cumulative expected rewards over time [26–28].

With respect to the PV system, the 's' is the observable state space vector comprising of PV measurements of voltage, current, and power, whereas 'a' corresponds to the duty cycle applied to the boost converter. In the simulation environment, the generated duty cycle is modeled as a continuous variable within the range [0.1, 0.9] but in a practical digital controller, this signal can be quantized according to the resolution of the pulse-width modulation, PWM hardware. Such quantization introduces only a negligible error relative to the MPPT precision requirements so that the continuous action assumption in simulation remains consistent with what can be achieved in a practical digital implementation. The reward 'r' is formulated to maximize power extraction and 'γ' ensures long-term efficiency in decision-making by the DRL agent. In the context of PV, long-term efficiency refers to maximizing cumulative energy output over time that accounts for challenges such as irradiance fluctuations. Whereas short-term actions could trap the system at local maxima or cause oscillations around the MPP. During training the DRL agent continuously observes the PV environment at each simulation time step to update its policy $\pi(s)$ with the objective of maximizing the cumulative expected reward. It can achieve stable MPPT, which refers to reliable convergence to the GMPP with minimal oscillations in duty cycle and power output. Achieving a stable MPPT is promoted through the reward structure and the choice of discount factor with the objective of maximizing the expected cumulative reward. This enables the agent to achieve stable MPPT, meaning it can converge reliably to the global maximum power point while minimizing oscillations around it and maintaining robustness under dynamic irradiance and load variations.

DRL is an extension of reinforcement learning that employs deep neural networks to approximate the policy and value function [29], enabling effective control in high-dimensional and nonlinear systems such as large-scale PV arrays. In this context, effectiveness refers to the controller's ability to track the MPP while maintaining stability under nonlinear conditions such as irradiance fluctuations and partial shading. High dimensionality arises from the system state space that includes electrical variables of PV voltage, current and power. Although the present study focuses on a 100 kW PV array with a single boost converter, the approach is scalable to larger systems where multiple converters and interacting subsystems increase both the state and action dimensionality. By demonstrating feasibility in the single-converter case, this work establishes a foundation for extending DRL MPPT approach to distributed converter architectures used in utility-scale PV plants. The DRL method is particularly suitable for MPPT in PV systems particularly for partial shading conditions, PSC because the deep neural networks allow the policy to adapt to nonlinear PV characteristics across diverse irradiance and temperature profiles without requiring an explicit system model. An example of DRL method is the Q-learning-based Global Flexible Power Point Tracking algorithm that helps to improve the performance of the PV system by dynamically adapting to changes due to PSC [30]. Unlike conventional MPPT approach, the DRL does not require prior PV array knowledge. Rather it explores the PV characteristics over time, thereby improving convergence and robustness.

There are various DRL algorithms that have been implemented for MPPT control in PV systems. These algorithms include Deep Q-Network, DQN [31], Proximal Policy Optimization, PPO [32], Soft Actor-Critic, SAC [33], Deep Deterministic Policy Gradient, DDPG [34] and a hybrid combination with other control techniques [35]. A comparison of DQN, DDPG and TD3 was also conducted in [36]. Based on the comparative study shown in Table 1, the DQN is the fastest in tracking the MPPT but its reliance on discretization can affect the duty-cycle resolution. DDPG provides the natural continuous control mapping and off-policy sample reuse that justify it as our baseline for fine duty-cycle control and it is also moderate in terms of computational complexity. The implementation of DRL for MPPT control under partial shading conditions have been shown to have faster response times and higher accuracy in locating the GMPP, compared to conventional methods. Although previous studies have shown the effectiveness of DRL over conventional MPPT methods, there is not much emphasis on the loading condition that better represents real-world PV system operation. The present study aims to fill this gap by providing a comparative analysis of DRL and conventional MPPT control techniques under non-uniform irradiance and varying load conditions that highlights the practical advantages of DRL in realistic operating scenarios.

The algorithm of the DDPG agent can be implemented in the Matlab environment [43] and this algorithm was adopted from the literature [44].

3. DRL MPPT Control Design for PV systems

In practical grid-connected PV plants, the boost converter typically interfaces with a DC–AC inverter stage, delivering power to the grid and local AC loads. Furthermore, in large PV arrays, it is common to employ multiple DC–DC converters [45] that can allow for independent MPPT control and mitigation of partial shading effects. This section presents the design of the DDPG – DRL agent system for the MPPT control of a *100 kW* PV boost converter system as shown in the schematic diagram for Figure 1. A single boost converter architecture is considered whereby the entire *100 kW* PV array feeds a single converter stage. This represents a simplified but challenging configuration because all panels share the same duty cycle, and the system must converge to a single operating point. Unlike distributed MPPT architectures or module-level power electronics with bypass diodes, this setup does not allow localized bypassing of shaded panels. As such, the optimization burden shifts entirely to the control strategy that makes it an appropriate benchmark problem for evaluating the effectiveness of DRL MPPT approach.

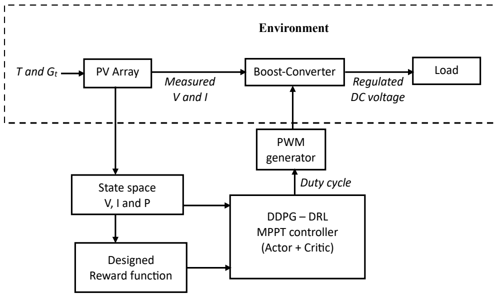

Figure 1: Schematic Design of the DDPG DRL MPPT control of PV system.

3.1. Environment Model description

The environmental model description is presented in Figure 1. The 100 kW PV system is designed based on the manufacturer ratings shown in Table 2 with 50 Modules connected in parallel and 10 modules connected in series and the corresponding P – V and I – V characteristics is shown in Figure 2. The output of the PV

Table 1: Comparative study of DRL Techniques.

Algorithm	Tracking Efficiency	Strengths	Limitations
DQN	Faster response than DDPG [37]; Outperforms PSO & GWO [38]	Fast convergence and Simple implementation	Requires discretization of action space (reduced precision)
DDPG	$<1\%$ deviation from theoretical GMPP [39]; Improved efficiency and ~ 0.8 convergence [40]	Works well in continuous state action space	Hyperparameter sensitive
TD3	Higher power output (up to +8.37% vs. DDPG at high irradiance) [36]	Most stable compared to DQN and DDPG and robust to noise	Computational complexity is high
SAC	Reduces oscillations compared to DDPG [41]	Smooth continuous control of duty cycle	Computational complexity is high (two critics + temperature tuning)
PPO	Demonstrated in PV systems with 96–98% tracking efficiency [42]; slower than DQN	Easy to implement, supports both discrete and continuous actions	Lower precision for fine MPPT than SAC and DDPG

arrays interfaced with a DC − DC boost converter with calculated values of *0.623 mH*, and *1888 μF* for the inductance and capacitance respectively. Considering Figure 1 again, an initial load resistance of *0.696* Ω is connected to this set-up to achieve resistance matching, $R = r$ for maximum power output. Where R and r are the load and internal resistances of the PV system respectively. By varying the load resistance starting with an initial value of *0.696* Ω, the performance of the system with respect to its power output can be evaluated.

Table 2: Data specification: Kyocera Solar KC200GT.

Parameter	Ratings
Maximum Power (W)	200.12
Voltage at Maximum Power (Vmp)	26.3 V
Current at Maximum Power (Imp)	7.61 A
Open Circuit Voltage (Voc)	32.9 V
Short circuit Current (Isc)	8.21 A
Total number of cells in Series (Ns)	54
Total number of cells in parallel (Np)	1

3.2. Model Verification

(a) PV Model

The PV array is implemented using the MATLAB/Simulink PV Array block, which is based on the single-diode equivalent circuit and parameterized using manufacturer datasheet values consistent with the NREL System Advisor Model (SAM) database. The model internally reproduces the I–V and P–V characteristics provided in the datasheet under standard test conditions. To ensure numerical reliability, the boost converter and DC-link dynamics are modeled using averaged continuous-time equations. The employed Simulink implementation is widely validated and extensively used in existing literature [40] for MPPT and DC–DC converter control studies.

