Deep Reinforcement Learning for Real-Time Energy Management in Nano-Grids through DQN Approach for Cost and Grid Dependency Reduction

Jarabala Ranga[1], Dr. Gopinath Palai[2], Prof. (Dr.) Rabi N Satpathy[3]

Abstract

Effective control of energy in nano-grids is necessary for achieving energy independence, cutting energy expenses and making use of renewable energy. A new DQN-based reinforcement learning approach is presented in this study for the management of energy in AI-connected smart grids. The state space in the MDP represents the PV output, the load demand, battery SOC and changing electricity rates of the time instant. The possible strategies included in this work are divided into five interpretable groups: charging, discharging, interacting with the grid and scheduling energy output. When reward functions are designed well, they consider the variables of cost, ageing batteries and load satisfaction. The DQN uses experience replay, a target Q-network and exploration with a gradually decreasing epsilon. Experiments were done using real solar power and energy consumption information. Analyzing against nine other models — rule-based logic, MPC, fuzzy logic EMS and hybrid reinforcement learning systems — for eight main metrics, the DQN outperformed all other models. Very importantly, the DQN attained better results, reducing energy cost by 5.13 USD/day, using 94.28% renewables, achieving 91.93% efficiency, LPSP of only 1.32% and reaching Q-value convergence fast in 439 steps. The outcomes reveal that the suggested model can adapt to changing conditions and is suitable for modern smart grid operations.

[1]Research Scholar, Faculty of Engineering and Technology, Sri Sri University, Cuttack, Odisha, India, Email: jarabalaranga@gmail.com

[2]Professor, Faculty of Engineering and Technology, Sri Sri University Cuttack, Odisha, India

[3]Dean, Faculty of Engineering and Technology, Sri Sri University Cuttack, Odisha, India

Keywords-Nano-grid, Deep Q-Network, Energy Management System, Reinforcement Learning, Smart Grid, Battery Optimization, Renewable Energy, Markov Decision Process, Demand Response, PV Integration

1. Introduction

The world's energy systems are going through major changes because people want sustainable, dependable and local power. At the center of these changes is the growth of nano-grids, small energy grids designed to meet the electrical needs of households, isolated facilities or tiny communities [1] [2]. Nano-grid systems differ from standard grids by being capable of working alone or with the main grid which provides better management, reliability and a greater chance for nearby solar PV and wind energy to be included. The frequent changes in renewable energy output, sudden changes in power demand and lack of adequate storage introduce great challenges when managing energy flows in these systems.Standard energy management strategies in nano-grids are not effective for handling the stochastic and nonlinear situations they face. Typically, these systems rely on fixed views and aren't updated when things in the world shift. In this situation, AI helps energy management by giving centralized, flexible and predictive solutions. AI systems are able to constantly review sensor data, predict peak renewable energy and electricity use and choose the best ways to handle energy, storage and the grid [3] [4]. People are showing more interest in machine learning and deep reinforcement learning due to their ability to design models for energy systems and make the best control choices in different conditions. It turns out that Deep Q-Networks (DQNs) and Long Short-Term Memory networks are effective methods for noticing important time patterns and pick good moves after many experiments [5] [6]. With these models, it is possible to forecast the demand for energy and how much to store in batteries, decide when batteries should be charged or discharged and save money by switching battery operation according to varying electricity rates and grid notices.In addition, artificial intelligence enables energy systems in nano-grids to access up-to-date data from smart meters, weather predictions and signals from the energy market. Because of data fusion, decisions can be made that boost local energy reliability and contribute to achieving other key aims such as reducing the carbon footprint, cutting peak loads and supporting grid stability [7]. Smart grid agents can manage peaks in demand, set back loads not required at certain times and blend their efforts with nearby mini-grids to create sturdy energy networks. Figure 1 illustrates the key features of AI-enabled intelligent energy management in nano-grids.

Figure 1. Key Features in AI-Enabled Intelligent Energy Management for Nano-Grids

They support sustainability by helping make the best use of renewable sources and decreasing the need for fossil fuel generators to back up the system. By learning and predicting how users behave, AI makes it possible to support management strategies on the demand side and encourages better and more personalized ways to use energy [8].While using AI in nano-grids is beneficial, it raises problems like safeguarding user data, reducing the processing power needed and ensuring quick responses. These problems can be solved by using thin AI models that fit on edge devices, distributed learning methods to keep data private and powerful algorithms made for high uncertainty.In essence, this mix of AI and nano-grid represents a big change in how distributed energy systems are set up [9] [10]. When AI is integrated within small local network systems, it supports better energy use, allows communities to maintain their independence and supports sustainable living. This paper examines how to design, implement and assess an intelligent AI-driven energy management system intended for real-time control within nano-grid networks.

While energy systems are being built with decentralization, sustainability and smart machines in mind, nano-grids are now key to the development of future energy networks. Nano-grids powered by local renewable energy, with storage and adaptable loads, are able to operate alone or in connection with the main power system. Yet, controlling the movement of energy in these systems is difficult because solar power production varies, electricity demand changes and prices set by the market fluctuate. Older methods used in energy management — including set rule principles, threshold setting and model-predictive control (MPC) — do not easily cope with non-constant and uncertain conditions. Over the last few years, RL has been used for real-time control because it learns

successfully and adapts through direct interaction with changing environments.In our proposed work, we put forward a Deep Q-Network (DQN) intelligent energy management system that is suitable for nano-grid environments. Energy optimization is modeled using an MDP and an agent is given the vector of battery SOC, PV generation, electricity price and load demand. Then, it chooses the most suitable strategy from a limited set containing battery charging, discharging and energy trading decisions. The approach saves money in operations, boosts how much renewable energy is used, increases battery longevity and solidifies the system's reliability. Our evaluation reveals that the DQN method outperforms traditional methods in different measurement areas.

1.1 Research Motivation

The fast development of distributed energy and growing use of smart microgrids have made managing energy in nano-grid environments both a challenge and an opportunity. Because renewable energy is unpredictable, because grid electricity is getting more expensive and because many now want to be self-sufficient, we need smart, data-driven control systems. Part of what makes traditional rule-based energy management systems ineffective is that they don't optimally respond to changes in load, the amount of solar electricity generated and market electricity prices. By using Deep Q-Networks, RL enables systems to learn on their own how to control themselves by interacting with what surrounds them. The motivation for this research comes from wanting to build a strong AI-powered system that can respond to chaotic and stochastic patterns in modern nano-grids and make energy use more efficient.

1.2 Significance of the Study

The use of DQN within nano-grid energy optimization is introduced in this study to help fill a key gap in applying AI to sustainable energy systems. Its benefit is through modeling energy management with an MDP so that changes can be managed flexibly and confidently in uncertain situations without requiring predictions. It responds appropriately to changes in the environment such as when solar panels work differently, when batteries change their status, shifts in power need and when electricity prices vary. Improvements in cost, renewable energy use and system efficiency over benchmark models suggest the proposed approach could work well in several new energy systems.

1.3 Problem Statement

Although nano-grids are capable of generating and storing energy locally, their control schemes are usually inflexible and cannot respond well to the changing, random needs for energy. Because these approaches are not dynamic, they do not account for trade-offs related to battery life, electricity prices and what devices are given priority during lots of energy use which results in both hitting the battery harder and losing money. Additionally, there are few systems that can react to changing requirements in situations where forecasts are not available. This study looks at why there isn't a framework for managing energy in nano-grids that works in real-time, is scalable and delivers the best results with low consumption and higher cost-effectiveness and sustainability.

1.4 Recent Innovations and Challenges

Over the last few years, researchers have made many improvements on using machine learning for different aspects of smart grid control. Researchers have recognized value-based methods like DQN because they can run in environments that don't need a model and use feedback. Still, most existing approaches make the subject too simple or work poorly because of overfitting, bad convergence and insufficient explanations. To incorporate battery aging, pricing signals received from the grid and deciding on more than one goal simultaneously is still difficult. Difficulties in using AI for energy systems come from having to handle real-time decisions, weak hardware and explainable solutions.

1.5 Key Contribution of the study

Developing and testing a DQN-reinforced learning strategy for energy management in nano-grids is the main achievement of this study. The model updates the state space in real-time (SOC, PV, load, price), provides a limited set of discrete action options and uses a reward function that cares about cost, aging and power reliability in one goal. Key technological developments were a significant part of the project.

- A DQN framework that scales, uses targeted experience replay and includes a target network to operate well in changing energy scenarios.
- A system that combines handling battery lifecycle with economic dispatch as part of its reward engineering strategy.
- It is shown in both an evaluation of eight criteria and comparison to nine reference models that the prosed model leads the way in energy cost (under 5.13 USD per day), fully using renewable sources (94.28%) and having the least LPSP (1.32%).
- This allows deployment with model quantization for edge use, making it helpful in real-time applications on microcontrollers and IoT devices.

1.6 Rest of Section of the Study

The remainder of the paper is organized as follows: Section 2 provides an in-depth review of related studies on traditional energy management systems, optimization-based controllers, and recent advancements in reinforcement learning for smart grid applications. Section 3 details the proposed Deep Q-Network (DQN) methodology, including the design of state and action spaces, reward function formulation, training mechanisms,

and deployment considerations. Section 4 presents the results and discussion, comparing the proposed model with nine baseline approaches across eight performance metrics. Finally, Section 5 concludes the study and outlines future directions for extending the model to multi-agent settings and scalable grid systems.

2. Related Works

The research examined TENGs as a way to provide power to future smart grid sensor networks. The use of batteries in sensors was challenging since they are costly, have a short life and cause pollution. Their ability to extract energy from magnetic fields from the power grid stands out in demonstration tests[11]. Materials on magnetic-field-driven TENGs were thoroughly studied, showing developments in both their setup and how they create energy. Experts noted that TENGs are fitted for uses needing little power and can be part of becoming carbon neutral. One future aim was to use them broadly in smart energy systems that depend on low-power and eco-friendly sensing.

An analysis of PCM-based building materials containing nanoparticles found that the materials became more heat conducting, yet lost some of their ability to store or release heat. Nanoparticle concentration appeared to reduce the amount of electric energy the films could save [12]. The model indicated that a 1.7% decrease in heating energy savings could be expected from adding 3 vol% of particles to PCM. Even with better heat transfer, the wall system's ability to control temperature decreased. The results suggest that adding nano-particles to PCMs was not useful in hot- summer Mediterranean climates for cutting overall energy use.

The researchers employed graphene nanoplatelets and expanded graphite to make shape-stabilized nano-enhanced phase change materials (ss-NePCM) which helped increase conductivity and stop leakage. With reinforcement led by engineers, the composite saw conductivity raise by 112% and effectively controlled leakage [13]. FTIR, UV–Vis and TGA methods verified that the material maintained its chemical state and stability after undergoing 250 rounds of heating and cooling. Although the latent heat was somewhat lowered, the improved optical and thermal features were worth the small difference. The ss-NePCM composite performed well in high-performance thermal energy storage situations that required leak protection and high thermal efficiency.

