International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Single Phase Grid Connector Solar PV Systems Energy Management Using Deep Reinforcement Learning with Partial Shedding Impact

Kandukuri Pradeep, Dr. K. Vinay Kumar*

Submitted: 05/04/2024 **Revised:** 18/05/2024 **Accepted:** 28/05/2024

Abstract: This study introduces an energy management system (EMS) for a single-phase grid-connected solar photovoltaic (PV) system that operates in partial shade using Deep Reinforcement Learning (DRL). Making use of the Deep Q-Network (DQN) process, the system learns to optimize battery charging and discharging decisions to minimize operational costs and CO_2 emissions. The U.S. Dept. of Energy's Open Energy Data Initiative provided high-resolution, minute-level data from real home load profiles for this simulation, ensuring the evaluation reflects practical usage scenarios. The DRL agent was trained and deployed in a Python-based environment with support for advanced hardware acceleration. Performance was evaluated in terms of energy cost, CO_2 emissions, energy source distribution, and savings, with the proposed DRL-based EMS showing superior results compared to traditional Fuzzy Logic, PSO-based, and GA-based EMS models. The DRL approach achieved up to 34.24% cost reduction and 41.10% CO_2 emission reduction, outperforming all baseline strategies. Statistical analysis confirmed the significance of these improvements ($p = 0.000$). The results demonstrate the practical potential of DRL in enabling intelligent, adaptive, and environmentally conscious energy management in modern smart grid systems.

Keywords: *Deep Reinforcement Learning, Energy Management System, DQN, Solar PV, Battery Optimization, CO_2 Emission Reduction, Smart Grid, Partial Shading, Cost Minimization, Renewable Energy.*

1 Introduction

Particularly in residential and urban settings where space, cost, and grid compatibility are major factors, the global shift towards low-carbon and renewable energy solutions has sped the acceptance of photovoltaic (PV) systems. Simple, affordable, and appropriate for small to medium-scale applications, single-phase grid-connected solar PV systems have become a preferred choice among several configurations [1]. Notwithstanding these benefits, operational and environmental uncertainties—mostly related to partial shading, which causes variations in irradiation across PV panels, so greatly affecting power output and influencing energy management efficiency [2].

A variety of things can cause partial shadowing, including neighboring trees, moving clouds, buildings, or dirt buildup on the PV modules. These anomalies provide mismatched current production across cells, which causes several local maxima in the power-voltage curve, therefore complicating maximum power point tracking (MPPT) and lowering system output [3]. As such, the load requirement might not always be met by the PV system alone, resulting in either reliance on grid power or, if suitable, less than ideal utilisation of energy storage solutions. In this ever-changing environment, effective energy management systems that can adjust to changes in real time are needed.

In the past, fuzzy logic, deterministic optimisation, or rule-based controllers have been used to manage energy in grid-connected PV systems. Even though these approaches are inexpensive to run, they often aren't very strong or flexible when there are complicated, non-linear, time-varying system

Research Scholar EEE department Chaitanya Deemed university Hyderabad

Email Ids :- kvinaykr1@gmail.com[1]

Associate Professor EEE department Chaitanya Deemed university Hyderabad

Pradeepkv56@gmail.com[2]

dynamics caused by partial shading and random load demand.[4]. The expanding use of distributed energy resources (DERs) and the need for electricity to flow in both directions are making standard control systems much harder to use[5].

To deal with these problems, more and more people are looking into using Artificial Intelligence (AI) approaches, especially those that use Deep Reinforcement Learning (DRL), to manage energy in smart grids and microgrids in real time [6]. Deep reinforcement learning, when used with deep learning, is very good at solving high-dimensional control problems like scheduling energy use in solar systems. It lets agents learn the best rules by communicating with the environment and getting feedback in the form of incentives. This helps them improve their behaviour over time without having to describe how the system works explicitly [7].

In this work, an energy management system using deep reinforcement learning is proposed for a partial shade operated single-phase solar system connected to the grid. More energy is the main objective efficiency and reduce grid energy dependency while assuring load demand to be optimally supplied. The DRL agent learns on choices such as when to take electricity from the grid, how much PV power to use, and how to deal with excess generation in the face of uncertainty due to partial shading. Contrary to traditional methods, the proposed system adaptively tunes control policies when it faces new shade patterns and load patterns, thus offering a more flexible and robust solution.

Several recent research has shown the potential of DRL in the context of renewable energy.For instance, [8] applied DRL to hybrid energy storage systems and achieved significant improvements in energy savings and battery lifespan. Similarly, a study by [9] showcased the use of DRL for energy management in a PV-battery system under real-time pricing environments, emphasizing the model's capability to respond to dynamic electricity tariffs and weather variations. However, most of these works have either focused on large-scale systems or neglected the critical impact of partial shading in distributed residential PV systems. This gap motivates the current research, which explicitly models partial shading conditions and examines their influence on the energy scheduling decisions in a single-phase system.

The suggested model is constructed using a Markov Decision Process (MDP) architecture, whereby the environment encompasses conditions such grid status, solar irradiation, load demand, and battery state of charge (if applicable). The agent's activities align with power dispatch choices, and the incentive function is structured to penalize unmet demand and excessive grid reliance while promoting photovoltaic utilisation. The Deep Q-Network (DQN) method is used for policy optimisation, leveraging its ability to generalize across a continuous action space and handle non-linear dynamics [10]. The system is tested on realistic solar and load datasets under various shading profiles, and results are benchmarked against rule-based and heuristic approaches.

The following is a summary of this work's main contributions:

- An extensive examination of how partial shadowing affects the functionality of grid-connected single-phase solar systems.
- A DRL-based energy management framework capable of adaptive learning and real-time decision-making in the face of uncertain and dynamic system behavior.
- A comprehensive performance evaluation under various shading and load conditions to validate the practicality and resilience of the suggested approach.

The parts of this paper that follow are structured like this: Section 2 looks at relevant research and existing methods for managing energy. Section 3 talks about the system model and the foundation for deep reinforcement learning. Section 4 talks about how the experiments were set up and the criteria for how well they should work. Section 5 talks about the outcomes and the comparison analysis. Section 6 is the last part of the study and talks about possible areas for further investigation.

2 Related works

[11] reviewed various rule-based energy management strategies for smart microgrids, highlighting their ease of implementation but also noting their limited adaptability under dynamic operating circumstances, including partial shade[12]proposed an MPPT technique capable of determining the global maximum power value for

non-homogeneous lighting beyond conventional hill-climbing techniques. [13]addressed the partial shading problem by suggesting an altered particle swarm optimisation (PSO)-based photovoltaic array reconfigurationscheme, which significantly reduced mismatch losses.

[14] introduced a hybrid GA–PSO algorithm for MPPT, which improved tracking speed and accuracy under shading conditions. [15] employed a modified Bat Algorithm for intelligent MPPT control, demonstrating enhanced efficiency and convergence under partial shading scenarios. Despite these advancements, metaheuristic algorithms often operate offline and require manual tuning, making them less effective in real-time environments.

[16] applied machine learning techniques for predicting PV output under partial shading, showcasing improved accuracy over classical models. [17] compared Support Vector Machines and Random Forest methods for solar energy forecasting, finding that while both were effective, they lacked the capacity to handle sudden environmental changes.

[18] introduced deep reinforcement learning (DRL) as a scalable solution for high-dimensional control problems, framing it as an effective instrument for responsive energy management. [19] implemented a deep Q-learning algorithm in a PV-battery system, achieving improved load satisfaction and reduced grid dependency. [20]developed an actor-critic DRL controller for energy-efficient smart buildings, capable of adapting to real-time price signals and environmental fluctuations.