(b) Boost Converter

The boost converter is designed to raise the voltage output of the PV array from an input voltage to a regulated output voltage. Its performance is dependent on the duty cycle i.e. $D \in ([0, 1])$, which is autonomously adjusted by the DDPG DRL agent. With respect to these present analyses, the boost converter component

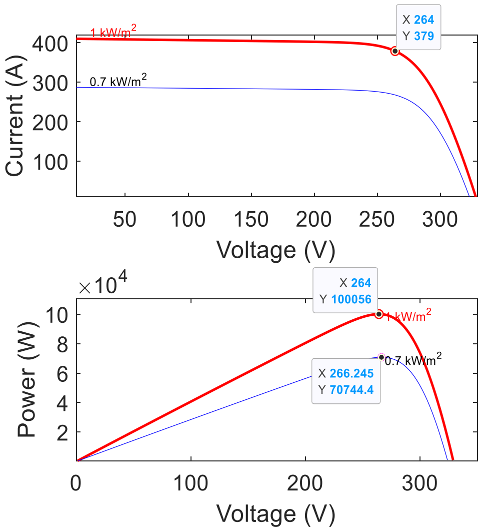

Figure 2: I–V and P–V characteristics for 100 kW PV system @ 1000, 700 W/m^2 irradiance levels.

sizing is obtained considering the input voltage of $263\ V$ and regulated output of $\sim 320\ V$, 5% ripple in inductor current, 1% ripple in regulated output voltage, 0.95 efficiency, η and switching frequency of $10\ kHz$ are assumed. Estimated values of the duty cycle, input current, inductance and capacitance for the boost converter are summarized in Table 3.

The inductance of $0.236\ mH$ and capacitance values of $1125\ \mu F$ are the theoretical values obtained based on standard calculations. These calculations do not consider possible effect of losses in the PV boost converter system. The simulated values used for the present study are $0.623\ mH$ and $1888\ \mu F$ for the inductance and capacitance respectively.

The proposed control architecture integrates a DDPG MPPT strategy into the DRL controller of the PV boost converter system comprising of the PV system and the DRL control loop. The PV system model, makes up the environment dynamics, consisting of PV array, boost converter, and load. Whereas the DRL control loop comprises of the state observer, reward function, and the DDPG agent. The measured variables obtained from the sensors are PV voltage and current. The irradiance and temperature Gt and T respectively are the inputs to the PV array. Therefore, observed states are based on the normalized values of the measured voltage, current and power as defined by Equation (3)

Table 3: Parameter Estimation for Boost Converter.

S/N	Description of parameter	Symbol	Unit	Formula	Value
1	Duty cycle	D	–	$1-\frac{V_{in}}{V_{out}}$	0.18
2	Input power	P_{in}	kW	$\frac{P_{out}}{\eta}$	105.26
3	Input current	I	A	$\frac{P_{in}}{V_{in}}$	400.23
4	Inductor size	L	mH	$\frac{V_{in}.D}{\Delta I_L.f_{sw}}$	0.238
5	Capacitor size	C	μF	$\frac{I_{out}.D}{\Delta V_{out}.f_{sw}}$	1125

$$s_t = \left[\frac{V_{pv}(t)}{V_{rated}}, \frac{I_{pv}(t)}{I_{rated}}, \frac{P_{pv}(t)}{P_{rated}}\right] \qquad (3)$$

where $V_{pv}(t)$, $I_{pv}(t)$ and $P_{pv}(t)$ are the instantaneous measured PV voltage, current and power respectively.

The duty cycle, D is modeled as a continuous action variable a_t, in the DDPG–DRL framework but practically it is constrained by converter operation limits as stated in Equation (4).

$$a_t = D \in [0.1, 0.9] \qquad (4)$$

and then sends it to the PWM signal generator that generates the PWM signal to control the boost converter's switch.

For a boost converter operation:

$$V_{out} = \frac{V_{in}}{1-D} \qquad (5)$$

Where V_{out} is the controlled output voltage while V_{in} is the PV voltage. Based on Equation (5), as D tends to 1, inductor current becomes very large and switch stress increases which can affect negatively the control sensitivity. As D tends to 0 there is no boost action and thus MPPT cannot be achieved. Thus, a good practice is to avoid extreme duty ratios of $D = 0$ meaning no boost or $D = 1$ meaning inductor saturation.

3.3. Reward design

The design of the reward function is critical for the effective learning of the DDPG agent and its operational performance. The reward function translates PV system objectives such as maximum power extraction and stability into quantitative evaluation signals. This shapes the DDPG agent's policy and influences its action selection. Previous studies have emphasized that in the context of PV MPPT, the reward design and algorithm selection are particularly critical due to the challenges that exist in PV cases. For example, the PV power–voltage curve is highly nonlinear, and it exhibits multiple local maxima under partial shading. If the reward function is poorly designed, the agent may converge to suboptimal local maxima or oscillate around the MPP. A fuzzified reward approach [46] was introduced to improve the mapping of continuous PV states into effective learning signals. In a similar direction, DDPG was employed for MPPT [47], demonstrating the advantages of continuous action spaces in providing smooth duty cycle control while reducing steady-state oscillations.

Based on the challenges identified in these previous studies such as the sensitivity of convergence to reward shaping in nonlinear P − V characteristics, this study develops a customized reward function as shown in Equation 6. The reward function explicitly incorporates PV array dimensions, instantaneous power, irradiance, and converter duty cycle. To address the issue of smooth and stable control under varying irradiance and partial shading conditions. A DDPG framework is adopted, leveraging on its continuous action space and actor–critic structure. This study develops a customized reward function based on Equation 6 that incorporates PV array dimensions, instantaneous power, irradiance, and converter duty cycle. It adopts a DDPG framework to ensure stable convergence under varying partial shading conditions.

The irradiance is not typically measured in practical PV system, but it is included in the reward function solely for normalization purposes. The role of irradiance, G_t is to provide a reference for the maximum theoretically available solar power, ensuring that the reward remains bounded between 0 and 1 regardless of irradiance level or array size. Importantly, the DDPG agent itself does not require irradiance as an input state but relies only on voltage, current and power measurements that are normally available in PV systems. In practical deployment, since irradiance is not usually measured the normalization factor can be replaced by alternative schemes. A simple option is to use a constant scaling term based on nominal maximum power under STC conditions [48] so that the reward remains bounded. This approach however ignores short-term irradiance variability and hence a more adaptive option is to use a moving-window estimate of peak power whereby the instantaneous extracted power is normalized by the maximum power observed within a recent time window [49].

The reward design, r_t is a function of both the measured PV system outputs and the agent's action explained in Equation (6). Specifically, the instantaneous power $P(t)$ and irradiance G_t are environment measurements, while the duty cycle $D(t)$ represents the control action. The first term of the reward normalizes the extracted power $P(t)$ against the theoretical maximum available solar power on the array surface. This ensures that the reward is scale-independent and bounded between 0 and 1. The second term, penalizes abrupt changes in the duty cycle between consecutive steps, thereby discouraging excessive switching activity. This promotes smooth control actions that help reduce converter stress and switching losses. It also helps to prevent oscillations around the MPP. Physically, the combined formulation captures the dual objectives of MPPT in achieving efficient power extraction and ensuring stable converter operation. The time index t refers to the discrete decision step of the DDPG agent at which the continuous duty cycle action is updated. In simulation, this decision interval is set equal to the environment sampling time that corresponds to the switching frequency of the boost converter. Thus, the fast PV dynamics can be captured while ensuring numerical stability of both the power electronic model and the learning process.

$$\begin{aligned} r_t &= f(s_t, a_t) \\ &= \frac{P(t)}{\max(\eta A G_t(t), \varepsilon)} - \lambda \left| D(t) - D(t-1) \right| \end{aligned} \tag{6}$$

where $P(t)$ is the instantaneous power at time t, G_t is the solar irradiance (W/m^2), $|D(t) - D(t-1)|$ is the absolute change in duty cycle. λ is the penalty weight on changes in duty cycle, η and A are the efficiency and total area of the PV array respectively. Also, ε (10^{-3}) is a small constant to avoid division by zero, especially when the solar irradiance $G_t = 0$. This term is only applicable for simulation purposes and has no physical interpretation. In the Simulink environment when $P(t)$ or G_t is exactly zero, certain blocks can produce NaN or zero-gradient issues for the DRL agent. To avoid this, a small offset $\varepsilon > 0$ is added. This does not represent meaningful energy or MPPT performance but ensures the simulation and learning process remain numerically stable. The weight λ is a hyperparameter that balances maximizing power output with control stability by penalizing large duty cycle variations.

