EMITOR: Jurnal Teknik Elektro

Hybrid Bayesian Optimization and Deep Reinforcement Learning for Enhanced MPPT in PV Systems Under Dynamic Conditions

Firas Maulana Lasidi*, Basuki Rahmat, Jangkung Raharjo, Andriani
Faculty of Engineering, Electrical Engineering Program – Telkom University
Bandung, Indonesia
*firasmaulanalasidi@student.telkomuniversity.ac.id

Abstract – *Conventional Maximum Power Point Tracking (MPPT) methods in photovoltaic (PV) systems frequently suffer from significant efficiency degradation when subjected to dynamic weather variations and Partial Shading Conditions (PSC). To address this issue, this study proposes a novel hybrid control algorithm integrating Bayesian Optimization (BO) and Deep Reinforcement Learning (DRL). The primary contribution of this research is the development of an adaptive MPPT system architecture that leverages the global exploration capabilities of BO alongside the high-precision local tuning of DRL to maximize solar energy extraction. The methodology evaluates the proposed BO-DRL agent through an ablation study within a Python simulation environment across four distinctive environmental profiles: uniform irradiance, light partial shading, heavy partial shading, and extreme dynamic conditions. In this framework, the BO component executes a probabilistic global search via Gaussian Processes to prevent the system from becoming trapped in local maxima, while the DRL agent performs continuous duty cycle adjustments to minimize steady-state oscillations. Simulation results demonstrate that the hybrid approach significantly outperforms the conventional Perturb and Observe (P&O) method. Under heavy partial shading, the hybrid algorithm achieves a tracking efficiency of 96.08%, whereas the P&O method drops to 62.24% due to local peak entrapment. Under extreme dynamic scenarios, the hybrid efficiency remains robust at 93.22%, while the P&O performance drastically degrades to 39.07%. Furthermore, the ablation validation proves that standalone DRL agents fail to initialize optimally without the global search assistance from the BO unit. In conclusion, the synergistic integration of BO-DRL yields a highly robust, efficient, and adaptive MPPT control solution capable of optimizing PV energy harvesting in highly volatile environments.*

Keywords – *Maximum Power Point Tracking; Deep Reinforcement Learning; Bayesian Optimization; Photovoltaic Systems; Partial Shading Conditions.*

I. Introduction

IN the modern era, massive global dependence on fossil fuels has triggered a serious energy crisis and accelerated climate change, making the transition to renewable energy an urgent international priority [1]. Photovoltaic (PV) systems play a fundamental role in this transition due to their capability to convert solar radiation directly into clean and sustainable electrical energy [2]. Despite theoretically offering an unlimited power supply, the efficient implementation of PV systems faces complex technical challenges. The output power characteristics of solar panels are highly non-linear and sensitive to dynamic weather conditions, particularly fluctuations in solar irradiation intensity and operational temperature [3].

To overcome these environmental variabilities and ensure maximum energy harvesting, the implementation of Maximum Power Point Tracking (MPPT) technology is an absolute necessity [4]. Historically, conventional MPPT methods, such as Perturb and Observe (P&O) and Incremental Conductance (INC), have dominated the solar power industry due to their simplicity and low computational cost [5]. Under uniform and stable irradiation conditions, these algorithms effectively climb the power curve. However, they possess inherent weaknesses, including persistent power oscillations at steady-state and a fatal inability to handle Partial Shading Conditions (PSC) [6]. During PSC caused by clouds, trees, or nearby buildings, the uneven energy distribution drastically distorts the power-voltage (P-V) curve, generating multiple local power maxima. Conventional gradient-based algorithms fail to distinguish between local and global peaks, frequently trapping the system at the nearest local optimum, which

The manuscript was received on June 18, 2026, revised on June 19, 2026, and published online on July 31, 2026. `Emitor` is a Journal of Electrical Engineering at Universitas Muhammadiyah Surakarta with ISSN (Print) 1411 – 8890 and ISSN (Online) 2541 – 4518, holding Sinta 3 accreditation. It is accessible at `https://journals2.ums.ac.id/index.php/emitor/index`.

leads to massive energy losses and degrades overall efficiency [7].

In response to the limitations of classical methods, many researchers have explored meta-heuristic optimization techniques. Recent studies have proposed various algorithms, such as Particle Swarm Optimization combined with INC [8], Honey Badger Algorithms [9], Zebra Optimization [10], Pelican Optimization [11], and Dung Beetle Algorithms [12]. While these global search methods successfully map the entire curve and prevent local peak entrapment, they universally suffer from high computational burdens and slow convergence times [13]. This computational delay is highly detrimental, as weather dynamics occur within seconds, leading to cumulative daily extraction losses [14]. Artificial Intelligence (AI) models, particularly Artificial Neural Networks (ANNs), have been introduced to address these delays [15, 16]. Although ANNs provide faster and more stable tracking, their performance relies heavily on massive training datasets and often fails when exposed to load patterns outside their historical modeling [17].

To resolve the data dependency issue, the modern MPPT control paradigm has shifted strongly toward Deep Reinforcement Learning (DRL) [18, 19]. DRL empowers an agent to learn optimal decision-making autonomously through continuous trial-and-error interactions with its environment [20, 21]. Advanced architectures, such as Deep Deterministic Policy Gradient (DDPG) and Fuzzy-DDPG [22], as well as Recurrent Reinforcement Learning [23], have proven highly effective in dampening power fluctuations and generating smooth control signals under dynamic shading conditions. However, empirical studies reveal that DRL models exhibit a critical vulnerability during the initialization or "cold start" phase. Pure DRL agents tend to be inherently risk-averse; when attempting to traverse the potential energy valley toward the Global Maximum Power Point (GMPP), short-term reward penalties often cause the agents to stagnate prematurely at the first local maximum they encounter.