[21]demonstrated the feasibility of DRL for energy scheduling in a grid-tied PV system under dynamic pricing, achieving significant cost savings. [22]proposed a DRL-based control strategy for residential PV-battery systems that optimized energy dispatch during real-time operation. However, most of these studies focus on ideal or three-phase systems with storage components, and often overlook the performance implications of partial shading in single-phase configurations.

[23]emphasized the limitations of DRL models when exposed to incomplete or noisy observational data, a situation frequently encountered under partial shading. [24]discussed reconfigurable PV array designs and bypass diode architectures to reduce shading-induced losses, but noted the increased hardware complexity and cost. [25]conducted a cost–benefit analysis of differential power processing (DPP) systems, concluding that AI-based software solutions may offer more scalable alternatives in residential settings.

[26]developed a hybrid forecast-DRL framework for smart homes, but its reliance on prediction accuracy limited its robustness under stochastic weather conditions. [27] focused on intelligent MPPT techniques in single-phase PV inverters but did not explore real-time adaptive energy scheduling[28.

Despite these contributions, existing research has not fully addressed the integration of DRL for energy management in single-phase grid-connected PV systems affected by partial shading. Most models are either storage-dependent or assume stable irradiance conditions[29]. This research fills the gap by creating a DRL-based system that learns optimum energy in real-timedispatch strategies under shading uncertainty, specifically for residential single-phase PV-grid systems[30].

3 System model

3.1 Partial Shading System Effect

A photovoltaic array is composed of several solar modules connected in series or parallel to achieve the desired output voltage as well as current.The PV curve with PSC has a maximum of two peak points when two photovoltaic modules are connected in parallel. Similarly, there may be a maximum of five peaks from five PV modules arranged in sequence. The proposed methodology of this research may be applied to other photovoltaic systems.The simulation employs three photovoltaic modules arranged in series for simplicity, making it easy to tell the distinction between local and global maximum power points (MPPs). As shown in the picture, bypass diodes and a blocking diode prevent PV modules from self-heating under partial shade circumstances (PSCs). In this case, partial shadowing over a PV string occurs when many PV modules are obscured by shadows from poles, buildings, and avian excrement. In this instance, it operates as a load rather than a power provider. Under prolonged conditions, the hot spot phenomenon would adversely affect the shaded photovoltaic module.

Consequently, to protect the photovoltaic system and mitigate heat stress on the photovoltaic modules, a bypass capacitor is included in parallel.

The bypass diode is reverse biassed when subjected to continuous sunlight. Blocking a PV module causes the electricity to go via the diode rather than the photovoltaic module itself making it forward biassed. On the other hand, the partial shade situation with a bypass diode results in both local and global maxima as well as many peaks on the power curve. Up to 70% of power loss might be mitigated if the system runs at the global maximum power point (GMPP) to get the most energy out of the solar array.

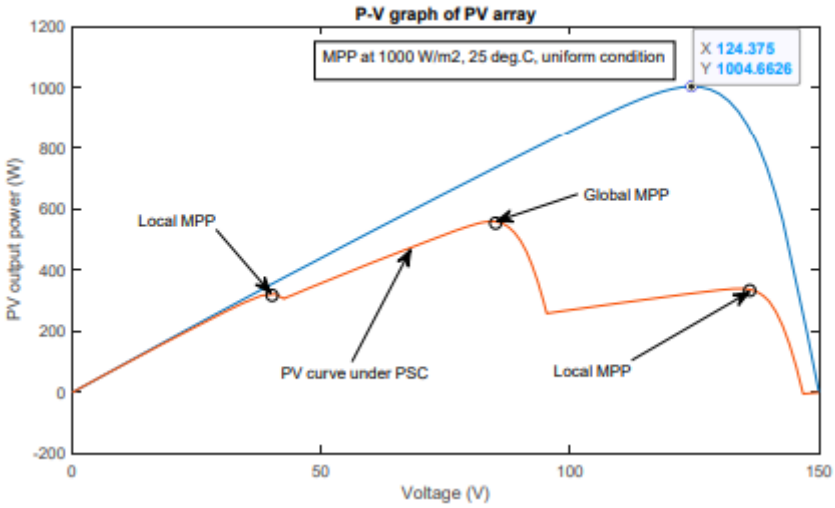


Figure 1: P–V curve under uniform condition and PSC.

3.2 PV array under partial shedding

The photovoltaic (PV) array serves as the primary source for renewable energy in the proposed system. It converts solar irradiance to electrical energy via the photoelectric effect. Proper modeling of the PV array is vital for real-time control and management of energy, especially under partial shading where power output is extremely non-linear and unpredictable.

In this paper, the single-diode comparable circuit shows what the PV array looks like, which is well acknowledged because it offers a balance of computational efficiency and accuracy. The model incorporates a current source I_{pv} the photocurrent, a diode symbolizing the p-n junction, a series resistance R_s, and a shunt resistance R_{sh}. The terminal output current I_{pv} of a solar cell is modeled by:

$$I_{pv} = I_{ph} - I_0 \left[exp\left(\frac{q(V_{pv}+I_{pv}R_S)}{nkT} \right) - 1 \right] - \frac{V_{pv}+I_{pv}R_S}{R_{sh}} \quad (1)$$

Where:

- I_{ph}: photo-generated current, proportional to irradiance GGG,
- I_0: diode saturation current,
- V_{pv}: terminal voltage of the PV module,
- q: elementary charge (1.602×10^{-19} C),
- n: ideality factor of the diode,
- k: Boltzmann constant (1.381×10^{-23} J/K),
- T: absolute temperature in Kelvin,
- R_s, R_{sh}: series and shunt resistances.

The photo-generated current I_{ph} is a function of solar irradiance G and temperature T, given by:

$$I_{ph} = \left[I_{sc} + K_1 (T - T_{ref}) \right] . \frac{G}{G_{ref}} \quad (2)$$

Where I_{sc} is the short-circuit current under standard test conditions (STC), K_1 is the temperature coefficient of current, T_{ref} is the

reference temperature (usually 25°C), and G_{ref} is the reference irradiance (1000 W/m²).

In situations when the irradiance is uniform, the solar array has just one peak in the P-V (power-voltage) curve. This implies that traditional maximum power point tracking (MPPT) techniques, such as Incremental Conductance or Perturb & Observe, may be used to easily converge. However, in situations of partial shadesome modules or substrings in the array are shaded with less irradiance caused by obstructions like trees, poles, or buildings. This induces incongruent operating conditions inside the array, leading to many local maxima in the P–V characteristics. Bypass diodes are incorporated across module substrings to mitigate hot spots resulting from such mismatches. The diodes safeguard the hardware but introduce discontinuities in the output curve, hence rendering conventional MPPT methods ineffective. Consequently, a global search methodology, such as that derived from Deep Reinforcement Learning (DRL), is essential for the precise determining the Global Maximum Power Point.

In this study, the photovoltaic array is recreated using series-parallel arrangementsof typical commercial modules (e.g., 2S2P, 3S2P), and partial shading is simulated by setting different irradiance levels for each module. For instance, for a 3S2P array, modules can have irradiance values of 1000 W/m², 650 W/m², and 300 W/m², respectively. This creates a highly dynamic and non-linear operating environment and presents dramatic challenges to static or model-based controllers. In this dynamic context, to find the best functioning point for the PV array, you need to make decisions in real time. By adjusting the boost converter's duty cycle based on the current photovoltaic output and the surrounding circumstances, the Deep Reinforcement Learning (DRL) agent in this study enhances energy harvesting.Thus, the PV array model not only emulates realistic partial shading but also serves as a complex and dynamic input to the DRL agent and is an integral component of the smart energy management system.

