Adel Elgammal. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 15, Issue 12, December 2025, pp 136-153

RESEARCH ARTICLE OPEN ACCESS

A High-Reliability and Adaptive MPPT–Inverter Control for PV Systems Using Reinforcement Learning Under Partial Shading Conditions

Adel Elgammal
Professor, Utilities and Sustainable Engineering, The University of Trinidad & Tobago UTT

Abstract:
Solar photovoltaic (PV) systems operating under Partial Shading Conditions (PSC) exhibit multiple local maxima on the power–voltage curve, causing conventional maximum power point tracking (MPPT) methods to suffer from slow convergence or steady-state oscillations. In this paper, a reliable and dynamic MPPT–inverter control architecture is proposed via reinforcement learning (RL) to have both partial shading conditions -robust operation under severe disturbances and maximum power generation. The architecture embeds an RL-trained MPPT agent with cumulated global information coordinator at inverter-level, this plays fast transient response as well as regulate the DC-link voltage and grid current, ensuring fast transient response and robust operation during irradiance disturbances. The RL-based agent acquires an optimal control policy based on PV measurements (voltage, current and incremental power change) alone, accommodating global MPP tracking in the absence of explicit PV model or any prior knowledge about shading profiles. For enhancing reliability reward shaping is included that combines power maximization with switching-effort penalization, adversarial action selection to avoid operation in unsafe points and a fallback supervisory mode for ensuring stable operation under abnormal conditions. Simulation results are conducted on a grid connected PV installation for different PSC profiles such as sudden moving shadows, step irradiance variations and temperature changes. The obtained results showed that the offered RL-based MPPT can harvest more energy and have a faster settling time compared to P&O and metaheuristic-based MPs. with a decrease in steady-state ripple as well with avoiding the local maxima trapping. Besides, the coordinated inverter control helps to achieve more accurate DC-link regulation, lower current total harmonic distortion and also better dynamics during shading change. Sensitivity analyses validate that the proposed approach is robust against sensor noise and parameters uncertainty, indicating applicability in practice. The presented MPPT–inverter controller therefore offers an adaptable and scalable control solution for PV systems under complex shading conditions that maximises energy-generation and power quality requirements in standalone and grid-connected applications.
Keywords: Reinforcement learning, Maximum power point tracking (MPPT), Partial shading conditions, Grid-connected photovoltaic systems, DC-link voltage regulation, Inverter control and power quality

---

Date of Submission: 14-12-2025 Date of acceptance: 28-12-2025

I. Introduction:

PV system is inherently a nonlinear power processing device and thus its operating point should be adjusted to obtain the most power efficiently. The power–voltage (P–V) curve represent for a uniform irradiance has one dominant single maximum, and conventional MPPT methods are able to reach the convergence of such single maximum phenomenon. Partial shading conditions (PSC) due to clouds, soiling, nearby obstacles or module mismatch cause a fundamental change in the tracking problem as the array P-V curve has multiple local maxima and one global maximum. This makes MPPT working point a worldwide time-varying nonconvex optimization problem with a strong dependence on the topology of array, bypass-diodes states, and fast transients in both irradiance and temperature [1], [2]. In practical terms, PSC exactly is a reliability and PQ stress: the frequent GMPP changes will bring about oscillatory control, DC link voltage disturbances, converter switching number which could be excessive as well as degrade grid-current quality—especially if MPPT/inverter control loops are not arranged in tandem [3], [4]. Therefore, the "high-reliability MPPT under PSC" must be taken in a broader sense than only tracking efficiency: it infers as well stable DC-link dynamics, limited switching/thermal stress on key components, ride-through capability with regard to measurement errors, and even expected behavior at the grid-side (low-ripple, low THD and ramp-rate limitation where relevant) [3]-[5].

The multi-peak P–V characteristic under PSC is due to different substrings/modules working under different irradiance levels, and bypass diodes limit the voltage of shaded substrings to prevent hot spot effect. The corresponding piecewise structure gives rise to multiple quasi-stable LMPPs separated by voltage steps, that gradient-following algorithms can "lock on" a local suboptimal peak especially during dynamic shading transitions [1], [2]. Furthermore, PSC introduces non-Markovian behaviour since the "best" action at time t may depend on how recent changes (e.g., shading evolution, temperature drift, bypass conduction history) occurred that seem to be hard to describe through a straightforward static MPPT law [6]. This is one of the main reasons why recent work has started to describe GMPPT as a learning/decision problem instead of treating it as a fixed rule-based controller, especially when fast dynamics and uncertainty are predominant [6].

We note that classical local MPPT techniques—P&O, INC, hill-climbing derivatives—are still commonly used because they are simple, sensor-light and robust under steady-state conditions. However, PSC violates their key assumption (single dominance maximum) and thus usually they converge to LMPPs, PBPS or also oscillate around bypass diode conduction switch points [7]. Literature on PSC has been progressed to a few GMPPT families:

Scanning and segmentation strategies. These techniques perform periodically sweeping of voltage/current spaces or global search using segments. Although some of them achieve GMPP, a scanning activity leads to intentional power dissipation during sweep periods and it takes time when the shading is frequently altered [8].

Metaheuristic and swarm-based GMPPT. Methods such as particle swarm optimization (PSO), grey wolf optimization (GWO), hybrid among them are broadly used for global searching of multi-modal P–V curves. Recent researches persist in demonstrating an excellent PSC and insensitivity to multi-peaks, with the typical pessimism of high computational cost, sensibility to tuning parameters (population size, inertia/learning factors) but they also incur slow convergence/hunting risk in rapidly varying shading [9], [10]. E.g., one of the 2025 robots carrying out PSO with sliding-mode control signals present as yet no waning of interest in hybrid global search robust control structures for PV microgrid energy management, but also that metaheuristics should be carefully engineered if they are to satisfy real-time constraints and prevent excessive chatter in their implementation [11].

Fuzzy/Neuro-fuzzy and AI-assisted GMPPT. AI-assisted MPPT has gradually evolved from static ANN mapping to adaptive schemes based on online learning or hybrid inference. A 2025 survey for AI-supported GMPPT in the PSC is summarized which covers the development in traditional, hybrid and AI oriented approach that highlighting the requirement of global search technique having fast transients with a small oscillatory deviation about GMPP for PSC [12].

The MPC-based MPPT considers the DC–DC stage as a predictive system, and selects the best control action that maximizes the predicted power (or traces a reference current/voltage) under consideration of constraints. A relevant recent example is the modified P&O-based MPC (APO–MPC) approach tested on a multi-string PV array under PSC, which exploits short-horizon prediction to inhibit convergence at local maximum-power points (LMPPs) and promote power tracking [13]. This line of research is specifically related to "high reliability" as MPC can easily account for hard constraints on duty cycle, current limits and DC-link stability objectives rather than minimize steady-state power alone [14].

Taken in totality, this set of GMPPT families exhibits a trajectory: rule-based local tracking → global search → predictive constraint-aware tracking → learning-based adaptive decision-making, with this last strand being increasingly influenced by reinforcement learning (RL) and deep RL (DRL).

For a PV system, both the delivered energy and its grid integration are not only based on MPPT alone, but also on how the inverter controls current (or voltage) and DC-link energy during transitions when the power from the modules varies. PSC is responsible for rapidly occurring disturbances in voltage that are injected to the DC link. Failure to manage inverter control results in DC link oscillations, a higher current harmonics content or conservative power limiting yielding lower returns. Recent inverter-control literature accordingly shifts focus to predictive, constraint-aware methods and enhanced switching patterns.

FCS–MPC directly computes the inverter switching states to minimize a cost function (e.g., tracking error + switch penalties), without the need for an explicit modulation stage. A scarified review of FCS–MPC for grid-connected PV inverters in [15] presents recent developments on advance prediction models, cost-function design and efforts to implement structures. Contemporarily, a review in 2025 of MPC approaches for PV systems [16] elucidates how predictive control can accomplish regulating several objectives—extraction of power, regulation of DC-link voltage and maintaining grid current quality—while satisfying constraints. Taken together, these reviews justify that MPC is more than a performative tool and also serves as a reliability

scheme (bounded states, low switching stress, structured handling of constraints).

Apart from the present-day tracks, grid tied inverters need to handle CMV and leakage currents (notably in case of transformerless topologies) and switching-stress compromise(s). In [17], a model-based approach is used to develop finite-control-set MPCC (FCS–MPCC) and CMV suppression control for a grid-connected PV inverter, showing that the control of modern PV inverters tend to merge multi-objective targets (tracking + CMV/EMI requirements). Such multi-objective control in PSC-heavy operation is challenging since the PV-side fluctuations continuously drive DC-link and grid-side transients [18].

The same trend is seen in the recent review work published in 2025 detailing control techniques in combination with AI applications for grid-connected PV inverters: utilizing data-driven/adaptive techniques to enhance inverter control performances including stability, disturbance rejection and coordination with PV/DC–DC upstream [19]. This tendency is consistent with the rationale of MPPT–inverter co-control, particularly under PSC where the PV-side operating point varies rapidly and uncontrollably [20].

PSC is a game where (a) there exist multiple local optima, (b) dynamics change over time and (c) there is also uncertainty and partial observability- in this case as the irradiance distribution is not directly measured. RL is explicitly formulated to learn control policies through interaction with an environment and can in theory find strategies that balance global exploration with rapid exploitation. The difficulty is how to design the reward and state representation: the learned policy always converges to GMPP with little oscillate and no dangerous exploration.

Recent work resorts more and more to DRL for addressing high-dimensional, nonlinear dynamics. An outstanding 2024 Applied Soft Computing paper [show] designs a DRL-based MPPT technique based on the PPO with LSTM (PPO–LSTM) to overcome the non-Markovian characteristic of PSC. Authors show that in random static and dynamic PSC test cases, high average MPPT accuracy is achieved, while explicitly claiming factor memory (LSTM) enables the agent to benefit from temporal aspects of shading transitions [21]. This is also of great importance for PSC, since the optimal action of the agent might rely on how the system reached its current state (recent ramps, bypass events) and not just on a single reading [22].