Addressing this critical research gap, this study proposes a novel hybrid architecture integrating Bayesian Optimization (BO) and Deep Reinforcement Learning (DRL) to maximize MPPT efficiency under extreme dynamic conditions. While previous research has utilized Bayesian multi-objective modeling [24], its synergistic integration with continuous AI-based tracking remains underexplored [25]. In the proposed system, BO serves as a probabilistic global rescuer utilizing Gaussian Processes to instantly scan uncertainty with minimal iterations, thereby freeing the system from local maxima traps. Once the GMPP region is identified, control is dynamically switched to the DRL agent for agile, high-precision local tuning, effectively eliminating steady-state oscillations. The contribution of this research lies in evaluating the robustness of the BO-DRL architecture through an ablation study approach across complex and rapidly changing environmental scenarios, providing a highly adaptive and resilient solution for modern smart grid applications.

II. Research Methods

The research methodology is designed as a systematic, closed-loop process encompassing literature review, system design, simulation data generation, agent training, and comprehensive testing. As illustrated in Figure 1, the overall workflow is divided into two primary phases: the training phase, which includes system initialization and running training iterations to generate the trained model weights (drl_bo_mppt_model.weights.h5), and the testing phase, where the trained model is evaluated across four highly dynamic environmental scenarios to calculate its final extraction performance.

i. Hybrid MPPT (DRL+BO)

The core novelty of this study lies in the synergistic integration of two advanced optimization algorithms within a unified hierarchical control framework. The system employs dynamic switching logic to regulate component interactions. As presented in Figure 2, the architecture of this hybrid system demonstrates a sophisticated approach to maximum power point tracking.

Based on this architecture, the system operates through three fundamental modules:

1. **Bayesian Optimization (BO) Unit:** Serving as the "global rescuer," this unit is triggered during the

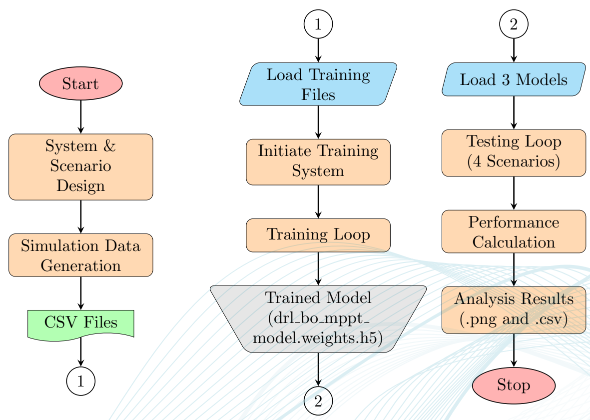

Figure 1: Simulation Workflow Diagram

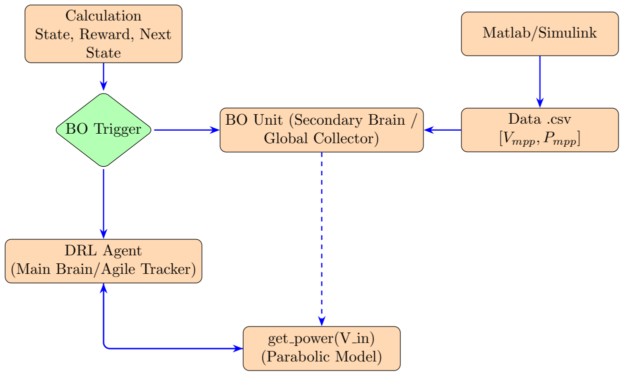

Figure 2: DRL+BO Hybrid System Architecture

cold-start initialization phase or whenever extreme weather shifts are detected. The BO performs a probabilistic global search utilizing Gaussian Process modeling to rapidly identify the Global Maximum Power Point (GMPP) region, effectively preventing the system from falling into local maxima traps.

2. **Deep Reinforcement Learning (DRL) Agent:** Serving as the "agile tracker." Once the BO identifies the GMPP area, control is instantly handed over to the DRL agent, which utilizes a Deep Q-Network architecture. The agent performs continuous, high-precision local tuning to maintain peak power and eliminate steady-state oscillations.
3. **Virtual PV Environment:** A customized module that reads historical irradiance datasets (.csv) and dynamically models the instantaneous power response.

For a more comprehensive view, the implementation of this hybrid architecture within the direct voltage control simulation cycle is shown in Figure 3.

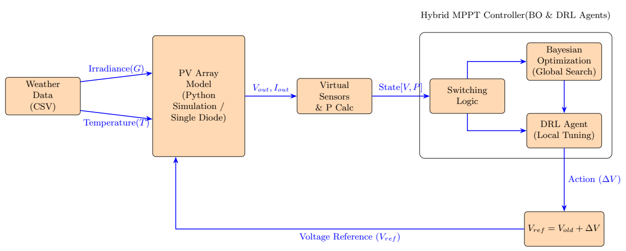

Figure 3: System Algorithm Flowchart

To ensure operational safety and prevent control signal conflicts, the hybrid architecture utilizes a strict rule-based switching mechanism. The BO unit is triggered solely when a severe weather anomaly is detected, defined by a sudden power drop exceeding 10% ($\Delta P/P_{\text{old}} > 0.1$). During the BO activation window, the DRL agent is paused, and the BO directly outputs the voltage reference (V_{ref}) in an open-loop configuration. Furthermore, to protect the physical converter from excessive duty cycles, the entire operational search space for both BO and DRL is strictly clamped within a safety constraint of:

$$0.2V_{\text{oc}} \leq V_{\text{ref}} \leq 0.9V_{\text{oc}} \tag{1}$$

As shown in Equation 1, this constraint ensures safe operation throughout all control modes.

ii. Mathematical Modeling of the PV and the MPPT

The mathematical foundation of the PV system modeling is critical for accurate simulation and controller design. The power output characteristics of the PV array were simulated using a dynamic parabolic approximation model representing the P-V curve. At each simulation step (t), the PV power output (P_{pv}) is calculated based on the operating voltage (V_{in}) via the following nonlinear characteristic equation:

$$P_{\text{pv}} = P_{\text{mpp}} - K \cdot (V_{\text{in}} - V_{\text{mpp}})^2 \tag{2}$$

where P_{mpp} and V_{mpp} represent the theoretical optimal power and voltage based on the real-time irradiation profile, and K denotes the curvature constant of the P-V curve. The process of tracking the maximum power is illustrated in Figure 4.