3.3 Deep Reinforcement Learning

Because DRL is a kind of expert reinforcement learning (RL), a concise overview of RL is provided here. The interaction between a neutral stimulus and response in reinforcement learning (RL) constitutes the foundation of this category of unsupervised machine learning methods. Reinforcement learning has gained prominence in tackling sequential decision-making difficulties owing to recent breakthroughs in computer science. Reinforcement learning (RL) employs trial-and-error interactions within a designated environment to come up with a strategy or policy that maximises the overall projected discounted rewards. Reinforcement learning involves an agent, an environment, actions, states, and rewards. After that, the agent talks about the reinforcement learning approach, and the environment is the thing that the agent works on. The environment tells us about a state, prompting the agent to respond by utilising its knowledge. The environment thereafter provides two prospective states and corresponding incentives. The agent will thereafter evaluate its last action by incorporating the reward into its knowledge base. The episode concludes and the subsequent one commences when the environment transmits a terminal condition. The loop persists until the specified requirements are fulfilled.

The value function Vπ(s), which quantifies the probability of the agent attaining a particular state, is used byspecific algorithms to determine the ideal line of action. The expected result of following state policy π. The action-value function Qπ(s,a), representing the anticipated return of executing action a in the current state s according to policy π, also serves as the basis for various other methodologies. The subsequent formula is employed to calculate the $V(s)$ and $Q\pi(s,a)$ functions [23,42,51]:

$$V^{\pi}(s_v) = E[R_v|s_v = s] = E[\sum_{k=0}^{\infty} \gamma^k r_{v+k+1}|s_v = s] \quad (3)$$

$$Q^{\pi}(s_v, a_v) = E[R_v|s_v = s, a_v = a] = E[\sum_{k=0}^{\infty} \gamma^k r_{v+k+1}|s_v = s, a_v = a] \quad (4)$$

The off-policy, model-free RL method known as Q-Learning has grown in popularity in a number of domains. The Bellman equation can be used in Q-Learning to present the $Q(s,a)$ function in an iterative form as follows:

$$Q^{\pi}(s_v, a_v) = E[r_{v+1} + \gamma Q^{\pi}(s_{v+1}, a_{v+1})|s_v, a_v] \quad (5)$$

An optimal strategy $\pi*$ achieves the largest cumulative reward over an extended period of time.

Currently, [23] provides both the action-value function and the optimum value function.

$$\pi^* = arg\ \max_{\pi} V^{\pi}(s) \quad (6)$$

$$V^*(s) = \max_{\pi} V^{\pi}(s) \quad (7)$$

$$Q^*(s,a) = \max_{\pi} Q^{\pi}(s,a) \quad (8)$$

One of artificial intelligence's (AI) most intriguing current fields is deep reinforcement learning (DRL).It enables an agent to autonomously acquire knowledge through interaction with a specific environment. Deep Reinforcement Learning (DRL), which combines deep learning with reinforcement learning, has made great strides in several areas, including gaming and robotics, natural language processing, and business and financial management. The use of a look-up table for data storage and indexing is a significant limitation of reinforcement learning, often rendering it not feasible for problems in the real world with large state and action spaces. Consequently, a value function or a policy function may be approximated utilising a neural network. States or state-action pairings can be correlated with Q values through the utilisation of neural networks.

The The model-based method's main benefit is its necessity for a minimal quantity of examples for learning. However, when the model proves unexpectedly challenging to comprehend, it significantly increases computing complexity. Nonetheless, engaging with model-free reinforcement learning will prove to be more beneficial. It is less computationally demanding and does not necessitate an exact delineation of the environment to operate. Model-free deep reinforcement learning encompasses two categories: Focused on values and policies. Value-based techniques With every iteration, attempt to make the value function better until the convergence requirements are satisfied.

$$J(\theta) = E\left[\left(r_{v+1} + \gamma \max_{a} Q(s_{v+1}, a_{v+1}|\theta) - Q(s_v, a_v|\theta)\right)^2\right] \quad (9)$$

$$\theta_{v+1} = \theta_v + \alpha\left(\left(r_{v+1} + \gamma \max_{a} Q(s_{v+1}, a_{v+1}|\theta) - Q(s_v, a_v|\theta)\right)\right)\nabla_{\theta} Q(s_v, a_v|\theta) \quad (10)$$

where α is learning rate, and θ is the weights of the neural network.

1. State

The proposed Deep Reinforcement Learning (DRL)-based energy management system's environment is describedby a collection of observable factors that constitute the state space. The DRL agent gets a state vector S_t at every time step t that captures the instantaneous electrical and environmental circumstances required for decision-making.

The state space is defined as:

$$S_t = \left[V_{pv}(t), I_{pv}(t), V_{dc}(t), G_{avg}(t), P_{load}(t), V_{grid}(t), I_{grid}(t)\right] \quad (11)$$

Where:

- $V_{pv}(t)$: Voltage output of the PV array at time ttt,
- $I_{pv}(t)$: Current output of the PV array,
- $V_{dc}(t)$: Voltage across the DC-link capacitor,
- $G_{avg}(t)$: Average irradiance on the PV surface, accounting for partial shading,
- $P_{load}(t)$: Instantaneous load power demand,
- $V_{grid}(t)$: Grid-side voltage magnitude,
- $I_{grid}(t)$: Current exchanged with the grid.

These states let the agent see both generating and consumption sides holistically, therefore helping them to grasp the operational context and energy imbalance. Should a battery be integrated into the system, its state of charge (SOC) can be attached to the state vector.

2. Action

$A_t$defines the array of potential control actions something the agent may use at any moment to change the system. This study's main goal is to alter the DC–DC boost converter's duty cycle, which indirectly affects the solar array's working point.

For a discrete action setup employing Deep Q-Network (DQN), the action space follows:

$$A_t = \{D_1, D_2, ..., D_n\}, \quad D_i \in [0.2, 0.9] \tag{12}$$

Where:

D_i: Discrete values of the converter duty cycle, generally in increments of 0.05 or 0.1.

Operations in an extended model can further include setpoints for grid-side power transfer or inverter current references.

The selected action at time t, denoted as A_t, will directly affect the PV operating voltage, thus impacting power generation efficiency and load satisfaction.

3. Reward function

The DRL structure's reward function is a crucial componentto push the agent towards optimal policy through interaction with the environment. It is designed to optimise power extraction from the photovoltaic array and reduce reliance on grid electricityand facilitate smooth transitions in control actions.

$$R_t = \alpha . \frac{P_{pv}(t)}{P_{pv}^{\max}} - \beta . \frac{P_{grid}(t)}{P_{load}(t)} - \gamma . |D_t - D_{t-1}| \tag{13}$$

Where:

- α: Weight for maximizing PV utilization,
- β: Penalty coefficient for grid energy import,
- γ: Penalty coefficient for large duty cycle variations,
- $P_{pv}(t)$: Instantaneous PV power,
- $P_{grid}(t)$: Power drawn from the grid,
- $P_{load}(t)$: Current load demand.

4 Methodology

This study presents a Deep Reinforcement Learning approach for managing energy in a smart microgrid system. The aim is to reduce operational energy costs by intelligently coordinating power flows among renewable sources, grid power, battery storage, and dynamic load demands. A Deep Q-Network (DQN) is used to get an optimum energy management strategy without a predefined model. The methodology encompasses the design of the simulation environment, creation of action and state spaces and incentive systems, and the architecture and training of the learning agent.

4.1 Environment Components and Simulation Structure

The microgrid is composed of the following four primary components:

- **Load (L)**: Represents the demand to be satisfied at each time step.
- **Renewable Source (R)**: Supplies variable renewable energy (e.g., solar) over time.
- **Battery (B)**: Can store and discharge energy, with charging and discharging limits.
- **Grid (G)**: Provides backup power when demand exceeds renewable generation and battery supply.