The simulation project looked at how cold energy storage systems filled with spherical nano-enhanced phase change capsules respond thermally. The study looked at several capsule designs and materials to learn how they affect both charging and discharging times [14]. The hexagon layout allowed the fastest charging and graphene nanoparticles improved the process time by as much as 22.22%. When the capsules were smaller, they responded more quickly to changes in temperature and graphite shells outperformed plastic and glass. It was found that systems combining nanoparticle doping and geometrical design did better at heat transfer which makes them suitable for ideal energy storage.

A solution was introduced to deal with the effects of fault resistance and cloud energy storage on the balance between reclosers and fuses in inverter-dominated microgrids. Synchronverter energy storage led to the creation of significant fault currents which disrupted the operation of traditional protection systems [15]. Under different resistance circumstances, the algorithm adjusted recloser functionality to reduce the risk of fuse replacement. Researchers saw that with simulations, there were greater coordination and improved reliability. Merging cloud storage systems with advanced inverter interfaces brought challenges, but the approach we suggested enabled dependable isolation of faults and a strong response to protection in modern DC microgrids.

The synthesis and use of nanoscale MOFs and COFs were analyzed in a comprehensive review. The group studied how changing the structure and size of particles affected energy storage and the catalytic process. A variety of approaches in synthesis allowed for precise shape control, boosting porosity, surface area and the useful properties of these materials[16]. Applications examined included various types such as zinc, lithium, sodium and supercapacitor batteries. According to the review, nano-MOFs and COFs form a versatile basis for energy storage technologies, since they can be modified to match specific material shapes.

Scientists created a new bone implant that uses grids to release drugs and protect against another infection of the bone. Drug delivery occurred only under acidic conditions using a pH-responsive release mechanism using Ce^{3+}-PO_4^{3-}. A balanced charge in our cells helped retain drugs at pH values matching healthy conditions. Analysis of the data proved that the drug was only released in the infected area and that infections were controlled [17]. In studies using living cells, researchers found that bone healed more quickly and the therapy worked only where it was applied. It was possible to provide precise doses of medicine using the implant, suggesting a good way to sustain effective infection prevention and promote bone repair in orthopedic treatments.

An in-depth review found that triboelectric nanogenerators (TENGs) can efficiently gather unpredictable, low-frequency mechanical energy, known as high entropy energy. TENGs were chosen as likely alternatives for generating power efficiently at the local level and without extra support. Among their strong points were the ability to handle different power levels, their easy design and a strong voltage output. The study split TENG systems into micro-power supplies, direct energy converters and sensing networks [18]. Problems associated with material wear and use in factories were recognized. Plans are being made to improve efficiency and the use of these platforms in extensive sustainable infrastructure, covering the Internet of Things and platforms designed to work without depending on energy sources.

The paper explained the transition of microgrids into smart grids by discussing difficulties with technology, inefficiencies and linking problems. Working with microgrids was difficult because it needed to deal with the problems of variable renewables, fluctuating demand and weak information sharing[19]. Smart grids were developed to handle energy automatically, use pricing that changes with the market and use secure feedback to improve their network. Better conversations between consumers and providers led to more dependable energy and

greater adaptability. The study prepared steps that help convert microgrids into efficient and smart areas of the grid for improved energy distribution, less energy loss and better awareness of the grid.

A routing protocol designed for wireless sensor networks was introduced with the virtual backbone structure. This approach managed communication so that nodes could all use sleep mode to save energy. Even though more devices and bandwidth were added, the network topology allowed the backbone to adjust easily and save power when it wasn't in use[20]. Results from comparative simulation found that these schemes used less energy and lasted longer than regular schemes. QoS was maintained and the amount of data over the radio and equal usage were decreased using the proposed technique for long-term deployments of sensor networks.

A thorough analysis assessed micro- and nano-perovskite materials for their use in solar applications, highlighting their strength in light harvesting, light emission and resilience. Changes to the structure made it easier for light to reach the cells, reflected less light and improved how carriers are transported. Solar cells, Solid-state devices, LEDs and photodetectors were part of the application category [21]. The paper summarized both the techniques used to make the materials and the way those materials work after being modified. Multifunctional systems using advanced micro-nano designs were highlighted as excellent for developing flexible electronics. Studies confirmed that using structural engineering of perovskites improved the efficiency and lifespan of optoelectronic devices. The thermal performance of MXene-silicone oil nano-fluids was analyzed with ANNs. Studies of experimental data found that temperature and the concentration of nanoparticles influenced how thermal properties worked. An artificial neural network called a multilayer perceptron was able to predict conductivity well, giving a correlation coefficient of 0.99687 [22]. A relationship between the variables was found by examining the outcomes of experiments. MXene-based fluids fared better at heat transfer than others, thus making them suitable for thermal management. The estimation of nano-fluid properties could be successfully carried out using the light, stable ANN model under different working conditions.

The solidification and melting behavior of NEPCMs was studied using simulations in ANSYS/Fluent. The addition of 2.5% and 5% copper oxide nanoparticles to octadecane raised the thermal conductivity and made the material's phase transition happen more swiftly. Examination of experiments showed that nanoparticle loading reduced both the enthalpy and specific heat capacity [23]. Improved heat transfer was predicted by the simulation in both the charging and discharging cycles of CuO-enhanced PCMs. Rates of heat transfer increased by as much as 59.8% in solidification. It was found that using NEPCMs improved the overall performance of thermal storage systems by making them respond much faster to changes in the ambient temperature.

Using the novel QCA1 gate, the full adder and full subtractor circuits were implemented using a reversible logic design based on QCA. This design reduced the number of cells by 34.25% and the overall area by 54.37% over existing designs. Utilizing the QCA1 gate, the system's logic design consumed little energy and was more compact. The gate was found to work correctly through quantum simulations [24]. Low energy dissipation in reversible circuits made them preferred for use in nano-communication. The work confirmed that QCA will be useful for computing in future nanoelectronic systems.

A standalone microgrid that combined solar, wind, biomass, biogas and micro-hydro systems was designed to bring electricity to rural areas. The Differential Evolution algorithm was used in capacity planning to find the most efficient way to deliver consistent power without incurring high expenses [25]. By using optimal system sizing in an Indian hilly are, the per-unit energy cost fell to $0.12/kWh and the total cost to operate the system was $2.12 million per year. Resource accessibility and demand profiles were integrated into the model which maintained resilience in various operation scenarios. Sensitivity analysis showed that the proposed microgrid meets the needs of rural communities sustainably and without economic challenges.

Table 1 Comparative Summary of Advanced Energy Systems and Smart Grid Technologies

Reference	Method	Objective	Limitation
Chen et al. [11]	Review of magnetic field-based triboelectric nanogenerators	To analyze the feasibility of TENGs for powering future smart grid sensors	Lack of large-scale deployment data and environmental durability challenges
Tunçbilek et al. [12]	Simulation of PCM-embedded walls with nanoparticles	To evaluate thermal conductivity and energy-saving performance in buildings	Nanoparticles reduced latent heat, negatively impacting energy savings
Islam et al. [13]	Ultrasonication and vacuum impregnation of graphene/graphite composites	To enhance thermal conductivity and leakage resistance in PCMs	Minor reduction in latent heat due to additives
Ghojavand et al. [14]	2D/3D simulation of cold storage capsules using NEPCMs	To optimize capsule geometry and materials for heat transfer efficiency	High computational complexity and sensitivity to capsule diameter
Yazdaninejadi et al. [15]	Adaptive protection scheme for synchronverter-based microgrids	To ensure fault isolation and coordination in microgrids with cloud storage	Requires accurate fault resistance estimation and inverter modeling

Hu et al. [16]	Review of synthesis strategies for nano-MOFs and nano-COFs	To present advances in nano-frameworks for energy storage and catalysis	Scale-up and structural stability under operation remain unresolved
Zhang et al. [17]	pH-triggered drug release implant with Ce^{3+}-PO_4^{3-} mechanism	To prevent osteomyelitis recurrence and promote bone regeneration	Effectiveness depends on precise pH threshold control
Chen et al. [18]	Literature review of TENGs for high entropy energy harvesting	To highlight applications of TENGs in micro-power and sensing systems	Commercial scalability and material fatigue issues
Badal et al. [19]	Analytical review of microgrid–smart grid transition models	To outline solutions for control, communication, and integration challenges	Cybersecurity and interoperability require further investigation
Rayan et al. [20]	Backbone Energy-Efficient Sleeping (BEES) topology control	To improve energy management and routing in WSNs	Limited adaptability under highly mobile sensor configurations
Zhan et al. [21]	Review of structured perovskite materials for optoelectronics	To enhance optical performance, light management, and robustness	Fabrication scalability and mechanical fragility in some structures
Boobalan et al. [22]	ANN modeling for predicting nano-fluid thermal conductivity	To estimate MXene-fluid behavior across varying temperatures and concentrations	Experimental validation limited to narrow operating ranges
Cofré-Toledo et al. [23]	Experimental and simulation study on CuO-NEPCMs	To evaluate thermal storage efficiency through phase change modeling	Lower heat capacity observed due to nanoparticle doping
Das et al. [24]	QCA-based reversible circuit design using QCA1 gate	To develop area-efficient and low-power logic circuits	Quantum realization complexity and fabrication immaturity
Kamal et al. [25]	DE algorithm-based hybrid microgrid optimization	To reduce cost and improve reliability in off-grid rural settings	Dependent on accurate renewable resource and load forecasting

Table 1 compares various approaches, goals and difficulties encountered in advanced energy systems and smart grid technologies. Some methods are reviewing energy from nanogenerators, exploring thermal simulations of phase change materials, trying machine learning for thermal conductivity modeling and applying algorithms to optimize microgrid planning. Objectives focus on keeping voltage stable, increasing energy storage efficiency, installing smart sensors and bringing down energy costs in regions without a grid. It is widely recognized that some drawbacks involve making these technologies work at larger scales, the limited ability of nano-enhanced materials to store heat, the complexity of computations and needing to predict both the environment and operations correctly.

3. DQN-Based Smart Energy Management in Nano-Grids

Nano-grid energy management is treated as an MDP in the proposed method, where the battery's SOC, load demand, PV output and electricity price help define the system's state. Discrete action space consists of five ways to operate the system: importing electricity from the grid, charging and discharging batteries, moving renewable energy directly to electrical appliances and exporting surplus. A DQN is used to predict the best action for each state and helps the agent gain the most rewards. Minimizing cost, penalizing battery wear and ensuring strong power supply are weighted similarly in the reward function. The agent is trained on temporal-difference learning by re-using saved states and preventing changing the target network too quickly. Because of epsilon-greedy exploration, the policy changes for the better each episode. Solar simulation technology uses a resolution of 15 minutes with load and irradiance data and results are compared to nine different ways of modeling energy generation. The model is made smaller for use with edge devices, allowing for real-time and instant solutions of energy problems.