The first term in Equation (6) is the normalized power term that is bounded between 0 and 1 and therefore does not require additional weight. With respect to the second term, the penalty is scaled through the parameter λ which is selected to be small enough so that power maximization remains the dominant objective. Nevertheless, it is large enough to discourage excessive duty cycle oscillations. λ is tuned empirically by comparing the magnitude of the two reward terms. This ensures that the DDPG agent prioritizes maximizing power while still being penalized for excessive switching. Sensitivity tests can be performed by varying λ within $[0.01, 0.1]$, and the chosen value corresponds to the best tradeoff between fast convergence and stable duty cycle trajectories.

The parameter η represents the nominal efficiency of the PV array and it is typically obtained from manufacturer specifications. The conversion efficiency varies with temperature, dust accumulation, and aging but in this study, the role of η in the reward function is only to normalize the instantaneous extracted PV power with respect to the theoretical maximum available solar power defined by $\max(\eta A G_t(t))$. This normalization provides a bounded reward rather than an exact physical quantity so that the learning process is not highly sensitive to variations in η. Moreover, this makes the reward a relative measure of utilization efficiency that is independent of the absolute array size or irradiance level. This provides a scale-free performance measure bounded between *0* and *1* that allows the DDPG agent to consistently evaluate its performance under different irradiance levels and array sizes.

This choice was made to facilitate a fair and consistent performance comparison under rapidly varying irradiance conditions during the training and evaluation phases, as G_t normalization ensures numerical stability of the reward signal and accelerates policy convergence in simulation. Also in the context of MPPT control, the reward formulation does not aim to optimize a specific numerical reward value but rather to guide the agent toward a physically meaningful control policy that maximizes harvested energy while maintaining system stability.

The DRL agent is trained offline prior to deployment and the reward parameters were selected based on empirical stability and convergence considerations. A sensitivity analysis was conducted during preliminary experiments by varying the weighting parameter λ within the practical range $0.01 \leq \lambda \leq 0.1$. Within this interval, no significant changes were observed in convergence behavior or steady-state power tracking accuracy. Consequently, the qualitative conclusions of the study remained unchanged whereas for λ values outside this range, the training process exhibited slower convergence and increased duty-cycle oscillations. Thus preventing reliable policy learning and rendering the resulting controller unsuitable for deployment in the PV system.

3.4. Training Repeatability and Variability

The primary objective of this study is the evaluation of closed-loop control performance after offline training rather than statistical benchmarking of learning dynamics. In this context, the agent is trained until convergence to a stable policy and then subsequently it is deployed and evaluated deterministically in the PV system environment. Once training converges, the learned policy is fixed and does not exhibit stochastic behavior during deployment. Preliminary repeated training runs with different random seeds produced policies with comparable steady-state MPPT efficiency and DC-link voltage regulation, indicating limited sensitivity to initialization. Therefore, for clarity and conciseness, only representative results are reported.

Hyperparameters shown in Table 4 were tuned empirically to ensure stable convergence while minimizing oscillatory behavior in the duty cycle. The actor and critic network architecture specify the number of hidden layers and neurons while the activation functions were used to model the policy and value functions. Learning parameters such as the discount factor, target smoothing, experience buffer length, and mini-batch size, define how the agent updates its networks during training. Exploration parameters describe the Gaussian noise added to the actions to encourage exploration, including its magnitude and decay rate. Finally, simulation and training settings specify the sample time, number of episodes, steps per episode, and criteria for saving the trained agent. Collectively, these parameters ensure that the training is reproducible, and that the agent achieves effective MPPT performance.

Table 4: DDPG Agent Training Hyperparameters.

Component	Parameter	Value
Actor Network	Hidden Layers	64, 64
	Activation Function	ReLU
	Output Layer	Sigmoid (duty cycle between 0 and 1)
Critic Network	State Path Hidden Layers	32, 32
	Action Path Hidden Layers	32
	Combination Layer	Addition + ReLU
Learning Parameters	Discount Factor (γ)	0.99
	Target Smooth Factor	0.001
	Experience Buffer Length	1,000,000
	Mini-batch Size	64
Exploration	Noise Type	Gaussian
	Noise Standard Deviation	0.1
	Noise Decay Rate	1×10^{-5}
Simulation /Training	Sample Time	1×10^{-5} s
	Max Episodes	200–1000
	Max Steps per Episode	2000

4. Analysis of DRL with Conventional MPPT

This section presents the DDPG DRL results for MPPT under uniform irradiance and non – uniform irradiation conditions. The evaluation was conducted using MATLAB R2024b and Simulink. VSS P&O [23] combined with a PID regulator is used as a benchmark to validate the design of the DRL MPPT control. Importantly, only the VSS P&O [23] technique was adopted from the literature while all PV array parameters, boost converter specifications, and controller parameters including PID gains were derived and tuned for the 100 kW Kyocera- PV system modeled in this study.

The training of the DDPG DRL agent is shown in Figure 3 and the agent's adaptive nature can be explained by the reward formulation and the choice of state variables. The reward is defined in terms of the PV power output, and the state vector includes parameters of PV voltage, current, and power. This allows the agent to learn the underlying system dynamics that govern maximum power extraction. Even though the agent was trained under uniform irradiance, these state variables continue to describe the system consistently under varying irradiance and load conditions. As a result of this, the trained agent can be deployed without retraining for any operating condition within the same PV system environment. Retraining is only required when the environment itself changes structurally in terms of the different PV sizing or converter configuration.

Episode reward is a measure of how much reward the agent has earned in each episode, increasing from approximately 0 to 10^5 over the first 20 episodes. Thereafter, it flattens out close to 9.5×10^5 after 50 episodes indicating reward saturation. This implies that the DDPG agent can track the maximum power point effectively. Average Reward is the mean of the instantaneous rewards across each episode, and it is an indication of reward stability. It initially starts low and then steadily converges to the episode reward $\sim$ 20, which is an indication of stable learning. The training results indicate that the DDPG agent consistently achieves high episode

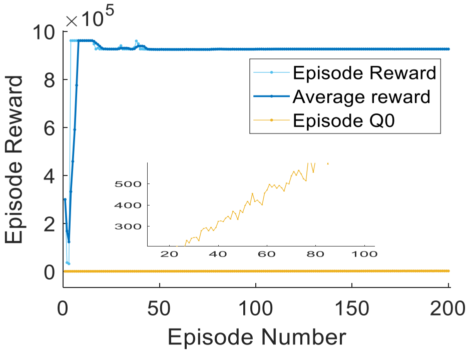

Figure 3: Training for DDPG control of PV boost converter.

rewards which are defined based on the instantaneous power of the PV system relative to its maximum power under uniform irradiation.

The critic network in DDPG estimates the function $Q(s, a)$ which represents the expected cumulative discounted reward from taking action 'a' in state 's'. In training logs, the notation *Episode* Q_0 is used to denote the value of $Q(s_0, a_0)$ which is the estimated expected return from the initial state at the start of each episode. Monitoring this quantity provides an indication of how the critic's estimation of overall episode performance evolves during training. It increases over time, suggesting that the DDPG agent is in a good state and likely to perform well. Both average reward and critic Q-values provide complementary insights. For instances, the average reward indicates the stability and convergence of learning, while critic Q-values reflect the quality of the learned policy.

Therefore, the learning curves are directly linked to physical MPPT performance through the reward formulation. Specifically, the episode reward is proportional to the instantaneous PV power which is normalized by the theoretical maximum power under the same irradiance condition. Therefore, an increase in episode reward directly reflects improved power tracking efficiency and closer operation to the MPPT and convergence of the episode reward directly indicates convergence of MPPT tracking efficiency and DC-link stability. Moreover, the reduction in reward variance across episodes indicates diminished duty-cycle oscillations and improved steady-state behavior of the boost converter.

Similarly, convergence of the Episode Q0 metric reflects stabilization of the expected long-term return which is manifested physically as steady MPPT operation and eventually regulated DC-bus voltage.

4.1. Environment Validation

The DRL agent is trained within a physics-based Simulink environment representing the PV–boost converter dynamics. The trained DRL DDPG agent is evaluated against a theoretical PV model under identical irradiance and load conditions and the converter dynamics were perturbed by varying the inductance and capacitance parameters by ±30% around their nominal values, as summarized in Table 5. The results demonstrate that the average harvested PV power and settling time remain consistent across all cases, indicating that the environment dynamics are robust to parameter variations and faithfully capture the underlying PV–converter behavior.