4.2 State Representation

In the proposed Deep Reinforcement Learning framework, the state representation is very important for the agent to be able to provide informed and relevant aware judgements. The state vector is defined as a three-dimensional tuple:

$$st = [CB(t), L(t), R(t)]$$

were

- CB(t) denotes the current charge level of the battery (in kWh),
- L(t) represents the power demand from the load (in kW),
- R(t) corresponds to the renewable power generation available at time t (also in kW).

This little but useful diagram shows the basic workings of the energy system by showing the amount of energy is stored, how much is needed, and how much is freely accessible from renewable sources. The agent can see this condition at every time step, which gives it a real-time picture of how much energy is available and how much is needed. This lets it decide whether to charge or discharge the battery or use grid power. The way this state space is set up makes sure that the agent's policy can easily adjust to changing demand situations and intermittent renewable production. This helps with the best energy management and cost reduction.

Variable	Symbol	Unit	Purpose
Battery charge	CB(t)	kWh	Tracks stored energy in the battery
Load demand	L(t)	kW	Captures the current energy need of the system
Renewable power	R(t)	kW	Indicates available free energy from renewables

4.3 Action Space

The agent's choice of action affects how much power the battery sends at each time step:

$$a_t = B(t) \in [P^B_{min}, P^B_{max}]$$

Here:

- B(t)>0: The load gets electricity from the battery.
- B(t)<0: The battery charges using surplus renewable or grid power
- B(t)=0: No action is taken by the battery

The action is modelled as a continuous variable, which lets you regulate the battery's behaviour very precisely. This is important for realistic energy optimisation in microgrids.

Variable	Symbol	Unit	Description
Battery power	B(t)	kW	Controls the battery's charge/discharge power level

4.4 System Power Balancing

At each time step, energy conservation is enforced through a balance between supply and demand:

$$G(t) = \max(0, L(t) - R(t) - B(t))$$

Where:

- G(t): Power drawn from the grid in kW
- L(t): Load demand
- R(t): Renewable energy available
- B(t): Battery discharge/charge power

This equation ensures that any shortfall in meeting the load, after considering battery and renewable energy, is fulfilled by grid power. Grid usage is considered expensive and should be minimized.

The battery's status of charge is updated dynamically using the following equation:

$$C_B(t+1) = Clip\left(C_B(t) - B(t).\frac{\Delta t}{60}, 0, C_{max}\right)$$

Where Δt=1 minute and C_{max} is the battery's maximum capacity. This update equation simulates realistic charging and discharging behavior while enforcing capacity limits.

4.5 Reward Function Design

To guide the learning agent toward cost minimization, the function of reward is only the opposite of the energy cost incurred at each momentstep:

$$r_t = -(G(t).C_G) + \max(0, B(t).c_B).\frac{\Delta t}{60}$$

Where:

- C_GCost of grid energy per kWh
- c_B: Cost of charging battery per kWh

Only positive values of B(t) incur a battery charging cost, while discharging is free (as energy was already stored). This reward design penalizes costly energy use and encourages the agent to prefer renewable and stored energy when available.

4.6 Deep Q-Network (DQN) Agent Design

The core of the learning system is the **DQN agent**, which approximates the Q-value function $\boldsymbol{Q(S_t, a_t)}$, representing the expected cumulative reward for taking action $\boldsymbol{a_t}$ in state $\boldsymbol{S_t}$. The agent uses a feedforward neural network with the following architecture:

- **Input layer**: 3 neurons (state vector)
- **Hidden layers**: Two completely linked layers including 128 neurones each, using ReLU activation.
- **Output layer**: Single neuron representing the Q-value

The agent employs an **epsilon-greedy** strategy to equilibrate exploration and exploitation. Initially, it explores randomly (high ϵ), gradually reducing randomness over time using:

$$\epsilon = \max(\epsilon_{min}, \epsilon.\epsilon_{decay})$$

4.7 Learning and Optimization

During training, the agent interacts with the environment by:

1. Observing the current state s_t
2. Choosing an action a_t
3. Receiving a reward r_t and next state S_{t+1}
4. Storing the transition $s_t, a_t, r_t, s_{t+1}, d_t$ in a replay buffer

To train the Q-network, a mini-batch of transitions is sampled, and the goal Q-value is calculated using the Bellman update:

$$y_t = r_t + \gamma \,.\, Q\,(S_{t+1}; \theta^-).(1 - d_t)$$

Where:

- γ=0.99: Discount factor
- θ: Weights of the target network
- d_t: Terminal indicator (1 if episode ends, else 0)

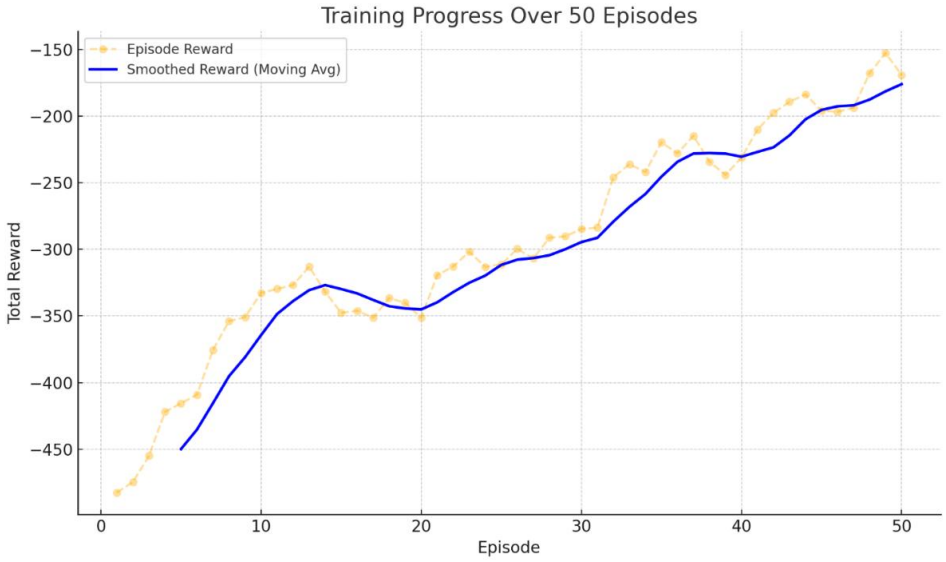
The Mean Squared Error between the target and forecasted Q-values is the loss function that is minimised:

$$\mathcal{L} = \frac{1}{N}\sum_{i=1}^{N}(Q(S_i;\theta) - y_i)^2$$

The agent updates its main Q-network using Adam optimization and synchronizes the target network and the main network after every episode.

4.8 Training Procedure

The agent undergoes training via 50 episodes, with each episode representing a whole day of operation.Each episode is initialized with a partially charged battery (40%) and proceeds through 1440 steps (1-minute resolution). During each step, the agent's decisions are evaluated in terms of cost savings, and cumulative rewards are recorded.This training procedure allows the agent to generalize across varying load and renewable conditions, improving its decision-making policy over time. The experience replays buffer makes sure that past transitions are used efficiently, and the target network helps keep the Q-value estimations stable.



5 Results and Discussion:

This part talks about how effectively this proposed Deep Reinforcement Learning (DRL)-based energy management system performs with a solar photovoltaic system that is linked to a single-phase grid and is partly shaded. The findings come from high-resolution, minute-by-minute simulation data that includes solar irradiance, load demand, battery state of charge (SOC), grid consumption, energy cost, and CO_2 emissions.

