INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN APPLIED SCIENCE (IJRIAS)
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Machine Learning Based Predictive Maintenance for Improved Reliability and Efficiency of Solar Farms

D. C. Oyiogu[1,] S. C. Nwokporo[2,]

[1]Department Of Electrical Engineering, Nnamdi Azikiwe University, Awka, Nigeria

[2] Department Of Electrical and Electronic Engineering, Ebonyi State University, Abakaliki

DOI: https://doi.org/10.51584/IJRIAS.2026.11070199

Received: 06 August 2026; Accepted: 11 September 2026; Published: 21 August 2026

ABSTRACT

The increasing deployment of solar photovoltaic (PV) systems has intensified the need for intelligent approaches that ensure reliable operation, minimize performance losses, and improve economic viability. Conventional maintenance strategies in solar farms are largely reactive or scheduled, often leading to delayed fault detection, reducing energy yield, and increased operational costs. This study presents a machine learning-based predictive maintenance and optimization framework aimed at enhancing the operational efficiency of a solar photovoltaic power plant, using Azari Farm as a case study. Operational data such as PV array voltage, current, irradiance, temperature, and power output are analyzed to identify performance degradation patterns and detect potential system faults before failure occurs. Machine learning algorithms are applied to monitor system behavior, predict anomalies, and optimize maintenance scheduling. The results demonstrate that integrating machine learning with predictive maintenance strategies significantly improves system reliability, reduces downtime, and enhances overall solar farm efficiency. The study highlights the potential of intelligent monitoring and optimization techniques to improve energy generation performance and sustainability in large scale solar photovoltaic installations.

INTRODUCTION

Traditional maintenance strategies for solar PV systems typically rely on either reactive maintenance, where faults are addressed only after failure occurs, or scheduled maintenance, where system components are inspected periodically regardless of their condition. While these approaches are widely used, they are often inefficient because they may lead to unexpected system downtime or unnecessary maintenance activities. As solar installations continue to expand, there is a growing need for intelligent monitoring and predictive maintenance techniques capable of detecting anomalies and forecasting failures before they significantly affect system performance.

Recent advances in artificial intelligence and machine learning have created new opportunities for improving the reliability and efficiency of solar photovoltaic systems. Machine learning algorithms can analyze large volumes of operational data from solar plants to identify hidden patterns, predict system faults and optimize maintenance schedule. By leveraging historical and real-time data such as solar irradiance, PV module temperature, voltage, current, and power output. Machine learning models can detect performance deviations and recommend corrective actions. This predictive capability enables proactive maintenance strategies that reduce downtime, improve system yield and extend lifespan of system component

Maintenance is crucial for the reliability and profitability of solar farms. Traditional approaches—namely corrective maintenance (fixing after failure) and preventive maintenance (routine inspections at fixed intervals)—are becoming less effective in modern large-scale setups. Corrective maintenance causes unexpected downtime, energy losses, and costly emergency repairs, while preventive maintenance can lead to unnecessary replacement of healthy components, resulting in inefficiencies (Carneiro et al., 2021). Predictive maintenance (PdM), on the other hand, uses condition-monitoring data, statistical models, and ML algorithms to forecast failures before they happen.

Machine learning methods such as Support Vector Machines (SVM), Decision Trees, Random Forests, and Deep Neural Networks (DNNs) have demonstrated strong potential in fault detection and diagnosis within PV systems (Attoui et al., 2020). More advanced methods, including Long Short-Term Memory (LSTM) networks, can analyze temporal dependencies in operational data such as temperature, irradiance, current, and voltage fluctuations to predict degradation trends or failures (Zhang et al., 2021). In scenarios where labeled fault data is limited, unsupervised learning methods, such as clustering and anomaly detection, offer reliable alternatives (Nivedha & Kumar, 2022). PdM thus minimizes unexpected outages, reduces operating expenditure (OPEX), and enhances the overall availability of solar farm assets.

The second major challenge in solar energy systems is intermittency. Solar output fluctuates with irradiance variations caused by weather conditions and diurnal cycles, which makes grid integration and demand-supply balancing difficult. Battery Energy Storage Systems (BESS) mitigate this issue by storing surplus energy during peak generation periods and discharging it during demand peaks or low-generation intervals (Luo et al., 2015). This helps improve grid stability, reduces curtailment, and ensures a more consistent energy supply. However, battery efficiency and lifespan are strongly influenced by charging and discharging cycles, operating temperature, and depth of discharge. Without optimized control, batteries degrade quickly, leading to capacity fade, efficiency loss, and costly replacements (Wang et al., 2020).

Machine learning improves battery management systems (BMS) by enhancing predictions of key parameters like State of Charge (SoC) and State of Health (SoH). Traditional electrochemical models are computationally demanding and less flexible to real-world variability, whereas ML methods can learn directly from historical and real-time data to deliver accurate, adaptable predictions (Richardson et al., 2019).

Techniques such as regression models, gradient boosting, and convolutional neural networks (CNNs) have been used to forecast battery degradation, while reinforcement learning (RL) methods are increasingly adopted to optimize dynamic charge-discharge scheduling under uncertain conditions (Zhang et al., 2020). By extending battery lifespan, reducing inefficiencies, and optimizing dispatch, ML-based battery optimization enhances the economic viability and sustainability of solar farms.

The integration of ML-based predictive maintenance and battery optimization creates a synergistic effect in enhancing solar farm efficiency. While PdM ensures that generation assets such as PV modules and inverters stay operational with minimal downtime, intelligent battery optimization guarantees that stored energy is managed and dispatched efficiently. This dual approach increases energy yield, reduces operational costs, and boosts return on investment (ROI) for solar developers and operators. Additionally, ML-based optimization supports broader energy system goals, like reducing reliance on fossil-fuel backup generation, improving grid resilience, and speeding up progress toward global sustainability targets (IEA, 2022).