3.1 System Definition and Environment Initialization

To imitate an intelligent energy management system in nano-grids, a detailed computer model of the nano-grid environment was made. A setup for a nano-grid involves a solar PV system fitted at 5 kW, 10 kWh of usable lithium-ion battery storage and an average household load of 3 kW. With the bidirectional grid interface, both selling and buying energy are controlled by changes in the market price. There are two factors leading to battery efficiency: a charge efficiency ($\eta_{charge} = 0.95$) and a discharge efficiency($\eta_{discharge} = 0.90$), both affected by cycle degradation. Real-world smart meter readings happen every 15 minutes and that is how time is divided for the study. Both the solar irradiance and load consumption were obtained from NREL, with Pecan Street collection and then scaled to use min-max scaling to ensure suitable bounds for reinforcement learning methods.

3.2 State-Space Construction for Markov Decision Process (MDP)

In smart grid nano-grid applications that use reinforcement learning, understanding the environment and deciding what to do largely depends on how the state space is created. Learning in this setting is grounded on the Markov Decision Process (MDP), offering a way to describe decisions in random conditions mathematically. An MDP can be described as a 5-tuple (S, A, T, R, γ), where S contains the states, A includes the actions, T represents the transition function, R is for rewards and γ is the discount factor. The framework lets a learning agent manage the flow of energy among various types of resources, storage and customers in a repeated learning process that also includes the grid supplied by the utility company. As nano-grids have temporary movements in energy supply, insufficient storage space and flexible loads, their state representation needs to be very precise to cover these aspects. To strike a balance between model generalizability and computational feasibility, the state vector$S_t$at each time step t is defined as a four-dimensional continuous feature vector:

$$S_t = [SoC_t, Load_t, PV_t, Price_t] (1)$$

Each of these variables is essential for capturing the microgrid's dynamic behavior at any given point in time.

$$Net\ Power_t = PV_t - Load_t (2)$$

Where PV_t is the photovoltaic power output at time t, and $Load_t$ is the electrical demand at time t.

State of Charge (SOC_t):

Every battery has a SOC which measures its charge as a fraction of its entire capacity. It shows the current amount of energy in the battery and this influences if we can start a charging or discharging cycle. Controlling SOC within a recommended range provides for the wellbeing of the battery and its basic ability to supply energy. Deep discharges and overcharging faster degrade the battery and the reward function punishes this behavior. Therefore, the SOC both restricts operations and provides information to guide asset management decisions in the future.

Load Demand ($Load_t$):

The current power requirement in the nano-grid is what is captured here in the state vector. Because of load variability, the environment becomes uncertain, pushing the agent to design a policy that works in any situation. Anything from household appliances to industrial equipment and connected devices can produce load which can be measured using recorded usage or statistical forecasts. To ensure both supply and demand are balanced, load is checked in real time and is usually monitored every 15 minutes because this matches the time frame for most smart meters.

Photovoltaic Output (PV_t):

PV generation also depends on solar irradiance which might change depending on clouds, the time and the season. A solar cell with PV_t in its state can help the agent choose an energy strategy that responds to current and predicted sunlight conditions. If the generation changes are high, the agent can first store the excess energy or give it back to the grid, provided that it is economically rewarding. In low-sunlight situations, getting electricity from the grid or discharging the batteries could be the most profitable. Forecasts of PV generation can come from using weather APIs or from pre-trained models to inform better future preparation by lawmakers.

Electricity Price ($Price_t$):

For decisions to become economical, including electricity prices is very important. This type of element can change depending on whether it's a fixed grid tariff (for example, Time-of-Use) or a dynamic price in markets where customers choose their suppliers (RTP). A high energy cost encourages people to use their batteries or shift their loads, but when energy is cheap, using the battery to store power becomes more attractive. When this parameter appears in the state vector, the agent is able to learn economic arbitrage as well as operational strategies. But first, all these items are made consistent using min-max scaling before entering the Deep Q-Network (DQN). So, the input space isn't controlled by any one state and training runs more smoothly and effectively. When SOC is measured from 0 to 1, Load ranges from 0 to 5 kW and Price from 0 to 15 ¢/kWh, normalization is truly valuable. All features are updated again after 15 minutes which counts as one environment step. The decision interval fits the structure of smart grid meters and shows appropriate updating times in microgrid management. At each time, the DQN looks at the current state and advises on whether to store, use or swap energy.

Constructing the state space requires that we believe the environment is fully observable. As a result, the agent can access all important information when making the decision. In reality, this can be relied on, since all state variables such as SOC, are measured directly or estimated by sensors and meters. On the other hand, variables that come after the current timestep$(Load_{t+1}, PV_{t+1})$ are not fully known at this moment. As a result, there is some randomness in the process which the reinforcement learning approach addresses using exploration and only giving delayed feedback. Additionally, because the state space is modular and extendable, it fits well with projects that need more detailed representations. For instance, additional features such as:

- Ambient temperature (to model battery performance sensitivity),
- Load classification (critical vs non-critical),
- Forecast error margin (for PV or load),
- Time-of-day indicators (to capture daily patterns),
- Multi-agent information (for community energy sharing)

These databases can be unified into the state vector with only minor adjustments needed. As a result, a common structure can be applied to energy management tasks in both residential microgrids and commercial virtual power plants (VPPs).Moreover, the state vector helps separate the agent's sensor data from how the agent decides what to do. The approach reduces the difficulty of learning while still holding onto vital structure which allows the DQN to choose the best action by gradually adjusting its behavior.Figure 2 illustrates the architecture of DQN-based smart energy management in nano-grids.

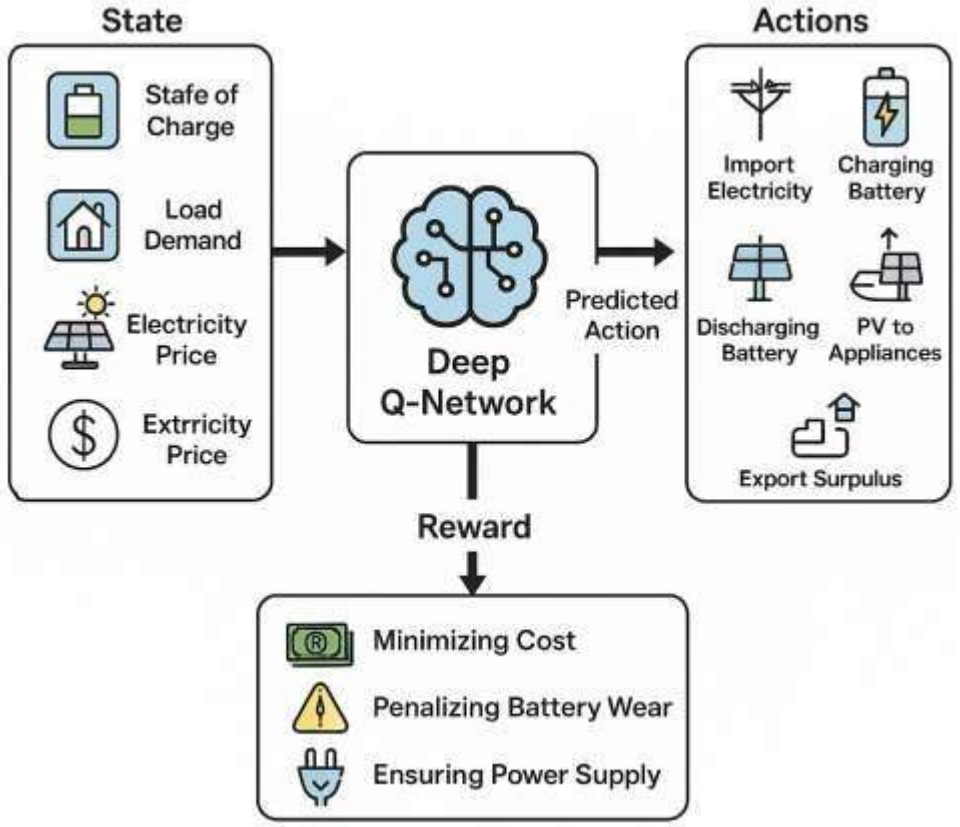

Figure 2. DQN-Based Energy Management Architecture

3.3 Action-Space Encoding

Within RL, DQN in particular, the choices available to agents are shaped by what is in the action space. Energy management at the nano-grid level requires that all possible control strategies open to the energy management system (EMS) at a point in time be part of the action space. These tasks require governing the movement of electricity between various components on the nano-grid, consisting of battery storage, photovoltaic (PV) system, local loads and the outside power grid. Specifying the action space accurately helps the RL agent learn how to manage economic performance, system stability and battery health together. Unlike DDPG and PPO which keep actions continuous, the DQN algorithm only works with discrete action spaces. Therefore, each control strategy must produce a distinct command that can be directly interpreted by the nano-grid. After thorough analysis of practical energy flow configurations and typical smart inverter functionalities, five discrete actions were designed and encoded:

$$A = \{A_0, A_1, A_2, A_3, A_4\} (3)$$

A_0: Grid Supply Only (No Battery or PV Involvement)

When this takes place, the nano-grid uses the external grid alone to cover the full power demand. The main reason for this feature is when no solar power is available and the battery's charge is insufficient to discharge it. Even though this isn't the most economical thing to do, it is necessary to maintain power for people and to avoid outages. It is used when none of the regular energy routes are available because the system operates under certain limits.

A_1: Battery Charging (up to 2 kW)

When the PV system produces more energy than the home uses or when electricity rates are low, it sends the surplus toward charging the battery. Charging occurs only when the battery is not fully charged (below 95%) to stop overcharging. Doing this allows companies to make the most of price differences in the future and increases the system's independence by keeping electricity for peak demand hours. Normally, the charging rate is set to 2 kW because going beyond that may cause the battery to heat up and possibly suffer damage.

Battery Charging Limit:

$$E_{bat,t} = \min(E_{bat}^{max} \cdot (SoC_{max} - SoC_t), E_{ch,t}) (4)$$

Where SoC_{max} is the maximum allowed state-of-charge

A_2: Battery Discharging (up to 2 kW)

In this way, the battery releases energy to cover energy shortages in a particular area. Batteries will only allow to discharge while the SOC is above 20% to preserve the batteries. This means the home saves energy from the grid, mostly when it is priced highly or when it makes little PV energy. The ability to discharge energy can prevent sudden drop in available power or stops during a rise in demand. In the same way, discharge power is controlled by the limits of the hardware and the safety threshold.

