© April 2025 | IJIRT | Volume 11 Issue 11 | ISSN: 2349-6002

Improving Solar Panel Efficiency with Predictive Alerts from IoT and Machine Learning

Gursharan Singh[1], Lakshay Singhal[2], Rahul Sharma[3], Haider Abbas[4]
[1,2,3,4] *Dr. Akhilesh Das Gupta Institute of Professional Studies*

***Abstract*—Solar energy is a cornerstone of the global transition toward sustainable power generation, yet the efficiency of solar panels in established installations often declines due to environmental and operational factors. This study presents a machine learning-driven approach to enhance the performance of existing solar infrastructure by predicting optimal maintenance schedules, particularly for panel cleaning. Utilizing real-time environmental data collected through IoT sensors—measuring temperature, humidity, dew point, and precipitation—several machine learning models were implemented to forecast efficiency drops and identify ideal cleaning windows. Models such as Recurrent Neural Networks (RNN), Random Forest, K-Nearest Neighbors (KNN), and Linear Regression were trained and evaluated using historical performance data and weather conditions. The results demonstrate that predictive modeling can significantly improve operational efficiency, reduce resource usage, and extend panel longevity. This approach enables a scalable and intelligent maintenance framework for solar farms, contributing to more sustainable and cost-effective energy generation.**

I. INTRODUCTION

The increasing demand for renewable energy sources has positioned solar power as a vital component of the global energy transition. With the rapid deployment of photovoltaic (PV) systems, especially in urban and semi-urban environments, the focus is now shifting from merely installing solar panels to optimizing their long-term performance. While new technologies continue to improve the efficiency of PV cells, many existing solar installations face the challenge of declining performance due to environmental and operational factors such as temperature fluctuations, humidity, dust accumulation, and precipitation.

Traditionally, enhancing solar panel efficiency involved hardware upgrades or manual maintenance routines. However, such solutions are often impractical or cost-prohibitive for already-installed systems. As a result, there is a growing need for intelligent, data-driven strategies that ensure sustained panel performance without altering existing infrastructure.

In this paper, we propose a predictive maintenance framework leveraging machine learning (ML) techniques and real-time environmental monitoring through Internet of Things (IoT) devices. Our system collects localized weather and performance data to train ML models capable of forecasting efficiency drops and triggering timely maintenance actions, especially panel cleaning. We explore and compare the effectiveness of four different ML models—Recurrent Neural Networks (RNN), Random Forest, K-Nearest Neighbors (KNN), and Linear Regression—in predicting the ideal time for cleaning based on multiple environmental parameters. Additionally, solar panel owners will receive proactive alerts via a connected mobile application whenever panels are underperforming or require repairs—often even before the scheduled maintenance—enabling quicker response times and preventing further efficiency loss.

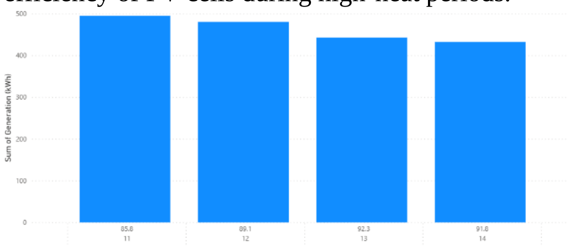
By shifting from reactive to predictive maintenance, this approach aims to improve energy yield, reduce unnecessary water and labor usage, and enable smarter, more sustainable management of solar panel installations.

II. ENVIRONMENTAL FACTORS AND HOW THEY AFFECT GENERATION

The performance of solar panels is closely tied to a variety of environmental factors that influence energy generation. In our previous study, real-time sensor data was collected to analyze the impact of four key environmental parameters: temperature, humidity, dew point, and precipitation. These factors significantly contribute to the fluctuation in solar energy output and, if not properly accounted for, can lead to decreased efficiency and operational reliability.

2.1 Temperature:

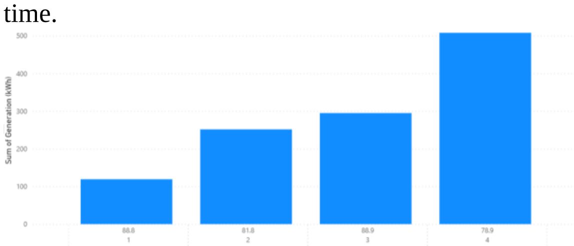
Temperature is a critical determinant of solar panel performance. While PV cells require sunlight to generate electricity, excessive heat can cause efficiency to decline beyond the standard test conditions (STC). Our analysis showed a consistent inverse relationship between panel temperature and power output, with energy generation dropping as temperature rose above 25°C. This phenomenon, known as the "hotspot effect," limits the conversion efficiency of PV cells during high-heat periods.