Table 5: Perturbation of Boost converter parameters.

Case	L	C	Average Power (kW)	Settling time (secs)
Nominal	L_0	C_0	96.91	0.00050
L−30%	$0.7\,L_0$	C_0	95.98	0.00042
L+30%	$1.3\,L_0$	C_0	96.73	0.00057
C−30%	L_0	$0.7\,C_0$	96.55	0.00044
C+30%	L_0	$1.3\,C_0$	96.46	0.00056

4.2. MPPT Control Performance Under Uniform Irradiation

This section presents a performance comparison between the proposed DDPG and VSS P&O PID controllers under uniform irradiance conditions of 1000 W/m^2 where the nominal PV parameters and MPP position are particularly reliable. This is to ensure that the results obtained reflect the inherent behavior of the controllers rather than model uncertainty. Subsequently, the analysis is extended to non-uniform irradiance patterns, where the reliability of manufacturer data is reduced and the MPP location is more uncertain, thereby allowing a more rigorous assessment of the DDPG agent's robustness under realistic conditions. The VSS P&O PID is used to validate the DDPG technique under uniform irradiance and resistance matching loads of $R = r = 0.697\ \Omega$. This setup is commonly adopted in MPPT studies to provide a controlled environment for comparing algorithms.

At 1000 $\mathrm{W/m^2}$, the nominal PV parameters and MPP position are particularly reliable, ensuring that the results obtained reflect the inherent behavior of the controllers rather than model uncertainty. Subsequently, the analysis is extended to non-uniform irradiance, where the reliability of manufacturer data is reduced and the MPP location is more uncertain, thereby allowing a more rigorous assessment of controller robustness under realistic conditions.

The result of this comparison is shown in Figure 4. The PV system is subjected to uniform irradiation of 1000 $\mathrm{W/m^2}$ and the performance of both controllers are evaluated in terms of maximum power output and tracking accuracy.

From the PV power results in Figure 4, both controllers were able to reach the theoretical maximum power point of approximately 100 kW, which confirms their capability to identify the MPP. The tracking error, TE was observed for both controller and it is defined as the relative deviation between the maximum power achieved by the controller and the theoretical maximum power at the corresponding irradiance as defined in Equation (7)

$$TE = \frac{P_{\text{measured}} - P_{\text{theoretical}}}{P_{\text{theoretical}}} \tag{7}$$

where P_{measured} is the measured PV power and $P_{\text{theoretical}}$ is the reference power. Under uniform irradiance of 1000 $\mathrm{W/m^2}$ it was observed that the PID controller achieved a lower tracking error of 5.58% compared to 6.24% for the DDPG agent. Also, it is observed that the DDPG agent exhibits a slight negative bias in voltage and power, behaving like a tight but uniform band that reflects the trade-off imposed by the reward function and critic approximation.

The slight reduction in steady-state power observed with the DRL controller reflects a reward - critic trade-off to suppress duty-cycle oscillations as well as prevent over-excitation of the DC–DC converter, which is not accounted for in conventional MPPT control designs.

4.3. MPPT control Under Non uniform irradiation

In this study, irradiance variation was implemented deterministically using a time-based shading function. whereby the nominal irradiance of 1000 $\mathrm{W/m^2}$ was reduced to 70% of its original value during the time interval $t = 0.15$ to $t = 0.30$s. This represents a temporary partial shading event with a defined magnitude and duration so that outside this interval, the irradiance remained constant. This deterministic formulation ensures that the temporal shape, shading depth, and rate of change are precisely defined and fully reproducible.

This section presents the testing of both controllers under non - uniform irradiance at the nominal matched resistance $R = r$. This provides a fair baseline for comparison where the PV parameters are most reliable. Once this reference performance was established, the controllers were further assessed under more challenging conditions of a combination of non-uniform irradiance and varying load resistances.

Figures 5, 6, 7 and 8 are used to evaluate the effect of non − uniform irradiance on the performance of both controllers for MPPT of the PV system. Table 6 presents a comparative evaluation of the proposed DDPG MPPT controller against the VSS P&O-PID controller. Four performance indices are reported: Average Power, Average Voltage and % Ripple for power and voltage at steady state. In this study, ripple (%) is defined as the peak-to-peak variation of a measured steady-state signal that is normalized by its mean value and calculated over a steady-state observation window following transient settling. This definition is applied consistently for all reported voltage and power ripple measurements.

At $R = r = 0.697\ \Omega$, corresponding to nominal power operation, the VSS P&O_PID con-

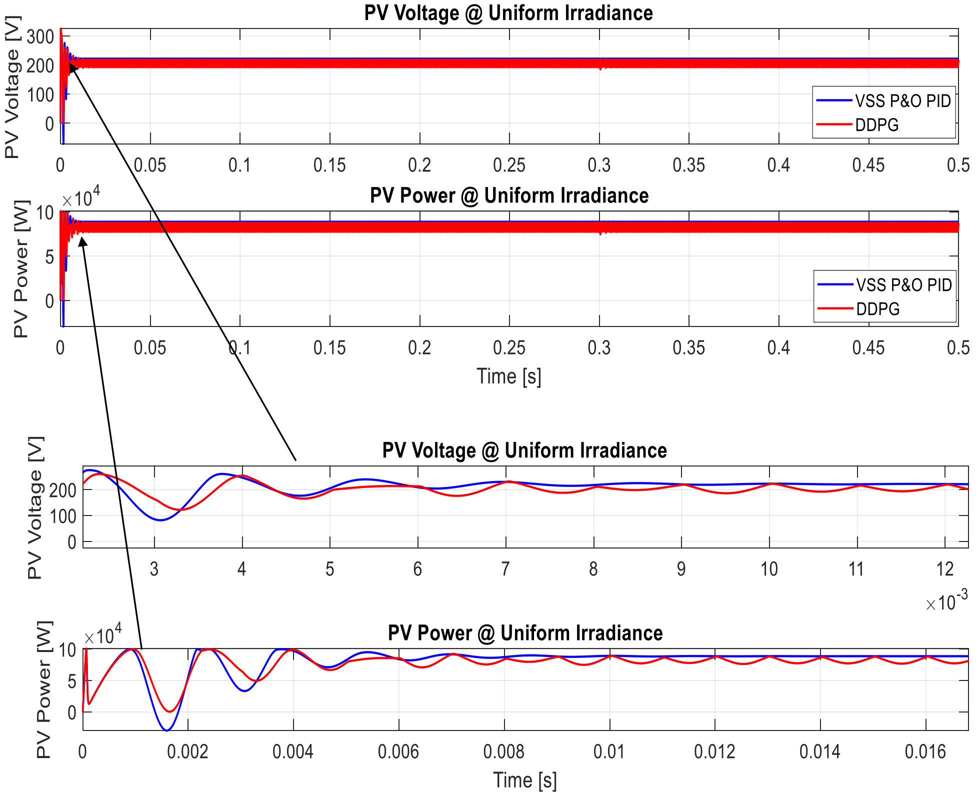

Figure 4: MPPT Comparison of DDPG and VSS P&O PID.

troller achieves slightly higher average power extraction compared to DDPG. However, both controllers exhibit relatively low ripple power and voltage with DDPG showing a larger fluctuation compared to VSS P&O_PID. At $R = 1\ \Omega$, DDPG demonstrates a better performance in the average power and voltage with substantially reduced % ripple. This indicates that at medium load conditions the DDPG policy successfully learns the optimal duty cycle adjustments with reduced ripple. At $R = 2 - 5\ \Omega$, both methods achieve comparable average power levels, but the DDPG controller consistently yields lower ripple levels.

This improvement is primarily due to the reward structure that is designed to penalize oscillatory behavior and large control deviations. As a result, the DDPG agent reduces the excitation of converter dynamics that would otherwise amplify ripple. In addition, at this medium load the slope of the P–V curve is moderate which indicates that small perturbations in duty cycle have limited effect on the extracted power.

Table 6: Comparative results of DDPG and VSS P&O PID control.