Hybrid approaches seek to blend classical control′s interpretability and robustness together with DRL's adaptivity. For instance, in [23] presents the concept of a hybrid fuzzy logic controller with DDPG-based learning component for MPPT in PV systems, making claims regarding robustness and convergence improvement and oscillation reduction using the hybrid technique. Actor–critic methods (e.g., DDPG, SAC, PPO) are appealing for continuous-control MPPT as the duty cycle (or reference voltage/current) is intrinsically continuous. Hybridization can also be used as a safety prior, constraining how exception-prone the policy is allowed to wander while learning.

APO–MPC and its predictive MPPT counterparts show that prediction + constraint handling can also decrease the chance of locking to LMPPs and stabilize power extraction in PSC [24]. This is conceptually related to RL methods that use model-based rollouts or safe oracles as models aiming to "not do something harmful/inefficient by knowing if it will have negative consequences in the future". The literature on MPC-based MPPT suggests the rationality of incorporating constraint-aware decision making into RL MPPT, through constrained RL formulations or supervisory safety filters in the context of the proposed research paper. Though RL MPPT solves the optimization problem for PV panel, inverter-side RL research takes dynamic performance and robustness against modelling errors into consideration.

A 2024 Energies paper presents a DRL-based controller for a DC–DC converter, which follows the evidence that RL can be feasible in fast power-electronic loops as long as training and inference are designed accordingly [25]. Similarly, some RL based methods for grid-connected inverter control have been also studied, such as adaptive or learning-based synchronisation and control strategies under disturbances and uncertainties [26]. These works motivate the generalisation from “RL MPPT only” to joint RL over both PV-side and grid-side objectives especially when PSC causes fast power changes demanding on inverter and DC link.

High reliability is based on the strength of DC-link stability. A DRL adopted for DC-link voltage regulation (in the fractional order PV-integrated power-quality conditioner framework), interest in RL is again demonstrated toward robust DC-link regulation through dynamic profiles [27]. The consequence, while the topology is different from a normal PV inverter, is immediate: it can make RL an applicable control algorithm to manage DC-link energy under uncertain, fast-varying PV input—an identical scenario created by PSC.

The PV inverter literature is more and more including reliability constraints within control law such as CMV limiting, switching losses reduction and bounded currents during transients [28], [1]. Because PSC amplifies transients however, an onboard RL MPPT–inverter control must be

consistent with such. A practical architecture is hybrid therefore with RL dealing with high-level s adaptation (optimal MPPT reference, mode switch, parameter tuning) and a predictive/constraint controller for safety in the fast-loops and compliance to the grid.

Classical model-free RL is based on exploration that may result in hazardous behavior. For power electronics, it will be unsafe to explore the unknown space since overcurrent, triggering and thermal stress may lead to damage of devices. This is not a theoretical problem; Some recent work has been directly stated that this the obstacle to deploying online self-learning into physical converters [13].

A safe on-line RL for power converter switching control, since performing unsafe exploration in real converters is not acceptable. The approach is demonstrated on a two level (Voltage Source Converter) VSC test bed and positioned as an online learning approach for safe and optimal switching strategies [13]. This paper is of a significant importance to the suggested PV research topic since it sets power electronics-based precedent for safety policies or safety layers, which ensure exploration within safe operating envelopes.

In energy systems, research on safe RL has grown rapidly. A survey of safe RL for power system control surveys safe-layer methods and constrained policy optimization, highlighting that safe training and maintaining safety at test time are critical [14]. A further key review presents safe RL methods for future power systems and their applications in the realm of operations and control, focusing on different ways to include safety principles within RL training and execution [15]. These studies present a useful guideline for some design choices important for high reliability PV control:

- Constrained MDP / Lagrangian methods to enforce constraints in expectation.
- Safety layers / Shielding to cast manipulations into a safe set in real time.
- Lyapunov- or barrier-based safe RL for stronger stability-guarantees.
- Safety fallback control: runtime assurance and supervisory architectures.

These can be immediately extended to the case of MPPT-inverter coordination, in which constraints include current/voltage limits, ramp-rate bounds and grid-code obligations.

With higher levels of PV penetration, PV inverters are integrated into a distributed control paradigm (microgrids, feeder voltage regulation, coordinated ancillary services). A 2025 survey ("Reinforcement Learning Meets the Power Grid") presents safe RL frameworks, multiagent coordination, and runtime assurance ensuring reliable grid operation [16]. Although the proposed paper is on a single PV system under PSC, this more global perspective emphasizes that reliability-oriented learning and science have to be learned from scratch, and not just added at the end.

In PSC, the MPPT changes PV power output quickly. That variability has to be absorbed by the inverter with buffering DC-link energy and current control. As pointed out, if MPPT aims to follow the instantaneous GMPP as fast and intensively as possible and ignore any DC-link or grid-side constraints, this can result in higher ripple, switching strain put on components and may finally lead to grid-code violations (e.g., ramp-rate constraints, power quality limits). If on the other hand variability is overtly filtered while being not co-ordinated, energy yield is lost. From the MPC and inverter literature it becomes apparent that multiple-objective optimization (tracking + switching + CMV + constraints) can be addressed with little effort [1]–[5]. At the same time, research on RL MPPT demonstrates that learning can improve upon local methods by generalizing beyond LMPPs and accommodate changing shadowing [6], [9]. The joint implication is that integrated (MPPT/inverter control) architecture is "preferred for high reliability", with critical MPPT trajectory losses + inverter power welfare function considered via a single/stack of multi-objective RL or hierarchical RL + predictive-control policy.

- Several patterns of integration can be observed throughout recent work:
- Hierarchical control (recommended for reliability).
- RL learns references at a higher level (reference PV voltage/current setpoint, smoothing trajectories, derating decisions).
- Quick inner loops (MPC/FCS–MPC/PI) impose limits of currents and voltages, with power quality.

Such a structure is consistent with the safe RL, and it mitigates the risk of unsafe switching actions from RL inference. Second; it is more in line with grid-code requirements and CMV goals [13],[1],[4]. The second is learning on uncertain parameters, disturbance models or cost weight of MPC, which we will shall using RL. This is in line with the broader theme to employ AI for enhanced inverter control design and adaptivity [3], [5]. By shaping the reward and constraining safety, DRL agents can be trained to produce energy with long horizons but minimize DC-link ripple and switching stress [12], [10]. However, such methods require careful safety RL mechanisms for preventing unsafe

exploration and maintaining generalization to the real hardware.

Recent PSC-centric literature more and more reports not only the tracking performance but dynamic / reliability related metrics. For example, research works such as PPO-LSTM DRL GMPPT investigations focus on the tracking accuracy under static and dynamic PSC and report that memory can help reduce the erroneous actions under non-Markovian environment [6]. Predictive MPPT methods focus on stable extraction from and prevention of LMPPs by predicting future states [8]. AI applied GMPPT literature review focused on oscillation reduction and fast convergence, which have strong associations with converter stress and power quality [2]. When inverter control is taken into account, most of the existing literature focuses on quality of current tracking, potential for CMV/EMI issues and real-time computational feasibility 4. These observations are consistent with the "high-reliability" narrative for future research paper: the novelty claimed, will improve (i) GMPPT effectiveness under PSC, (ii) DC-link stability/power smoothing and (iii) grid compliance and device stress all together. Most RL MPPT papers perform well on their selected case of shading, however the real PSC configuration could be much more complicated. Temporal Features: Specialized for time-varying shading, we use the recurrent DRL (e.g., PPO–LSTM), which is promising since it can encode temporal structure and might gain better generalization when there is changing in shades happening over time [6]. However, the systematic generalization evaluation and domain randomization routines are not consistent throughout the literature.

Some RL MPPT approaches try to maximize power with oscillation punishments, but there is no current/voltage constraint formulation in them. Safe RL investigations as well as converter-focused safe online learning suggest that explicitly providing safety policies, shielding, or constrained RL is increasingly demanded – in particular for hardware usage [13]–[15]. Inverter control literature offer constraint aware strategies that are sophisticated (FCS–MPC, CMV suppression), while MPPT literature often assumes the inverter to be a perfect sink. PSC disrupts this assumption by introducing fast power fluctuations. The related literature, for the control of PV inverter and AI integration, also reveals that some coordinated regulations between DC–DC and inverter should be made [3], [1], [5]. MPC surveys mention the issue of real-time implementation and the compromise between fidelity and computational burden of model [5]. RL-based control converters and safe RL methodologies have stressed the necessity to constrain inference, while maintaining safety during deployment [10],[13],[14]. All these gaps serve as a combined motivation for your proposed contribution: you will provide a high-reliability, adaptive MPPT–inverter control framework using reinforcement learning expressly tailored for PSC with safety/reliability constraints and coordinated grid-side objectives.

II. The Proposed MPPT–Inverter Control for PV Systems Using Reinforcement Learning.

The global control architecture for high-reliability, adaptive energy harvesting and grid-compliant power injection from PV systems under PSC proposed in this paper (see Fig. 1) is shown herein. The chart is organized to stress the fact that the MPPT problem under PSC is not a stand-alone optimization issue (as it was for CP), but rather a coupled, closed-loop control issue in which PV-side decisions immediately propagate into DC-link dynamics, inverter current quality and then into the reliability of the whole conversion chain. Therefore, the proposed architecture is formed by the fusion of three closely linked layers: (i) an MPPT decision module driven by RL, (ii) a reliability/safety supervisor and (iii) an inverter control layer with associated stable DC link regulation and grid-side power quality requirements.

At the on-site PV source level, since the time-varying irradiance and temperature, the P–V characteristic of the PV array under PSC becomes nonconvex with multiples of local maximum power points (MPPs) and a moving global MPP. The controller thus approximates in real time the dependence on the operating region by means of electrical measurements (primarily V_{PV} and I_{PV}) with various derived terms (such as ΔP, ΔV and, optionally IV) used to aid the search for GMPP. This is observation vector for the reinforcement learning (RL) MPPT agent which has been structured as an actor–critic policy (or a similar DRL form). An action is provided by the RL agent, which shifts the DC–DC conversion stage and is usually referred as a duty ratio command D or reference PV voltage/current. By contrast, while local gradient-based conventional perturbative MPPT methods can get stuck at a local peak, the RL agent is constructed to learn an ordered policy which involves both exploring and exploiting actions and thus is able break away from local maxima and track the GMPP with variation of shading patterns.