Graph showing how rising temperature leads to decreased solar generation due to hotspot effects.

2.2 Humidity:

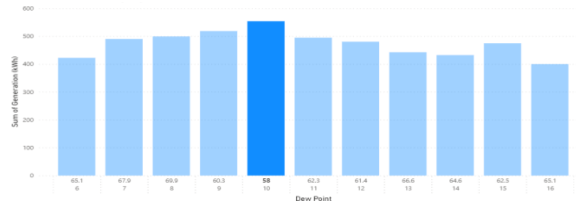
High humidity levels were found to contribute to condensation on the surface of the solar panels. This not only reduces the amount of sunlight reaching the PV cells but also promotes the growth of mold and algae, which can create long-term performance issues. Though the impact of humidity is not as immediate or drastic as temperature, it remains a relevant factor over time.



Graph showing the effect of humidity on solar generation—high humidity slightly reduces output but has less impact than extreme temperatures.

2.3 Dew Point:

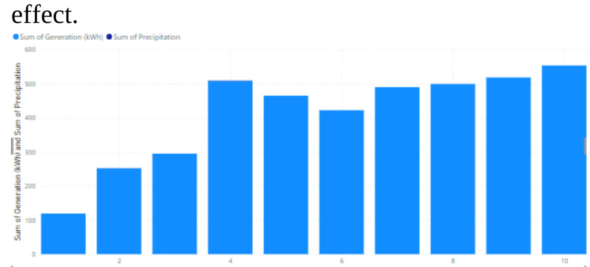
The dew point provides insight into the likelihood of condensation forming on panel surfaces. When the dew point nears the actual air temperature, conditions are ideal for dew formation, which can obscure sunlight and reduce efficiency. This metric, derived from temperature and humidity readings, was crucial in identifying periods where preemptive cleaning or monitoring was needed.



Graph showing how higher dew point levels lead to dew formation, which lowers solar generation efficiency.

2.4 Precipitation:

Rainfall has a dual impact on solar panel efficiency. During precipitation, generation decreases due to reduced sunlight penetration from cloud cover. However, rain also serves as a natural cleaning agent, removing accumulated dust and debris from the panels. Our data revealed improved panel performance in the days following rainfall, validating its cleaning effect.



Graph showing rain's short-term drop, long-term gain.

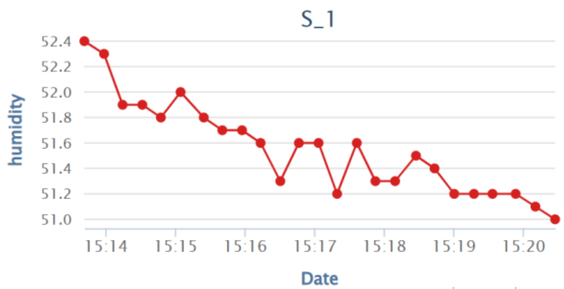
III. DATA COLLECTION AND IMPLEMENTATION CHALLENGES

To build a predictive model capable of improving solar panel efficiency, a robust data infrastructure was essential. Our setup combined hardware for environmental sensing, cloud-based data storage, and real-time monitoring, forming the backbone of the machine learning pipeline. However, this process was not without its technical and operational hurdles.

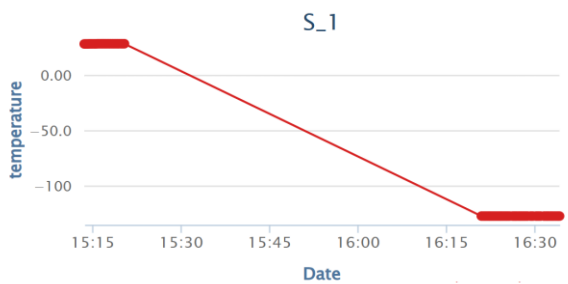
The data collection framework was implemented on a rooftop solar installation at our institute. To monitor environmental conditions, we used an ESP32 microcontroller integrated with a DHT11/22 sensor to collect real-time data on temperature and humidity. The microcontroller continuously transmitted this data to Firebase, a cloud-based platform chosen for its scalability and ease of integration with real-time analytics.

To enrich our dataset, we incorporated external weather data from APIs such as Weather Underground, which provided additional variables like dew point and precipitation forecasts. Concurrently, the energy generation data—including voltage, current, and power output—was obtained directly from the solar charge controller and matched with environmental data over time.