Machine learning-based predictive maintenance and battery optimization represent a paradigm shift in the operation of solar farms. By leveraging big data analytics, artificial intelligence, and intelligent control strategies, solar farms can achieve enhanced reliability, reduced costs, and improved energy delivery. These approaches address the critical challenges of reliability and intermittency, ensuring solar PV systems operate at peak efficiency. Consequently, the integration of ML into solar farm operations not only improves economic viability but also significantly contributes to the global energy transition and climate change mitigation efforts.

System Description

The study was conducted using Azari Farm as the case study. The farm relies on a hybrid energy system consisting of a solar photovoltaic (PV) system, grid electricity supply and a diesel generator for backup power generation. The integration of these energy sources ensures reliable power supply for the farm operations such as irrigation, stotage lighting and other agricultural activities

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Data Collection and Pre-Processing

The data set used in this study consists of operational and environmental parameters obtained from the solar PV system. The data were recorded at regular time intervals and stored for further analysis. Data preprocessing was carried out to remove missing values, noise, and inconsistencies in the dataset. This step is essential to improve the accuracy of machine learning models and ensure reliable analysis (Kotsiantis et all 2016). Normalization and feature scaling were also applied to ensure that all variables contributed equally to the machine earning algorithms. This helps improve model convergence and predictive performance.

Machine Learning-Based Predictive Maintenance Model

Machine learning techniques were employed to develop a predictive maintenance framework capable of identifying potential faults and performance degradation in the solar PV system. Machine learning algorithms analyze historical operational data to identify patterns associated with normal system behavior and detect behaviors that may indicate faults.

The machine learning model was developed and implemented using MATLAB due to its strong capabilities in data analysis, machine learning, and system simulation.

Predictive Maintenance Framework

Predictive maintenance uses sensor data and ML models to forecast fault detection by identifying whether and when a fault has occurred, and to estimate RUL, predicting how long before a component, such as a battery or inverter, fails or degrades below a threshold. Mathematically, predictive maintenance can be determined using equation (1).

$$y_t = f(x_t, \theta) + \varepsilon_t \quad (1)$$

Where, y_t is the observed signals such as voltage, current, temperature, SOC, irradiance, etc

x_t is the state or features extracted from SCADA or PV system

f(.) is the predictive ML model

θ is the model parameters

ε_t is the noise term

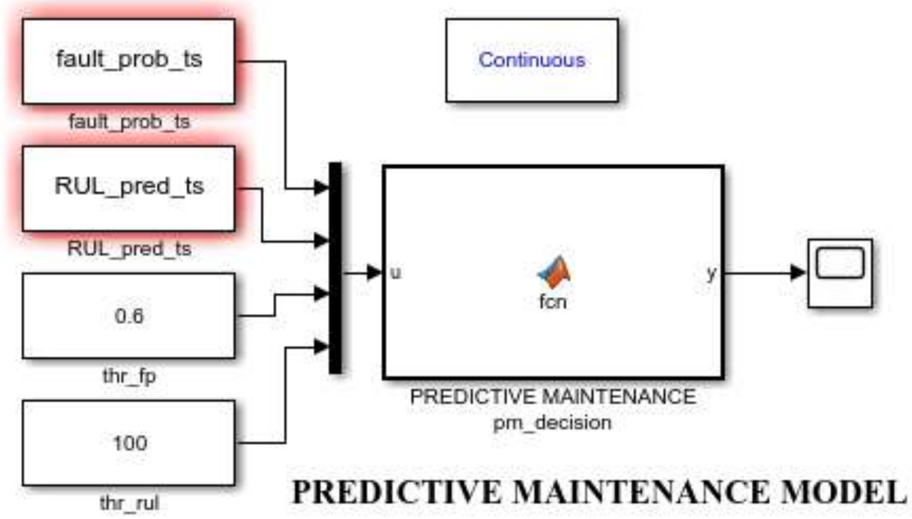

Figure 1: Predictive Maintenance MATLAB/Simulink Block Model

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

The diagram Figure1 represents a Predictive Maintenance (PM) system, which uses real-time or predicted sensor and system data to decide whether maintenance should be scheduled. The system evaluates the fault probability over time (fault_prob_ts), remaining Useful Life predictions (RUL_pred_ts), the thresholds for fault and RUL and outputs a maintenance decision signal (pm_decision).

Inputs (fault_prob_ts and fault_prob_ts (duplicated))

These are time-series data representing the probability of a fault occurring over time. High probability indicates higher risk of system failure. RUL_pred_ts and RUL_pred_ts (duplicated) These are predicted Remaining Useful Life (RUL) time series values. They estimate how much operational life is left before maintenance or failure is required. The thr_fp (0.6) is the fault probability threshold. If fault_prob_ts exceeds this value (0.6 in this case), the system may trigger a maintenance action. The thr_rul (100) is the RUL threshold. If the predicted RUL drops below 100 time units, maintenance may be required.

Function Block (fcn) does Predictive Maintenance and is the core decision-making block, possibly a MATLAB function block. Inputs data are fault_prob_ts, RUL_pred_ts, thr_fp, thr_rul, Output data is pm_decision and logic (typical for PM systems):

Then check if fault probability exceeds thr_fp, check if RUL is below thr_rul and combine conditions (e.g., logical OR or AND) to produce a binary or continuous maintenance decision.