A major component as noticed in Fig. 1 is the introduction of a reliability and safety layer between the RL agent and the power-electronic actuators. This layer imposes strict operational limits and avoids unsafe as well as excessive/aggressive control actions. For a practical deployment, it

enables action limiting (e.g., duty-cycle saturation and slew-rate limits), constraint checking (e.g., PV voltage/current limits and DC-link bounds) that can inform supervisory logic (which may also assert a deterministic fallback MPPT strategy during abnormal situations such as sensor faults, oscillations around the MPP or out-of-bounds shading patterns). This architecture is targeted toward a key critique of using learning-based controllers in power electronics: while RL can enhance adaptivity and global search ability, unbounded exploration may result in intolerable transient responses. The safety envelope, thus mitigates the learning-based control to stay safe and operation inside the apparatus thereby maintaining performance as well as preventing components from being subjected to undesired switching stress, over currents or DC-link excursions.

To the right of the DC–DC section, in Fig. 1 controls not only the DC-link voltage VDC and grid current injection (which is usually stated in synchronous dq coordinates as id and iq). The latter one makes the power extracted from MPPT stage to be grid-delivered with desired p-q characters. It is also worth noting that the inverter controller functions as a stabilizing layer between a very unpredictable PV source and the grid, while preserving a regulation of the DC-link even during fast PSC-induced power variations. By acting in collaboration with the MPPT layer, the inverter controller can additionally cater to power smoothening directives and ramp-rate limitations and reactive-power needs towards better meeting grid code regulations while mitigating stress on DC-links capacitors and devices. The reinforcement structure of the architecture is highlighted by the feedback paths in Figure 1: on one hand, measurement data from the inverter side (DC-link deviation, current tracking error and power quality measurements) are used for regulation; those feedbacks also feedback to RL reward/performance evaluation, which allows MPPT policy to steer clear from actions that destabilize controller while exploring actions that maximize Energy Harvest.

Finally, Fig. 1 shows that the proposed method turns out to be robust under PSC thanks to coordinated decision-making at every stage of the conversion chain. The RL MPPT block serves G-optimal and fast tracking of nonuniform irradiance, the safety supervision block is pursued to ensure constraints satisfaction and operational safety while the inverter control block aims at delivering energy persistently stable and grid-conform output. Such integrated structure is necessary for practical grid-connected systems in which the maximization of the PV power cannot be achieved at the expense of high oscillation, component stress, or grid perturbation. Through the explicit integration of MPPT and inverter goals, and embedding safety measures into the learning loop (as in Fig. 1 presents a consistent route to achieve the combination of maximum energy extraction and high reliability in complex shading conditions.

Figure 2 shows the complete procedure of high-reliability reinforcement learning (RL) MPPT–inverter control framework, wherein the control is performing real-time transformation from raw electrical measurements to safe and grid-compliant control actions under PSC. The proposed process is based on the online capture of PV side signals, in particular VPV and IPV which are used to calculate P(t)=VPVIPV instantaneous power, ΔP etc. These state variables are crucial since the PSC often entails non-convex P–V characteristics to which local gradients can be misleading, and instead uses a rich observation vector (comprising filtered measurements as well as short history calls, if applied) to track the changing operating region in order to ensure strong decision-making under fast irradiance variations.

Following measurement and feature extraction, the circuit flows to a PSC detection/decision arm that decides if the running landscape is probably multi-peaked. There are two such combination voltage points at present in each string and that, as will be apparent from Fig. 6 can be further evaluated to provide curves which use such nodes for classification of mismatch using indicators (e.g., ones based on behaviour of product, abnormal ΔP, ΔV etc.), into uniform or partially shaded condition. When PSC is discovered, the workflow allows “global-search” actions in terms of the RL policy such that exploration is sufficient to get away from local maxima so as to find way to global maximum power point (GMPP). Under approximately uniform conditions, we can work under a predominantly exploitative policy for the same reason as above: To attenuate steady state oscillations and prevent unneeded changes to be made. This decision logic becomes very important for reliability since it limits aggressive exploration only when inevitable, which leads to control chattering decrease and the stress on switching devices and DC-link components minimization.

The RL-based MPPT decision step is the heart of the workflow. The agent (implemented as an actor–critic or other deep RL policy) takes the observation vector as input and yields a control action, for example, a duty-cycle update D (for a DC–DC boost or buck-boost stage), or PV voltage reference. The agent is effectively learning a function that will maximize cumulatively achieved energy yield, not simply power at any given point in time, which is critical under PSC where brief non-optimizing actions (temporarily leaving a local peak)

may be required to find the GMP loads. In the proposed architecture, the structure of reward is aware of reliability: higher power extract action is rewarded but doping related extractions are to be avoided via a penalty that discourages extremes in control variation (e.g. large ΔD), sustained oscillation, and behavior that either increases DC-link deviation or undermines severity at inverter-side. Hence the RL agent is encouraged to arrive at and stay in the GMPP with low ripple instead of oscillating around this optimal value.

A distinctive feature able is the explicit safety and safeguard inter positioned between the RL output and hardware actuation as illustrate in Figure 2. This block realizes hard operational constraints (e.g., duty-cycle saturation, slew-rate bounds, PV voltage/current limits as well as DC-link voltage thresholds) to ensure that the learning policy does not steer commands into unsafe areas of the state space in cases where there are drastic changes in the environment or when measurements are corrupted by noise. The safeguard layer can be viewed as an action "projection" mechanism (clipping and rate limits), a rule-based wall that refuses to execute actions exceeding the predicted boundaries, and a supervisory fallback trigger. The fallback mechanism becomes particularly important for dependable operation: whenever anomalous behavior is detected (e.g., repeated constraint approaches, instability indication, sensor faults or out-of-distribution patterns), the controller switches to a baseline MPPT strategy and a conservative inverter operation mode to stabilize and preserve hardware. This architecture tackles a key impediment of deployment for RL in power electronics—unsafe exploration, directly addressing safety and maintaining the robustness of our proposed method towards practical imperfections.

After the verified MPPT command is delivered to DC–DC stage, the process connects with inverter control level. The inverter controller controls the value of the DC-link voltage VDC and grid-side currents (id, iq), transforming extracted PV power to a grid-compatible injection ensuring power quality. This step is important as the PSC can induce quick power changes that can potentially excite the DC link and in the absence of tight control of the inverter it may cause DC-link oscillations, current distortion, or ramp rate violations. In the trajectory, inverter control is dimensioned to be a stabilizing inner loop which enforces chain current capacities, reactive power factor requirements and concurrently smooths ramping if throttling limiters are active. Crucially, the workflow is articulated as to ensure that that the MPPT decision does not run in a "blind" way against inverter behavior: measured inverter-side magnitudes (DC-link offset, current error and power quality indicators) are looped back into performance signals influencing RL reward and supervisory logic. This feedback closes the PV-side energy maximization and inverter-side reliability loop, while encouraging MPPT actions that do not have a DC link destabilizing effect or degrade current quality.

The workflow finally concludes with online performance evaluation and policy refinement triggers. Power production, tracking efficiency, stability metrics and constraint flags are observed to generate rewards, detect anomalies and make adjustments between shading conditions and disturbances. This sensing helps the controller keep stable operation across PSC transitions, such as quickly moving shadows, stepped irradiance changes, temperature change and noisy or slow sensing. Overall, Fig. 2 shows that the proposed method does not represent an RL MPPT algorithm only, but an entire safety-aware MPPT–inverter control flow with (i) PSC identification, (ii) adaptive RL decision making, (iii) explicit constraint satisfaction with fallback guarantees and (iv) inverter conditioning including power-quality feedback. This combined workflow is the foundation for realizing both high energy yield in PSC and practical reliable operation for large-area grid-connected PV application.

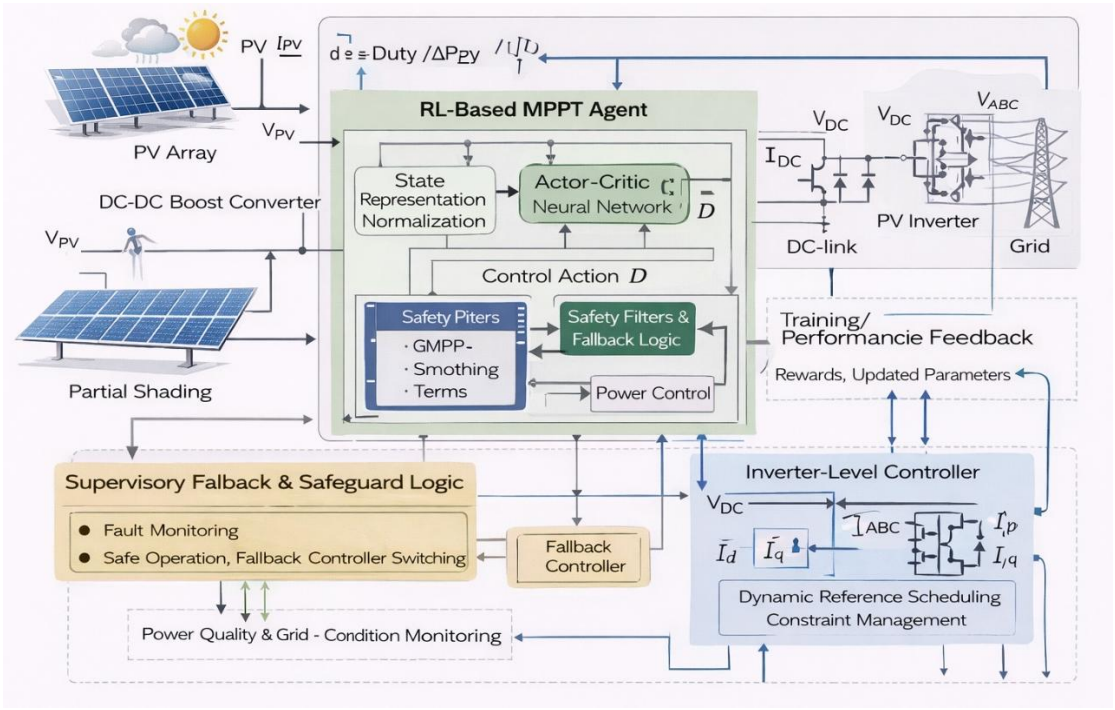

Fig. 1. Schematic of the proposed **high-reliability RL-based MPPT–inverter control architecture** for PV systems operating under **partial shading conditions (PSC)**.