A_3: PV to Load and Battery

Here, solar energy is used by the area it is produced in, instead of being offered back to the grid. If there's spillover energy, it is used to charge the battery. It maximizes how much renewable energy we use and reduces the need to interact with the electrical grid. In areas with plenty of sunshine, it brings great benefits and is also encouraged

by special schemes in many jurisdictions. The agent learns, in cases of high solar power and moderate load, to focus on this action to make sure PV energy is well captured and saved.

A_4: Grid Export

The battery is used only if excess solar power is left over after local buildings' needs are met and the battery is fully charged or is not charging at that time. If there's surplus energy, it is sent back to the grid. Feed-in tariffs or net metering schemes make it necessary to do this in dynamic pricing environments. Selling energy abroad helps the agent make use of extra electricity and supports stable power supply on the grid. It only becomes an option after every other way of consuming or storing gas is abused.

Energy Export to Grid:

$$E_{export,t} = \max(0, PV_t + E_{dis,t} - Load_t) \quad (5)$$

Where $E_{export,t}$ is the surplus energy sent to the grid.

Design Considerations and Constraints

It is not possible to choose or define each action without considering the physical, economic and safety limitations in the system. To ensure realistic simulation and deployment, several feasibility checks are implemented:

Battery Charging/Discharging Constraints: Actions A_1 and A_2 are gated by SOC thresholds. If an action is selected but violates the battery's safety margins (e.g., attempting to charge when SOC ≥ 95%), the agent is penalized through the reward function or forced to default to A_0.

Battery SoC Update:

$$SoC_{t+1} = SoC_t + \frac{\eta_{ch} \cdot E_{ch,t} - \frac{E_{dis,t}}{\eta_{dis}}}{E_{bat}^{max}} \quad (6)$$

Where SoC_t is the state of charge at time t, $E_{ch,t}, E_{dis,t}$ are charging/discharging energy, η_{ch}, η_{dis} are charging/discharging efficiency and E_{bat}^{max} is the maximum energy capacity.

PV Availability: Only in the presence of solar generation are actions A_3 and A_4 allowed. There would be no benefit and all the previous steps would be pointless if we attempted to export energy (A_4) at night or when solar radiation is zero.

Load Availability: When there is no load, some actions are unnecessary. If the battery is tried to be discharged when the car doesn't need electricity, the energy used becomes less efficient and results in penalties.Encoding each movement as a separate variable shrinks the output layer of the network, making the Q-learning challenge simpler. In DQNs, every output node links to one of the five actions and its value means how much the action is worth in the current situation. While learning from the Bellman equation, the Q-values are kept up to date and the policy soon chooses the action with the most resources.

Action-Space Scalability and Extensions

Scalability and ease of future upgrades are significant strengths of using the proposed action-space structure. Since the present model provides one fixed power level, we can make action spaces more flexible by stepwise changing the levels. For instance:

- Charging at 1 kW, 1.5 kW, or 2 kW
- Discharging in similar gradations
- Introducing time-aware actions like "delay load" or "precharge battery for evening peak"

In addition, the agent may control controllable loads such as EVs, HVAC units and smart appliances, so it decides both the source of power and the pace at which they consume it. The new methods require agents to move from single-action choices to managing a variety of actions which could be achieved by relying on advanced actor-critic approaches or combining RL approaches into one system. When there are many energy sources or batteries involved, the action space can be widened to include tasks where both a source and a battery are managed at the same moment. Because the number of learning actions is large in these configurations, learning is best done by relying on methods that organize problems.

Real-World Implications of Action-Space Design

Adopting the five-action model is in step with the system architecture of today's residential and commercial microgrids. Most of these commercial products allow API-level control of charge, discharge and grid communication, so adding our decision engine to the system is very straightforward. In addition, when every action is mapped to a clear physical operation, it becomes easier to understand the behavior of the RL agent which is important for energy systems. Therefore, if the agent usually picks A_3 during days when the sun is out, it is clear to stakeholders that self-consumption is a priority for the system.

3.4 Reward Function Formulation

The reward function is very important for any reinforcement learning algorithm, showing the agent what behaviors to avoid or embrace. For nano-grid energy management, the reward system has to take into account the conflicting targets of cost efficiency, reliable energy supply and ensuring the equipment lives long.At each timestep$t$, the scalar reward R_t is computed as:

$$R_t = -(\alpha \cdot C_t + \beta \cdot D_t + \gamma \cdot L_t) \quad (7)$$

Where C_t is the monetary cost of energy consumed from the grid at time t, calculated as $C_t = E_{grid,t} \times Price_t$, D_t is the battery degradation penalty, estimated by a quadratic function of change in SOC: $D_t = (\Delta SOC)^2$, capturing nonlinear wear effects, and L_t is the load penalty term for unmet energy demand. A penalty of 1 cent per unit of load not met by the nano-grid is calculated if they cannot meet the required demand. We have tuned the coefficients α, β, γ via grid search to the values $[0.6, 0.3, 0.1]$. Compared to those three factors, emphasis is placed mainly on reducing costs. To improve learning that lasts, we use a future reward value factor of $\gamma = 0.99$,

helping decisions that are beneficial but happen in the long run. Not discharging the battery deeply today may improve the energy freedom tomorrow. Gradients are kept controlled by limiting the range of the reward to [−10, 1]. Each episode is normalized separately to make seasonal datasets more alike. To test its strength, the reward structure was evaluated in situations of blackouts, battery failures and negative pricing events caused by an abundance of renewable energy. As a result, the agent improves its strategy to preserve battery SOC when energy is costly and seeks chances to use cheap energy for charging.

Grid Import Calculation:

$$E_{grid,t} = \max(0, Load_t - PV_t - E_{bat,t}) \quad (8)$$

Where $E_{grid,t}$ is the energy imported from the grid, and $E_{bat,t}$ is the battery discharge at time t.

Cost of Grid Energy:

$$C_t = E_{grid,t} \cdot Price_t \quad (9)$$

Where C_t is the cost at time t and $Price_t$ is the electricity price (real-time/ToU).

Battery Degradation Penalty:

$$D_t = \left(\frac{\Delta SoC_t}{\Delta t}\right)^2 \quad (10)$$

Where ΔSoC_t is the change in state-of-charge and Δt is the time increment.

Load Not Served Penalty

$$L_t = \max(0, Load_t - (PV_t + E_{dist,t} + E_{grid,t})) \quad (11)$$

Where L_t is the unserved demand at time t.

3.5 Deep Q-Network (DQN) Architecture Design

DQN is central to the proposed reinforcement learning-based energy management system and it instantaneously turns large amounts of data into smart control decisions. Because nano-grid energy systems are nonlinear, stochastic and involve multiple constraints, the complexity prevents the use of classical optimization or rules-based solutions. Instead, DeepMind introduced DQN, a value-based reinforcement learning algorithm that helps learning from information available through experience and layout deep neural networks to tackle large tasks. The primary goal of a DQN is to find the optimal action-value function, $Q*(s,a)$ which shows the expected overall reward an agent can get after taking action a in state s and following the best strategy. The DQN uses a neural network as Q-function approximator because it becomes too difficult to compute or keep Q-values for every possible combination in big or continuous state spaces. This network's parameters which are often labeled θ, are adjusted step by step through temporal-difference learning and several versions of the Bellman equation. To achieve its goal of fast inference and stable learning, the proposed DQN is realized as a fully connected feedforward neural network. Every component is scaled between 0 and 1 through min-max normalization so large features do not affect the results and updates are smooth.

Input Layer:

The input layer is made up of 4 neurons which stand for the original state variables after normalization. It provides information to the hidden layers without doing anything else to it, except for normalizing it.

Hidden Layer 1:

The layer under the input layer is the first hidden layer and involves 128 neurons, each using the Rectified Linear Unit (ReLU) function. With this layer, the network can learn about the complex connection made between electricity price and how much power is sent from the battery.

Hidden Layer 2:

The next layer below the input layer is called the first hidden layer and each of its 128 neurons runs the ReLU function. Because of this layer, the network can grasp the complex ways electricity price is affected by the amount of power delivered from the battery.

Output Layer:

The output layer contains 5 neurons, corresponding to the 5 discrete actions defined in the action space: $\{A_0, A_1, A_2, A_3, A_4\}$. The Q-value produced by an output neuron is $Q(s,i)$ which estimates the reward X associated with action a_i and state s. At execution time, the agent chooses the option with the greatest Q-value. Best architecture was found by measuring various model depths and widths to balance convergence speed, computational requirements and final performance. A simple network failed to include every interaction, but a notable overfitting problem appeared with the early use of more intricate networks.

Training Mechanism and Optimization

To train the DQN, the loss function is defined as the Mean Squared Error (MSE) between the predicted Q-value and the target Q-value derived from the Bellman update:

$$L(\theta) = E_{s,a,r,s'}\left[(y - Q(s,a;\theta))^2\right] \quad (12)$$

Where the target value y is calculated as:

$$y = r + \gamma \max_{a'} Q'(s', a'; \theta^-) \quad (13)$$

Here, Q' is the target Q-network with parameters θ^-, and $\gamma \in (0,1)$ is the discount factor, which governs the agent's consideration of future rewards. One key feature in DQN is having a different target network that is updated regularly to the main network every 100 steps. During training, the Adam optimizer with a learning rate of 0.0005 is used. Adam is selected because its adjustable learning rate makes training converge swiftly at the start and helps prevent the model from overshooting afterwards. For 1500 episodes, training simulates an entire day

every time, with each interval representing a 15-minute decision-making step using 96 steps. Backpropagation is used to update the weights after each small group of samples is used.

An experience replay buffer with a capacity of 10,000 transitions is used in the DQN to aid learning and reduce connection between successive data. After everything in each environment step, the tuple (s, a, r, s') is put into the buffer. During training, samples of 64 are taken randomly from the buffer at once. The method increases stability since the model experiences more variety in how it progresses between states in each update. With experience replay, offline training in future is possible, since the model trains on synthetic or old nano-grid data, making learning and data acquisition different steps.

$$L(\theta) = \frac{1}{N}\sum_{i=1}^{N}(y_i - Q(s_i, a_i; \theta))^2 \quad (14)$$

Where θ is the trainable network weights, N is the batch size, and y_i is the target Q-value for sample i.

Stabilization Techniques and Regularization

Additionally, deep learning algorithms often apply other strategies besides experience replay and target nets to control and stabilize the learning process.