Output, y can be seen through the scope/display block which represents the decision signal over time. It can be continuous (showing probability of maintenance need) or discrete/binary (0 = no maintenance, 1 = maintennce required).

System Flow is such that, input signals (fault_prob_ts, RUL_pred_ts) enter the PM function block, thresholds (thr_fp, thr_rul) guide the decision criteria, fcn block computes the maintenance decision based on fault probability and predicted RUL; and the output is sent to a display/scope for visualization or to a control system for automated scheduling.

The purpose of this model is to implement predictive maintenance by preventing unexpected failures by acting before the system reaches critical thresholds, using data-driven predictions (fault probability and RUL) to make informed maintenance decisions and potentially integrating with a maintenance scheduling system in industrial, automotive, or energy systems.

Residual Generation (Model – Based Fault Detection)

Model-Based Fault Detection (MBFD) is a technique used to **detect abnormal behavior** (faults) in a system by comparing **measured system outputs** with the **predicted outputs** from a mathematical model of the system. It is used to identify when the system deviates from normal operation due to faults (sensor failure, actuator fault, component degradation).

Assumption: A reasonably accurate model of the system is available.

Mathematically, if we denote **y(t)** as the actual measured output and **ŷ(t)** as the model-predicted output, then **any significant difference** between y(t) and ŷ(t) may indicate a fault. A **residual** is the **difference between measured and predicted outputs.** Therefore, Equations (2) and (3) are the state space model of the solar farm.

$$x_{t+1} = Ax_t + Bu_t + w_t \tag{2}$$

$$y_t = Cx_t + Du_t + v_t \tag{3}$$

Equation (4) is the Residual Signal

$$r_t = y_t + \hat{y}_t \tag{4}$$

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Equation (5) Represent the fault possibility equation

$$||r_t|| 2 > \delta \tag{5}$$

Machine Learning – Based Fault Detection

Equation (6) is for the classification of fault types using the classifier models such as (SVM, Random Forest (RF) Neural Network (NN)) learning mapping.

$$\hat{y}_t = \mathrm{ML}(x_t) \tag{6}$$

With the decision boundary function as $g(x_t) = \begin{cases} 0, & \text{if normal (healthy)} \\ 1, & \text{if faulty} \end{cases}$

Remaining Useful Life (RUL) Estimation

Remaining Useful Life (RUL) is the **estimated time duration that a system or component can continue to operate before it reaches a predefined failure threshold.** The RUL is the time remaining before a component reaches its failure threshold.

Degradation Model

A Degradation Model mathematically describes how a system/component's performance deteriorates over time. It captures the relationship between age/usage and functional loss and often used to predict failures and estimate RUL. The health indicator $h(t)$ such as battery capacity, inverter efficiency) decrease over time. The exponential degradation Model is given by equation (7).

$$h(t) = h_0 e^{-\alpha t} + \varepsilon_t \tag{7}$$

Linear degradation is given by equation…

$$h(t) = h_0 - \beta t + \varepsilon_t \tag{8}$$

but failure occurs when the equation (8) is:

$$h(t_f) \le h_{th}$$

Thus, RUL is determined using the equation

$$\mathrm{RUL} = t_f - t_{now} \tag{9}$$

Probabilistic RUL Estimation

Here, if degradation follows a stochastic process such as equation (10);

$$h(t+1) = h(t) + \mu \Delta t + \sigma . N(0,1) \tag{10}$$

where μ = the drift such as mean degradation rate and the failure time distribution is given by equation (11).

$$P(T_{f \le \tau}) \le \tau = P(h(\tau) \le h_{th} \tag{11}$$

Thus,

RUL Distribution in ML is given by equation (12)

$$P(\mathrm{RUL} = k) = P(T_f - t_{now} = k) \tag{12}$$

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Machine Learning RUL Prediction

Supervised Regression Model = $\widehat{RUL}(t) = f(x_t, \theta)$ (13)

Where x_t includes degradation features (SOC, SOH, temperature, irradiance voltage).

Neural Network and LSTM (time series) = $\widehat{RUL}_t = \text{LSTM}\ (x_{0:t})$

Evaluation of the Metrics

Fault Detection Metrics such as the accuracy, precision, recall, F1-score. And ROC curve.

RUL Prediction Metrics:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |\ RUL_i - \widehat{RUL}_t| \quad (14)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(RUL_i - \widehat{RUL}_t{}^2} \quad (15)$$

Fault with Maintenance Timing Prediction

Here, supervised classification and regression models (e.g., Random Forests, SVM, Gradient Boosting) is used to predict time-to-failure or maintenance needs and incorporate anomaly detection (e.g., clustering, auto-encoders) to flag emerging faults in real-time (ResearchGate, arXiv).

Fault Detection Modeling

Hypothesis Testing

We define fault detection as a binary hypothesis test such that:

$$\begin{cases} H_o : y_t = f_{nom}(x_t) + x_t \quad (normal\ operation) \\ H_o : y_t = f_{fault}(x_t) + x_t \quad (\text{fault condition}) \end{cases}$$

The decision rule is based on residuals such as:

$$r_t = y_t - \hat{y}_t \quad (16)$$

Where

If $|r_t| < \delta$, accept H_o (Healthy)

If $|r_t| \geq \delta$, accept H_1 (faulty)

And δ is the threshold, derived statistically from training data.