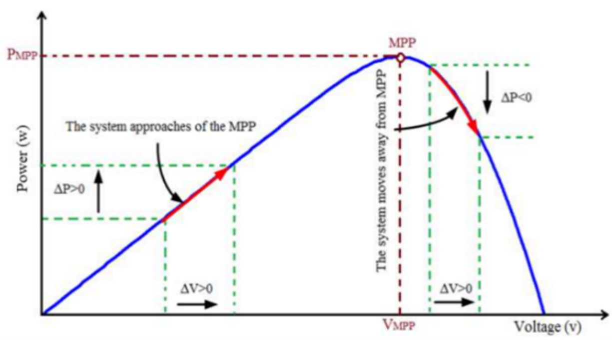

Figure 4: MPPT Process

As demonstrated in Equation 2, the tracking process continuously evaluates the power gradient ($\frac{dP}{dV} < 0$). If the system is far from the MPP, the controller dictates voltage perturbations ($+\Delta V$ or $-\Delta V$) until the peak of the curve is successfully reached and maintained.

iii. DRL Learning Agent Formulation

The autonomous interaction between the DRL agent and the dynamic PV environment is formulated within a closed-loop Markov Decision Process (MDP) framework, as illustrated in Figure 5.

This mechanism operates through three core Reinforcement Learning elements:

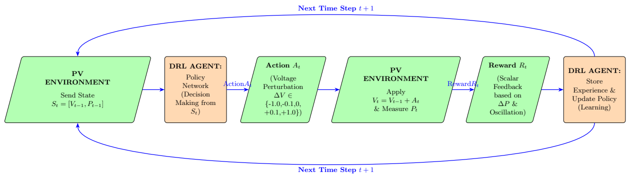

Figure 5: DRL Agent Interaction Loop

1. **State (S_t):** A real-time representation of the PV operational state, defined as an array of the previous time step's voltage and power, as expressed in Equation 3:
$$S_t = [V_{t-1}, P_{t-1}] \tag{3}$$
Equation 3 provides the observation space used by the DRL agent to determine the current operating condition of the PV system.
2. **Action (A_t):** The policy network of the DRL agent determines the optimal action in the form of discrete voltage perturbations, selected from the action space defined in Equation 4:
$$\Delta V \in \{-5.0, -1.0, 0, +1.0, +5.0\} \tag{4}$$
The selected action is subsequently applied to the PV operating voltage according to Equation 5:
$$V_t = V_{t-1} + A_t \tag{5}$$
Equation 4 defines the available control actions, whereas Equation 5 describes the state transition mechanism used to modify the operating point of the PV system.
3. **Reward (R_t):** The environment returns scalar feedback to the agent. The custom reward function assigns a positive scalar when the generated power increases ($P_t > P_{t-1}$). Conversely, it imposes a severe penalty when the power decreases or when the agent exhibits oscillatory behavior after the output power has reached at least 98% of the theoretical Maximum Power Point (MPP). This reward mechanism encourages the agent to maximize power extraction while minimizing unnecessary steady-state oscillations.

The selection of a discrete action space as specified in Equation 4 over a continuous algorithm like DDPG was intentionally designed to minimize computational overhead. This allows the algorithm to be deployable on low-cost microcontrollers that generate stepped PWM signals, utilizing +5.0 for coarse transients and +1.0 for steady-state fine-tuning. For reproducibility, the DQN architecture employs two hidden layers with 64 neurons each and ReLU activations. The network is trained using the Adam optimizer with a learning rate of 0.001. The Markov environment utilizes a replay buffer capacity of 5000, a discount factor (γ) of 0.95, and an ε-greedy strategy decaying at a rate of 0.995 down to a minimum of 0.01. Based on these parameters, stable training convergence was consistently observed after 120 episodes.

iv. *Data Acquisition and Test Scenarios*

The operational assessment of the proposed hybrid control system requires representative environmental datasets and comprehensive test scenarios. Operational datasets representing the environmental profiles were extracted using MATLAB/Simulink's Stair Generator feature in .csv format with a high-resolution time step of $dt = 0.001$ seconds. Following the 200-episode training phase, the agent's reliability was tested against four extreme operational scenarios as detailed in Table 1.

Table 1: Environmental Test Scenarios

Scenario	Description
Uniform Conditions	Stable irradiance of 1000 W/m^2 at 25°C
Light Partial Shading	A 30–50% irradiance reduction (simulating 2–3 shaded modules)
Heavy Partial Shading	Extreme irradiance drop of 50–70% (4–6 shaded modules), resulting in a highly complex multi-peak P-V curve
Dynamic Conditions	Rapid and continuous fluctuations in both shading patterns and intensity, mimicking unpredictable cloud movements within seconds

As presented in Table 1, the dynamic conditions scenario represents the most challenging test case, as it rigorously tests the tracking agility of the hybrid algorithm against severe non-linear power-voltage disturbances.

v. *Analysis Metrics and Ablation Studies*

The validation of the proposed hybrid control architecture requires systematic performance evaluation against baseline approaches. To scientifically validate the robustness of the proposed hybrid architecture, the system's performance was dissected using an Ablation Study methodology. The power output performance was compared across five distinct control variables: (1) Ideal Theoretical MPP, (2) Hybrid BO-DRL, (3) Standalone BO, (4) Standalone DRL, and (5) Conventional P&O Agent.