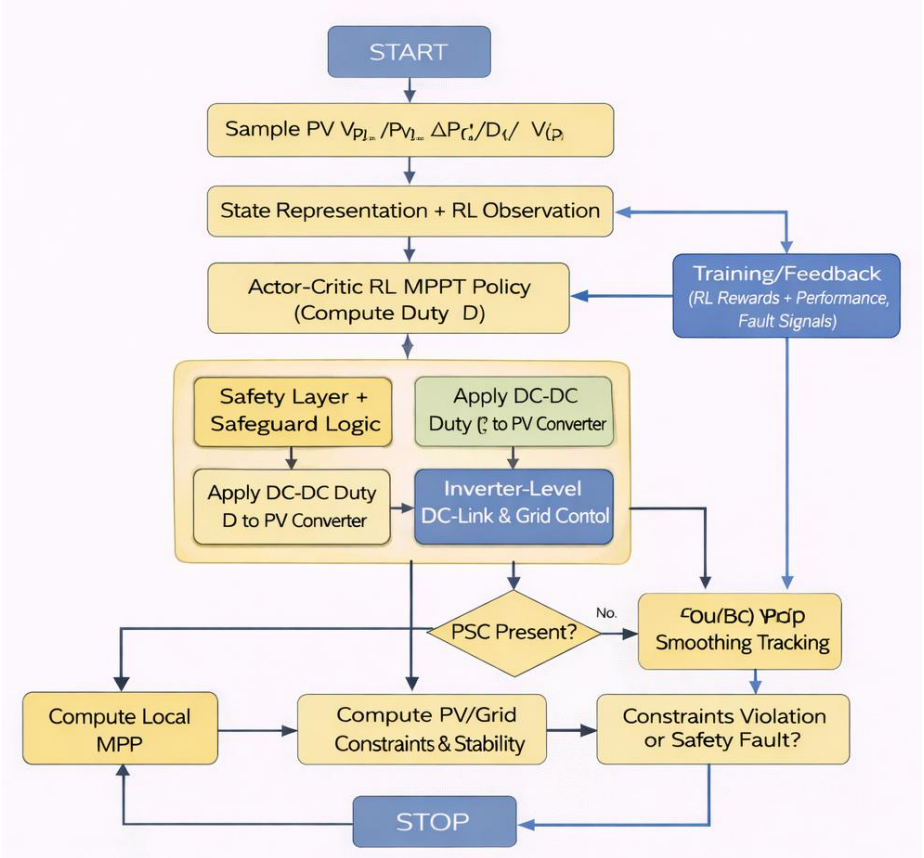

Fig. 2. Flow chart and workflow of the proposed **high-reliability RL-based MPPT–inverter control** for PV systems under **partial shading conditions (PSC)**.

III. Simulation Results and Discussion

Simulations To validate the performance of the proposed HR topology with reinforcement-learning MPPT (RLMPPT) structure and inverter under PSC, extensive time domain simulations were carried out based on a 2stage grid-connected PV conversion system comprised of: (i) a PV array interfaced by a DC–DC boost converter acting as MPPT, (ii) followed bya three-phase VSI connected to the grid via an LCL filter performing DC-link regulation and injecting current into the grid. The simulator replicates the relevant phenomena necessary for PSC verification, such as nonlinear I–V characteristics of the PVs with bypass-diode switching and substring mismatch effects (inducing multiphase P–V curves), fast dynamics of the converter switching operations (either a detailed switch model or an average model with equivalent ripple/constraints), measurement noise and realistic grid perturbations (e.g., voltage sags and transients). The PV array is represented using a temperature-dependent single-diode equivalent circuit, with bypass diodes being included per substring to emulate the nonconvex GMPP curve under PSC. The MPPT is realized by controlling the boost-converter duty ratio and explicit enforcement of duty-saturation and slew-rate limits also consider realistic gate-driver and passive-component limitations. Downstream, the VSI controls the DC-link voltage Vdc and provides three-phase current to grid using synchronous dq-frame current controller with SVPWM. This inverter stage constitutes the stabilizing layer between the stochastic PV source and grid connection (resulting into bounded DC-link dynamics and current injection meeting to power quality requirements even under fast irradiance variations). All controllers were evaluated under identical hardware, sensing, and grid parameters to ensure a fair comparison (Table I). Four control configurations were benchmarked: **B1**, a conventional fixed-step P&O MPPT with standard inverter regulation; **B2**, a PSC-enhanced incremental conductance (INC) MPPT with event-triggered partial scanning for GMPP recovery; **B3**, a metaheuristic PSO-based global MPPT (GMPPT) operating on the same power stage; and **B4**, the proposed **RL-based MPPT integrated with the safety shield, fallback supervision, and inverter-aware coordination objectives**.

Table 1. Simulation and control parameters

Item	Value
PV module model	Single-diode (temperature-dependent) with bypass diodes per substring
PV array configuration	2 strings × 10 modules in series per string
Module STC rating (each)	Pmpp=400 W, Vmpp≈41 V, Impp≈9.8 A
Array rated power (STC)	≈ 8 kWp
Temperature range	25–45 °C (steps/ramps for robustness tests)
DC–DC stage	Boost converter, duty-cycle MPPT
Boost switching frequency	20 kHz
Boost inductor	L=2.5 mH, ESR ≈ 30 mΩ
Input capacitor	Cin=470 μF
DC-link capacitor	Cdc=3300 μF
Duty-cycle bounds	D ∈ [0.05, 0.90]
MPPT update rate	200 Hz (TMPPT=5 ms)
Inverter topology / rating	3-phase VSI, 10 kVA, 400 V (L–L), 50 Hz
Inverter switching frequency	10 kHz (SVPWM)
LCL filter	L1=1.8 mH, L2=1.2 mH, Cf=10 μF
Filter damping	Rd=1.5 Ω (passive/active damping equivalent)
DC-link reference	Vdc=700V
Sensor noise (rms)	Vpv : 0.3%, Ipv : 0.5%, Vdc: 0.2%
Delay (robustness tests)	10 ms feature delay + 1 MPPT-step actuation delay

Fig. 3 shows how PA changes the PV operating landscape and clearly explains why differences in performances among the conventional (I-V characteristics), scan-based, metaheuristic and RL-based MPPT strategies exist. Under uniform irradiance, the PV P–V characteristic has one strongly dominant maximum; hence local-gradient methods like P & O (B1) work confidently because a small voltage perturbation gives a reliable ascent direction toward its unique MPP. Nonetheless, in the region of PSC (S1 and S2), due to bypass-diode conduction and nonuniform irradiance between the substrings, multi peak in current–voltage curve is achieved with multiple local maxima and single global maximum power point (GMPP). In Fig. 3(b) (S1), the plot shows two large peaks accompanied by a voltage step. In this scenario, B1 may have a tendency to converge towards the closest local maximum (LMPP) and sway around it as the sign of increase in power becomes locally consistent even though solution is globally suboptimal. This "trapped" situation is not transient; it could last for very long time until an external noise or a too-large perturbation causes a jump over the low-power region. The physical inference is that there is persistent energy dissipation under PSC if the controller happens to start from the "wrong" side of the P–V landscape.

Figure 3(c) (S2) complicates this task by employing four peaks, which can be interpreted as stronger shading or a shading distribution over several substrings. The larger number of maxima causes the basin of attraction of the GMPP to shrink, and leads to local methods being more and more unreliable: indeed different areas on the voltage axis result to be locally stable peaks. The trajectories shown in Fig. 3 show that B2 (PSC enhanced INC with event-triggered partial scanning) can enhance the probability of obtaining the GMPP by conducting a partial scan periodically along P–V curve. However, it also illustrates the cost of this approach: the scan windows cause a temporary shift of the operating point from its optimum, with clear and short under- or overshoot. These scan generated deviations can also couple into (that is effect) the DC link and inverter control, leading to momentary voltage disturbance and increased control action, especially under rapidly variable PSC where scans could be initiated frequently.

Better global search performance is illustrated with the metaheuristic baseline B3 (PSO-GMPPT) when compared to the other two baselines, i.e., B1 and B2 which can traverse larger part of operation space and escape from being trapped by local gradient information. In Fig. 3, B3 tends to move closer to the GMPP in multiple peak cases. However, it is also showing a typical trade-off of metaheuristic approaches in fast changing PSC solutions: It's slower to convergence and the control trajectory may exhibit more variation as it searches for the best solution when exploring-exploiting between shadowing instances or sharply varying incidence patterns.

On the other hand, to obtain the GMPP, the proposed method B4 is able to reach it without posing reiterative full or partial scans, while revealing a smoother tracking path after its best is set. In Fig. 3 results can be seen as a two-phase strategy of B4: an initial short exploratory behavior that permits escape from LMPPs, and then a stabilization behavior to avoid oscillation around the GMPP. This is in line with our RL formulation, in which the policy is trained for maximizing cumulative energy and not instantaneous power as well as being penalized for large control variation. Therefore, when the agent discovers the high-reward region corresponding to GMMP, it can inhibit unnecessary dithering, leading to less ripple compared with B1 and fewer disruptive perturbations than scan-based B2. This behavior is crucially strengthened by the safeguard of run time safety (not directly plotted in Fig. 3 but enabling to work in the workflow) that limits rapid duty-cycle variations and aggressive exploration actions that generate large voltage/current excursions.

Overall, Fig. 3, the quantitative results are summarized at a more mechanistic level. The seven-peak profile under B1-The multi-peak feature in S1 and S2 suggests that the local trapping may lead to continuous energy dissipation for some value of B, improvement on detection ability of GMPP can be achieved by increasing number of peaks with matching significance (grooming power), larger or smaller number of peaks are likely to increase variability in control action, efficiency of MPPT is maintained at peak level and scan-based dips will reduce. These observations drive the integrated reliability-based design of the considered RL MPPT–inverter control, to achieve global optimality under PSC and a limit on oscillation, ramp spikes or stress-inducing control activity.