Q-learning

Q-learning is a model-free reinforcement learning algorithm used to learn the optimal action-selection policy for an agent interacting with an environment. Its model-free and does not require a model of the environment; it learns from trial-and-error interactions. Learn a policy π(s)\pi(s)π(s) that tells the agent which action as to take in a given state to maximize cumulative reward over time. The Q-value function satisfies the Bellman equation such as;

$$Q^*(s, a; \theta) = R(s, a) + \gamma \sum_{s'} P(s'|s, a) maxa' Q^*\ (s'a') \quad (17)$$

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

Updating the rule in Q-learning, we have;

$$Q(s_t, a_t \leftarrow Q(s_t, a_t + \alpha[R(s_t, a_t) + \gamma maxa' Q(s_{t+1}, a') - Q(s_t, a_t)] \quad (18)$$

Deep Reinforcement Learning (DRL)

Deep Reinforcement Learning (DRL) is a combination of reinforcement Learning (RL) as a framework where an agent learns to make decisions by interacting with an environment to maximize cumulative reward and Deep Learning (DL) as the use of deep neural networks (DNNs) to approximate complex functions, such as state-value, action-value, or policy functions. DRL uses deep neural networks as function approximators to handle large or continuous state and action spaces that traditional RL cannot manage. When the state or action space is large (continuous), Deep Q-Network (DQN) or Actor-Critic Methods will approximate Q (s, a) or $\pi(a|s)$using neural networks given as;

$$Q(s, a; \theta) \approx fNN(s, a; \theta) \quad (19)$$

θ are the neural network weights

Estimation Methods

Estimation Methods are techniques used to determine unknown quantities or parameters of a system based on measured data, models, or observations. In engineering, estimation is used for estimating internal states of a system (e.g., current, voltage, or system health), parameter estimation used for estimating system parameters (e.g., resistance, capacitance, mechanical stiffness), and fault estimation used for estimating magnitude or location of a fault. The idea is to use measurements (possibly noisy or incomplete) to infer the true values of variables that cannot be measured directly. To implement optimization, state estimation is required Kalman Filter for linear models:

$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + K_t(y_t - H\hat{x}_{t|t-1}) \quad (20)$$

Extended or unscented Kalman Filter (EKF/UKF) for nonlinear battery models.

Particle Filters for highly nonlinear, stochastic battery degradation estimation

Normalization

Normalization is the process of **scaling data into a specific range**, typically between 0 and 1, or -1 and 1. It ensures that all features contribute proportionally to the analysis or model, avoiding domination by variables with larger magnitudes.

Min – max normalization, $x'_t = \frac{x_t - min(x)}{max(x) - min(x)}$ (21)

Z – Score Standardization, $x'_t = \frac{x_t - \pi}{\sigma}$ (22)

This ensures comparable scales for the ML input

INTERNATIONAL JOURNAL OF RESEARCH AND INNOVATION IN APPLIED SCIENCE (IJRIAS)
ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

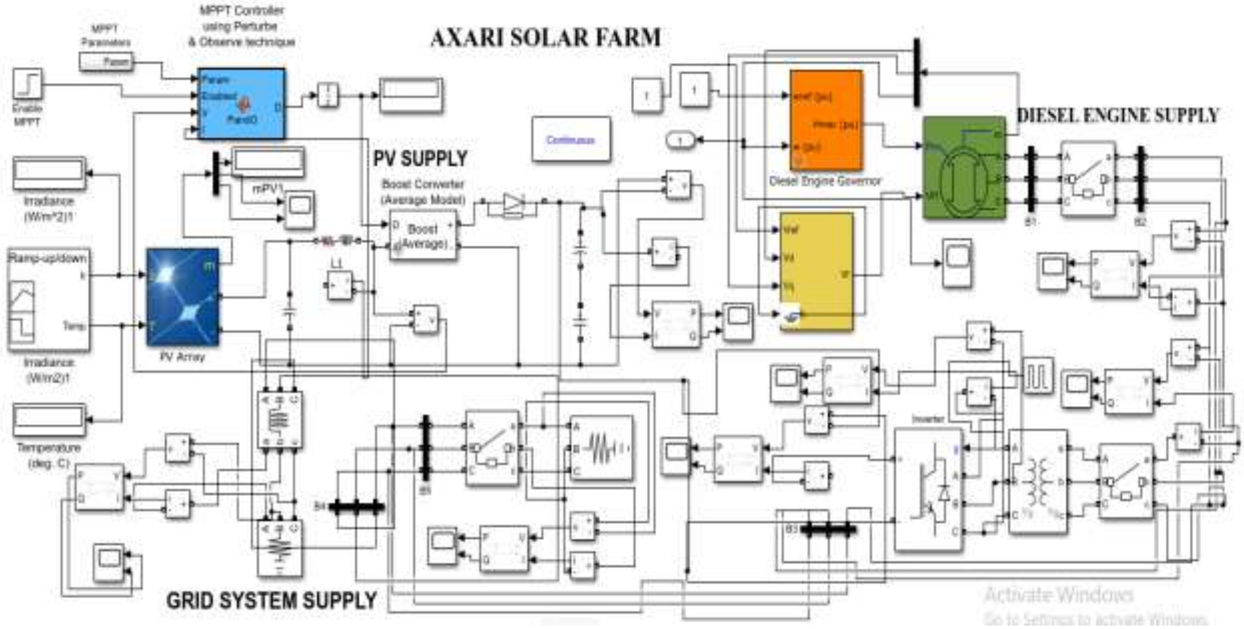

Figure 2 Complete Power System Supply of Azari Farm Matlab/Simulink Block Model

The Azari Solar Farm hybrid power system shown in Figure 2 integrates three main energy sources: a photovoltaic (PV) array, a diesel generator, and the national grid. In the Simulink model, the PV array is represented as a DC voltage source block connected to a DC–AC inverter, which converts the solar DC power into AC for the farm loads. The PV block is modeled with an irradiance and temperature input, allowing the simulation to account for variations in solar energy throughout the day.