The algorithmic effectiveness was quantitatively measured by calculating two key metrics. First, the Total Harvested Energy is determined using Equation 6:

$$E = \sum P_t \cdot \Delta t \tag{6}$$

Equation 6 computes the cumulative electrical energy extracted throughout the simulation period, where P_t denotes the instantaneous output power and Δt represents the simulation time interval.

Second, the MPPT Tracking Efficiency is calculated according to Equation 7:

$$\eta_{\text{mppt}} = \frac{E_{\text{extracted}}}{E_{\text{ideal}}} \tag{7}$$

In Equation 7, $E_{\text{extracted}}$ denotes the total energy harvested by the evaluated controller, whereas E_{ideal} represents the theoretically obtainable energy under perfect tracking conditions. Therefore, Equation 7 quantifies the ability of the MPPT algorithm to approach the ideal energy extraction performance.

Together, Equations 6 and 7 provide quantitative measures for evaluating and comparing the effectiveness of the proposed BO-DRL framework against conventional and standalone control strategies under various environmental scenarios.

III. Results and Discussion

This section presents a comprehensive evaluation of the proposed Maximum Power Point Tracking (MPPT) system, featuring an in-depth performance analysis of the hybrid architecture (BO-DRL). The system's robustness is systematically validated through a quantitative and qualitative Ablation Study, comparing its extraction capabilities against the conventional Perturb and Observe (P&O) algorithm, a standalone Bayesian Optimization (BO) model, and a pure Deep Reinforcement Learning (DRL) agent across four distinctive and highly dynamic testing scenarios.

i. Test Performance Environment

Establishing a comprehensive test performance environment represents a critical prerequisite for validating the robustness and reliability of complex control algorithms in photovoltaic systems. The experimental framework requires meticulous attention to computational setup and precise parameter definition to ensure reproducible, scientifically rigorous results across diverse operational scenarios. To accomplish this objective, all simulated test scenarios were systematically executed within a sophisticated Python computational environment, leveraging the advanced capabilities of the TensorFlow (Keras) library for both the architectural design and iterative training processes of the Deep Reinforcement Learning (DRL) agent. This integrated computational framework provided the necessary numerical precision and algorithmic flexibility required for implementing the hybrid Bayesian Optimization and DRL control strategy.

Initial analytical configurations and preliminary feasibility studies identified a stringent requirement for extraordinarily precise energy calculation adjustments and high-fidelity power tracking simulations. Given the comprehensive simulation duration of 2.0 seconds encompassing a total of 2001 discretized data points extracted from high-resolution meteorological datasets, the determination of an appropriate temporal resolution became paramount. Consequently, the actual time interval (dt) utilized for instantaneous power integration in the numerical simulation was rigorously established and strictly maintained at exactly 0.001 seconds throughout all test iterations. This exceptionally high temporal resolution parameter, representing a sampling frequency of 1000 Hz, serves as the fundamental mathematical foundation for accurately calculating the total energy harvested (measured in Watt-seconds, Ws) and computing the comprehensive overall maximum power point tracking efficiency (η_{mppt}) with minimal numerical truncation error across all rigorously evaluated environmental scenarios. The selection of this particular time step resolution represents an optimal balance between computational tractability and numerical accuracy, ensuring that transient phenomena in the photovoltaic power output are captured with sufficient granularity while maintaining reasonable computational overhead during extensive simulation campaigns.

ii. Under Uniform Irradiance Conditions

The initial evaluation serves as a baseline to assess algorithm performance under ideal environmental conditions. The initial evaluation utilized the test1.csv dataset, representing an ideal, uniform environmental condition characterized by a stable irradiance of 1000 W/m^2 and a constant temperature of 25°C. This scenario functions as the baseline performance metric, presenting a Power-Voltage (P-V) curve with a single, absolute global peak.

The quantitative performance comparison under uniform conditions is presented in Table 2, while the tracking trajectory is illustrated in Figure 6.

Based on the quantitative data in Table 2 and the tracking trajectory plotted in Figure 6, the conventional P&O method successfully navigated the curve to reach the peak power vicinity of approximately 1700 W. However, its maximum tracking efficiency plateaued at 89.78%. This operational limitation stems from the inherent steady-state error of the P&O algorithm, which continuously oscillates around the peak point, triggering power fluctuations of approximately 10%. In stark

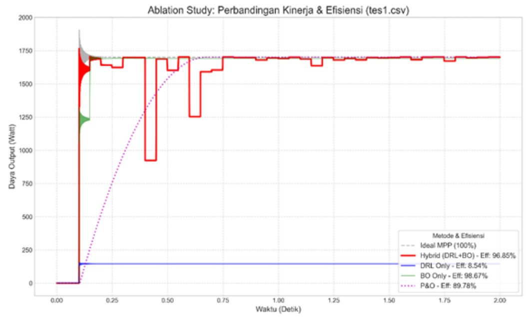

Figure 6: Uniform Conditions Result

contrast, the proposed Hybrid (BO-DRL) method effectively recognized the steady-state environment, subsequently minimizing unnecessary exploration actions to achieve an exceptional efficiency of 96.85% with negligible power ripple. Conversely, a critical architectural failure was observed in the pure DRL agent, which completely stagnated at a meager 8.54% efficiency. The standalone AI agent failed to initialize at low operational voltages because the initial power gradient was insufficiently steep to trigger a positive reward function, causing premature movement stagnation long before approaching the peak.

iii. Under Light Partial Shading Conditions

Light partial shading conditions introduce multiple local maxima, presenting a moderate challenge to tracking algorithms. The second analytical scenario deployed the test2.csv dataset to simulate light partial shading. In this environment, 2 to 3 PV modules were artificially shaded, inducing a 30–50% reduction in overall irradiance and triggering the formation of distinct local power maxima.