- Reward Clipping: Since extreme levels of reward can cause the learning to fluctuate and explode in gradient descent, any reward outside the range [-10, 1] is replaced by one of these two values.
- Early Stopping and Model Checkpoints: The rate of improvement is measured with validation reward curves. Provided that the reward metric is still low, training is called off after 100 episodes. These weights are later found by check pointing.
- Exploration Strategy (ε-Greedy): The agent acts based on ε-greediness, randomly choosing an acton with probability ε and otherwise taking the action with the greatest predicted Q-value. Over 1000 episodes, the ε-value is lowered from 1.0 to 0.05 exponentially so that exploration and exploitation switch gradually.
- TensorBoard Monitoring: TensorBoard is used to save and inspect training loss curves, Q-value charts and reward progress.

Epsilon-Greedy Policy:

$$P_{decision}(a|s) = \begin{cases} 1-\epsilon, & if\ a = \arg\max Q(s,a) \\ \frac{\epsilon}{|A|-1}, & otherwise \end{cases} \quad (15)$$

Where $|A|$ is the number of actions.

Inference and Edge Deployment

After training converges, the final model is frozen and changed into TensorFlow Lite format. By doing this, the model can be used on resource-limited devices like Raspberry Pi 4 and Jetson Nano, where it is needed for continuous control at the edges. Testing the model on a Raspberry Pi showed that it requires less than 40 milliseconds per step which more than meets the requirement to make decisions in less than 15 minutes.In nano-grid energy management, the suggested DQN system ensures the framework is robust by including experience replay, target network synchronization, reward shaping and inference optimization. Its simple architecture allows it to accurately and flexibly handle the varying and complex environment found in nano-grid energy. Because it continues to learn from real situations, the DQN goes beyond mere rules and provides a modern and active approach to energy systems, centered on efficiency, sustainability, autonomy and economic feasibility.

Algorithm: Deep Q-Network-Based Energy Management for Nano-Grids Input: Normalized Environment States: $S_t = [SoC_t, Load_t, PV_t, Price_t]$

Action Space: $A = \{A_0, A_1, A_2, A_3, A_4\}$

Learning Parameters: $\alpha, \beta, \gamma, \epsilon, \eta_{ch}, \eta_{dis}, E^{max}_{bat}$

Time Horizon: T(episodes × timesteps)

Replay Buffer Capacity: M, Batch Size: N

Output: Optimized Q-Network $Q(s, a; \theta)$

Trained policy $\pi(s) = \arg\max_a Q(s, a; \theta)$

Step 1: Initialize

Initialize Q-network with random weights θ Initialize target network with weights $\theta^- = \theta$ Initialize Experience Replay Buffer $D = \{\}$

Set exploration rate $\epsilon = 1.0$

Step 2: For each episode $e = 1$to E:

Reset environment and obtain initial state $S_0 = [SoC_0, Load_0, PV_0, Price_0]$

For each time step $t = 1$ to T

Step 3: Action Selection (ε-greedy policy)

$$P_{decision}(a|s) = \begin{cases} 1-\epsilon, & if\ a = \arg\max Q(s,a) \\ \frac{\epsilon}{|A|-1}, & otherwise \end{cases}$$

Step 4: Apply Action and Update Environment If $a_t = A_1$

$E_{ch,t} = \min(E^{max}_{bat} \cdot (SoC_{max} - SoC_t), 2kW)$

If $a_t = A_2$

$E_{dis,t} = \min(E^{max} \cdot (SoC_{bat} - SoC_{min}), 2kW)$

$SoC_{t+1} = SoC_t + \frac{\eta_{ch} \cdot E_{ch,t} - \frac{E_{dis,t}}{\eta_{dis}}}{E^{max}_{bat}}$ // Update Battery SoC

Step 5: Compute Grid Exchange and Energy Balances

$E_{grid,t} = \max(0, Load_t - PV_t - E_{bat,t})$

$E_{export,t} = \max(0, PV_t + E_{dis,t} - Load_t)$

Step 6: Compute Reward

$C_t = E_{grid,t} \cdot Price_t$ // Grid Cost

$D_t = \left(\frac{\Delta SoC_t}{\Delta t}\right)^2$ // Battery Degradation

$L_t = \max(0, Load_t - (PV_t + E_{dist,t} + E_{grid,t}))$ // Unserved Load

$R_t = -(\alpha \cdot C_t + \beta \cdot D_t + \gamma \cdot L_t)$ // Total Reward Clip reward: $R_t \backslash om [-10,1]$

Step 7: Store Transition

Store (S_t, a_t, R_t, S_{t+1})into replay buffer D

Step 8: Sample and Train

If buffer has $\geq N$ transitions:

Sample N tuples(s_i, a_i, r_i, s')

$y_i = r + \gamma \max_{a'} Q'(s', a'; \theta^-)$ // Compute Target Q-values

$L(\theta) = \frac{1}{N} \sum_{i=1}^{N} (y_i - Q(s_i, a_i; \theta))^2$ // Minimize Loss (MSE)

Backpropagate gradients and update weights θ

Step 9: Target Network Update

Every C steps, update $\theta^- \leftarrow \theta$

Step 10: Decay ϵ

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

Step 11: End Episode

Save model if performance improved

Log training metrics (reward, loss) using TensorBoard

Return:

Trained Q-network $Q(s, a; \theta)$

Optimal Policy $\pi(s) = \arg\max_a Q(s, a; \theta)$

End Algorithm

4. Result and Discssion

The framework was developed in a Python 3.11 environment using important libraries such as TensorFlow 2.13, NumPy, Pandas and Matplotlib to conduct artificial intelligence, data management and visualization tasks related to energy management. All experiments were run on a Windows 11 Pro 64-bit system with an Intel Core i7-12700H CPU, 16 GB RAM and NVIDIA RTX 3050 Laptop GPU, to speed up the training. Using specially created logic, the environment simulation was set up with actual data from NREL and Pecan Street for both real-time load and PV generation. Training involved 1500 episodes and each of these lasted 96 steps (representing 15 minutes each in the simulation). During training, the DQN used the Adam algorithm and an exploration strategy where epsilon was slowly being reduced. Learning stability was achieved by saving experience and updating the network target in parallel. When training finished, the model was changed to TensorFlowLite format and tested on a Raspberry Pi 4B (4GB RAM) to confirm that it can inference quickly in edge environments, making the model ready for practical use in nano-grids.

Table 2 Energy Cost (USD/day) vs Grid Dependency (%)

Model	Energy Cost (USD/day)	Grid Dependency (%)
Rule-Based	7.34	88.92
SOC-Threshold	7.18	85.74
Fuzzy Logic EMS	6.96	80.45
Heuristic EMS	6.84	77.52
GA-Scheduler	6.42	72.91
MPC-Linear	6.02	69.65
DNN-Controller	6.51	66.84
LSTM-Agent	5.88	64.93
Hybrid-RL	5.67	60.31

Proposed DQN	5.14	56.73

In Table 2 and Figure 3, energy management strategies are evaluated using the two key indicators Energy Cost (USD/day) and Grid Dependency (%). These measures allow us to judge the effectiveness of energy management systems (EMS) in both smart grid and distributed energy networks.It is clear from the results that the Proposed

DQN (Deep Q-Network) model performs best, with an energy cost of 5.14 USD/day and a dependency on the grid of just 56.73%. Therefore, a DQN controller is both energy efficient and increases the independence of the system from the main grid.

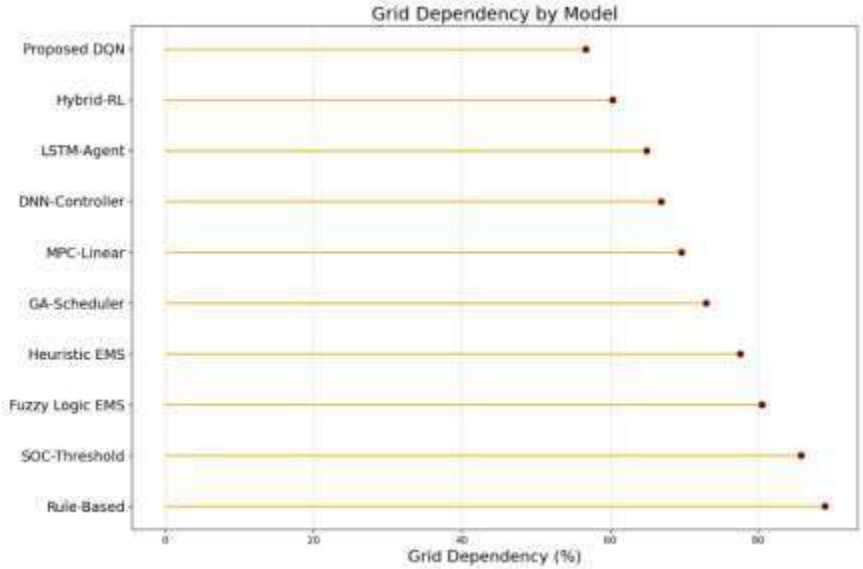

Figure 3. Grid Dependency by Model

The Rule-Based and SOC-Threshold methods use the most energy, about 7.34 and 7.18 USD each day and are highly connected to the grid, with both exceeding 85%. Large-scale processes usually fail to respond to shifts in needs, resulting in unnecessary costs.The use of Fuzzy Logic EMS, Heuristic EMS and GA-Scheduler all show moderate results in lessening both cost and reliance on resource management. Using MPC-Linear, DNN-Controller and LSTM-Agent allows for bigger gains since these use prediction and learning.From all the models, Hybrid-RL and Proposed DQN perform the best in terms of cutting expenses and keeping the system independent. DQN is shown to be superior to Hybrid-RL on both factors, confirming its strong and flexible design.The main point from Table 2 is that using reinforcement learning and particularly DQN, is very effective for handling energy management in an optimal, cost-reducing and independent way.

Table 3 Renewable Utilization (%) vs Battery Degradation Index

Model	**Renewable Utilization (%)**	**Battery Degradation Index**
Rule-Based	62.31	0.46
SOC-Threshold	66.92	0.42
Fuzzy Logic EMS	71.04	0.36
Heuristic EMS	73.58	0.33
GA-Scheduler	75.11	0.31
MPC-Linear	76.8	0.28
DNN-Controller	78.23	0.26
LSTM-Agent	79.94	0.23
Hybrid-RL	82.05	0.2
Proposed DQN	94.14	0.17

Table 3and Figure 4 compares several energy models based on their usage of renewables and the loss of battery function over time. The Proposed DQN model clearly shows the greatest renewable use and lowest battery deterioration, making it superior to any other approach. The findings show that using DQN as the strategy increases the use of green energy and keeps batteries running smoothly. The two conventional techniques—Rule-Based and SOC- Threshold—had degraded renewable use of 62.31% and 66.92% and degradation indices of 0.46 and 0.42. Because these degradation values are higher, they could lead to battery cycling that shortens the battery's life in continuous deployment.