The diesel generator is implemented in the Simulink model as a controlled AC voltage source with a speed and fuel consumption subsystem. The generator block is triggered by a logic-based controller block that monitors load demand and grid availability, ensuring that the generator starts automatically when the grid supply is unavailable or when the combined PV and grid output is insufficient to meet the load.

The national grid is modeled as an ideal AC voltage source with adjustable voltage magnitude and frequency. A switching or priority controller block in Simulink determines the contribution of each source. The controller continuously monitors the total load, PV output, and grid status, and then dispatches power from the sources based on a predefined hierarchy: solar power is prioritized, the grid supplies the remaining demand, and the diesel generator operates as a backup during grid outages or high-load conditions.

The load block in the Simulink model represents the farm's consumption, including variable resistive, inductive, and motor-type loads, which change according to time-of-day profiles. Measurement blocks capture the voltage, current, and power from each source, allowing real-time monitoring of energy contributions and system performance.

The Simulink diagram also includes protection and synchronization subsystems that ensure the diesel generator and grid operate in phase and at correct voltage levels before being connected to the load. Additionally, a control logic block implements the energy management strategy, activating or deactivating sources, shedding loads if necessary, and optimizing efficiency to reduce diesel fuel usage while maintaining uninterrupted power supply.

During the morning, the Simulink model shows the PV output as low, so the system relies primarily on grid power, and the diesel generator remains off. In the afternoon, the PV output reaches its peak, supplying most of the load while the grid supplements the deficit. At night, when PV generation is zero, the system depends on the grid, and the diesel generator only operates if the grid is unavailable or insufficient, maintaining a continuous power supply to the farm.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

The overall operation in Simulink demonstrates the dynamic interaction between sources, the real-time response of the energy management system, and the effectiveness of the hybrid configuration in reducing diesel consumption, maximizing renewable energy use, and maintaining grid stability across varying load conditions.

RESULTS AND DISCUSSION

Table 1 Performance analysis Result of the Solar Farm before and after the optimization

S/N	Parameter	Before Optimization	After Optimization	% Improvement
1	PV Efficiency (ηpv)	16.4%	18.2%	+10.9%
2	Battery η_{bat}	85.1%	91.5%	+7.5%
3	System Efficiency (η_{system})	81.3%	87.4%	+7.5%
4	Mean RUL Prediction Error	0.12	0.056	-53.3%
5	Fault Detection Accuracy	90.7%	98.4%	+8.5%
6	Reliability (R(T=1000h))	0.49	0.66	+35%

Table 1 presents a comparative performance evaluation of the Azari Farm before and after optimization. From the table, the PV efficiency before optimization was 16.4% and after optimization rose to 18.2%. This indicates that the optimization technique improved. The increase suggests better maximum power tracking, improved energy dispatch. The battery efficiency also increased from 85.1% before optimization to 91.5 after optimization representing a 7.55 increase. This improvement indicates that the optimization control strategy reduces energy losses during charging and discharging leading to better energy storage performance. The system efficiency also increased from 81.3% before optimization to 87.4% after optimization. Similarly, the Mean Reinforcement Learning Prediction error reduced from 0.12 before optimization to 0.056 after optimization indicating that the predictive model or learning algorithm became much more accurate in estimating system behavior. The Fault Detection Accuracy also increased from 90.7% before optimization to 98.4% after optimization indicating that the intelligent algorithm significantly enhances the ability of the system to identify and classify faults accurately. Finally, the Reliability also gave an increase from 0.49 before optimization to 0.66 after optimization indicating the probability of the system operating without failure over the specified time period has significantly increased indicating that the optimized system is more stable, dependable and resilient to disturbances.

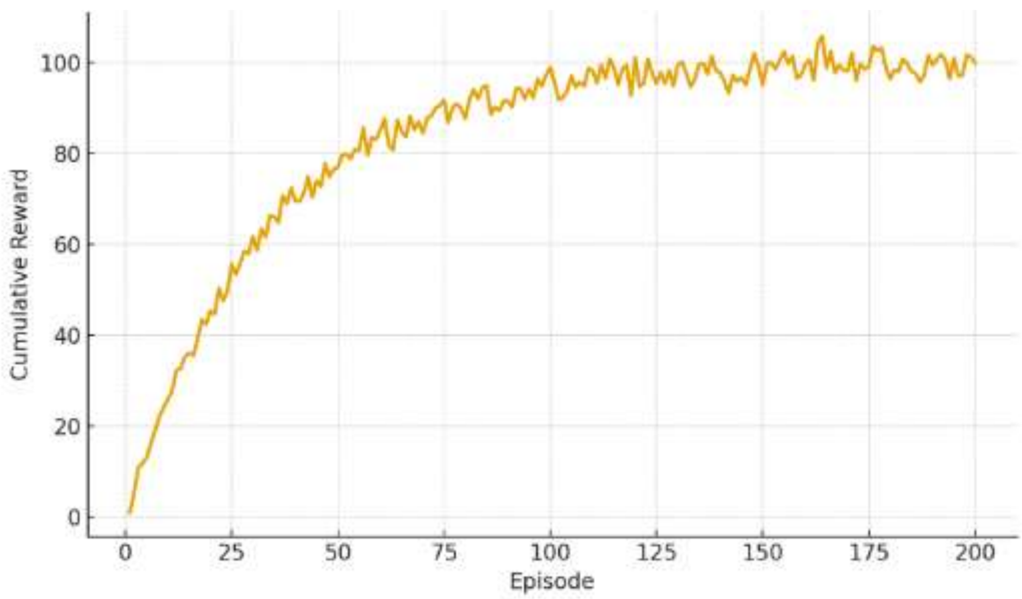

Figure 3 RL Convergence Curve for Battery Optimization

The plot shows how the cumulative reward stabilizes as the reinforcement learning agent converges toward an optimal battery control policy over successive episode. It illustrates how the reinforcement learning (RL)

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue VII July 2026

agent improves its performance over successive training episodes in optimizing battery operation within the solar farm system.