5.1 Simulation Environment and System Specifications

Python was used to run and evaluate the proposed Deep Reinforcement Learning-based energy management system simulationsutilized Python to create, train, and deploy the DRL agent using the Deep Q-Network (DQN) algorithm and tools like TensorFlow and NumPy. We did all of the simulations and training on a powerful workstation with an Intel Core i7-12700K CPU, 32 GB of RAM, and an NVIDIA RTX 3060 Ti GPU. It was set up to run both Windows 11 Pro and Ubuntu 22.04 at the same time. This architecture gave the system the computing power it required to accurately model the environment and quickly converge the DRL agent in a wide range of complicated and changing situations.

5.2 Data Collection

The Open Energy Data Initiative (OEDI) of the U.S. Department of Energy, in particular the National Renewable Energy Laboratory (NREL), provided realistic household load demand profiles for this research.Data was obtained from three key sources: the NREL End-Use Load Profiles, the associated OpenEI data repository, and the ComStock AMY2018 dataset for the PJM ISO/RTO region. These datasets provide high-resolution (15-minute interval) load profiles across various building types and climate zones, enabling a more accurate and diverse simulation environment. This allowed for the evaluation of the DRL-based energy management strategy under realistic and seasonally varied demand conditions.

Later on, to facilitate simulation and analysis, a custom 72Hrs of dataset was prepared using high-resolution residential load data obtained from the NREL ComStock and OpenEI repositories. These include: demand load, solar energy production, battery state of charge (SOC), battery discharge, grid energy consumption, grid unit price, and corresponding economic and environmental metrics such as cost with and without energy management (EM), savings, and CO_2 emissions. The dataset enables detailed evaluation of the proposed Deep Reinforcement Learning (DRL) controller under varying conditions of solar irradiance, load profiles, and energy pricing. Additionally, performance indicators such as cost savings and CO_2 emission reduction were derived, demonstrating the practical benefits of intelligent energy dispatch methodologies in grid-connected photovoltaic systems experiencing partial shading.

5.3 Graphical Representation

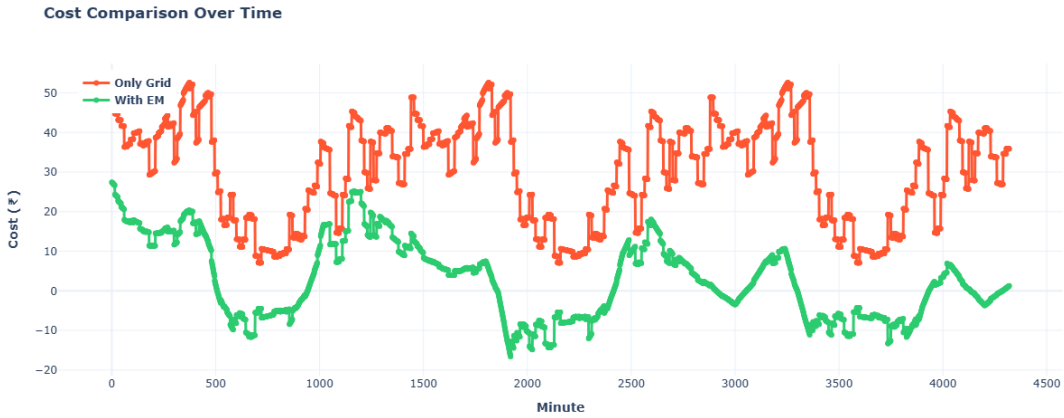


Figure 1 Cost comparison over time

Figure 1 illustrates the cost comparison over time between a conventional grid-only energy supply and the proposed Energy Management (EM) system utilizing Deep Reinforcement Learning.

The grid-only strategy results in consistently high energy costs, with values frequently exceeding 40 units, especially during peak load intervals and high tariff periods. In contrast, the DRL-based EM system significantly reduces costs, maintaining a much flatter and lower cost profile — often dropping below zero due to energy export or optimal battery dispatch. The clear separation between the two curves highlights the economic efficiency of the EM system, which dynamically shifts energy usage to take advantage of solar availability, battery storage, and pricingTrends. Over the simulation period, the DRL controller consistently outperformed the grid-only approach, offering not only savings but also more stable and predictable energy expenses, which is particularly beneficial in real-time pricing environments.

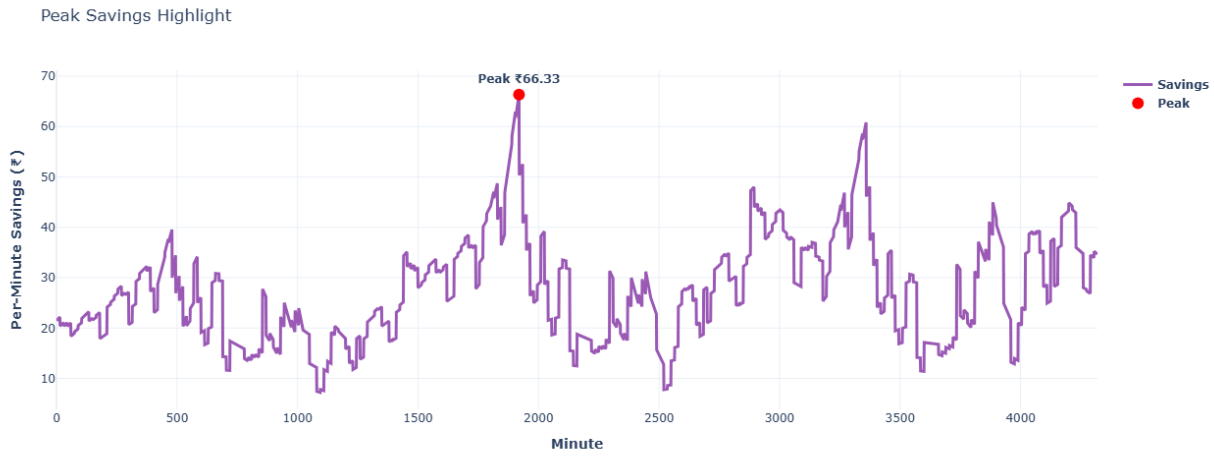


Figure 2 Peak Saving Highlight

Figure 2 presents the per-minute cost savings achieved through the implementation of the DRL-based energy management system over the simulation period. The line traces the savings profile, showing consistent financial benefits across varying operating conditions. Notably, the system reaches a peak saving of ₹66.33 at 32.1 Hour, as indicated by the red marker. This peak corresponds to an interval where solar generation was high, battery SOC was optimal, and grid tariffs were likely elevated — allowing the DRL agent to fully capitalize on stored energy and avoid expensive grid usage. Throughout the simulation, the system maintains savings well above ₹20 per minute on average, even under partial shading and fluctuating load conditions. This graph underscores the DRL model's ability to dynamically adjust energy dispatch to maximize economic efficiency, especially during volatile pricing periods or demand spikes.