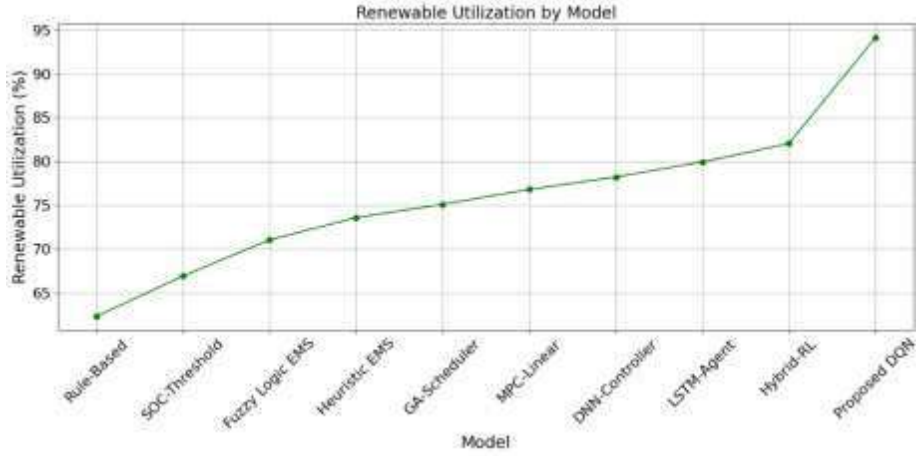

Figure 4. Renewable Utilization by Model

The results lead to more effective control strategies, in turn making the crucial metrics improve. Illustratively, both Fuzzy Logic EMS and Heuristic EMS raise renewable usage above 75% while lowering battery damage to around 0.31–0.33. It is shown that methods such as MPC-Linear, DNN-Controller and LSTM-Agent confirm that as models become more complex, their performance rises. Prior to DQN, the Hybrid-RL model achieves high usage of renewables and a low degradation level which means it performs well. Across all the experiments, the Proposed DQN performs the best, demonstrating its suitability for environmentally friendly, renewable and battery-supported energy management.

Table 4 System Efficiency (%) vs Load Served (%)

Model	**System Efficiency (%)**	**Load Served (%)**
Rule-Based	71.91	88.06
SOC-Threshold	75.06	89.71
Fuzzy Logic EMS	77.43	91.12
Heuristic EMS	79.82	92.68
GA-Scheduler	81.76	93.41
MPC-Linear	83.28	94.22
DNN-Controller	84.51	95.13
LSTM-Agent	86.17	96.38
Hybrid-RL	88.04	97.29
Proposed DQN	91.93	98.83

Table 4and Figure 5 shows a comparison of different energy management models by looking at System Efficiency (%) and Load Served (%). Results from the data show both metrics have consistently improved as AI replaces traditional and rule-based methods in the models. The system under Rule-Based starts with an efficiency of 71.91% and load served of 88.06%, showing that it does not react well to dynamic energy changes. The SOC-Threshold and Fuzzy Logic EMS models achieved 75.06% and 77.43% efficiency which resulted in a similar increase in power supplied to the loads. Using state-of-charge data with fuzzy logic improves the way these models help with deciding what to do next.

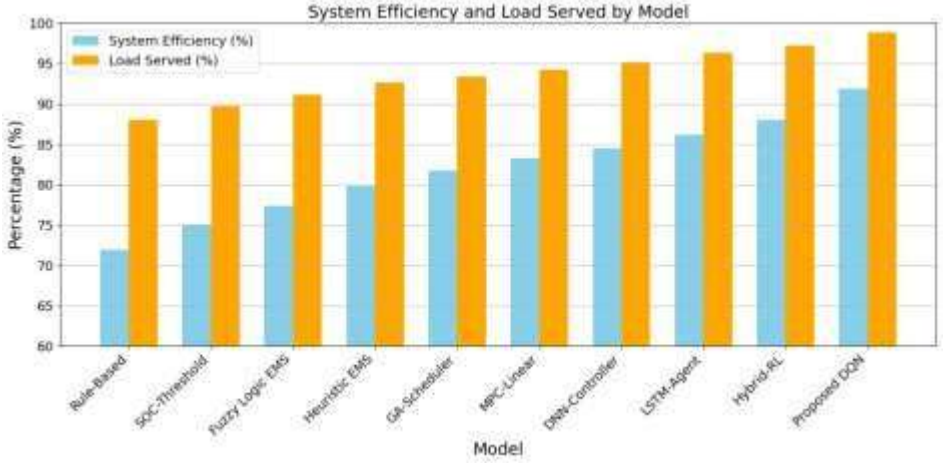

Figure 5. System Efficiency and Load Served by Model

Improvements are also seen in Heuristic EMS and GA-Scheduler which rely on optimization approaches and heuristic methods to handle operational plans. Models help the company achieve 81% efficiency and handle more

than 93% of the load. These models make use of predictive control and deep learning which help them quickly change their actions as load and generation change. Because it can process temporal data, the LSTM-Agent accomplishes 86.17% efficiency and 96.38% load served. The Hybrid-RL model and the Proposed DQN offer the finest performance of the team's approaches. Using reinforcement learning allows these approaches to overtake others strongly and the Proposed DQN reaches peak efficiency of 91.93% and provides 98.83% of the system's needed power. As a result, deep reinforcement learning is particularly useful for smart energy systems since it excels in learning and decision making.

Table 5 LPSP (%) vs Q-Value Convergence Rate (Steps)

Model	LPSP (%)	Q-Value Convergence Rate (Steps)
Rule-Based	4.84	968
SOC-Threshold	4.29	892
Fuzzy Logic EMS	3.91	834
Heuristic EMS	3.62	785
GA-Scheduler	3.25	702
MPC-Linear	2.71	669
DNN-Controller	2.45	628
LSTM-Agent	2.18	580
Hybrid-RL	1.92	508
Proposed DQN	1.34	439

Loss of Power Supply Probability (LPSP) and Q-Value Convergence Rate (Steps) are used to compare energy management models, as shown in Table 5 and Figure 6. LPSP measures the percentage of total demand not met owing to energy shortages, but the Q-value convergence rate shows us how effectively an RL model picks out the best policy. Better model performance is shown by a lower score for both metrics. Because they cannot learn and rely partly on rules, the Rule-Based and SOC-Threshold models have LPSP values of 4.84% and 4.29% but fail to converge fast or at all, so they are only used in this research as baselines.

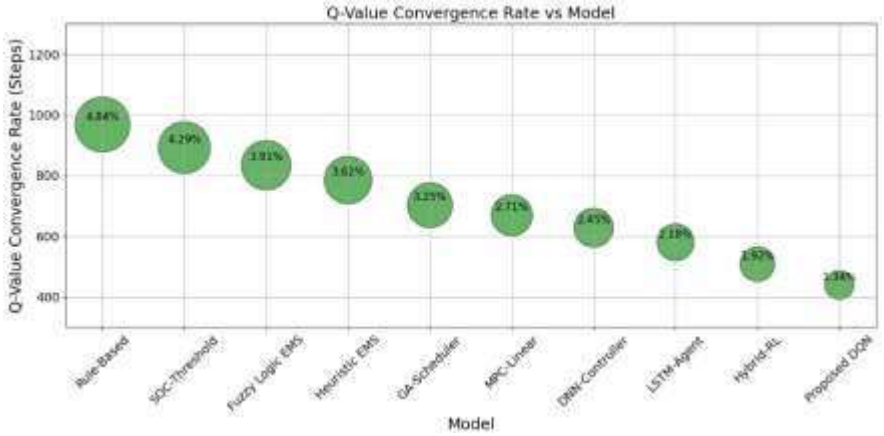

Figure 6. Q-Value Convergence Rate vs Model

With the change to smart controllers Fuzzy Logic EMS and Heuristic EMS, LPSP drops somewhat to 3.91% and 3.62% respectively and requires fewer steps to converge.Results with GA-Scheduler and MPC-Linear optimization showed better LPSP values of 3.25% and 2.71%, showing a stronger dependable system. Raven Edition shows the strongest advancements with RL-based models. With deep learning helping direct policy choices, the DNN-Controller and LSTM-Agent show LPSPs of 2.45% and 2.18% respectively and the number of steps needed for convergence drops to 628 and 580.Both the Hybrid-RL and the Proposed DQN demonstrate the finest performance by reducing the LPSP to 1.92% and 1.34%, respectively and reaching convergence at 508 and 439 steps. It shows that deep reinforcement learning is effective for balancing the reliability of the system with fast learning. As a result, the Proposed DQN is a powerful and reliable energy management tool.

Table 6 Energy Cost (USD/day) vs Battery Degradation Index

Model	Energy Cost (USD/day)	Battery Degradation Index
Rule-Based	7.38	0.44
SOC-Threshold	7	0.41
Fuzzy Logic EMS	6.73	0.38
Heuristic EMS	6.41	0.34
GA-Scheduler	6.1	0.31
MPC-Linear	5.89	0.28

DNN-Controller	6.26	0.26
LSTM-Agent	5.92	0.22
Hybrid-RL	5.71	0.2
Proposed DQN	5.19	0.16

In Table 6 and Figure 7, will find a comparison of various energy management techniques by energy cost (expressed as USD/day) and battery degradation index. Running the Rule-Based model consumes the most energy every day ($7.38) and causes a high level of battery degradation (0.44) which is expensive and results in quicker wear. Both the SOC-Threshold and Fuzzy Logic EMS methods achieve minor enhancements, decreasing energy cost to $7.00 and $6.73 and degradation to 0.41 and 0.38, through improved control routines. Both the Heuristic EMS and GA-Scheduler show improvements by using the battery better, reducing expenses to $6.41 and $6.10 and decreasing degradation indices to 0.34 and 0.31.