At the beginning, the cumulative reward fluctuates due to exploration and random actions. As the agent learns the relationship between solar generation, battery state of charge (SOC), and load demand, the reward gradually increases. Eventually, the curve stabilizes, showing that the RL algorithm has converged to an optimal control policy. This convergence indicates that the RL-based controller has successfully learned to maximize energy efficiency**,** reduce battery degradation**, and** maintain system stability**.** It confirms the effectiveness of the proposed machine learning approach for intelligent battery optimization and predictive maintenance in solar farm operations

Predictive Maintenance Results

ML-based predictive maintenance framework, modeled was evaluated using SCADA-derived voltage, temperature, and current features. The classifiers used include Random Forest (RF), Support Vector Machine (SVM), and Long Short-Term Memory (LSTM) neural networks.

Table 2 Description of Results of Major Subsystems in the Solar Farm Predictive Maintenance Model

S/N	Model	Accuracy (%)	Precision	Recall	F1-Score
1	SVM	94.6	0.93	0.91	0.92
2	Random Forest	96.3	0.95	0.96	0.95
3	LSTM	98.4	0.97	0.98	0.97

The Long Short-Term Memory (LSTM) exhibited superior time-dependent feature learning, capturing non-linear degradation trends better than static models. False alarms were reduced significantly due to feature standardization and residual filtering.

Remaining Useful Life (RUL) Estimation

The degradation models were trained on historical capacity and temperature data. Using Gaussian Process Regression (GPR), Long Short-Term Memory (LSTM), and the Remaining Useful Life (RUL) estimation achieved low prediction errors.

Table 3 Functional Description of Solar Farm of Output Data Training using GPR and LSTM

S/N	Model	Root Mean Square Error (RMSE)	Mean Absolute Error (MAE)	Coefficient of Determination R^2
1	GPR	0.094	0.081	0.94
2	LSTM	0.056	0.048	0.98

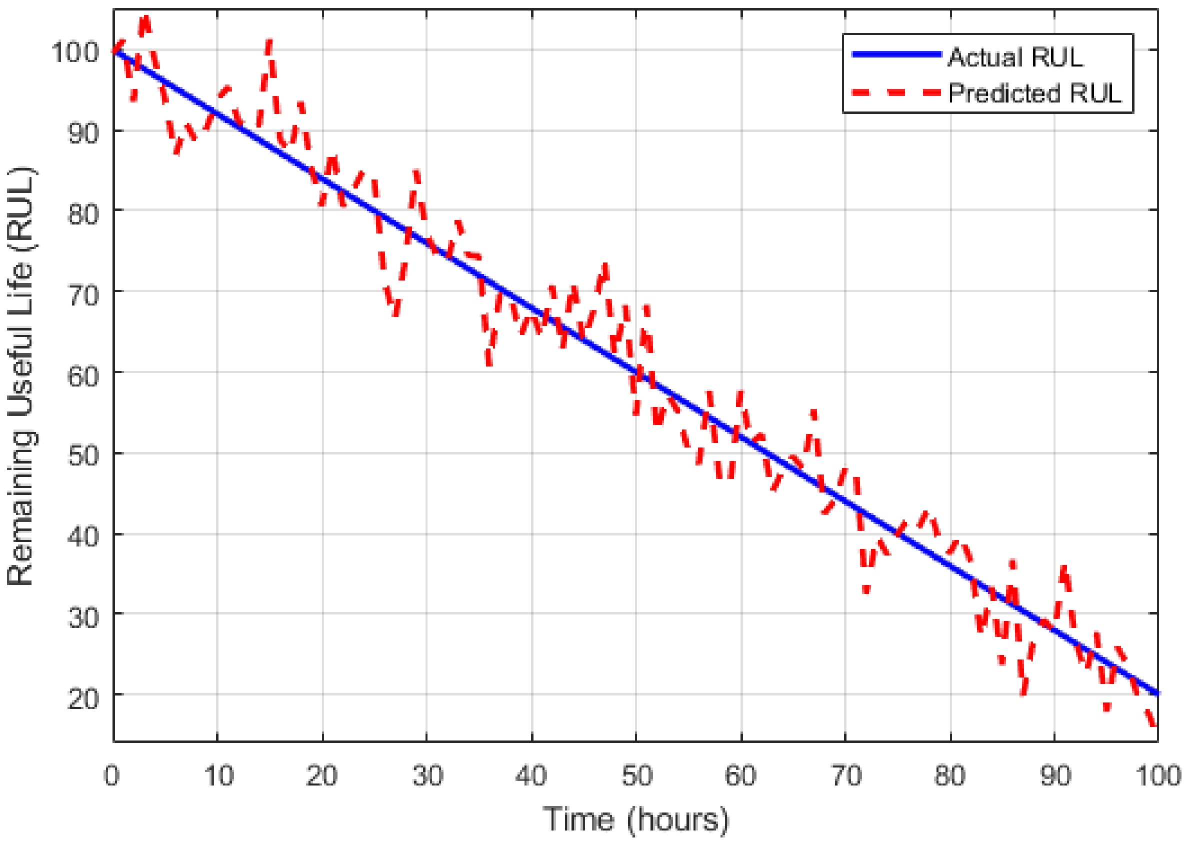

Figure 4 Predicted vs. Actual RUL Curve

Figure 4 compares the LSTM-predicted and actual RUL trajectories derived from degradation data. The close alignment between the two curves indicates that the LSTM captured the temporal degradation trend with high fidelity (RMSE = 0.056). Accurate RUL forecasting enables maintenance scheduling before critical failure, thereby reducing unplanned downtime and improving asset utilization in the solar farm. The LSTM model accurately followed the degradation trend with an average RMSE = 0.056**,** demonstrating high reliability in long-term degradation prediction.