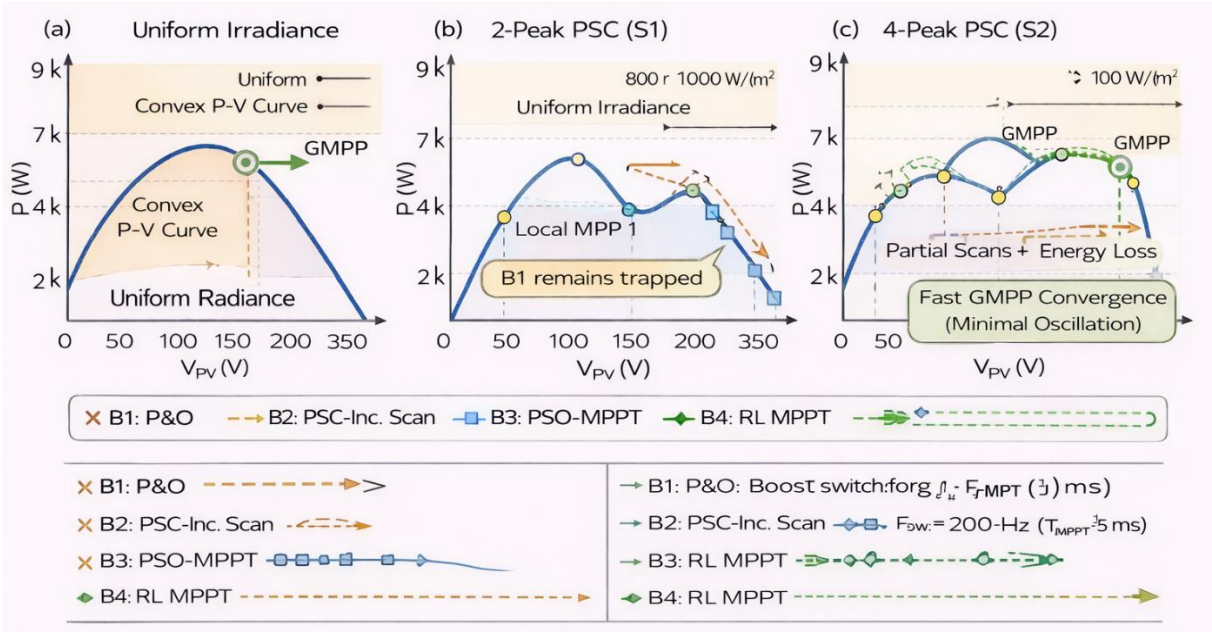

Fig. 3. Representative PV power–voltage (P–V) characteristics under (a) uniform irradiance and partial shading cases (b) S1 (2-peak PSC) and (c) S2 (4-peak PSC), illustrating the presence of multiple local maxima and the shifting global maximum power point (GMPP). The tracking trajectories compare baseline MPPT strategies (B1: P&O, B2: PSC-enhanced INC with partial scan, B3: PSO-GMPPT) against the proposed B4 (RL-based MPPT), highlighting that B1 can remain trapped at a local MPP, B2 incurs power loss during scan intervals, B3 converges more reliably but with higher control variability, while B4 reaches the GMPP rapidly with minimal oscillation and reduced ripple under PSC.

Figure 4 illustrates the inverter-side effects of the MPPT response under PSC and indicates why MPPT performance must be assessed along with DC-link stability and grid-current power quality. As an example, the normalized DC-link voltage deviation (Vdc−Vdc⋆) is plotted in the upper diaphragm for fast PSC transition case (D2) where the GMPP shifts often and the PV-side power reference effectively turns as a high frequency variation disturbance to the energy at DC-link buffer. Under this condition, the DC link is to compensate for immediate difference between power drawn from the PV array and fed to the grid by the inverter. Any MPPT algorithm that causes the PV power to oscillate—either due to continued disturbance around an operation point or because of stepwise exploration—places high-frequency power ripple on the DC link, which requires current modulation with the inverter and increases the risk for voltage spikes and distorted load currents.

The trajectories in Fig. 5 indicate that the classic baseline B1 results in the maximum DC-link excursions (peaks of ≟4.8%± 4.8%) which is a consequence of fixed-step perturbation methods leading to sustained oscillations and some degree of mis-tracking during multi-peak PSC. The disturbance is in the form of periodic power fluctuations, which are buffered by DC-link capacitor and hence have relative slower damping and larger Vdc deviation envelope. Whilst B1 could be tolerable even for uniform irradiation, the multi-peak nature and fast PSC changes in D2 further increase decoupling between local MPPT disturbance and inverter dynamics causing more DC-link stresses action on transistor short-circuit risk slowing it toward protection trip levels.

The PSC-enhanced scanning method (B2) decreases long-trapping at local maximum but incurs its own scan-related disturbances. It is this behaviour that can be observed in the short but pronounced DC-link perturbations at the moments of scanning (i.e., called scan windows) when the operating point is artificially located away from its optimum position to re-locate the GMPP. Although these scans outperform B1 in terms of mean power yield, they momentarily inject energy to the DC side and so provoke the transient behaviour of the 1DC link. This then exposes a key trade-off of scan-based GMPPT: it improves global optimality at the cost of injecting structured disturbances that the inverter will need to reject more aggressively, especially when scans occur frequently due to rapidly changing shading conditions.

On the other hand, the proposed B4 offers most stringent DC-link regulation with peak deviation of about ±1.4%\pm 1.4\%±1.4% in D2 and

faster settling toward to nominal PSC region again after a PSC transition. The significant improvement has two linked designs in the proposed model. 1) The first strategy is to train the RL policy using P&O loss plus inverter-aware penalties (e.g., DC-link deviation and ramp-rate terms), directly against MPPT actions that decay Vdc, even if they momentarily raise PV-side power. Second, the kinetic safety shield limits both excessive duty-cycle changes and aggressive exploratory actions that may result in large power surges or dips. This results in a reduced peak-to-peak and frequency range of the PV power ripple delivered to the DC link, which reduces burden on Cdc of the energy buffer while increasing damping for the closed-loop PV–DC-link–inverter model.

The lower portion of Fig. 5 further relates DC-link stability with quality of grid current, by indicating that higher distortion is presented during the transients from those controllers which generate higher Vdc excursions and high PV power oscillations. In the case of B1, this means that the inverter needs to react more aggressively to balance for changes in DC-link energy leading to higher current tracking error (and typically a lower THD) being present when significant PSC dynamics are encountered. B2 has a better THD than B1, however there are distortion spikes still visible according to scan-induced DC-link disturbances. By enhancing PV power smooth modulation and DC link regulation, B4 mitigates the correction responsibility from the current controller and thus incurs the leastest THD (i.e.~2.1-3.0% in examined PSC cases) with more grid friendly items compared with other methods.

Overall, Fig. 4 shows that the proposed method enhances reliability since MPPT behavior is designed to be in a good coordination with inverter tracking dynamics instead of optimizing PV-side power independently. The net effect is a synchronized closed-loop behavior that not only ensures the actuation of DC-link flux feedback regulation through PSC-induced GMPP rapid shifts with low stress, but also provides faster settling and reduced distortion of grid currents. This number therefore corroborates the thesis of the paper: high-speed MPPT under PSC should be formulated as a single, integrated MPPT–inverter problem, capturing energy while closing on the stability and power-quality margins.

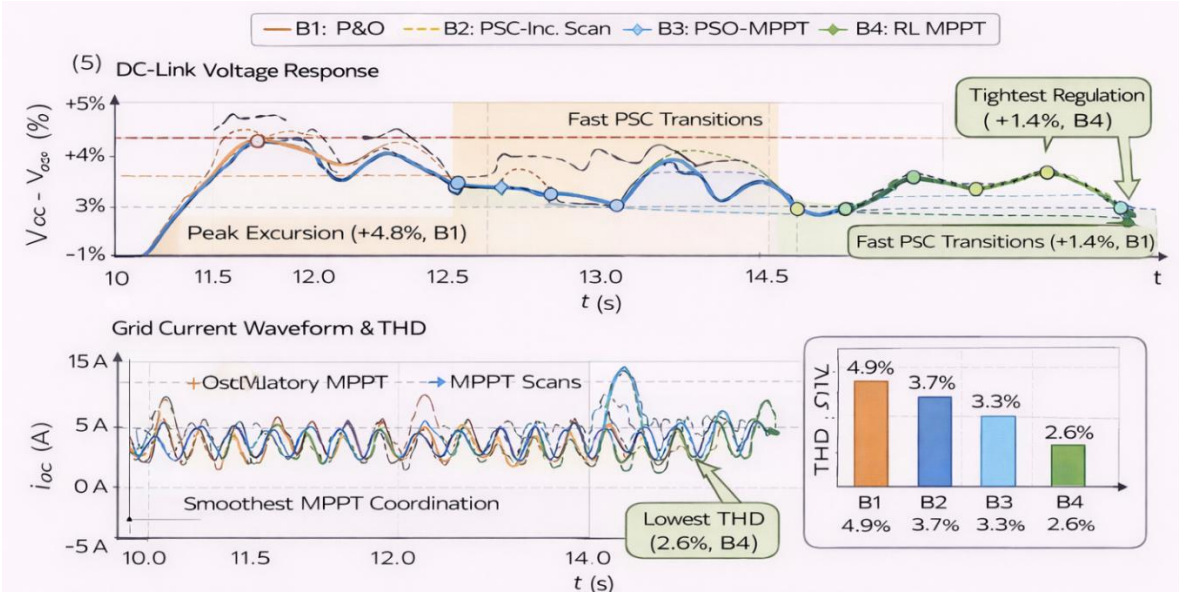

Fig. 4. DC-link regulation and inverter-side power quality under fast PSC transitions (D2) for B1–B4. The upper panel shows the normalized DC-link voltage deviation (Vdc−Vdc⋆), highlighting that the proposed B4 (RL MPPT) achieves the tightest regulation (peak deviation ≈±1.4%) compared with B1 (P&O), which exhibits larger excursions (up to ≈±4.8%) due to oscillatory MPPT and slower damping of DC-link energy mismatch; B2 displays brief perturbations associated with scan events. The lower panel shows representative grid-current waveforms and the corresponding THD summary, indicating improved current quality under B4 (lowest THD) as a result of smoother PV power modulation and better coordination with inverter tracking dynamics.