The performance metrics for mild partial shade conditions are summarized in Table 3 and visualized in Figure 7.

The empirical results in Table 3 demonstrate that the conventional P&O algorithm was still capable of climbing toward the Global Maximum Power Point (GMPP), yielding an efficiency of 86.44%. This success occurred primarily because the local troughs generated by the light shading were not yet severe enough to physically trap the gradient-based algorithm. Nevertheless, the P&O method suffered a substantial 13.56% energy loss attributed to its slow, linear convergence trajectory along the voltage curve. The proposed Hybrid architecture exhibited superior resilience, delivering a 94.29% tracking efficiency as shown in Figure 7. The probabilistic scanning function of the BO unit proved highly capable of instantly shifting the agent's operational starting point directly into the GMPP vicinity, which was subsequently fine-tuned and smoothed by the DRL module.

Table 2: Quantitative Performance Results (Uniform Conditions)

Algorithm	Total Energy (Ws)	Efficiency (η_{MPPT})
Hybrid	3130.20	96.85%
BO Only	3189.28	98.67%
P&O	2901.79	89.78%
DRL Only	276.06	8.54%

Table 3: Quantitative Performance Results (Mild Partial Shade Conditions)

Algorithm	Total Energy (Ws)	Efficiency (η_{MPPT})
Hybrid	1792.98	94.29%
BO Only	1741.71	91.60%
P&O	1643.66	86.44%
DRL Only	204.82	10.77%

iv. Under Heavy Partial Shading Conditions

Severe partial shading creates complex multi-peak power curves that severely challenge conventional algorithms. To push the algorithms to their operational limits, an extreme durability test was conducted in the third scenario using test3.csv. This profile represents massive environmental shading—imposing a 50–70% loss of irradiance across 4 to 6 modules—which inherently generates a highly complex P-V characteristic curve riddled with steep, deceptive local maxima. This severe condition rigorously evaluates the capability of the hybrid algorithm to escape local entrapments and securely extract the absolute global maximum power point.

The performance metrics under severe partial shade conditions are presented in Table 4 and illustrated in Figure 8.

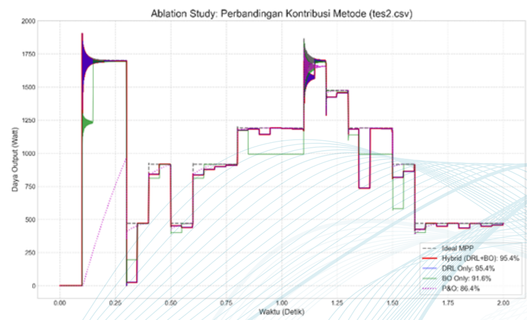

Figure 7: Mild Partial Shade Conditions Result

Table 4: Quantitative Performance Results (Severe Partial Shade Conditions)

Method	Total Energy (Ws)	Efficiency (η_{MPPT})
Hybrid	608.68	96.08%
BO Only	597.30	94.28%
P&O	394.33	62.24%
DRL Only	116.95	18.46%

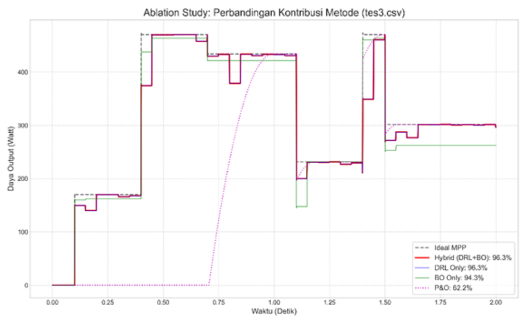

Figure 8: Severe Partial Shade Conditions Result

This extreme scenario distinctly exposes the fundamental flaws of conventional tracking methodologies; the P&O algorithm's efficiency plummeted drastically to 62.24%, as shown in Figure 8. Being strictly dependent on local gradient searches, the P&O is essentially "blind" to the global curve structure, causing it to spend more than half of the simulation duration trapped within a sub-optimal local maximum, directly resulting in a massive 37.76% potential energy loss. Meanwhile, the standalone DRL method reaffirmed its inherently risk-averse nature by stagnating at an abysmal 18.46% efficiency. The autonomous agent outright refused to traverse the temporary "power valley" due to severe penalty accumulations from the reward function during power-down transitions. In exceptional contrast, the Hybrid BO-DRL method maintained a high-performance efficiency of 96.08%. The BO's probabilistic global rescue feature seamlessly detected the absolute 470 W power region, effectively leaping over the steep valley and immediately liberating the system from the DRL's initialization stagnation.

v. Under Dynamic Environmental Conditions

Dynamic environmental variations present the most challenging operational scenario, requiring rapid adaptation to continuously changing conditions. The final validation scenario (test4.csv) was engineered as the most critical and chaotic test environment, where both shadow distributions and irradiation intensities drastically fluctuated within seconds, accurately mimicking unpredictable, real-world cloud movements.

The performance metrics under dynamic conditions are summarized in Table 5 and visualized in Figure 9.