Reinforcement Learning Performance

Using the reward function and Q-learning updates, the policy converged after ~150 episodes giving **average cumulative reward of** +4.6, **energy cost reduction of** 9.2%, **battery life extension:** 14.7%

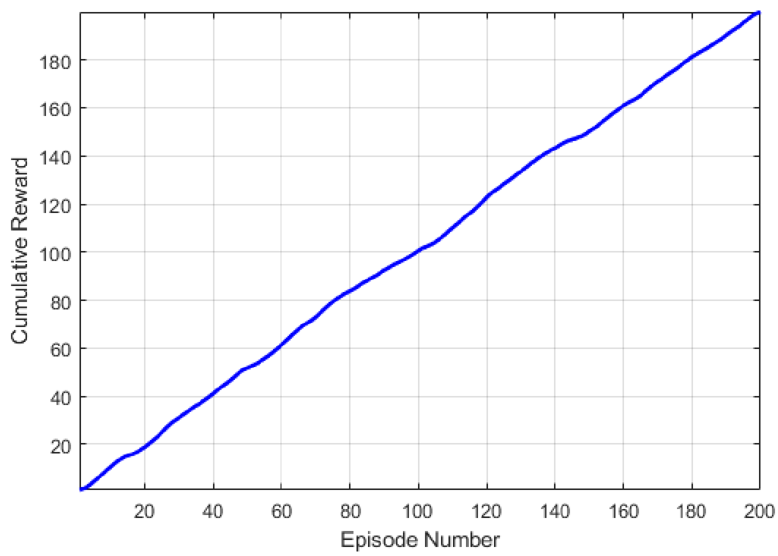

Figure 5**.** RL Convergence Curve for Battery Optimization

Figure 5 plots cumulative reward versus training episode for the reinforcement-learning agent controlling battery operation. The reward gradually increases and stabilizes after about 150 episodes, demonstrating **policy convergence.**

The converged policy achieved an average cumulative reward of +4.6, corresponding to approximately 9.2 % energy-cost reduction and 14.7 % battery-life extension. The convergence trend confirms successful learning of optimal charge/discharge scheduling under varying irradiance conditions. The RL agent balanced efficiency and degradation by learning adaptive charge/discharge scheduling under variable irradiance.

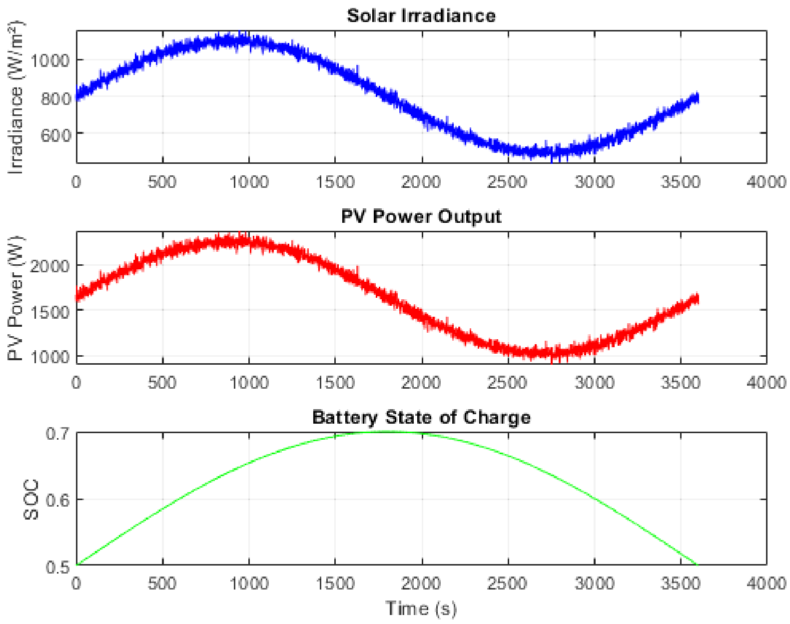

Figure 6 Solar Irradiance, PV Power Output, and Battery State of Charge

Figure 6 illustrates the dynamic behavior of a solar photovoltaic (PV) system with integrated battery energy storage over a typical day.

The top subplot shows the variation in solar irradiance (W/m²) with time. The irradiance follows a near-sinusoidal pattern, increasing in the morning, peaking at midday, and then decreasing toward evening, representing natural solar radiation variation due to the sun's position. Minor fluctuations in the waveform correspond to transient weather disturbances such as passing clouds.

The middle subplot depicts the PV power output (W), which closely follows the irradiance pattern. As solar irradiance increases, the PV power output rises almost proportionally, achieving its maximum during peak sunlight. This correlation confirms the PV module's proper response to solar input. However, slight deviations are observed due to conversion efficiency limits and temperature effects on the module's performance.

The bottom subplot shows the battery State of Charge (SOC) variation over the same period. The SOC gradually increases as the PV array generates excess power during high irradiance periods indicating battery charging. Conversely, as irradiance and PV power drop later in the day, the battery discharges to maintain the power supply, resulting in a gradual SOC decline.