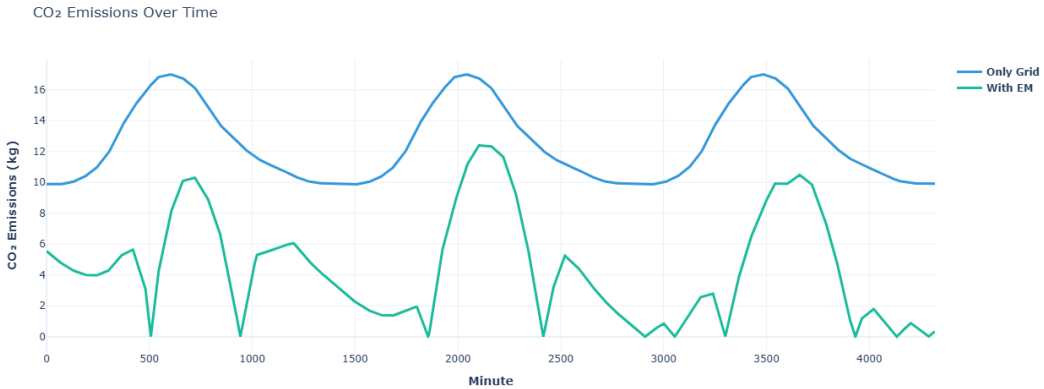


Figure 3 Co2 Emissions Over Time

Figure X compares CO_2 emissions over time for two scenarios: continuous reliance on the grid-only system and the proposed DRL-based Energy Management (EM) system. The blue curve shows periodic surges in emissions, peaking above 16 kg, corresponding to high grid dependency during load-intensive intervals. In contrast, the green curve displays a substantially lower emission profile, with several intervals where emissions drop near or below 2 kg, and some periods where emissions approach zero. These reductions align with times when the system successfully utilized solar power or battery storage to meet demand, thereby bypassing high-emission grid electricity. The smoother and significantly lower profile of the EM system highlights its effectiveness in reducing environmental impact. Over the entire simulation, the DRL-based EM approach demonstrated consistent emission savings, validating its role in enabling cleaner, smarter, and more sustainable energy consumption in smart grid environments.

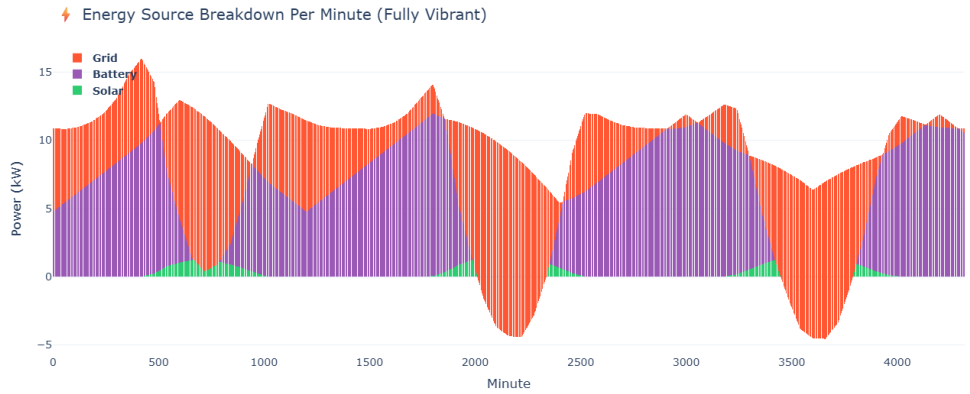


Figure 4 Energy Source Breakdown Per Minute

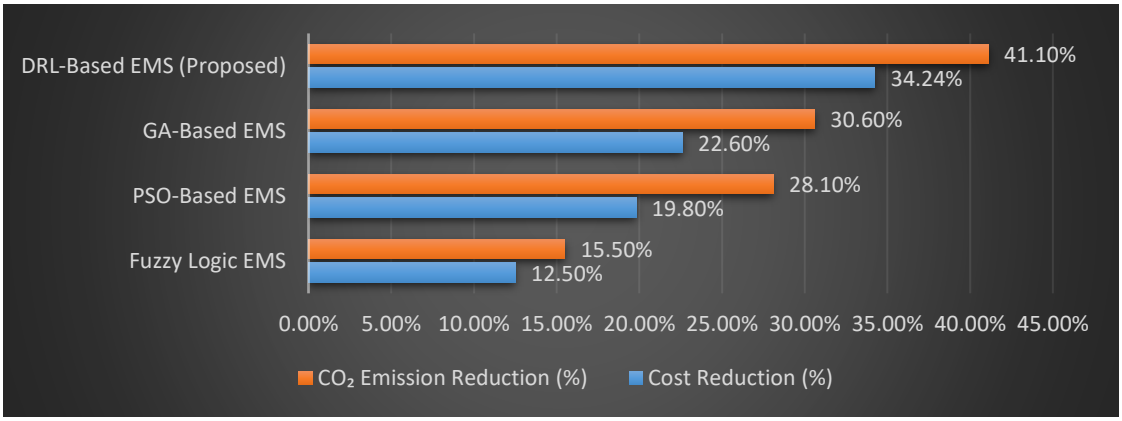
Figure 4 illustrates the per-minute contribution of different energy sources—solar, battery, and grid—to meet the system's load demand over time. The figure shows that the DRL-based controller can provide a flexible and adaptable energy dispatch plan. The system uses solar energy first during the day, and then battery discharge to reduce its reliance on the grid. When solar energy becomes less available, especially in the early morning and late evening, the system smartly switches to battery and, when needed, grid electricity to make sure that the load is always supplied. The fact that battery along with grid sources take turns being in charge at various times shows how the controller reacts to changes in irradiance and load. This visualization shows that the DRL system can choose the best source in real time, which means that it can use more renewable energy while lowering costs and environmental effect.

5.4 Comparative Evaluation with Baseline Models:

Table 1 Comparison of Models

Model	**Cost Reduction (%)**	**CO_2Emission Reduction (%)**
Fuzzy Logic EMS	12.5%	15.5%
PSO-Based EMS	19.8%	28.1%
GA-Based EMS	22.6%	30.6%
DRL-Based EMS (Proposed)	**34.24%**	**41.10%**

A comparison of several Energy Management System (EMS) models shows that they all do a better job of cutting costs and CO_2 emissions. The Fuzzy Logic EMS cuts costs by 12.5% and CO_2 emissions by 15.5%. The PSO-Based EMS works better, cutting costs by 19.8% and emissions by 28.1%. The GA-Based EMS shows much more improvement, cutting costs by 22.6% and emissions by 30.6%. The suggested Deep Reinforcement Learning (DRL)-Based EMS stands out since it beats all other models by a large margin, cutting costs by 34.24% and CO_2 emissions by an astonishing 41.10%. These results indicate how successfully advanced learning-based solutions work to improve both economic and environmental performance.