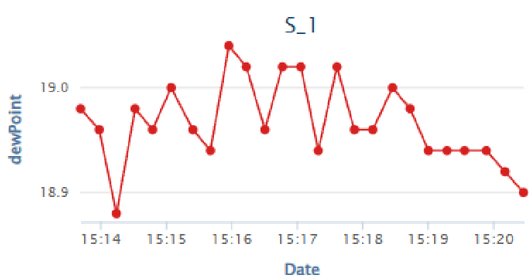
These graphs depict how environmental variables varied during a sampling window:



Humidity fluctuations across time, with a general downward trend.(Humidity over Time)



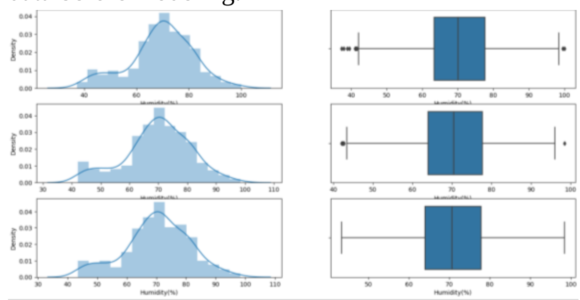
Illustrates a sharp and likely erroneous drop in temperature, indicating either faulty sensor readings or connectivity issues. (Temperature over Time)



the dew point variation, showing short-term fluctuations that impact condensation on panel surfaces. (Dew Point over TIme)

These graphs formed the basis for identifying environmental impacts on solar generation and guided feature engineering for the machine learning models.

Additionally, to ensure model accuracy, data cleaning was a critical step in our preprocessing pipeline. We applied outlier detection and removal techniques using statistical visualizations such as box plots and KDE (Kernel Density Estimation) graphs. Below graph shows the distribution and outlier ranges of humidity data before modeling.



Humidity Distribution and Outlier Detection

This blend of sensor data, forecasts, and filtering ensured a clean, reliable dataset for model training.

IV. MACHINE LEARNING ALGORITHMS

To build a predictive system capable of identifying underperformance in solar panel generation, four machine learning algorithms were implemented: K-Nearest Neighbors (KNN), Random Forest (RF), Linear Regression (LR), and Recurrent Neural Networks (RNN). Each model was trained on a dataset combining environmental parameters—temperature, humidity, dew point, and precipitation—with solar energy generation values. The effectiveness of each model was evaluated using Mean Squared Error (MSE).

4.1 Random Forest:

The Random Forest model emerged as the top performer among the four. Its ensemble-based architecture allowed it to handle complex, non-linear relationships between input features and solar power output. One of its key advantages was the ability to resist overfitting while still capturing detailed interactions across variables. Additionally, it provided insights into feature importance, helping us identify which environmental factors had the greatest impact on energy generation. Due to its balance of performance, stability, and interpretability, this model was ultimately chosen for integration into the mobile alert system.

4.2 Linear Regression:

Linear Regression, though simple, demonstrated surprisingly strong results. It performed close to the Random Forest model and offered the added benefit of complete transparency in how input features influenced predictions. Its mathematical simplicity made it a reliable baseline model and an excellent tool for validating insights obtained from more complex algorithms. Despite its assumption of linear relationships, it effectively captured the core trends in the data.

4.3 K-Nearest Neighbors:

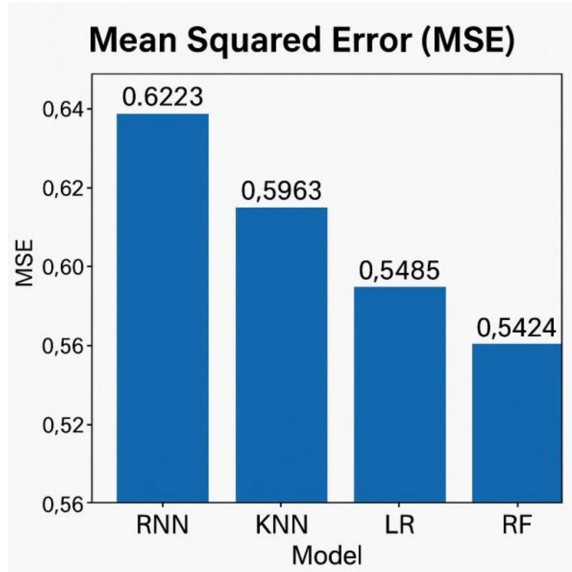
KNN offered reasonable performance, especially in stable environmental conditions. By predicting values based on the closest historical observations, it worked well when patterns repeated. However, its performance suffered in the presence of outliers or rapid fluctuations in weather, making it less dependable for edge-case predictions. Its sensitivity to the density and distribution of data limited its overall robustness.