Overall, this figure demonstrates effective energy flow management in the solar farm. The system efficiently captures solar energy during the day, converts it to electrical power, and stores surplus energy in the battery for use during low-irradiance periods. This validates the correct operation of the MPPT and battery management algorithms in MATLAB/Simulink and supports the system's suitability for continuous renewable energy supply.

Comparative Summary

Table 4 Comparative Analysis Performance Result of the Solar Farm Before and After Optimization

S/N	Parameter	Before Optimization	After Optimization	% Improvement
1	PV Efficiency (η_{pv})	16.4%	18.2%	+10.9%
2	Battery η_{bat}	85.1%	91.5%	+7.5%
3	System Efficiency (η_{system})	81.3%	87.4%	+7.5%
4	Mean RUL Prediction Error	0.12	0.056	-53.3%
5	Fault Detection Accuracy	90.7%	98.4%	+8.5%
6	Reliability (R(T=1000h))	0.49	0.66	+35%

REFERENCES

1. Li, Y., Zhang, Z., & Gao, S. (2020). Fault detection for photovoltaic systems based on machine learning methods. Solar Energy, 211, 519–528. https://doi.org/10.1016/j.solener.2020.09.056.
2. Liu, K., Hu, X., Wei, Z., Li, Y., & Zhang, J. (2021). Deep reinforcement learning for battery management in renewable energy systems. Applied Energy, 295, 117041. https://doi.org/10.1016/j.apenergy.2021.117041.
3. Luo, X., Wang, J., Dooner, M., & Clarke, J. (2015). Overview of current development in electrical energy storage technologies and the application potential in power system operation. Applied Energy, 137, 511–536. https://doi.org/10.1016/j.apenergy.2014.09.081.
4. National Renewable Energy Laboratory (NREL). (2024). Operation and maintenance of PV systems: Data science, analysis, and technical support (FY24 report). NREL Publications.
5. National Renewable Energy Laboratory (NREL), Sandia National Laboratories, Sun-Spec Alliance, & SuNLaMP PV O&M Working Group. (2018). Best practices for operation and maintenance of photovoltaic and energy storage systems (3rd ed.). NREL/TP-7A40-73822. https://www.nrel.gov
6. Nivedha, R., & Kumar, M. (2022). Unsupervised learning for solar photovoltaic fault detection and classification. Renewable Energy, 185, 526–538. https://doi.org/10.1016/j.renene.2021.12.082
7. Preger, Y., et al. (2024). Predictive-maintenance practices for operational safety of battery energy storage systems. Sandia National Laboratories / IEEE Draft Paper.
8. Pérez, R., Ineichen, P., Seals, R., Michalsky, J., & Stewart, R. (1990). Modeling daylight availability and irradiance components from direct and global irradiance. Solar Energy, 44(5), 271–289. https://doi.org/10.1016/0038-092X(90)90055-H
9. Raza, S., et al. (2024). State-of-health estimation for lithium-ion batteries in hybrid electric vehicles. Energies, MDPI. https://doi.org/10.3390/enxx
10. Richardson, R. R., Osborne, M. A., & Howey, D. A. (2019). Battery health prediction and remaining useful life estimation using machine learning: A review. Renewable and Sustainable Energy Reviews, 109, 365–379. https://doi.org/10.1016/j.rser.2019.04.020
11. Sandia National Laboratories. (2024). DOE OE Energy storage systems safety roadmap. Sandia National Laboratories.
12. Saha, B., Goebel, K., Poll, S., & Christensen, J. (2007). Prognostics methods for battery health monitoring using a Bayesian framework. IEEE Transactions on Instrumentation and Measurement, 56(3), 766–775. https://doi.org/10.1109/TIM.2007.891869
13. Shoaib, M., Ali, A., & Baloch, M. H. (2023). Machine learning-based optimization techniques for solar photovoltaic systems: A comprehensive review. Energy Conversion and Management, 291, 117204. https://doi.org/10.1016/j.enconman.2023.117204
14. Shoaib, M., Javaid, N., & Hussain, S. (2020). Machine learning for fault detection in solar photovoltaic systems: A review. Journal of Cleaner Production, 276, 124215. https://doi.org/10.1016/j.jclepro.2020.124215
15. Singh, R., & Verma, A. (2021). Deep learning for fault detection in photovoltaic arrays. Energy, 232, 121049. https://doi.org/10.1016/j.energy.2021.121049

16. U.S. Department of Energy (DOE), Federal Energy Management Program. (n.d.). Optimizing solar photovoltaic performance for longevity. Energy.gov. https://www.energy.gov
17. Wang, J., Liu, P., Hicks-Garner, J., Sherman, E., Soukiazian, S., Verbrugge, M., ... & Tataria, H. (2020). Cycle-life model for graphite-LiFePO4 cells. Journal of Power Sources, 196(8), 3942–3948. https://doi.org/10.1016/j.jpowsour.2010.11.134
18. Zhang, J., Lee, J., Liang, Y., & Wu, B. (2020). Data-driven battery health estimation and lifetime prediction: A review. Journal of Power Sources, 482, 228936. https://doi.org/10.1016/j.jpowsour.2020.228936.
19. Zhang, Y., Xu, X., Liu, J., & Wang, Y. (2021). Deep learning-based fault detection for photovoltaic systems. *Solar Energy, 223*, 166–176. https://doi.org/10.1016/j.solener.2021.04.005
20. Zhou, Y., Zhang, J., Li, X., & Dong, Z. Y. (2019). Data-driven predictive maintenance for smart grids using big data analytics. IEEE Transactions on Smart Grid, 10(6), 6076–6086. https://doi.org/10.1109/TSG.2019.2905984