MPPT controller	Average Power (W)	Average Voltage (V)	Ripple power (%)	Ripple Voltage (%)
R = r = 0.697 Ω				
DDPG	67,769	190.70	2.81	1.41
VSS P&O_PID	74,447	199.85	$\ll 1$	$\ll 1$
R = 1 Ω				
DDPG	82,766	286.69	1.55	0.78
VSS P&O_PID	79,960	281.90	$\ll 1$	$\ll 1$
R = 2 Ω				
DDPG	51,554	320.53	0.83	0.41
VSS P&O_PID	47,539	307.67	$\ll 1$	$\ll 1$
R = 5 Ω				
DDPG	22,481	334.63	0.44	0.22
VSS P&O_PID	20,907	321.86	$\ll 1$	$\ll 1$
R = 10 Ω				
DDPG	11,579	339.58	0.36	0.18
VSS P&O_PID	11,138	330.34	$\ll 1$	$\ll 1$

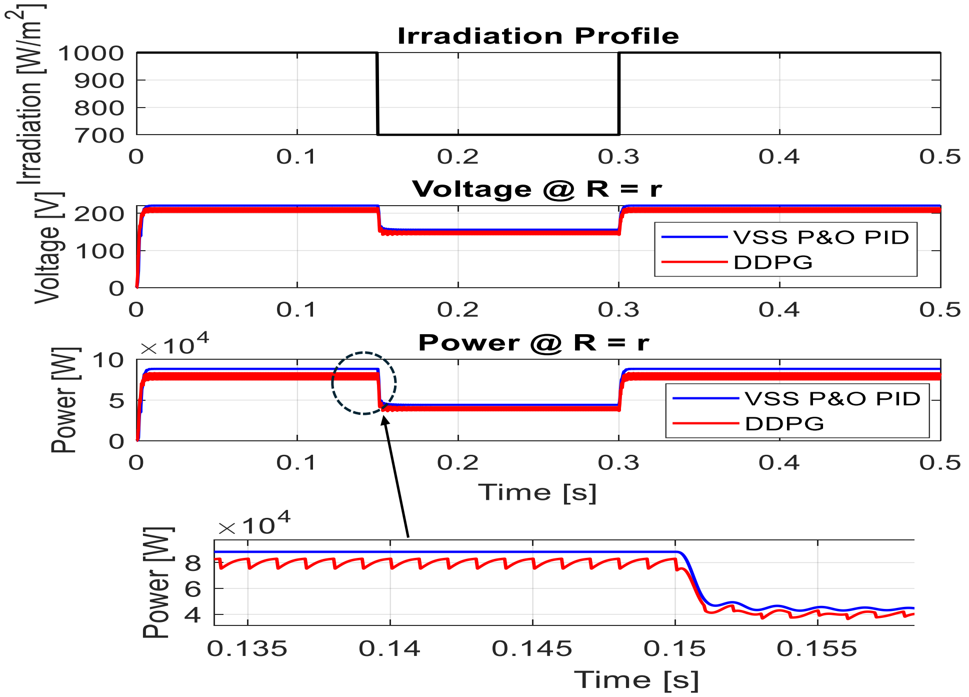

Figure 5: Comparison of DDPG and VSS P&O PID @ $R = r$.

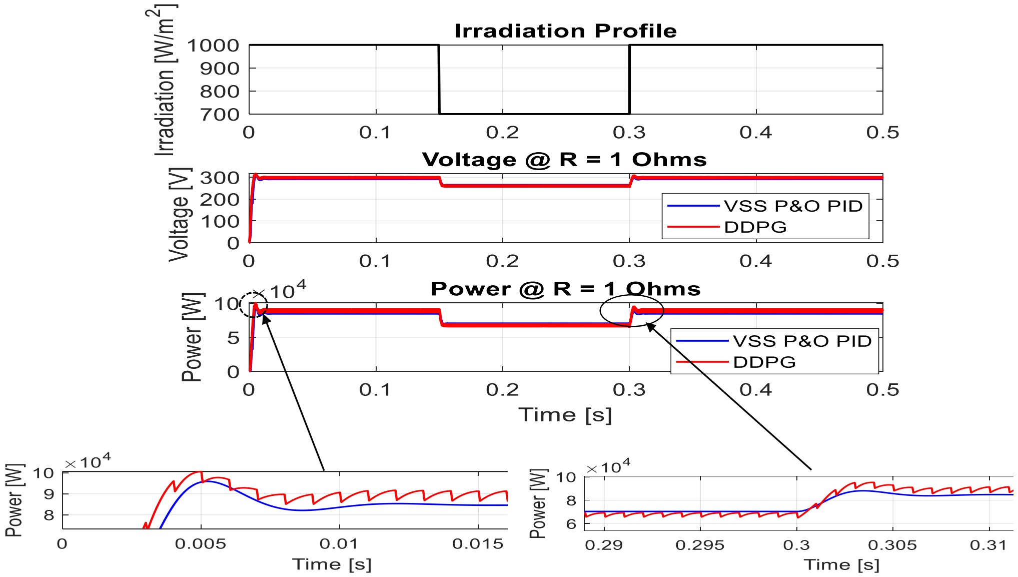

Figure 6: Comparison of DDPG and VSS P&O PID @ $R = 1\ \Omega$.

These factors allow the DDPG policy to maintain stable operation with reduced ripple while still tracking the maximum power point. By contrast, the VSS P&O controller adjusts the

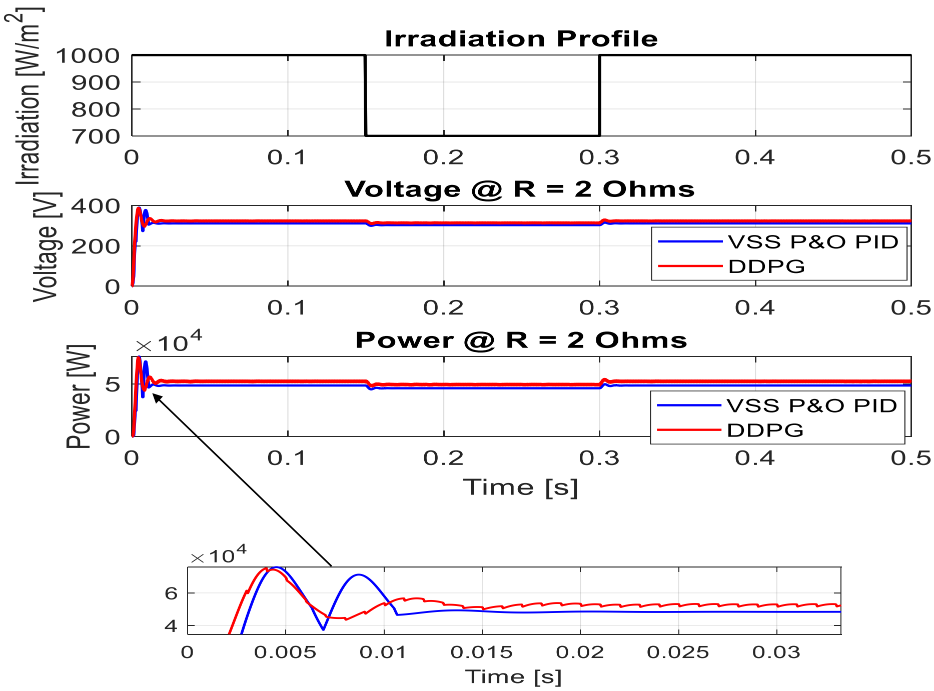

Figure 7: Comparison of DDPG and VSS P&O PID @ $R = 2\ \Omega$.