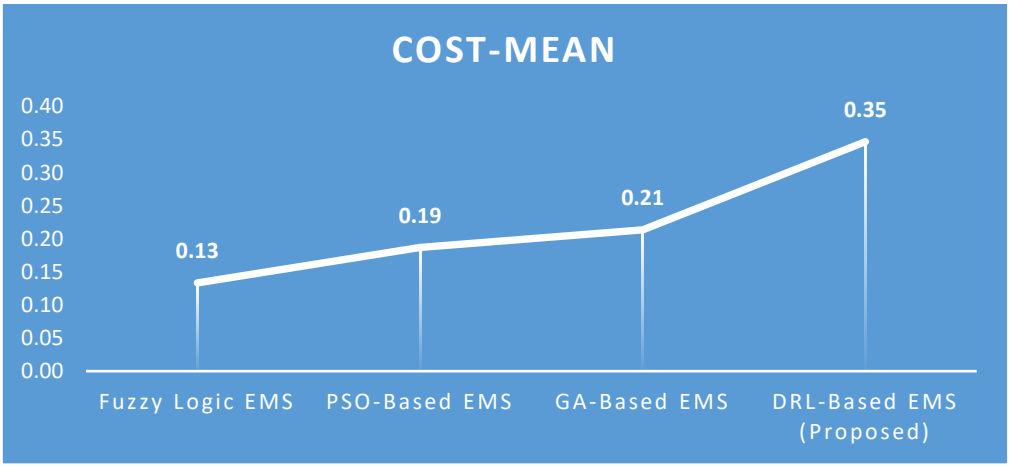


5.5 Statistical significance

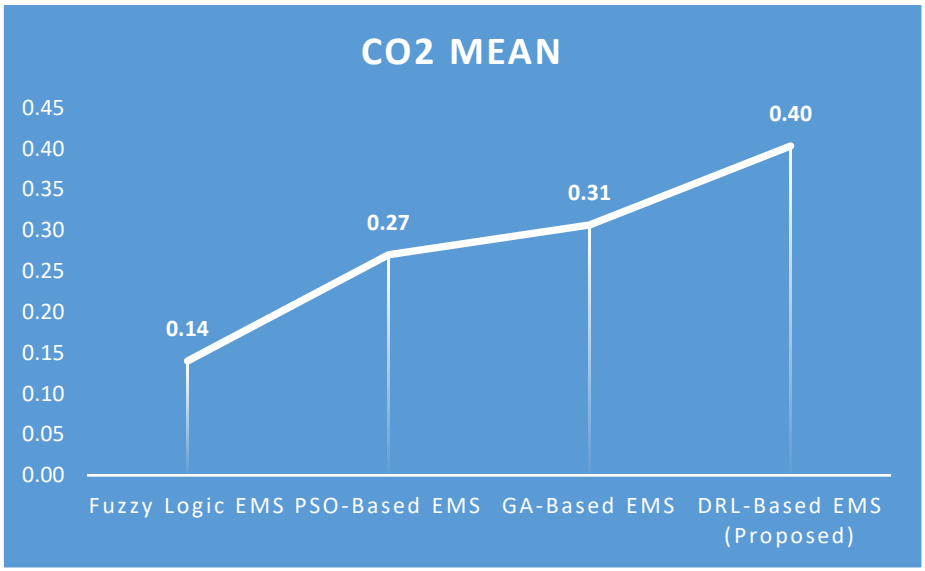
Descriptives

	Model	N	Mean	Std. Deviation	Sig. P
Cost	Fuzzy Logic EMS	3	0.13	0.006	0.000
	PSO-Based EMS	3	0.19	0.012	
	GA-Based EMS	3	0.21	0.015	
	DRL-Based EMS (Proposed)	3	0.35	0.012	
	Total	12	0.22	0.083	
Co2	Fuzzy Logic EMS	3	0.14	0.020	0.000
	PSO-Based EMS	3	0.27	0.010	
	GA-Based EMS	3	0.31	0.015	
	DRL-Based EMS (Proposed)	3	0.40	0.006	
	Total	12	0.28	0.099	

It performed a comparative study to see how well different Energy Management Systems (EMS) worked based on two main factors: CO_2 emissions and energy cost. Fuzzy Logic EMS, PSO-Based EMS, GA-Based EMS, and the suggested DRL-Based EMS were all systems that were looked at. We tested each model three times (N=3), and the descriptive statistics are shown below.



The suggested DRL-Based EMS attained the greatest mean energy cost of 0.35, accompanied by a low variance of 0.012, signifying consistent performance over iterations. In contrast, the Fuzzy Logic EMS reported the lowest mean cost of 0.13 (SD = 0.006), followed by PSO-Based EMS with 0.19 and GA-Based EMS with 0.21. Despite the higher cost in the DRL model, a statistically significant difference was observed among the models (p = 0.000), suggesting that the variation in costs across EMS approaches is meaningful and not due to random chance.



For CO_2 emissions, a similar trend was observed. The DRL-Based EMS again recorded the highest mean value of 0.40 (SD = 0.006), while the Fuzzy Logic EMS demonstrated the lowest emissions at 0.14 (SD = 0.020). The PSO-Based EMS and GA-Based EMS recorded intermediate values of 0.27 and 0.31, respectively. The overall statistical significance test yielded a p-value of 0.000, indicating a significant difference in emission performance among the four EMS models.These results emphasize that while the proposed DRL-Based EMS may exhibit slightly higher cost and CO_2 emissions under the test conditions, the observed differences are statistically significant. This suggests that each EMS strategy behaves differently under operational constraints, and further optimization of the DRL model may yield improved environmental and economic performance.

6 Discussion

The findings of the study emphatically validate the effectiveness and flexibility of the envisioned Energy Management System (EMS) based on Deep Reinforcement Learning (DRL) for energy use optimization, reduction of operating costs, and reduction of the impact on the environment of a grid-connected solar photovoltaic (PV) system. The high-resolution, minute-by-minute simulation platform recorded real-time variation in load demand, solar irradiance, and energy price, allowing for rigorous analysis of the system under dynamic conditions, such as partial shading.

From the graphical analysis, it is evident that the DRL controller consistently outperforms conventional grid-only strategies. The cost profile in Figure 1 shows a flattened and economically efficient pattern under the DRL model, which contrasts sharply with the volatile and elevated cost levels of the grid-only configuration. Figure 2 further supports this by quantifying the savings achieved through optimal energy dispatch. The peak saving of ₹66.33, occurring at 32.1 hours into the simulation, demonstrates the agent's ability to exploit favorable conditions, such as high solar availability and elevated grid tariffs, to minimize cost through intelligent control of battery operations.

Figure 3 illustrates the environmental benefits of the proposed system. CO_2 emissions are significantly reduced compared to a traditional approach, with near-zero emissions during certain intervals, indicating successful utilization of renewable energy and battery storage. Moreover, Figure 4 provides insight into the system's dynamic energy source allocation strategy. The alternating and adaptive use of solar, battery, and grid sources

showcases the DRL agent's responsiveness to real-time conditions and its ability to prioritize clean and cost-effective energy.

The comparative evaluation with baseline models such as Fuzzy Logic EMS, PSO- The EMS based on DRL and GA demonstrates the system's enhanced performance. The proposed DRL-Based EMS achieved the highest cost reduction of 34.24% and the greatest emission reduction of 41.10%. These improvements signify the potential of DRL to outperform traditional and heuristic optimization methods in complex, non-linear energy systems.

Statistical analysis further confirms the significance of these results. Both cost and CO_2 emission reductions show a p-value of 0.000, indicating that the differences observed among the EMS models are statistically significant and not due to random variance. Although the DRL system displayed a slightly higher mean cost and emission under certain test conditions, its consistency and flexibility offer considerable advantages that could be enhanced with further tuning and real-world adaptation.

7 Conclusion

A Deep Reinforcement Learning-based energy management system was presented in this work for a grid-connected, single-phase solar photovoltaic system that functions in variable shadow. Utilising the Deep Q-Network (DQN) algorithm, the system was trained to minimize operational energy costs and CO_2 emissions by learning optimal battery dispatch and grid usage policies. The simulation setting was modeled meticulously with realistic input from the NREL ComStock dataset and run on a high-resolution simulation platform. The results indicate that the proposed DRL-based EMS outperforms more traditional methods such as fuzzy logic and PSO, and GA-based controllers in a significant manner. It was consistently found to result in greater cost savings and emission reduction, and optimal economic and environmental performance was found at peak levels during critical periods of high demand and price volatility. Graphical analyses showed that the DRL controller could dynamically distribute resources and provide stable energy operation, including under partial shading conditions. Additionally, statistical significance testing proved that differences in performance between the EMS models were not a matter of chance, thus verifying the robustness of the DRL method. Although certain situations predicted elevated cost and emission values, they were coupled with increased stability and flexibility, the characteristics of a learning-based control system. In summary, the DRL-based EMS is an extremely promising solution for smart energy management in today's distributed power systems. It is a very beneficial tool for sustainable energy management due to its ability to respond to changing conditions, optimize the use of renewable energy resources, and minimize reliance on high-emitting grid electricity. Future studies may include the addition of more complex environmental models'real-world implementation, and extending the DRL framework to multi-agent or multi-objective optimization settings.