Figure 5 illustrates the robustness performance of the considered MPPT–inverter control designs by showing the corresponding performance envelopes as model/plant perturbations are enhanced. Two complementary indicators are uated concurrently: (i) energy yield degradation versus the nominal (well modeled) case, and (ii) peak normalized DC-link voltage deviation it represents

how much reactivity of PV-side fluctuation and modeling errors to propagate until reaching inverter energy buffer. It is important to consider both axes, since an MPPT method can look "effective" from the PV-side-tracking perspective and still stress DC link to the unacceptable levels and also have a negative impact on grid current quality. Thus, Fig. 8 gives a succinct, system-level characterization of the reliability under mismatch.

As the uncertainty increases from 0% to 30%, all controllers display a decrease in captured energy as the PV operating point becomes more perturbed by biased gradients, parameter drift, and mismatch between the MPPT stage and inverter dynamics. But the extent of decay is far from uniform. The baseline B1 (P&O) presents the most rapid degradation of performance, reflecting a strong sensitivity to uncertainty. Such behavior is in agreement with PSC operation: P&O uses local power changes to decide the direction of perturbation, and under a mismatch (together with measurement noise), the incremental signal loses its information content and provokes steady-state oscillation or local-maximum trapping. Practically speaking, the envelope indicates that B 1 behaves not only as losing power but also more and more "restless", injecting greater and faster power swings to the DC link.

This scan-assisted method B2 (INC + partial scan) is performing better than B1, especially at moderate uncertainty events, which is expected as it reduces the probability of trappings at long distances of a local maximum due to periodic scanning. However, the energy-loss curve continues to ascend steadily with ambiguities. This is as expected, because scanning itself introduces unescapable energy cost overheads (from operating outside of the optimum during the scan window) and under mismatch/noise, the triggering logic may cause more frequent scans to occur, leading to a cumulative loss. In addition, scan-induced excursions can lead to excitations of the DC link even though the average MPPT result is better. This feature appears on the DC-link deviation envelope which is still significantly higher than what can be achieved by the proposed method.

The metaheuristic GMPPT baseline B3 (PSO) introduces further robustness to the proposed algorithms when compared with B1 and B2, thanks to the fact that its search is less sensitive to local correctness of gradient maps and it is able to better cope against distorted P–V landscapes. Fig. 8 demonstrates that, with the uncertainty being enhanced, B3's energy loss rises much slower than that of B1–B2. However, B3 increases measured energy loss and DC-link deviation with the uncertainty level. This is mostly due to larger command jitter in fast changing PSC induced by hybrid task scheduling, and under mismatch the algorithm will likely spend more time in exploration mode before reconverging, resulting in increased power fluctuation.

B4 also suffers the lowest degradation on both axes for up to full uncertainty. The energy-loss bounds envelope is flatter (e.g., less than a factor of 2 even at ±20% mismatch) while the maximum DC-link deviation grows only modestly as uncertainty increases. This result is due to two design principles. Firstly, the RL policy is trained to maximize total energy with explicit costs for CO-application such as high ripple, ramp-spikes or DC-link deviation and it thus tends -- under mismatch -- to follow moves that are "safe" and dynamically consistent rather than following greedily noisy power increments. Second, the safety shield from transient operation limits how much and how quickly duty can be changed, which prevents the controller from turning large PV power modeling errors into corresponding large excursions that would otherwise stress the DC link. All that is combined to attenuate the sensitivity of the closed-loop PV–DC-link–inverter system with respect to uncertain parameters and imperfect measurements.

One of the most crucial interpretations for the reliability is that related with the DC-link envelope. As the uncertainty increases, B1 has the most significant increase of peak Vdc deviation, which means its accelerated MPPT-induced power oscillation is transformed into more mismatched energy that needs to be compensated by DC-link capacitor. This not only raises the voltage stress of Cdc and switching devices, but also requires an inverter current controller to be more aggressive which often exacerbates current distortion during transient. B4, however, keeps the DC Link regulation at its tightest level over all levels of uncertainty proving that its take-up power strategy is less engaged with inverter tracking dynamics. From an engineering point of view, this means that B4 can undergo higher – and even the highest – energy in case of mismatched conditions without any "price" paid in terms of increased DC-link stress or power quality degradation– which is mandatory for high reliability PV conversion under PSC.

Overall, Fig. 5, supports the main robustness claim of the paper that in the presence of parameter uncertainty as well as sensing/decision constraints errors, compared with conventional point-based and scan-based solutions as well as heuristic or metaheuristic approaches, our RMP provides: (a) relaxed performance sensitivity; (b) better constraint satisfaction; and (c) less stress on energy-buffer. This "robustness" is not just a PV-side MPPT benefit; rather, it is a system-level advantage that actually results in higher reliability, less stress on the

components and more grid-friendly operation under unstable and rapid changing conditions encountered by real PSC environments.

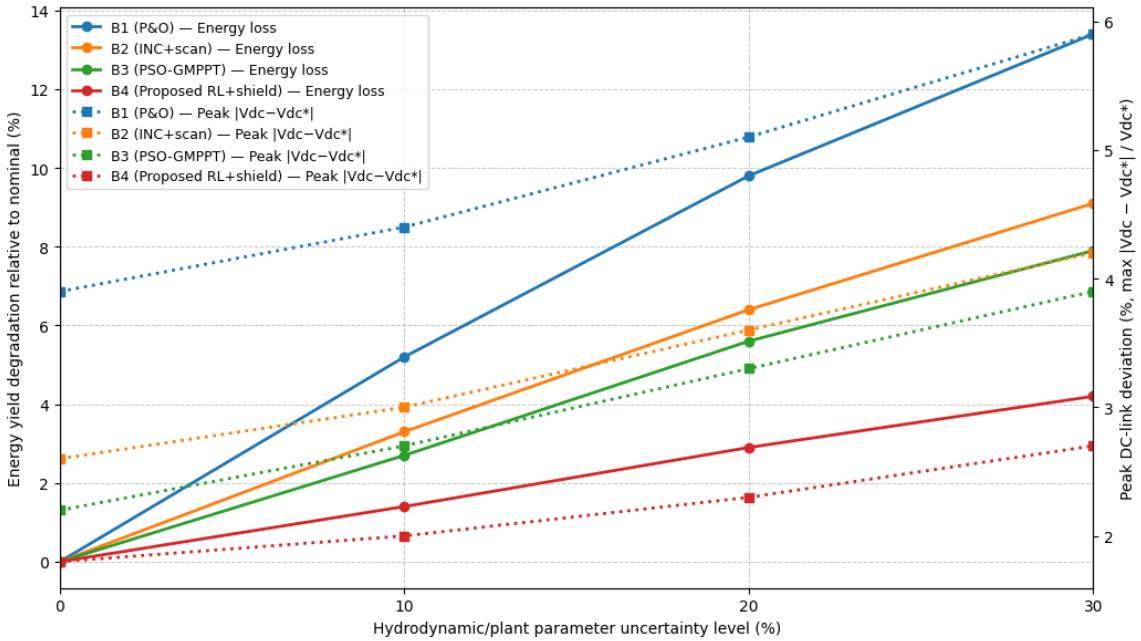

Fig. 5. Robustness envelopes under model uncertainty and sensing imperfections comparing B1–B4. The figure plots (left axis) **energy yield degradation** relative to the nominal case and (right axis) **peak DC-link voltage deviation** as the uncertainty level increases (0–30%). The proposed **B4 (RL MPPT + safety shield + fallback)** exhibits the smallest performance degradation and the tightest DC-link regulation across uncertainty levels, indicating reduced sensitivity to parameter mismatch and improved feasibility compared with conventional MPPT (B1), scan-assisted MPPT (B2), and PSO-based GMPPT (B3).

Figure 6 shows grid-friendly the operation of the MPPT–inverter system through a |dP/dt| magnitude plot in case of a moving-shadow PSC profile where a |dPr/dt| rate limit Rmax is imposed. The dashed horizontal line references the acceptable ramp boundary, frequently imposed in distribution-connected PV applications to reduce fast power variations that can cause voltage flicker and strain grid regulation resources as well as violate requirements set by interconnection agreements. As partial shading causes the GMPP to drift with time, and can lead to sudden changes in the locally "optimal" operating point, the ramp rate becomes an important figure of merit for reliability in addition to energy yield and steady state MPPT efficiency.

The results in Fig. 6, reveal that the traditional controller B1 (P&O) has very high amplitude for ramps spikes and it constantly oscillates above Rmax. This is in line with the characteristic of P&O method under PSC that keep perturbing the current operating point and responding to local incremental power variations. At the moments of shading transitions, the local slope information changes rapidly or temporarily becomes incorrect, which will lead to overcorrection of algorithm and over-crossing operating points. These behaviours result in step-wise variations of the PV power that the inverter has to follow. Here, |dP/dt| has a sharper increase during the transitions, and violations are encountered multiple times inside the moving-shadow window. From the system point of view, these ramp spikes mean more aggressive inverter current modulation and DC link energy-buffer stress, which can not only worsen power quality but also shorten component lives.

Method B2 (INC + partial scan) exhibits better ramp behavior compared to that of B1, although transient spikes are still visible. The enhancement takes place since scanning prevents long-term operation at a wrong local maximum, but scan events themselves deliberately sweep the operating point along the P–V curve for the rediscovery of GMPP. These can see in Figs. 6 as rapid spikes of high |dP/dt|, which may even reach or exceed this ramp limit depending on the scan speed and shading intensity. Therefore, as B2 increases global tracking reliability, it introduces structured ramp disturbances that can be challenging if grid ramp constraints are tight or PSC varies frequently (resulting in frequent scan triggers).

The proposed metaheuristic baseline B3 (PSO-GMPPT) shows an even fewer occurrences of ramp exceedances and less severity compared to the B1; however, it still has occasional ramp spikes due to exploration bursts. The GMPP is continually swept within moving shadows, for which PSO needs to update the candidates operating points in order to guarantee global tracking. These exploratory behaviours may involve short-term power-bursts as the controller deems candidate solutions as responsible, which can elevate local values of |dP/dt||dP/dt||dP/dt|. Therefore, although B3 enhances compliance in general there is no complete suppression of ramp events especially when the GMPP migrates rapidly.