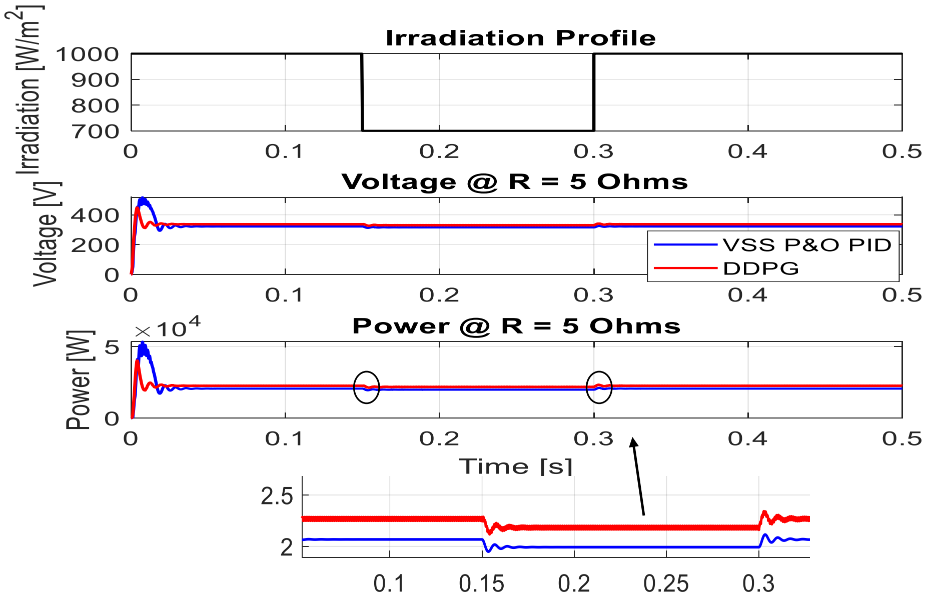

Figure 8: Comparison of DDPG and VSS P&O PID @ $R = 5\ \Omega$.

perturbation size based on the local slope of the P–V curve reducing oscillations near the MPP. At light current, $R = 10\ \Omega$, DDPG again achieves higher average power compared to VSS P&O_PID with substantially lower and ripple. Although the DDPG controller exhibits slightly higher steady-state ripple ≈ 0.1–3%, it provides superior performance in the power and voltage for medium to high resistance ranges. This trade off in the DDPG is due to its exploratory and stochastic nature. In contrast the PID maintains negligible ripple $\ll 1\%$ but at the cost of failing to track the GMPP. The simulation results demonstrate that the DDPG controller achieves higher load power and voltage with reduced ripple compared to VSS P&O – PID controllers across varying load conditions. The observed reduction in load voltage and power ripple directly reflects the PV-side duty-cycle adjustments by the DDPG controller which in turn minimizes ripple on the load power and voltage.

Table 7: Comparative Metrics of DDPG and VSS P&O.

MPPT Controller	Energy (kWh)	Dynamic Efficiency (%)	MRMSEP (%)
***R* = 0.697**			
DDPG	8.79	89.15	18.87
VSS_P&O	9.21	93.42	14.48
***R* = 1**			
DDPG	9.12	92.50	8.62
VSS_P&O	8.82	89.42	12.12

To complement instantaneous power and ripple-based assessments, the cumulative harvested energy, dynamic efficiency, DE and mean root mean square error of power (MRMSEP) is adopted as performance metric. The Energy metric (kWh) is obtained by integrating the delivered load power over the entire operating profile, thereby capturing both steady-state MPPT accuracy and transient tracking behavior under dynamic conditions while the MRMSEP and % DE are calculated based on Equations (8) and (9) respectively.

$$MRMSEP = \sqrt{\text{mean}\left((P_{\max} - P_{\text{avg}})^2\right)} \quad (8)$$

$$\%\,DE = \frac{\text{mean}\,P_{\text{avg}}}{\text{mean}\,P_{\max}} \times 100\% \quad (9)$$

Where P_{avg} and P_{max} are the instantaneous average and maximum power respectively. As shown in Table 7, the energy-based evaluation, dynamic efficiency reveals that although both controllers can achieve comparable instantaneous power levels near the MPP. However, when the load is increased from the impedance-matched condition ($R = 0.697\ \Omega$) to a more realistic operating load ($R = 1\ \Omega$), the DRL DDPG controller exhibits superior energy harvesting performance, improved dynamic efficiency and lower mean root mean square error power (MRMSEP) compared to the conventional VSS P&O method. This behavior highlights the enhanced adaptability of the DRL controller to non-ideal and varying load conditions, where classical perturbative methods become more sensitive to operating-point shifts.

5. Validation of DRL With AI based - MPPT

To further evaluate the effectiveness and generality of the proposed DRL MPPT scheme, a comparative study was conducted with a hybrid particle swarm optimization–neural network (PSO–NN) MPPT controller [50]. The PSO–NN approach represents a strong intelligent optimization baseline, combining global search capability with nonlinear function approximation, and has been shown to perform well under uniform irradiation conditions.

5.1. Description of the PSO–NN MPPT Setup

To ensure a fair and rigorous comparison with intelligent optimization-based MPPT methods, a PSO–NN-based controller was implemented following the structure reported in the referenced literature. The neural network was trained offline using a particle swarm optimization (PSO) algorithm as shown in Figure 9 to minimize the MPPT tracking error under varying operating conditions for the 100kW photovoltaic system.

The PSO algorithm was initialized with a population of candidate weight vectors and

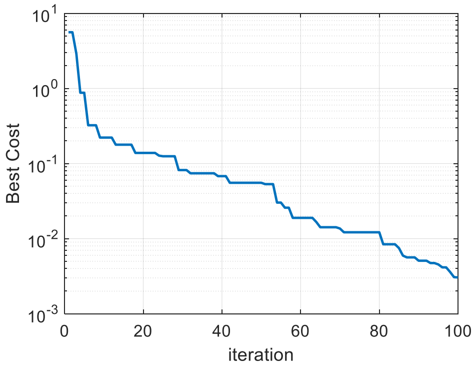

Figure 9: Trained PSO – NN for 100 kW PV system.

each particle's fitness was evaluated based on a power-tracking cost function. The tracking cost function is the normalized error between the extracted PV power and the estimated maximum available power. The optimization process iteratively updated particle positions and velocities until convergence was achieved.

Figure 9 illustrates the evolution of the PSO cost function over iterations for the 100 kW PV system. It can be observed that the swarm converges monotonically reaching a best cost value of approximately 0.0038 after 100 iterations. This low final cost indicates effective training of the neural network and reliable convergence of the PSO–NN MPPT controller.

5.2. Performance Under fast changing Irradiance

Under fast changing irradiance conditions, both the proposed DDPG DRL controller and the PSO–NN scheme achieved effective MPPT as shown in Figure 10. The PSO–NN method exhibits a marginally higher instantaneous PV power output compared to the proposed DRL scheme. This behavior is expected, as the PSO–NN controller is explicitly designed to maximize PV-side power which does not incorporate the DC-bus voltage constraints whereas, the DRL formulation explicitly balances power extraction with DC-link voltage regulation. As a result, a slight reduction in peak power is observed by the DRL method and this reflects a deliberate trade-off to ensure stable and safe system operation. This slight reduction in PV power is therefore not a disadvantage when compared with other approaches like the VSS P&O and the PSO – NN MPPT techniques.

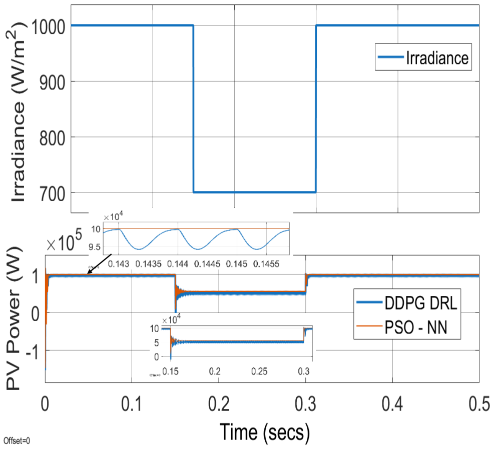

Figure 10: Validation of DRL with PSO – NN under uniform conditions.

5.3. Performance on DC Voltage Bus Regulation

To assess robustness under more realistic operating scenarios, both controllers were subjected to step-changing irradiance profiles. Under these conditions, the PSO–NN-based MPPT exhibits pronounced DC-bus voltage deviations as irradiance decreases and the absence of an explicit voltage regulation mechanism in the PSO–NN formulation leads to excess power injection into the DC-link capacitor, causing the DC-bus voltage to rise significantly above its rated value. In contrast, the proposed DRL MPPT controller can maintain the DC-bus voltage tightly regulated around its designed reference value of approx. 320 V_{dc} despite rapid irradiance fluctuations. Consequently, the DRL agent adjusts its power extraction behavior in real time to prevent overcharging of the DC-bus capacitance and at the same time preserves near-optimal MPPT performance.

Figure 11 presents the effect of fast changing irradiance on the DC bus voltage supplying a resistive load of 3 Ω. From these results, it can be observed that the DC voltage is sustained up to a much lower irradiance level of 400 $\mathrm{W/m^2}$. The comparison demonstrates that for realistic operating conditions involving fast environmental transients the DC-bus voltage stability can be sustained using DRL approach.