4.4 Recurrent Neural Network:

The Recurrent Neural Network (RNN) was designed to leverage time-series relationships within the data. While theoretically well-suited for this type of task, the model underperformed compared to others. This was largely due to the limited size and temporal range of our dataset, which constrained the RNN's ability to learn long-term dependencies effectively. Despite this, RNN remains promising for future versions of the system, especially with access to larger, more diverse datasets over extended periods.

V. RESULTS AND OUTCOMES

The integration of machine learning into our solar monitoring framework produced impactful results, both in predictive performance and practical application. Our goal—to forecast underperformance in solar panels and issue timely maintenance alerts—was successfully realized through a combination of sensor-based data collection and model-driven decision making.



All four machine learning models contributed to the system's development and were evaluated using Mean Squared Error (MSE) as the primary metric. The Recurrent Neural Network (RNN) resulted in an MSE of 0.6223, indicating challenges in capturing accurate temporal trends due to the limited size of our dataset. The K-Nearest Neighbors (KNN) model performed slightly better with an MSE of 0.5963, while Linear Regression (LR) improved upon that with an MSE of 0.5485. However, the best performance was achieved by the Random Forest (RF) model, which yielded the lowest MSE of 0.5424, highlighting its strength in handling non-linear interactions between environmental variables and power output.

Owing to its superior accuracy and stability, the Random Forest model was selected for integration into the live system. Once deployed, it was used to continuously analyze real-time sensor data and forecast drops in panel efficiency. When predicted output fell below optimal thresholds, the system generated maintenance alerts—typically sent via the connected mobile application—prompting users to take preventive action such as panel cleaning or inspection. These alerts were often triggered one to two days before a noticeable drop in generation, giving users valuable lead time and preventing energy losses. Overall, the results validated the feasibility of using ML for predictive maintenance in solar infrastructure. The system not only improved energy efficiency but also minimized unnecessary maintenance efforts, conserved resources like water, and allowed for better planning of servicing routines. Its low-cost design using IoT hardware and open-source platforms makes it highly scalable for broader adoption in both residential and commercial solar installations.

REFERENCES

[1] Almonacid, F., Rus, C., Hontoria, L., & Pérez, P. J. (2011). Calculation of the energy provided by a photovoltaic generator. Comparative study: Conventional methods vs. artificial neural networks. Energy, 36(1), 375–384. https://doi.org/10.1016/j.energy.2010.10.024

[2] Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.

[3] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

[4] DHT11 Temperature and Humidity Sensor Specifications. (n.d.). Retrieved from https://components101.com/sensors/dht11-temperature-sensor

[5] ESP32 Datasheet. (n.d.). Espressif Systems. Retrieved from https://www.espressif.com/en/products/socs/esp32/resources

[6] Muhammad, M., Hussain, I., & Choi, Y. (2019). Prediction of photovoltaic power output using machine learning. Energies, 12(2), 219. https://doi.org/10.3390/en12020219

[7] Niu, S., Tang, Y., & Hu, H. (2014). Photovoltaic energy forecasting using a statistical model. Energies, 7(5), 2711–2730. https://doi.org/10.3390/en7052711

[8] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

[9] Weather Underground API Documentation. (n.d.). Retrieved from https://www.weatherunderground.com/weather/api/

[10] Zhang, Y., & Wang, J. (2015). Time series forecasting using a hybrid ARIMA and RNN model. Neurocomputing, 144, 535–545. https://doi.org/10.1016/j.neucom.2014.07.081

[11] Wen, L., Zhou, K., Yang, S., & Lu, X. (2019). Optimal load dispatch of community integrated energy systems based on deep learning. Energy, 171, 1051–1065. https://doi.org/10.1016/j.energy.2019.01.096

[12] Bacher, P., Madsen, H., & Nielsen, H. A. (2009). Online short-term solar power forecasting. Solar Energy, 83(10), 1772–1783. https://doi.org/10.1016/j.solener.2009.05.016

[13] Google Firebase Documentation. (n.d.). Retrieved from https://firebase.google.com/docs

[14] Yuan, J., Luo, W., Zhang, H., & Ma, T. (2022). Performance analysis and degradation prediction of photovoltaic modules using ML algorithms. Energy Reports, 8, 2099–2110. https://doi.org/10.1016/j.egyr.2022.01.043