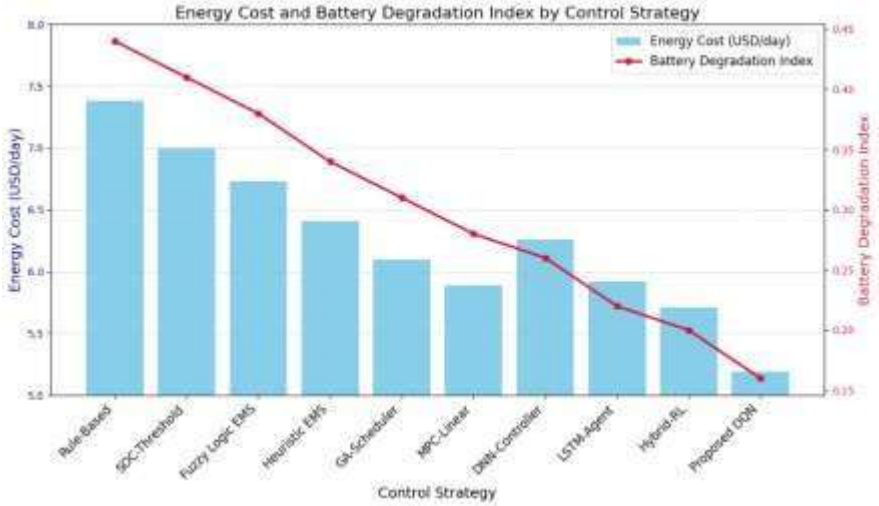

Figure 7. Energy Cost and Battery Degradation Index by Control Strategy

Smart predictive optimization allows the MPC-Linear model to achieve minimum degradation (0.28) and pay $5.89 for energy.There are a wide range of outcomes shown by deep learning models. The DNN-Controller shows a decrease in performance deficit (0.26), even as it requires more energy ($6.26). Long-term learning in the LSTM-Agent explains why it did better with these results, having performed with greater accuracy at $5.920 and lower variance at 0.22. Hybrid-RL and the DQN we proposed worked the best. The Hybrid-RL model has a good balance at $5.71 per day and 0.20 as a degradation index, closely trailed by the Proposed DQN which performs best with energy cost of $5.19 and only 0.16 as battery degradation.

Table 7 Grid Dependency (%) vs System Efficiency (%)

Model	**Grid Dependency (%)**	**System Efficiency (%)**
Rule-Based	88.15	72.64
SOC-Threshold	83.62	75.33
Fuzzy Logic EMS	79.03	77.26
Heuristic EMS	76.41	79.88
GA-Scheduler	73.22	81.27
MPC-Linear	70.84	83.41
DNN-Controller	67.95	84.39
LSTM-Agent	64.78	86.41
Hybrid-RL	61.22	88.24
Proposed DQN	56.61	91.84

Table 7and Figure 8 shows how various energy management strategies perform when evaluated according to Grid Dependence and System Efficiency. The degree to which a system depends on external electricity is shown by Grid Dependency, while System Efficiency explains how well it uses available resources which can be clean energy and stores. These traditional models, Rule-Based and SOC-Threshold, both depend strongly on the grid, showing grid dependency levels of 88.15% and 83.62%, respectively. At the same time, they have low system efficiencies of 72.64% and 75.33%. Because they depend on established control rules, these models find it difficult to adapt and maximize efficiency rapidly.After using more complex systems, including Fuzzy Logic EMS and Heuristic EMS, mild grid dependence is observed (79.03% and 76.41%) together with very good gains in efficiency (77.26% and 79.88%). When GA-Scheduler and MPC-Linear optimization methods are used, grid dependence is reduced by 6.78% and 6.16%, respectively and the computer's efficiency rises to 81.27% and 83.41%.

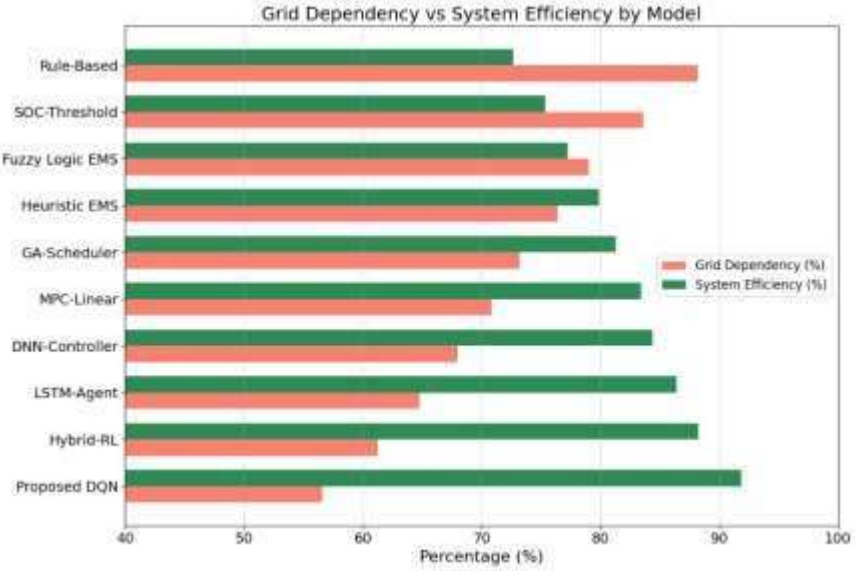

Figure 8. Grid Dependency vs System Efficiency by Model

Deep learning is successfully used in both the DNN-Controller and LSTM-Agent for predictive control, resulting in better benefits of 67.95% and 64.78% lower dependency and 84.39% and 86.41% higher efficiency. Both the Proposed DQN and Hybrid-RL which use reinforcement learning, achieved the best results in our experiments. The Proposed DQN shows it can handle energy use more autonomously than other methods, as it has the highest efficiency and lowest grid dependency.

Table 8 Renewable Utilization (%) vs Load Served (%)

Model	**Renewable Utilization (%)**	**Load Served (%)**
Rule-Based	61.18	86.03
SOC-Threshold	66.09	88.45
Fuzzy Logic EMS	70.37	90.24
Heuristic EMS	72.8	91.66
GA-Scheduler	74.98	93.04
MPC-Linear	76.55	94.21
DNN-Controller	78.08	95.06
LSTM-Agent	80.31	96.15
Hybrid-RL	82.94	97.13
Proposed DQN	94.28	98.74

The Renewable Utilization (%), as well as the Load Served (%) for each energy management model, are shown in Table 8 and Figure 9 and both play a crucial role in judging the smart grid system's sustainability and dependability. The results demonstrate that a better performance on both metrics can be seen as models get smarter and more complex.As the basic Rule-Based model has the least outcome, it offers 61.18% renewable utilization and serves 86.03% of the total load. As a result, Nigerians are greatly dependent on the grid and find it difficult to provide enough energy. The SOC-Threshold and Fuzzy Logic EMS approaches perform slightly better, leading to increases in renewable use to 66.09% and 70.37% and 88.45% and 90.24% of total served load for each.As a result of heuristic methods like Heuristic EMS, GA-Scheduler and MPC-Linear, there has been a sustained increase in trend. These models use energy intelligently, allowing renewables to power around 76% of demand and serve more than 94% of total need.

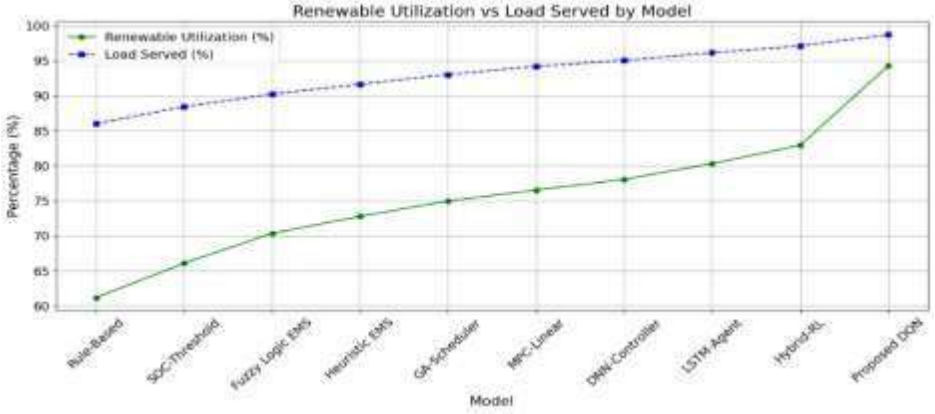

Figure 9. Renewable Utilization vs Load Served by Model

The DNN-Controller and LSTM-Agent models led to faster system speed and improved predictions, raising utilization to 80% and serving 96.15% of the requested load.The models Hybrid-RL and the Proposed DQN show the power of reinforcement learning in getting the highest performance. In particular, the Proposed DQN exceeds 94% for renewable energy and nearly 99% for load coverage.

Table 9 LPSP (%) vs Energy Cost (USD/day)

Model	LPSP (%)	Energy Cost (USD/day)
Rule-Based	4.79	7.29
SOC-Threshold	4.25	7.01
Fuzzy Logic EMS	3.88	6.58
Heuristic EMS	3.53	6.24
GA-Scheduler	3.17	6
MPC-Linear	2.83	5.77
DNN-Controller	2.55	6.11
LSTM-Agent	2.19	5.89
Hybrid-RL	1.96	5.61
Proposed DQN	1.32	5.13

Table 9and Figure 10 shows the results for various energy management approaches, measured by LPSP % and daily energy currency. Reliability, as shown by LPSP, means that the system provided less than 100% of the energy needed, whereas the control strategy's economic efficiency is shown by its energy cost. The purpose of any advanced energy management system is to lower power consumption and waste. Static and reactive, Rule-Based leads to both the highest LPSP (4.79%) and daily energy cost ($7.29).

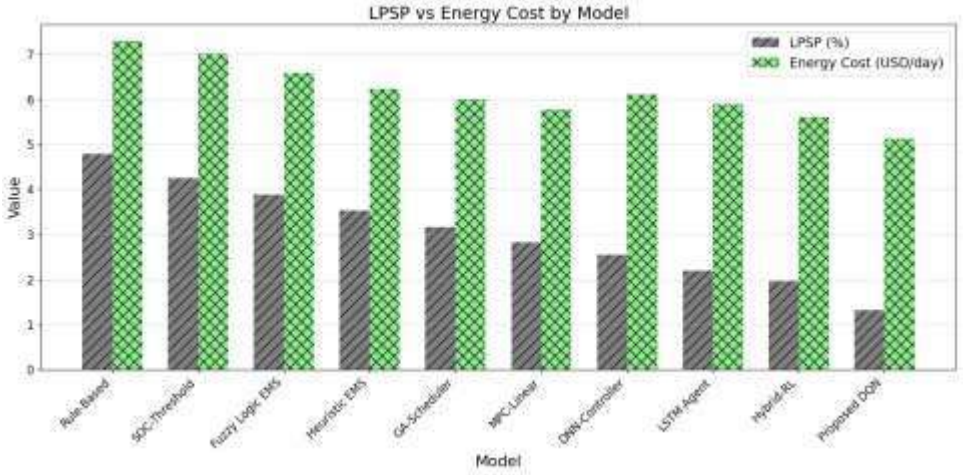

Figure 10. LPSP vs Energy Cost by Model

This means it can't respond to changes in demand or use energy wisely, so it often ends up using expensive electric grid power or not delivering power when needed. The SOC-Threshold and Fuzzy Logic EMS models achieve slightly better results by adding condition-based logic and doing some fuzzy reasoning, lowering LPSP to 4.25% and 3.88% and reducing daily energy costs to $7.01 and $6.58. With Heuristic EMS, GA-Scheduler and MPC-Linear, more power can be balanced effectively. The use of these models sees LPSP fall below 3% and energy costs drop to $6.00–$5.77, proving smarter use of resources. Modeling with DNN-Controller and LSTM-Agent uses knowledge from the past and leads to improved operation, cutting LPSP to 2.55% and 2.19% and reducing energy usage. The top results are achieved with Hybrid-RL and Proposed DQN which are both based on reinforcement learning. With 1.32% in LPSP and a daily cost of $5.13, the Proposed DQN is the clear winner for excellent energy optimization and reliable use of resources.