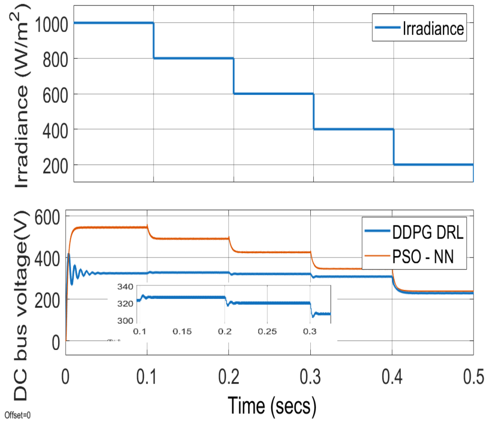

Figure 11: Performance on the DC bus voltage under fast changing irradiance.

5.4. Real-Time Applicability and Computational Considerations

In this study, the DDPG- DRL MPPT controller is trained offline using simulation data and so does not involve any online learning or optimization during real-time operation. Once trained, the controller generates the duty-cycle command through a single forward pass of a compact neural network, consisting of two hidden layers with a limited number of neurons.

As a result, the online inference complexity is significantly lower than that of population or iterative MPPT methods such as Particle Swarm Optimization (PSO) or Genetic Algorithms (GA) which require repeated fitness evaluations at each control step. Furthermore, the reward function and critic network are used only during the training phase and do not contribute to computational overhead during deployment.

Although explicit hardware timing benchmarks are not considered in this study, the employed network architecture is well within the computational capabilities of commonly used embedded platforms such as DSPs or real-time controllers for DC–DC converter control, where sampling periods in the microsecond to millisecond range are typical.

6. Conclusion

In this study, a 100 kW photovoltaic (PV) boost converter system was used as the benchmark to evaluate the performance of a DDPG – DRL controller in comparison with a conventionalVSS P&O PID controller under both uniform and non-uniform irradiance conditions. Simulation results indicate that both controllers provide reliable and stable operation with the conventional marginally higher than the DRL under nominal conditions. However, its reliance on continuous perturbations limits its adaptability during rapid irradiance or load variations, leading to its degrading performance.

In contrast, the proposed DDPG – DRL controller exploits a learned control policy that generalizes across a wide range of operating conditions without requiring persistent perturbations. As a result, it achieves smoother control trajectories, reduced power ripple, and improved robustness to parameter mismatches and measurement uncertainties.

The results further demonstrate the flexibility of the DRL framework in learning MPPT behavior that implicitly integrates desirable characteristics of traditional P&O and PID approaches within a unified control policy. Unlike heuristic or hybrid methods, the DRL controller operates as a system-level optimizer, jointly addressing MPPT and DC-link voltage regulation within a single learning objective. This eliminates the need for auxiliary voltage control loops and enables stable operation under both irradiance and load disturbances.

Comparatively, while alternative strategies such as PSO–NN-based controllers may achieve higher instantaneous power extraction under specific conditions, they do not inherently guarantee stable or safe operation in practical

DC-connected photovoltaic systems. The proposed DRL approach prioritizes overall system stability and operational feasibility alongside MPPT performance.

Future work will focus on hardware-in-the-loop and experimental implementation to further validate the effectiveness of the proposed DRL MPPT controller.

Acknowledgement

I wish to acknowledge DR. Akuro Big-Alabo for his encouragement and support. His scrutiny and proof reading of the work also helped to shape the content of this work. I wish to acknowledge the anonymous reviewers that also helped to improve the quality of the work. Finally, I wish to acknowledge the Alexander von Humboldt Foundation for the sponsorship of my research fellowship, of which this work forms a part.

Declaration of conflicting interest

The author declares that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability statement

All necessary data have been included in the manuscript.

References

[1] Hamed H Pourasl, Reza Vatankhah Barenji, and Vahid M Khojastehnezhad. Solar energy status in the world: A comprehensive review. *Energy Reports*, 10:3474–3493, 2023.

[2] Zuhair Alaas, ZMS Elbarbary, Alireza Rezvani, Binh Nguyen Le, et al. Analysis and enhancement of mppt technique to increase accuracy and speed in photovoltaic systems under different conditions. *Optik*, 289:171208, 2023.

[3] Ankita Saxena, Calum Brown, Almut Arneth, and Mark Rounsevell. Advanced photovoltaic technology can reduce land requirements and climate impact on energy generation. *Commun. Earth & Environ.*, 5(1):586, 2024.

[4] CH Hussaian Basha and C Rani. Different conventional and soft computing mppt techniques for solar pv systems with high step-up boost converters: A comprehensive analysis. *Energies*, 13(2):371, 2020.

[5] Marcelo Gradella Villalva, Jonas Rafael Gazoli, and Ernesto Ruppert Filho. Comprehensive approach to modeling and simulation of photovoltaic arrays. *IEEE Trans. on power electronics*, 24(5):1198–1208, 2009.

[6] Pallavi Bharadwaj, Kunal Narayan Chaudhury, and Vinod John. Sequential optimization for pv panel parameter estimation. *IEEE J. Photovoltaics*, 6(5):1261–1268, 2016.

[7] Lyu Guanghua, Farah Andleeb Siddiqui, Muhammad Mohsin Aman, Syed Hadi Hussain Shah, Aqsa Ali, Arsalan Muhammad Soomar, and Shoaib Shaikh. Improved maximum power point tracking algorithms by using numerical analysis techniques for photovoltaic systems. *Results engineering*, 21:101740, 2024.

[8] Pawan Kumar Pathak, Anil Kumar Yadav, and PA Alvi. A state-of-the-art review on shading mitigation techniques in solar photovoltaics via meta-heuristic approach. *Neural Comput. Appl.*, 34(1):171–209, 2022.

[9] Mohammed Hamouda Ali, Mohammad Zakaria, and Sally El-Tawab. A comprehensive study of recent maximum power point tracking techniques for photovoltaic systems. *Sci. Reports*, 15(1):14269, 2025.

[10] Saurabh Thakran, Jaspreet Singh, Rachan Garg, and Priya Mahajan. Implementation of p&o algorithm for mppt in spv system. In *2018 International Conference on Power Energy, Environment and Intelligent Control (PEEIC)*, pages 242–245. IEEE, 2018.

[11] GUIZA Dhaouadi, OUNNAS Djamel, SOUFI Youcef, and Chenikhe Salah. Implementation of incremental conductance based mppt algorithm for photovoltaic system. In *2019 4th International Conference on Power Electronics and their Applications (ICPEA)*, pages 1–5. IEEE, 2019.

[12] Kifayat Ullah, Muhammad Ishaq, Fairouz Tchier, Hijaz Ahmad, and Zubair Ahmad. Fuzzy-based maximum power point tracking (mppt) control system for photovoltaic power generation system. *Results Eng.*, 20:101466, 2023.

[13] Yingquan Zou, Fei Yan, Xiaomin Wang, and Jiyong Zhang. An efficient fuzzy logic control algorithm for photovoltaic maximum power point tracking under partial shading condition. *J. Frankl. Inst.*, 357(6):3135–3149, 2020.

[14] Hong Li, Duo Yang, Wenzhe Su, Jinhu Lü, and Xinghuo Yu. An overall distribution particle swarm optimization mppt algorithm for photovoltaic system under partial shading. *IEEE Trans. on Ind. Electron.*, 66(1):265–275, 2019.

[15] Jothi Swaroopan NM et al. Mppt of solar pv systems using pso memetic algorithm considering the effect of change in tilt angle. *Sci. Reports*, 15(1):1–17, 2025.

[16] Prakash Kumar, Gaurav Jain, and Dheeraj Kumar Palwalia. Genetic algorithm based maximum power tracking in solar power generation. In *2015 International Conference on Power and Advanced Control Engineering (ICPACE)*, pages 1–6. IEEE, 2015.

[17] Stefan Daraban, Dorin Petreus, and Cristina Morel. A novel mppt (maximum power point tracking) algorithm based on a modified genetic algorithm specialized on tracking the global maximum power point in photovoltaic systems affected by partial shading. *Energy*, 74:374–388, 2014.

[18] Kok Soon Tey, Saad Mekhilef, Mehdi Seyedmahmoudian, Ben Horan, Amanullah Than Oo, and Alex Stojcevski. Improved differential evolution-based mppt algorithm using sepic for pv systems under partial shading conditions and load variation. *IEEE Trans. on Ind. Informatics*, 14(10):4322–4333, 2018.