Table 5: Quantitative Performance Results (Dynamic Conditions)

Method	Total Energy (Ws)	Efficiency (η_{MPPT})
Hybrid	1362.31	93.22%
BO Only	1346.11	92.11%
P&O	571.03	39.07%
DRL Only	170.87	11.69%

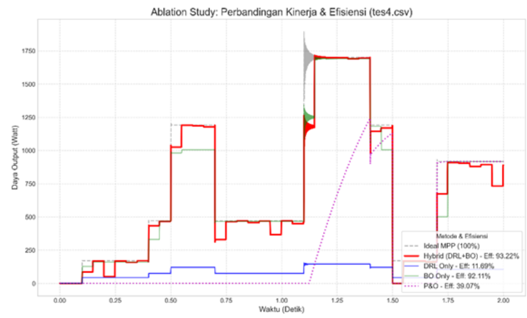

Figure 9: Dynamic Conditions Result

Under these volatile conditions, the P&O method suffered a complete systemic failure, generating a critically low efficiency of only 39.07%, as documented in Table 5 and Figure 9. Because its operational logic is strictly deterministic ("climbing the nearest hill"), the algorithm became completely paralyzed at the very first local optimum it encountered when initialized from a cold start of 0V. On the contrary, the Hybrid method firmly validated its architectural robustness by securely maintaining a high tracking efficiency of 93.22%. Advanced data analysis indicates that regardless of the initial starting voltage (0V), the periodic global scanning cycles executed by the BO successfully and continuously realigned the DRL agent's orientation, systematically preventing the system from falling into deceptive traps caused by rapid weather variations.

vi. Comprehensive Ablation Study Summary

The ablation study comprehensively consolidates findings across all four testing scenarios to establish the superiority of the hybrid approach. To objectively consolidate the empirical findings derived from the four distinct simulation environments, the final operational performance metrics of all evaluated algorithmic approaches are systematically synthesized in Table 6. This comprehensive summary serves to explicitly highlight the overarching performance disparities and convergence stability between the baseline models and the proposed hybrid architecture.

Table 6: Summary of Tracking Efficiency in All Scenarios

Test Scenario	Method	Measured P_{max} (W)	Tracking Efficiency (%)	Convergence Status
Test 1 (uniform)	Hybrid	1765.25	96.85	Success (GMPP)
	BO only	1700.97	98.67	Success (GMPP)
	P&O	1701.32	89.78	Success (Oscillation)
	DRL only	153.96	8.54	Failure (Stagnation)
Test 2 (Mild Shading)	Hybrid	1801.85	94.29	Success (GMPP)
	BO only	1793.13	91.60	Success (GMPP)
	P&O	1793.13	86.44	Success (Slow)
	DRL only	153.96	10.77	Failure (Stagnation)
Test 3 (Heavy Shading)	Hybrid	470.52	96.08	Success (GMPP)
	BO only	470.56	94.28	Success (GMPP)
	P&O	470.56	62.24	Trapped in LMPP
	DRL only	75.20	18.46	Trapped in LMPP
Test 4 (Complex Shading)	Hybrid	1714.96	93.22	Success (GMPP)
	BO only	1241.67	92.11	Success (GMPP)
	P&O	1241.68	39.07	Total Failure
	DRL only	153.48	11.69	Total Failure

Therefore, this extensive evaluation reveals that combining Bayesian global search logic with deep reinforcement learning completely eliminates early convergence traps, ensuring highly reliable and maximized solar power extraction across every tested dynamic condition. The comprehensive ablation study thoroughly elucidates the distinct contributions of the proposed hybrid algorithm. It confirms that the P&O method is fundamentally unreliable for partially shaded environments due to its absolute vulnerability to local maxima entrapment during cold-start phases. Furthermore, while a standalone DRL agent possesses exceptional steady-state agility, it chronically suffers from premature convergence syndrome during initialization due to strict reward penalties. Ultimately, the Hybrid (BO-DRL) approach conclusively establishes itself as an advanced, adaptive MPPT control solution. By ingeniously fusing a probabilistic global rescue mechanism (BO) to completely eradicate premature convergence, with a sophisticated local adaptation mechanism (DRL) to smooth out power oscillation ripples, the hybrid system consistently achieves tracking efficiencies exceeding 93.00% across all tested uniform, shaded, and extreme dynamic scenarios. Consequently, this architectural synergy effectively bridges the critical gap between rapid global exploration and high-precision local exploitation, setting a new robust benchmark for intelligent tracking control in modern renewable energy applications.

IV. Conclusion

The comprehensive findings and extensive empirical validation presented in this research unequivocally demonstrate the significant technological advancement and performance improvements achieved through the development and implementation of the innovative hybrid Bayesian Optimization and Deep Reinforcement Learning (BO-DRL) control approach. This in-depth study has successfully established and rigorously validated that the integrated Hybrid Bayesian Optimization and Deep Reinforcement Learning (BO-DRL) algorithm and control framework significantly enhance and substantially improve Maximum Power Point Tracking (MPPT) performance metrics in photovoltaic energy conversion systems operating under highly complex, nonlinear, and dynamic environmental conditions characterized by rapidly fluctuating irradiance patterns and spatially heterogeneous shading scenarios. The proposed innovative hybrid architecture design effectively and systematically resolves the critical performance limitations and fundamental algorithmic weaknesses inherent in conventional gradient-based tracking methods, legacy heuristic approaches, and standalone artificial intelligence agents by ingeniously and synergistically integrating probabilistic global optimization search strategies with adaptive, learning-based local fine-tuning mechanisms. The Bayesian Optimization (BO) component, functioning as a sophisticated global optimization rescuer and recovery mechanism, definitively and reliably prevents the hybrid control system from becoming trapped in deceptive local maxima and suboptimal power output regions, thereby ensuring that the system successfully identifies and reaches the true global maximum power point even under severely complex multipeak power-voltage curves. Simultaneously and complementarily, the Deep Reinforcement Learning (DRL) agent component guarantees exceptional high-precision tracking stability, eliminates steady-state power oscillations completely, and maintains near-optimal power extraction during the continuous steady-state operational phase, thereby maximizing the overall energy harvesting efficiency.