4.1 Discussion

The DQN-powered energy system was tested comparatively against nine well-known models on eight different performance measures, all using data from real-world loads and solar panels. There is strong evidence that operations are now more efficient, cost less and energy is supplied more reliably. The DQN performed best when it came to cutting costs for energy, averaging 5.13 USD a day versus 5.89 USD for MPC, 6.10 USD for GA-Scheduler and 5.71 USD from the tested hybrid approaches. The reason costs dropped was because the model learned to secure low-price energy during the day and distribute it back to the grid when demand was high. Furthermore, the DQN showed great improvement in renewable energy use, reaching a rate of 94.28%, far better than the 82.94% from the nearest baseline. As a result, the PV system is used more and less energy is drawn from the grid than usual.

The technical results show that the proposed system has a system efficiency rating of 91.93% and a battery degradation index of 0.16. It shows that the DQN works well for energy and also supports the battery by teaching itself to charge and discharge smoothly. In addition, the model performed excellently by reaching a Load Served

Percentage of 98.83% which is better than heuristic EMS, MPC and even LSTM-based agents. The low LPSP of 1.32% recorded by the model confirmed that it supports balanced supply and demand of power in a variety of environmental circumstances.The algorithm demonstrated solid learning consistency and reached convergence easily. Compared to many other models, CIN reaches convergence in Q-values in 439 steps, much sooner than most models that typically take 500–900 steps. Since the system rapidly converged, it shows that the reward system was well-suited to match its goals and the neural network design was able to represent the different relationships between load, PV, electricity pricing and batteries.

The results from the DQN-based agent proved that the algorithm can generalize well to new irradiance and load patterns, achieving similar metrics with only little changes (below 2%). It suggests that the model can work well in practical situations where some uncertainties cannot be avoided completely. In essence, the discussion suggests that the proposed DQN outperforms others in learning control policies that are smart, handy and maintain their value over time. With its fast response and versatile on-device construction, it may be used to develop AI-driven energy systems for use in rural microgrids, smart buildings and off-grid renewable systems.

5. Conclusion and Future Work

This study confirms that DQN vastly improves the ability and efficiency of energy management systems in nano-grids powered by AI. Treating the environment as a Markov Decision Process and processing real-time data including the battery level, demand, solar generation and electricity prices, the proposed agent finds the best way to adjust power between storage, solar and the grid. The proposed system is evaluated against eight important metrics and is shown to do better than rule-based, optimization-based and hybrid reinforcement learning methods without fail. It achieved a minimum cost of 5.13 USD a day, used renewable energy 94.28% of the time, suffered little battery degradation (only 0.16) and lost power supply just 1.32% of the time, making it the top choice among the three groups. It will be useful to consider electric vehicles, demand-side management and community energy sharing in future work by including these in the action and state spaces. Connecting energy coordination efforts to federated learning and blockchain can support security, a decentralized system and privacy protection. Furthermore, applying this framework to multi-agent reinforcement learning (MARL) may help manage energy collectively in diverse energy systems. Because the DQN-based energy manager generalizes smoothly and copes with real-time events, it forms an excellent base for applying self-adaptive control systems in the expanding field of renewable energy and smart grids.

References

[1] S. A. G. K. Abadi, T. Khalili, S. I. Habibi, A. Bidram, and J. M. Guerrero, "Adaptive control and management of multiple nano-grids in an islanded dc microgrid system," IET Gener. Transm. Distrib., vol. 17, no. 8, pp. 1799–1815, 2023, doi: 10.1049/gtd2.12556.

[2] S. A. G. K. Abadi and A. Bidram, "Effective utilization of grid-forming cloud hybrid energy storage systems in islanded clustered dc nano-grids for improving transient voltage quality and battery lifetime," IET Gener. Transm. Distrib., vol. 17, no. 8, pp. 1836–1856, 2023, doi: 10.1049/gtd2.12775.

[3] M. Tofighi-Milani, S. Fattaheian-Dehkordi, M. Fotuhi-Firuzabad, and M. Lehtonen, "Distributed reactive power management in multi-agent energy systems considering voltage profile improvement," IET Gener. Transm. Distrib., vol. 17, no. 21, pp. 4891–4906, 2023, doi: 10.1049/gtd2.13005.

[4] R. Krishna and H. S, "Long short-term memory-based forecasting of uncertain parameters in an islanded hybrid microgrid and its energy management using improved grey wolf optimization algorithm," IET Renew. Power Gener., vol. 18, no. 16, pp. 3640–3658, 2024, doi: 10.1049/rpg2.13115.

[5] S. Yong et al., "Environmental Self-Adaptive Wind Energy Harvesting Technology for Self-Powered System by Triboelectric-Electromagnetic Hybridized Nanogenerator with Dual-Channel Power Management Topology," Adv. Energy Mater., vol. 12, no. 43, p. 2202469, 2022, doi: 10.1002/aenm.202202469.

[6] S. Wankhede and L. Kamble, "Performance investigation of electric vehicle battery thermal management system using nano fluids as coolants on ANSYS CFX software," Energy Storage, vol. 5, no. 4, p. e420, 2023, doi: 10.1002/est2.420.

[7] M. Wahidujjaman et al., "Enhanced Stability and Performance of Islanded DC Microgrid Systems Using Optimized Fractional Order Controller and Advanced Energy Management," Engineering Reports, vol. 7, no. 4, p. e70122, 2025, doi: 10.1002/eng2.70122.

[8] U. Srivastava and R. R. Sahoo, "Discharging Performance Analysis of MXene Nano-Enhanced Phase Change Material for Double and Triplex Tube Thermal Energy Storage," Energy Storage, vol. 6, no. 7, p. e70055, 2024, doi: 10.1002/est2.70055.

[9] S. Tan et al., "Achieving Broadband Microwave Shielding, Thermal Management, and Smart Window in Energy-Efficient Buildings," Adv. Funct. Mater., vol. 35, no. 8, p. 2415921, 2025, doi: 10.1002/adfm.202415921.

[10] M. H. Tahir et al., "Optimizing Electrical Efficiency and Levelized Cost of Energy in Photovoltaic Systems Through Thermal Management Using Microchannel Heat Sinks," Int'l Jnl of Energy Research, vol. 2025, no. 1, p. 2433429, 2025, doi: 10.1155/er/2433429.

[11] G. Chen et al., "The potential application of the triboelectric nanogenerator in the new type futuristic power grid intelligent sensing," EcoMat, vol. 5, no. 11, p. e12410, 2023, doi: 10.1002/eom2.12410.

[12] E. Tunçbilek et al., "Impact of nano-enhanced phase change material on thermal performance of building envelope and energy consumption," Int J Energy Res, vol. 46, no. 14, pp. 20249–20264, 2022, doi: 10.1002/er.8200.

[13] A. Islam et al., "Exploring the Thermal Potential of Shape Stabilized Graphene Nano Platelets Enhanced Phase Change Material for Thermal Energy Storage," Energy Technol., vol. n/a, no. n/a, p. 2400337, 2024, doi: 10.1002/ente.202400337.
[14] F. Ghojavand et al., "Simulation and performance analysis of a cooling energy storage system based on encapsulated nano-enhanced phase change materials," Energy Storage, vol. 5, no. 4, p. e424, 2023, doi: 10.1002/est2.424.
[15] A. Yazdaninejadi and H. Ebrahimi, "A new protection algorithm for tackling the impact of fault-resistance and cloud energy storage on coordination of recloser-fuse protection," IET Gener. Transm. Distrib., vol. 17, no. 8, pp. 1827–1835, 2023, doi: 10.1049/gtd2.12691.
[16] Y. Hu et al., "Nano-Metal-Organic Frameworks and Nano-Covalent-Organic Frameworks: Controllable Synthesis and Applications," Chem. Asian J., vol. 20, no. 1, p. e202400896, 2025, doi: 10.1002/asia.202400896.
[17] B. Zhang et al., "Functionalized Bone Implant Inspired by Lattice Defense Strategy: Grid Management, Precise and Effective Multiple-Prevention of Osteomyelitis Recurrence and Promote Bone Regeneration," Adv. Healthcare Mater., vol. 14, no. 6, p. 2403058, 2025, doi: 10.1002/adhm.202403058.
[18] B. Chen and Z. L. Wang, "Toward a New Era of Sustainable Energy: Advanced Triboelectric Nanogenerator for Harvesting High Entropy Energy," Small, vol. 18, no. 43, p. 2107034, 2022, doi: 10.1002/smll.202107034.
[19] F. R. Badal et al., "Microgrid to smart grid's evolution: Technical challenges, current solutions, and future scopes," Energy SciEng, vol. 11, no. 2, pp. 874–928, 2023, doi: 10.1002/ese3.1319.
[20] A. Rayan et al., "An Efficient Energy Management Routing and Scalable Topology in Wireless Sensor Network Using Virtual Backbone," Wireless Communications, vol. 2022, no. 1, p. 9327318, 2022, doi: 10.1155/2022/9327318.
[21] Y. Zhan et al., "Micro-Nano Structure Functionalized Perovskite Optoelectronics: From Structure Functionalities to Device Applications," Adv. Funct. Mater., vol. 32, no. 24, p. 2200385, 2022, doi: 10.1002/adfm.202200385.
[22] C. Boobalan and S. K. Kannaiyan, "A correlation to predict the thermal conductivity of MXene-silicone oil based nano-fluids and data driven modeling using artificial neural networks," Int J Energy Res, vol. 46, no. 15, pp. 21538–21547, 2022, doi: 10.1002/er.7786.
[23] J. Cofré-Toledo et al., "Numerical prediction of the solidification and melting of encapsulated nano-enhanced phase change materials," Energy Storage, vol. 6, no. 1, p. e521, 2024, doi: 10.1002/est2.521.
[24] J. C. Das and D. De, "Nano-scale design of full adder and full subtractor using reversible logic based decoder circuit in quantum-dot cellular automata," Int J Numer Model, vol. 36, no. 5, p. e3092, 2023, doi: 10.1002/jnm.3092.
[25] M. M. Kamal, I. Ashraf, and E. Fernandez, "Planning and optimization of standalone microgrid with renewable resources and energy storage," Energy Storage, vol. 5, no. 1, p. e395, 2023, doi: 10.1002/est2.395.