[19] Yuan Gao, Songda Wang, Tomislav Dragicevic, Patrick Wheeler, and Pericle Zanchetta. Artificial intelligence techniques for enhancing the performance of controllers in power converter-based systems—an overview. *IEEE Open J. Ind. Appl.*, 4:366–375, 2023.

[20] Prabhat Ranjan Bana, Salvatore D'Arco, and Mohammad Amin. Ann-based robust current controller for single-stage grid-connected pv with embedded improved mppt scheme. *IEEE Access*, 2024.

[21] Umair Younas, Ahmet Afsin Kulaksiz, and Zunaib Ali. Deep learning stack lstm based mppt control of dual stage 100 kwp grid-tied solar pv system. *IEEE Access*, 12:77555–77574, 2024.

[22] Stephen Bassi Joseph, Emmanuel Gbenga Dada, Afeez Abidemi, David Opeoluwa Oyewola, and Ban Mohammed Khammas. Metaheuristic algorithms for pid controller parameters tuning: Review, approaches and open problems. *Heliyon*, 8(5), 2022.

[23] Xavier Serrano-Guerrero, José González-Romero, Xavier Cárdenas-Carangui, and Guillermo Escrivá-Escrivá. Improved variable step size p&o mppt algorithm for pv systems. In *2016 51st International Universities Power Engineering Conference (UPEC)*, pages 1–6. IEEE, 2016.

[24] Abdelkadir Belhadj Djilali, Elhadj Bounadja, Adil Yahdou, Habib Benbouhenni, ZMS Elbarbary, Ilhami Colak, and Saad F Al-Gahtani. Enhanced variable step sizes perturb and observe mppt control to reduce energy loss in photovoltaic systems. *Sci. Reports*, 15(1):11700, 2025.

[25] Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming.* John Wiley & Sons, 1994.

[26] P Kofinas, Stefanos Doltsinis, AI Dounis, and GA Vouros. A reinforcement learning approach for mppt control method of photovoltaic sources. *Renew. Energy*, 108:461–473, 2017.

[27] ATD Perera and Parameswaran Kamalaruban. Applications of reinforcement learning in energy systems. *Renew. Sustain. Energy Rev.*, 137:110618, 2021.

[28] Dajr Alfred, Dariusz Czarkowski, and Jiaxin Teng. Reinforcement learning-based control of a power electronic converter. *Mathematics*, 12(5):671, 2024.

[29] Richard S Sutton, Andrew G Barto, et al. *Introduction to reinforcement learning*, volume 135. MIT press Cambridge, 1998.

[30] Emmanouil Lioudakis and Eftichios Koutroulis. Global flexible power point tracking based on reinforcement learning for partially shaded pv arrays. *IEEE J. Emerg. Sel. Top. Ind. Electron.*, 2024.

[31] Grace Muriithi and Sunetra Chowdhury. Deep q-network application for optimal energy management in a grid-tied solar pv-battery microgrid. *The J. Eng.*, 2022(4):422–441, 2022.

[32] Dejun Ning, Xihui Chen, Jiyan Chen, Tao Meng, Biao Xu, and Huai Zhang. Ppomixclip: An energy scheduling algorithm for low-carbon parks. *Energy Reports*, 12:4195–4207, 2024.

[33] Sampson E Nwachukwu, Komla A Folly, and Kehinde O Awodele. Soft actor-critic-based mppt control of solar pv systems under partial shading conditions. *IEEE Open Access J. Power Energy*, 2025.

[34] Weng Ho Yew, Chien Fat Chau, Ahmad Wafi Mahmood Zuhdi, Wan Syakirah Wan Abdullah, Weng Kean Yew, and Nowshad Amin. Investigating the performance of deep reinforcement learning-based mppt algorithm under partial shading condition. In *2023 IEEE Regional Symposium on Micro and Nanoelectronics (RSM)*, pages 9–12. IEEE, 2023.

[35] Diana Ortiz-Munoz, David Luviano-Cruz, Luis A Perez-Dominguez, Alma G Rodriguez-Ramirez, and Francesco Garcia-Luna. Hybrid fuzzy–ddpg approach for efficient mppt in partially shaded photovoltaic panels. *Appl. Sci.*, 15(9):4869, 2025.

[36] Jayandi Panggabean, Nana Sutisna, Infall Syafalni, and Trio Adiono. Comparison of mppt based on deep reinforcement learning by dqn, ddpg and td3. In *2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 261–266. IEEE, 2023.

[37] Bao Chau Phan, Ying-Chih Lai, and Chin E Lin. A deep reinforcement learning-based mppt control for pv systems under partial shading condition. *Sensors*, 20(11):3039, 2020.

[38] Wenkai Pan, Chenggang Cui, and Hui Chen. Research on photovoltaic mppt technique based on deep reinforcement learning under varying irradiance levels. In *2023 8th International Conference on Power and Renewable Energy (ICPRE)*, pages 1794–1799. IEEE, 2023.

[39] Luis Avila, Mariano De Paula, Maximiliano Trimboli, and Ignacio Carlucho. Deep reinforcement learning approach for mppt control of partially shaded pv systems in smart grids. *Appl. Soft Comput.*, 97:106711, 2020.

[40] Radhika Guntupalli, M Sudhakaran, et al. Modeling & implementation of drla based partially shaded solar system integration with 3-ϕ conventional grid using constant current controller. *Heliyon*, 8(6), 2022.

[41] Sampson E Nwachukwu, Komla A Folly, and Kehinde O Awodele. A comparative

study between soft actor-critic (sac) and deep deterministic policy gradient (ddpg) algorithms for solar pv mppt control under partial shading conditions. *IEEE Access*, 2025.

[42] Anis Ur Rehman, Zia Ullah, Hasan Saeed Qazi, Hany M Hasanien, and Haris M Khalid. Reinforcement learning-driven proximal policy optimization-based voltage control for pv and wt integrated power system. *Renew. Energy*, 227:120590, 2024.

[43] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.

[44] The MathWorks, Inc. Matlab r2022b (version 9.13). https://de.mathworks.com/help/reinforcementlearning/ref/rl.agent.rlddpgagent.html, 2022. Accessed: Jul. 19, 2025.

[45] N Prabaharan and K Palanisamy. Analysis and integration of multilevel inverter configuration with boost converters in a photovoltaic system. *Energy Convers. Manag.*, 128:327–342, 2016.

[46] Yaduvir Singh and Nitai Pal. Reinforcement learning with fuzzified reward approach for mppt control of pv systems. *Sustain. Energy Technol. Assessments*, 48:101665, 2021.

[47] Eneko Artetxe, Jokin Uralde, Oscar Barambones, Isidro Calvo, and Imanol Martin. Maximum power point tracker controller for solar photovoltaic based on reinforcement learning agent with a digital twin. *Mathematics*, 11(9):2166, 2023.

[48] José R Angulo, Brando X Calsi, Luis A Conde, Jorge A Guerra, Emilio Muñoz, Juan de la Casa, and Jan A Töfflinger. Estimation of the effective nominal power of a photovoltaic generator under non-ideal operating conditions. *Sol. Energy*, 231:784–792, 2022.

[49] Enrica Scolari, Fabrizio Sossan, and Mario Paolone. Photovoltaic-model-based solar irradiance estimators: Performance comparison and application to maximum power forecasting. *IEEE Trans. on Sustain. Energy*, 9(1):35–44, 2017.

[50] Sadeq D Al-Majidi, Maysam F Abbod, and Hamed S Al-Raweshidy. A particle swarm optimisation-trained feedforward neural network for predicting the maximum power point of a photovoltaic array. *Eng. Appl. Artif. Intell.*, 92:103688, 2020.

About Authors

Big-Alabo Ameze is an Alexandra Humboldt research fellow at the E. ON Energy research Centre, Germany. She is also a Senior lecturer at the University of Port Harcourt, Nigeria. Her areas of specialization include Power Systems, Renewable Energy, and Thermoelectricity. She obtained her B.Eng and M.Eng (2008) at the University of Port Harcourt, Nigeria. Msc in Advanced Control and Systems Engineering, (2011) at the University of Manchester, UK and PhD in Electronic and Electrical Engr (2017) University of Glasgow, Scotland, UK. Her research interest include applications of deep reinforcement learning in renewable energy systems. She can be contacted at email: ameze.big-alabo@eonerc.rwth-aachen.de.