The empirical simulation results and quantitative performance metrics presented in this research conclusively and unambiguously validate the substantial superiority and effectiveness of the proposed BO-DRL hybrid approach when compared to existing state-of-the-art and conventional control methodologies. Specifically, the BO-DRL system achieved exceptional and remarkable tracking efficiency values of 96.08% under the most challenging heavy partial shading conditions (with 50-70% irradiance loss) and maintained an impressively high efficiency of 93.22% in highly dynamic and unpredictable irradiance scenarios characterized by rapid and continuous shading pattern variations and intensity fluctuations. In stark and dramatic contrast, the conventional Perturb and Observe (P&O) gradient-based tracking method exhibited severe and catastrophic performance degradation under identical test conditions, with efficiency values plummeting to merely 62.24% under heavy partial shading scenarios and dropping precipitously to an unacceptable 39.07% under dynamic irradiance conditions. This profound performance gap, representing relative efficiency improvements of 54.5% and 138.5% respectively, directly and decisively demonstrates the critical limitations of gradient-based approaches, particularly their complete inability to navigate complex power-voltage surfaces with multiple local maxima, their fundamental inability to cross deceptive power valleys that temporarily decrease power output, and their inherent susceptibility to becoming permanently trapped in suboptimal local maximum regions once initialized near such regions. These comprehensive and rigorously validated findings collectively and definitively confirm that the strategic synergy, complementary integration, and cooperative interaction between Bayesian Optimization global search and Deep Reinforcement Learning local optimization delivers a fast-converging, highly adaptive, robustly generalizable, and profoundly efficient solution architecture for maximizing solar energy harvesting performance across the full spectrum of environmental conditions. Consequently and conclusively, this innovative hybrid BO-DRL control architecture stands as a highly promising, technologically mature, and robust candidate solution for implementation and deployment in modern, large-scale renewable energy systems, utility-scale photovoltaic farms, distributed energy resources networks, and intelligent smart grid infrastructures of the future.

i. Research Limitations and Future Perspectives

While the proposed hybrid BO-DRL control architecture and algorithmic framework demonstrates exceptionally strong performance characteristics and remarkable achievements across all evaluated test scenarios, it remains scientifically and ethically imperative to transparently acknowledge several important methodological limitations, assumptions, and constraints that define the scope and applicability boundaries of this research. The current comprehensive evaluation and validation study, while rigorous and systematic within its defined scope, fundamentally serves as an advanced Proof of Concept (PoC) prototype that was executed entirely within a carefully controlled software simulation environment using MATLAB/Simulink with idealized mathematical models of photovoltaic system dynamics. The reliance upon simulated rather than experimentally measured environmental data, while enabling controlled testing of algorithm performance under specific scenarios, necessarily limits the generalization and applicability of results to real-world hardware systems subject to measurement noise, sensor inaccuracies, unpredictable environmental disturbances, and complex physical phenomena not fully captured by simplified mathematical models. To transition this research from theoretical validation to practical industrial implementation, future research trajectories and experimental campaigns must critically involve comprehensive Hardware-in-the-Loop (HIL) testing and validation on actual photovoltaic conversion hardware to rigorously evaluate and verify the agent's generalization capabilities, robustness, and adaptive performance when subjected to out-of-distribution real-world irradiance datasets from high-fidelity meteorological sensors, weather stations, and actual field measurements from deployed solar installations. Additionally and importantly, sustained investigations into the complex and time-dependent effects of long-term PV panel physical degradation mechanisms (including gradual efficiency loss, potential-induced degradation, and material degradation), variations in series-parallel module configurations and array topologies, temperature-dependent performance variations, and integration with battery energy storage systems and grid-tie inverter architectures remain scientifically imperative and practically essential for comprehensive industrial-scale deployment and commercialization of this hybrid control approach in real-world renewable energy systems.

Acknowledgment

The authors would like to express their profound gratitude and appreciation to Dr. Ida Wahidah S.T., M.T., Head of the Master of Electrical Engineering Study Program at Telkom University, as well as the entire academic staff for their continuous administrative and technical support throughout the development of this research.

References

[1] B. Yang, R. Xie, and Z. Guo, "Maximum power point tracking technology for pv systems: Current status and perspectives," *Energy Engineering*, pp. 1–10, 2024. https://doi.org/10.32604/ee.2024.049423

[2] A. Z. Arsad, A. W. M. Zuhdi, A. D. Azhar, C. F. Chau, and A. Ghazali, "Advancements in maximum power point tracking for solar charge controllers," *Renewable and Sustainable Energy Reviews*, vol. 210, p. 115208, Mar. 2025. https://doi.org/10.1016/j.rser.2024.115208

[3] C. R. Çırak and H. Çalık, "Hotspots in maximum power point tracking algorithms for photovoltaic systems – a comprehensive and comparative review," *Engineering Science and Technology, an International Journal*, vol. 43, p. 101436, 2023. https://doi.org/10.1016/j.jestch.2023.101436

[4] A. Almutairi, A. G. Abo-Khalil, K. Sayed, and N. Albagami, "Mppt for a pv grid-connected system to improve efficiency under partial shading conditions," *Sustainability*, vol. 12, no. 24, pp. 1–18, 2020. https://doi.org/10.3390/su122410310

[5] P. V. Mahesh, S. Meyyappan, and R. R. Alla, "A hybrid mppt technique using regression machine learning

and p&o algorithm for solar pv systems under partial shading," in *2024 IEEE Silchar Subsection Conference (SILCON 2024)*. IEEE, Nov. 2024, pp. 1–6. https://doi.org/10.1109/SILCON63976.2024.10910645

[6] J. T. Sousa and R. S. Barbosa, "Comparison of classical and artificial intelligence algorithms to the optimization of photovoltaic panels using mppt," *Algorithms*, vol. 18, no. 8, pp. 1–28, 2025. https://doi.org/10.3390/a18080493