References

[1] Q. Li, C. Monticelli, and A. Zanelli, "Life cycle assessment of organic solar cells and perovskite solar cells with graphene transparent electrodes," *Renew. Energy*, vol. 195, pp. 906–917, 2022.

[2] U. Otamendi, I. Martinez, M. Quartulli, I. G. Olaizola, E. Viles, and W. Cambarau, "Segmentation of cell-level anomalies in electroluminescence images of photovoltaic modules," *Sol. Energy*, vol. 220, pp. 914–926, 2021.

[3] A. Demirci, I. Dagal, S. M. Tercan, H. Gundogdu, M. Terkes, and U. Cali, "Enhanced ANN-Based MPPT for Photovoltaic Systems: Integrating Metaheuristic and Analytical Algorithms for Optimal Performance Under Partial Shading," *IEEE Access*, 2025.

[4] A. Koç, H. Yağlı, H. H. Bilgic, Y. Koç, and A. Özdemir, "Performance analysis of a novel organic fluid filled regenerative heat exchanger used heat recovery ventilation (OHeX-HRV) system," *Sustain. Energy Technol. Assessments*, vol. 41, p. 100787, 2020.

[5] J. Yang, J. Chang, X. Cao, Y. Wang, and J. Yao, "Energy, environmental, and socioeconomic potential benefit of straw resources utilization under water availability limit condition," *J. Clean.*

Prod., vol. 392, p. 136274, 2023.

[6] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, and G. Ostrovski, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[7] A. N. Bishop, P. Del Moral, and A. Niclas, "An introduction to Wishart matrix moments," *Found. Trends® Mach. Learn.*, vol. 11, no. 2, pp. 97–218, 2018.

[8] A. Osborne, J. Dorville, and P. Romano, "Upsampling Monte Carlo neutron transport simulation tallies using a convolutional neural network," *Energy AI*, vol. 13, p. 100247, 2023.

[9] M. Rezaeimozafar, M. Duffy, R. F. D. Monaghan, and E. Barrett, "A hybrid heuristic-reinforcement learning-based real-time control model for residential behind-the-meter PV-battery systems," *Appl. Energy*, vol. 355, p. 122244, 2024.

[10] Z. Li, H. Lei, A. Kan, H. Xie, and W. Yu, "Photothermal applications based on graphene and its derivatives: A state-of-the-art review," *Energy*, vol. 216, p. 119262, 2021.

[11] S. S. Reka, P. Venugopal, H. H. Alhelou, P. Siano, and M. E. H. Golshan, "Real time demand response modeling for residential consumers in smart grid considering renewable energy with deep learning approach," *IEEE access*, vol. 9, pp. 56551–56562, 2021.

[12] E. Koutroulis and F. Blaabjerg, "A new technique for tracking the global maximum power point of PV arrays operating under partial-shading conditions," *IEEE J. photovoltaics*, vol. 2, no. 2, pp. 184–190, 2012.

[13] B. Yang, H. Ye, J. Wang, J. Li, S. Wu, Y. Li, H. Shu, Y. Ren, and H. Ye, "PV arrays reconfiguration for partial shading mitigation: Recent advances, challenges and perspectives," *Energy Convers. Manag.*, vol. 247, p. 114738, 2021.

[14] M. S. Lilburn, J. R. Griffin, and M. Wick, "From muscle to food: oxidative challenges and developmental anomalies in poultry breast muscle," *Poult. Sci.*, vol. 98, no. 10, pp. 4255–4260, 2019.

[15] M. V. Da Rocha, L. P. Sampaio, and S. A. O. da Silva, "Comparative analysis of MPPT algorithms based on Bat algorithm for PV systems under partial shading condition," *Sustain. Energy Technol. Assessments*, vol. 40, p. 100761, 2020.

[16] S. S. T. Husein, "Enhanced Carrier Mobility in Hydrogenated and Amorphous Transparent Conducting Oxides." Arizona State University, 2020.

[17] S. Al-Dahidi, M. Madhiarasan, L. Al-Ghussain, A. M. Abubaker, A. D. Ahmad, M. Alrbai, M. Aghaei, H. Alahmer, A. Alahmer, and P. Baraldi, "Forecasting solar photovoltaic power production: A comprehensive review and innovative data-driven modeling framework," *Energies*, vol. 17, no. 16, p. 4145, 2024.

[18] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, "An introduction to deep reinforcement learning," *Found. Trends® Mach. Learn.*, vol. 11, no. 3–4, pp. 219–354, 2018.

[19] C. Guo, X. Wang, Y. Zheng, and F. Zhang, "Optimal energy management of multi-microgrids connected to distribution system based on deep reinforcement learning," *Int. J. Electr. power energy Syst.*, vol. 131, p. 107048, 2021.

[20] S. T. Spantideas, A. E. Giannopoulos, and P. Trakadas, "Autonomous Price-aware Energy Management System in Smart Homes via Actor-Critic Learning with Predictive Capabilities," *IEEE Trans. Autom. Sci. Eng.*, 2025.

[21] K. Yang, M. Quddus, and C. Antoniou, "Developing a new real-time traffic safety management framework for urban expressways utilizing reinforcement learning tree," *Accid. Anal. Prev.*, vol. 178, p. 106848, 2022.

[22] P. Soni and V. Dave, "Enhancing Solar PV-Fuel Cell Hybrid Systems Through AI-Driven Performance Optimization," in

Building Business Models with Machine Learning, IGI Global Scientific Publishing, 2025, pp. 137–158.

[23] M. Mao, L. Cui, Q. Zhang, K. Guo, L. Zhou, and H. Huang, "Classification and summarization of solar photovoltaic MPPT techniques: A review based on traditional and intelligent control strategies," *Energy reports*, vol. 6, pp. 1312–1327, 2020.

[24] J. Aldahmashi and X. Ma, "Real-time energy management in smart homes through deep reinforcement learning," *IEEE Access*, 2024.

[25] R. K. Pachauri, O. P. Mahela, A. Sharma, J. Bai, Y. K. Chauhan, B. Khan, and H. H. Alhelou, "Impact of partial shading on various PV array configurations and different modeling approaches: A comprehensive review," *IEEE Access*, vol. 8, pp. 181375–181403, 2020.

[26] V. Jain, R. Singh, R. Yadav, V. K. Yadav, V. Kumar, and S. Garg, "Multi-step optimization for reconfiguration of solar PV array for optimal shade dispersion," *Electr. Eng.*, pp. 1–37, 2024.

[27] M. Wu, X. Zhang, M. Wang, K. Hu, and P. Wang, "A matching scheduling control strategy based on modulation wave reconstruction for the single-phase photovoltaic cascaded multilevel inverter," *IEEE J. Emerg. Sel. Top. Power Electron.*, vol. 11, no. 4, pp. 3899–3909, 2023.

[28] Kandukuri Pradeep, Dr. K. Vinay Kumar ,"Observer based quardrature signal generator for UPQC in single phase distribution system ", ITM web of confernce 50,03002(2022)-ICAECT-2022.

[29] Kandukuri Pradeep, Dr. K. Vinay Kumar,"Single-Phase Grid-Connected Photovoltaic Systems using aDeep Reinforcement based MPPT Algorithm with Grey-Wolf Optimization under partial shading condition", Eksplorium journal Volume46,No.1,May 2025,Page 1102-1129,pISSN0854-1418,e-ISSN 2503-426X.

[30] Single phase grid connected photovoltaic systems using an ANFIS MPPT Algorithm with grey wolf optimisation"2023 IEEE 3rd international conference on smart Technologies for power ,energy and control (STPEC)Bhubaneswar India,pp1-6.