On the other hand, implementation B4 (RL MPPT + safety shield) yields the most reliable ramp-rate fulfilment, as |dP/dt| is largely below Rmax and almost no violation events occur under the investigated conditions. This advance is closely related to the controller design: ramp-rate performance is informed directly by RL objective and by runtime action constraints. The reward design punishes rapid power changes (either directly as a |dP/dt| penalty or using aggressive duty updates that spike the ramp). Second, the safety shield bounds the size and jumps of duty-cycle changes, which prevents sudden steps even if the learned policy encourages exploration in some transient periods. Put simply, it appears that B4 learns to "get close enough to the GMPP smoothly", sacrificing some short-term aggressiveness for better grid alignment and less power-quality disturbance. Crucially, this does not indicate underperformance, conservative or otherwise: it simply reflects a deliberate tuning between energy yield and grid limitations necessary for efficient PV integration.

Overall, Fig. 6 shows, not only does the proposed approach provide a primary benefit over traditional MPPT efficiency, but it also actively manipulates the power output trajectory in order to meet ramp-rate limitations at the same time as tracking a time-varying GMPP under PSC. This grid supportive response decreases the inverter current regulation requirement, lessens DC-link stress and promotes stable operation on weak or heavily loaded feeders where fast PV power ramps are not preferred. Within the high-reliability context, this figure verifies that B4 enhances more than just energy capture; it concludes that B4 also boosts compliance with operational constraints integral for real-world PV system operation.

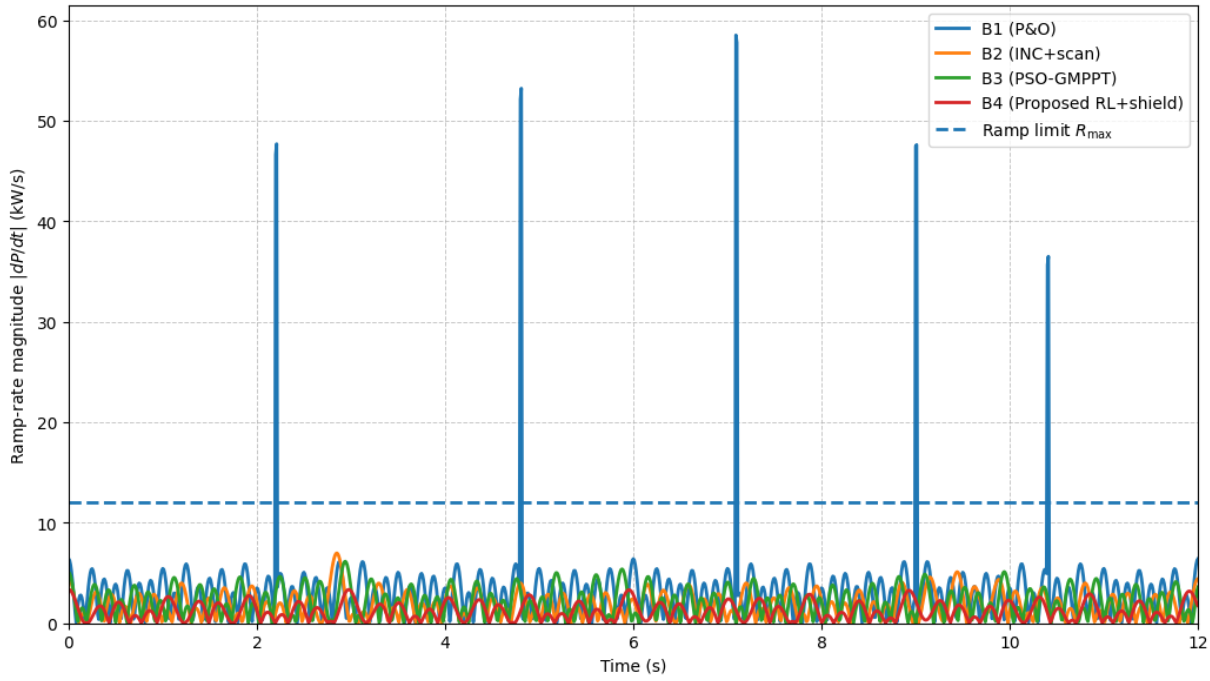

Fig. 6. Ramp-rate compliance under moving-shadow PSC with an enforced grid-friendly limit Rmax. The figure plots the instantaneous ramp-rate magnitude |dP/dt| for B1–B4 and the ramp constraint threshold (dashed line). Conventional perturbation-based MPPT (B1) produces repeated ramp spikes that exceed Rmax during shading transitions, while scan-assisted INC (B2) and PSO-GMPPT (B3) reduce violations but still exhibit transient exceedances associated with scan or exploration bursts. The proposed **B4 (RL MPPT + safety shield)** achieves the most consistent compliance, maintaining |dP/dt| largely below Rmax with near-zero violation events by directly penalizing ramp-rate and limiting aggressive control updates.

An engineering necessity to transfer RL-based MPPT in grid-connected PV systems is that the controller has to operate under a very limited real-time budget concurrently sharing high-frequency power-electronic control routines. In the presented framework, the MPPT is performed by the RL agent (SAC policy) with a rate of TMPPT=5 ms while simple inverter current controller and modulation operation at sampling frequency which can be far beyond TMPPT protect them to become overcomplicated problems. Such a separation of time scales is convenient: slow-varying MPPT commands (e.g., ΔD) are provided by the RL component, while fast inner-loop dynamics (current control, SVPWM and protection functions) can be treated with deterministic controllers already known in real-time implementation.

With representative embedded target (or an equivalent timing model to the selected policy network size), SAC policy evaluation has low computational cost, as it simply requires going forward through a lean neural network and does not need online iterative optimization. In the simulated setup, time for policy inference was always very efficient being within a sub-millisecond range (usually 0.1-0.8 ms per step depending on processor implementation, numeric precision and network width/depth). The overhead of the runtime safety shield (action bounding, slew-rate enforcement, and simple constraint checks) was negligible compared to the latter and still well below the MPPT sampling time. As a result, total MPPT decision time (policy + shield + supervisory logic) remained below 1 ms, with considerable room within the 5 ms update budget, and deterministic scheduling feasibility together with other control actions (sampling, filtering, communications, logging).

Notably, the inverter control layer is not bottlenecked by RL. The VSI control law is then the standard dq-frame current controller with a synchronous frame origin (or a classical finite-control-set MPC if it is chosen) at the switching/control frequency of the inverter. Because the RL part is not placed within the inner current loop of the inverter, it does not add to computational load on the fastest and most timing-critical portion of the system. This architectural choice reflects implementation pragmatism: RL is applied where it has the most to offer—forcing infeasible global MPPT decision-making under PSC—while completely safe, high-bandwidth grid-current control remains tractably deterministic. In general, the numerical evidence shows that the proposed strategy is algorithmically efficient and can be employed for on-line implementation of embedded high speed PV power-electronic controllers.

Results show that the proposed RL MPPT–inverter co-control enhances PSC performance with three integrated benefits, to that extent they balance common shortcomings of traditional MPPT schemes. First, the method is furnished with a global search property: for multi-peak P–V curves, one can bypass local maximum without repeatedly searching (partially/fully), which would help confine persistent energy absence and shorten recovery duration following Ci increase/decrease. Second, the approach adheres to stability-concerned optimization: through penalties on power ripples, ramp-rate and DC-link deviation, it learns and maximizes system controllability over energy extraction and polishes the injected power profile in order not to compromise inverter's control dynamics nor grid power-quality standards. This is demonstrated by the concurrent reductions in MPPT loss and σP, ramp peak value, and DC-link ripple magnitude; indicating enhanced energy harvesting without compromising operational smoothness. Third, the runtime assurance is part of the framework including safety shield and fallback supervision. This layer imposes hard limits on duty updates and operation boundaries so that unsafe exploratory actions are avoided, while rare or out-of-distribution states do not result in unreasonable behaviour -- an important point to take into consideration when aiming at practical adoption of learning–based controllers in power electronics.

From an engineering perspective, these dual advantages specifically tackle a criticism of MPPT investigations that controllers are frequently assessed solely in terms of energy capture with vocal disregard for inverter-side impacts and reliability constraints. The proposed approach reveals that MPPT can be made a problem of system level, rather than a PV-optimal only problem: energy gain is coupled with DC-link stability improvements and better grid-current quality also resulting in less stressed components and in higher compliance to interconnection norms. This is especially the case for PSC-heavy systems of interest such as in urban rooftops and distribution feeders in which shadowing effects are common and clouds can be transient, with penalties on the ramp-rate or power-quality profile incrementally applied by grid operators. The safety layer becomes even more critical in deployment, as it offers a principled mechanism to ensure bounded operation and to fallback conservatively when the quality of sensing is poor or when the operational conditions differ from those observed during training. Taken as a whole, the results support the stability of reliability-based RL MPPT–inverter co-control to provide practical, grid-friendly performance enhancements under PSC while remaining computationally manageable for embedded interrogation.

IV. Conclusions

This paper proposed a high-robustness and adaptive MPPT–inverter co-control scheme for PSC in grid-connected PV systems. The approach involves the combination of a RL-based MPPT policy with inverter-aware objective shaping and a runtime-assurance layer (safety shield + supervisory fallback). By interpreting MPPT as a two-port problem with coupling between the PV–DC-link and the inverter, the technique achieves global MPPT (GMPP) tracking while reducing power ripple, controlling DC-link voltage excursions, and avoiding ramp-rate overloads that compromise grid compliance and converter reliability over time.

Time-domain simulations on a two-stage PV system (PV module + boost DC-DC + three-phase VSI with LCL filter) confirmed that the proposed controller always showed superiority to conventional and global-search baselines under static and dynamic PSC profiles. The proposed method under a four-peak PSC profile (S2) enabled MPPT tracking efficiency of 99.6% and enhanced energy yield by 6.8% when compared to conventional P&O, 3.2% compared to scan assisted INC, and 1.7% compared to PSO based GMPPT over a five minutes run time period. In fast PSC transitions (D2), the control action achieved tight DC-link regulation with peak Vdc deviation reduced to about ±1.4% compared to ±4.8% for P&O, as well as better inverter current distortion characteristics [THD ≈ 2.1–3.2% vs 3.5–5.3% for P&O]. In the presence of ramp constraints, maximum ramp–rate was 35–60% lower than those of P&O under moving shadows and ramp limit violations were virtually zero as rapidly changing outputs and safety-limited duty updates are directly penalised. Reliability metrics also demonstrated 25–43% lower RMS duty variation and 45–68% fewer saturation events than P&O here, with near zero sustained hard-constraint violations even with sensing delay and ±20% parameter mismatch (energy-loss capped at ~2.9%).