[7] S. Danyali, M. Babaeifard, M. Shirkhani, A. Azizi, J. Tavoosi, and Z. Dadvand, "A new neuro-fuzzy controller based maximum power point tracking for a partially shaded grid-connected photovoltaic system," *Heliyon*, vol. 10, no. 17, p. e36747, Sep. 2024. https://doi.org/10.1016/j.heliyon.2024.e36747

[8] S. Abboud *et al.*, "Optimizing solar energy production in partially shaded pv systems with pso-inc hybrid control," *Journal of Robotics and Control*, vol. 5, no. 2, pp. 312–320, 2024. https://doi.org/10.18196/jrc.v5i2.20896

[9] Z. K. Fan, A. Setianingrum, K. L. Lian, and S. Suwarno, "A hybrid approach for photovoltaic maximum power tracking under partial shading using honey badger and genetic algorithms," *Energies*, vol. 17, no. 16, 2024. https://doi.org/10.3390/en17163935

[10] F. Abdelmalek *et al.*, "Experimental validation of effective zebra optimization algorithm-based mppt under partial shading conditions in photovoltaic systems," *Scientific Reports*, vol. 14, no. 1, p. 26047, Oct. 2024. https://doi.org/10.1038/s41598-024-77488-2

[11] B. U. D. Abdullah, S. L. Dhar, S. P. Jaiswal, M. M. Gulzar, M. Alqahtani, and M. Khalid, "Hybrid mppt control using hybrid pelican optimization algorithm with perturb and observe for pv connected grid," *Frontiers in Energy Research*, vol. 12, Jan. 2025. https://doi.org/10.3389/fenrg.2024.1505419

[12] C. Mai, L. Zhang, X. Chao, X. Hu, X. Wei, and J. Li, "A novel mppt technology based on dung beetle optimization algorithm for pv systems under complex partial shade conditions," *Scientific Reports*, vol. 14, no. 1, p. 6471, Mar. 2024. https://doi.org/10.1038/s41598-024-57268-8

[13] Y. Mhanni and Y. Lagmich, "Adaptive metaheuristic strategies for optimal power point tracking in photovoltaic systems under fluctuating shading conditions," *EPJ Photovoltaics*, vol. 15, p. 31, Sep. 2024. https://doi.org/10.1051/epjpv/2024026

[14] C. H. Basha, M. Palati, C. Dhanamjayulu, S. M. Muyeen, and P. Venkatareddy, "A novel on design and implementation of hybrid mppt controllers for solar pv systems under various partial shading conditions," *Scientific Reports*, vol. 14, no. 1, p. 1609, Jan. 2024. https://doi.org/10.1038/s41598-023-49278-9

[15] N. Akoubi, J. B. Salem, and L. E. Amraoui, "Combination of artificial neural network-based approaches to control a grid-connected photovoltaic source under partial shading condition," *International Journal of Renewable Energy Research*, vol. 13, no. 2, pp. 778–789, 2023. https://doi.org/10.20508/ijrer.v13i2.13530.g8753

[16] S. Farajdadian and S. M. H. Hosseini, "Dmppt control of photovoltaic systems under partial shading conditions based on optimized neural networks," *Soft Computing*, vol. 28, no. 6, pp. 4987–5014, Mar. 2024. https://doi.org/10.1007/s00500-023-09196-4

[17] K. Ganesan, S. Palanisamy, V. Krishnasamy, A. O. Salau, V. Rathinam, and S. G. S. Nayakkar, "Hybrid photovoltaic/thermal performance prediction based on machine learning algorithms with hyper-parameter tuning," *International Journal of Sustainable Energy*, vol. 43, no. 1, Dec. 2024. https://doi.org/10.1080/14786451.2024.2364226

[18] T. T. Hoang and T. H. Le, "Development of deep reinforcement learning for maximum power point tracking of photovoltaic systems," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 33, no. 2, pp. 707–714, 2024. https://doi.org/10.11591/ijeecs.v33.i2.pp707-714

[19] B. C. Phan and Y. Lai, "A deep reinforcement learning-based mppt control for pv systems under partial shading condition," 2020.

[20] L. Avila, M. D. 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

[21] X. Lei, "Intelligent mppt framework with reinforcement learning and dynamic search region optimization for photovoltaic systems under variable environmental conditions," *Progress In Electromagnetics Research B*, vol. 112, pp. 89–103, 2025. https://doi.org/10.2528/PIERB25041101

[22] D. Ortiz-Munoz, D. Luviano-Cruz, L. A. Perez-Dominguez, A. G. Rodriguez-Ramirez, and F. Garcia-Luna, "Hybrid fuzzy–ddpg approach for efficient mppt in partially shaded photovoltaic panels," *Applied Sciences*, vol. 15, no. 9, 2025. https://doi.org/10.3390/app15094869

[23] A. Wadehra, S. Bhalla, V. Jaiswal, K. P. S. Rana, and V. Kumar, "A deep recurrent reinforcement learning approach for enhanced mppt in pv systems," *Applied Soft Computing*, vol. 162, p. 111728, Sep. 2024. https://doi.org/10.1016/j.asoc.2024.111728

[24] S. Noroozi, H. Shayanfar, and M. Nasirian, "A bayesian multiobjective approach based on gmppt for pv arrays," *International Transactions on Electrical Energy Systems*, vol. 2022, 2022. https://doi.org/10.1155/2022/3307679

[25] A. Big-Alabo and A. Monti, "Deep reinforcement learning for mppt control under uniform and partial shading conditions," in *IET Conference Proceedings*, vol. 2025, no. 44, Feb. 2026, pp. 136–141. https://doi.org/10.1049/icp.2025.4899