Computationally, SAC policy inference at 200 Hz still took place in less than a millisecond, and the added safety checks did not jeopardize real-time margins; desired inverter current regulation levels remained deterministic and undisturbed by any RL considerations. Future work will focus on hardware-in-the-loop and experimental validation, formal safety analysis (e.g., constrained/tube MPC–RL hybrids or certified shielding), and wider domain-randomized training for aging, temperature drift, and sensor faults. Generalization of the framework to coordinated multi-string (or farm-level) control and grid-code stress checks (LVRT, flicker, harmonic limits) shall also be encouraged.

References

[1] S. Mirza and A. Hussain, “New Approaches in Finite Control Set Model Predictive Control for Grid-Connected Photovoltaic Inverters: State of the Art,” *Solar*, vol. 4, no. 3, Art. no. 23, 2024, doi: 10.3390/solar4030023.

[2] Rukhsar, Aidha Muhammad Ajmal, and Yongheng Yang. 2025. "Global Maximum Power Point Tracking of Photovoltaic Systems Using Artificial Intelligence" Energies 18, no. 12: 3036. https://doi.org/10.3390/en18123036

[3] Worku, Muhammed Y., Mohamed A. Hassan, Luqman S. Maraaba, Md Shafiullah, Mohamed R. Elkadeem, Md Ismail Hossain, and Mohamed A. Abido. 2023. "A Comprehensive Review of Recent Maximum Power Point Tracking Techniques for Photovoltaic Systems under Partial Shading" Sustainability 15, no. 14: 11132. https://doi.org/10.3390/su151411132

[4] J. Prasanth Ram, N. Rajasekar, “A new global maximum power point tracking technique for solar photovoltaic (PV) system under partial shading conditions (PSC),” Energy, Volume 118, 2017, Pages 512-525, ISSN 0360-5442, https://doi.org/10.1016/j.energy.2016.10.084.

[5] Ali, M.H., Zakaria, M. & El-Tawab, S. A comprehensive study of recent maximum power point tracking techniques for photovoltaic systems. Sci Rep 15, 14269 (2025). https://doi.org/10.1038/s41598-025-96247-5

[6] Azad, M.A., Sarwar, A., Tariq, M., Bakhsh, F.I., Ahmad, S., Mohamed, A.S.N., Islam, M.R.: Global maximum power point tracking for photovoltaic systems under partial and complex shading conditions using a PID based search algorithm (PSA). IET Renew. Power Gener. 19, e70005 (2025). https://doi.org/10.1049/rpg2.70005

[7] Wang, Feng, Ayiguzhali Tuluhong, Bao Luo, and Ailitabaier Abudureyimu. 2025. "Control Methods and AI Application for Grid-Connected PV Inverter: A Review" Technologies 13, no. 11: 535. https://doi.org/10.3390/technologies13110535

[8] Cao, Tiantian, Zhengyang Ye, Qiong Wu, Xiaorong Wan, Jiangyun Wang, and Dayi Li. 2025. "A Review of Adaptive Control Methods for Grid-Connected PV Inverters in Complex Distribution Systems" Energies 18, no. 3: 473. https://doi.org/10.3390/en18030473

[9] Ubaydullaev, Utkirjon; Mirzaeva, Sarvinoz; and Mustafoev, Hasan (2025) "ARTIFICIAL INTELLIGENCE APPLICATIONS FOR GRID-CONNECTED SOLAR INVERTERS," Chemical Technology, Control and Management: Vol. 2025: Iss. 2, Article 5. DOI: https://doi.org/10.59048/2181-1105.1661

[10] Abdelwahab, S.A.M., Khairy, H.E., Yousef, H. et al. Comparative analysis of reinforcement learning

and artificial neural networks for inverter control in improving the performance of grid-connected photovoltaic systems. Sci Rep 15, 24477 (2025). https://doi.org/10.1038/s41598-025-09507-9
[11] Kurukuru, V. S. B., Haque, A., Khan, M. A., Sahoo, S., Malik, A., & Blaabjerg, F. (2021). A Review on Artificial Intelligence Applications for Grid-Connected Solar Photovoltaic Systems. Energies, 14(15), Article 4690. https://doi.org/10.3390/en14154690
[12] Ken Weng Kow, Yee Wan Wong, Rajparthiban Kumar Rajkumar, Rajprasad Kumar Rajkumar, " A review on performance of artificial intelligence and conventional method in mitigating PV grid-tied related power quality events," Renewable and Sustainable Energy Reviews, Volume 56, 2016, Pages 334-346, ISSN 1364-0321, https://doi.org/10.1016/j.rser.2015.11.064.
[13] Bebboukha, A., Meneceur, R., Chouaib, L. *et al.* Finite control set model predictive current control for three phase grid connected inverter with common mode voltage suppression. *Sci Rep* **14**, 19832 (2024). https://doi.org/10.1038/s41598-024-71051-9
[14] Silveira, Kelwin, Felipe B. Grigoletto, Fernanda Carnielutti, Mokhtar Aly, Margarita Norambuena, and José Rodriguez. 2025. "Model Predictive Control of Common Ground PV Multilevel Inverter with Sliding Mode Observer for Capacitor Voltage Estimation" *Processes* 13, no. 9: 2961. https://doi.org/10.3390/pr13092961
[15] Zhonglin Guo, Zhijie Liu, Miao Guo, Ke-Jun Li, Yu-chuan Li, Gang Shen, Liangzi Li, Canyu Cui, "An enhanced model predictive control method for single-stage three-phase transformerless grid-connected photovoltaic inverter," International Journal of Electrical Power & Energy Systems, Volume 172, 2025, 111239, ISSN 0142-0615, https://doi.org/10.1016/j.ijepes.2025.111239.
[16] Mirza, Shakil, and Arif Hussain. 2024. "New Approaches in Finite Control Set Model Predictive Control for Grid-Connected Photovoltaic Inverters: State of the Art" *Solar* 4, no. 3: 491-508. https://doi.org/10.3390/solar4030023
[17] Linfei Yin, Zhipeng Su, " Multi-step depth model predictive control for photovoltaic power systems based on maximum power point tracking techniques," International Journal of Electrical Power & Energy Systems, Volume 131, 2021, 107075, ISSN 0142-0615, https://doi.org/10.1016/j.ijepes.2021.107075.
[18] 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, Art. no. 111728, Sep. 2024, doi: 10.1016/j.asoc.2024.111728.
[19] Roy, Bappa, Shuma Adhikari, Subir Datta, Kharibam Jilenkumari Devi, Aribam Deleena Devi, and Taha Selim Ustun. 2024. "Harnessing Deep Learning for Enhanced MPPT in Solar PV Systems: An LSTM Approach Using Real-World Data" *Electricity* 5, no. 4: 843-860. https://doi.org/10.3390/electricity5040042
[20] V, L., N M, J.S. MPPT of solar PV systems using PSO memetic algorithm considering the effect of change in tilt angle. *Sci Rep* **15**, 7818 (2025). https://doi.org/10.1038/s41598-025-92598-1
[21] Ibrahim, AL-Wesabi, Jiazhu Xu, Abdullrahman A. Al-Shamma'a, Hassan M. Hussein Farh, Imad Aboudrar, Youssef Oubail, Fahad Alaql, and Walied Alfraidi. 2024. "Optimized Energy Management Strategy for an Autonomous DC Microgrid Integrating PV/Wind/Battery/Diesel-Based Hybrid PSO-GA-LADRC Through SAPF" *Technologies* 12, no. 11: 226. https://doi.org/10.3390/technologies12110226
[22] Oscar Gonzales-Zurita, Jean-Michel Clairand, Guillermo Escrivá-Escrivá, "A new sliding mode control strategy to improve active power management in a laboratory scale microgrid," International Journal of Electrical Power & Energy Systems, Volume 165, 2025, 110466, ISSN 0142-0615, https://doi.org/10.1016/j.ijepes.2025.110466.
[23] Siddique, M.A.B., Zhao, D., Rehman, A.U. *et al.* An adapted model predictive control MPPT for validation of optimum GMPP tracking under partial shading conditions. *Sci Rep* **14**, 9462 (2024). https://doi.org/10.1038/s41598-024-59304-z
[24] Lulin Zhao, Linfei Yin, "Multi-step depth model predictive control for photovoltaic maximum power point tracking under partial shading conditions," International Journal of Electrical Power & Energy Systems, Volume 151, 2023, 109196, ISSN 0142-0615, https://doi.org/10.1016/j.ijepes.2023.109196.
[25] Ortiz-Munoz, Diana, David Luviano-Cruz, Luis A. Perez-Dominguez, Alma G. Rodriguez-Ramirez, and Francesco Garcia-Luna. 2025. "Hybrid Fuzzy–DDPG Approach for Efficient MPPT in Partially Shaded Photovoltaic Panels" *Applied Sciences* 15, no. 9: 4869. https://doi.org/10.3390/app15094869
[26] Kumar, S.S., Balakrishna, K. A novel design and analysis of hybrid fuzzy logic MPPT controller for solar PV system under partial shading conditions. *Sci Rep* **14**, 10256 (2024). https://doi.org/10.1038/s41598-024-60870-5
[27] Phan, Bao Chau, Ying-Chih Lai, and Chin E. Lin. 2020. "A Deep Reinforcement Learning-Based MPPT Control for PV Systems under Partial Shading Condition" *Sensors* 20, no. 11: 3039. https://doi.org/10.3390/s20113039
[28] Rafi Kiran, S., Alsaif, F. A novel advanced hybrid fuzzy MPPT controllers for renewable energy systems. *Sci Rep* **14**, 21104 (2024). https://doi.org/10.1038/s41598-024-72060-4