

Machine Learning-Based Optimization of Photovoltaic Panel Cleaning Schedules in Arid Environments: A Case Study of Riyadh, Saudi Arabia

Lyu Guanghua, Syed Hadi Hussain Shah *, Rizwan Arshad, Ni Jiahua

Power China Huadong Engineering Corporation Limited, China

Corresponding author: shah_shh@hdec.com

Received: 17 June 2025, Revised:19 January 2026, Accepted: 31 March 2026, Published: 01 April 2026

KEY WORDS	ABSTRACT
Photovoltaic (PV) Systems Dust Accumulation Machine Learning (ML) Models XGBoost Dynamic Cleaning Schedules	Photovoltaic solar power systems in arid regions offer tremendous energy potential due to high solar irradiance. This study proposes a machine learning framework for predicting the effects of pollution on photovoltaic panels. Environmental data were collected from NASA, the Solar Photovoltaic Information System (PVGIS), and local air quality stations. The XGBoost, Random Forest, and LSTM machine learning models were used in the analysis. The results show that the XGBoost model outperformed the other models with a coefficient of determination (R^2) of 0.93, thanks to its ability to accurately predict energy loss due to pollution. The cost-benefit analysis demonstrated that biweekly cleaning maximizes energy recovery and reduces operating costs by up to 15% compared to fixed schedules. The proposed framework integrates real-time dust prediction and dynamic scheduling, providing a scalable solution for solar farms in arid environments. This approach improves the efficiency of photovoltaic solar energy systems, reduces water and labor waste, and supports the deployment of sustainable solar energy in dust-prone areas.

1. Introduction

Photovoltaic (PV) solar power generation is affected by numerous environmental and operational factors. The expected energy yield reduction in photovoltaic systems can be assessed based on factors including geographical location, atmospheric dust levels, wind speed and direction, variations in Relative Humidity (RH), temperature, precipitation rates, and dust events. The tilt angle of photovoltaic panels is essential in reducing dust accumulation, as steeper angles enhance the natural cleansing effects of wind and rain. Solar energy is becoming increasingly recognized as one of the most feasible and sustainable ways to address the world's energy crisis. The global installed PV capacity reached more than 1 terawatt (TW) in 2022, meeting almost 2% of the world's power demand [1]. After wind and hydropower, photovoltaics (PV) are the third most capable renewable energy source worldwide [2-4]. Saudi Arabia, the biggest producer and exporter of oil in the world, has a lot of potential for solar energy because

of its location on the Arabian Peninsula. By 2032, the nation's renewable energy sector, headed by Ka-care, had installed 54 gigawatts of solar power, falling short of its goal of 41 GW. In contrast to the global construction of 100 GW of solar power in 2017, Saudi Arabia declared plans in 2018 to install 200 GW by 2030 [5]. In line with the objective of producing 50% of electricity from renewable sources by 2030, the National Renewable Energy Program (NREP) plans to work with ACWA Power to implement 11.8 GW of renewable energy projects by 2025. Despite the abundance of sunlight, the effectiveness of PV systems in desert settings faces unique problems, one of which is dust deposition on the PV panel surface [6-7]. PV systems produce less energy as a result of this condition, which is called soiling. Soiling can have a significant effect on PV performance in areas with regular dust storms, such as Riyadh, NEOM, and Jeddah, resulting in energy losses of up to 20–30%.

Dirt deposition on solar panels can have a variety of repercussions, including reduced sunlight reaching

cells, shadowing, and increased temperature, all of which diminish the solar system's overall efficiency. According to research, the amount of dust on solar panels can have an impact on their power output, open-circuit voltage, and short-circuit current. Experiments in the field have shown that the maximum power of solar panels can drop by about 34% when dust levels rise. Furthermore, on a daily, weekly, and monthly basis, the combination of dust and sunshine can reduce PV performance by 6.24%, 11.8%, and 18.74%, respectively. This is because the solar panel only receives a portion of the light's intensity, which lowers the amount of energy that enters the panel. Dust's effects on solar panel manufacturing and possible effects on solar panel efficiency must thus be addressed. Existing PV cleaning procedures in dry locations such as Saudi Arabia are not optimized. Real-time environmental circumstances are not taken into consideration by the cleaning schedules that are frequently based on predefined intervals or triggered by performance degradation. This results in either seldom cleaning, which wastes energy, or extensive cleaning, which raises operating costs. Predictive algorithms that can suggest more effective cleaning schedules based on historical or real-time environmental data are desperately needed in order to reduce energy losses and operating expenses.

The aim of this study is to develop a machine learning-based algorithm for Riyadh, Saudi Arabia, that utilizes environmental data to create optimal PV panel cleaning schedules. The data for this study are taken from a mix of reliable global and local datasets, i.e., NASA's Power data, the Photovoltaic Geographical Information System (PVGIS), the Air Quality Index (AQI), and local climate stations that monitor dust and weather conditions. The analysis is conducted using a range of machine learning models, including XGBoost, Random Forest, and LSTM, to predict PV performance and recommend cleaning schedules based on these predictions. This study provides a machine learning-based framework to predict PV energy losses due to dust accumulation in arid environments. This study also provides the Integration of environmental and air quality data to optimize PV panel cleaning schedules. The Cost-benefit analysis is performed to determine economically optimal cleaning intervals for large-scale PV installations. The study provides validation of the proposed framework across multiple regions with varying dust patterns.

2. Literature review

2.1 Impact of Soiling on PV Performance

Soiling is the accumulation of material on light-collecting surfaces in solar power systems. Accumulated contaminants can block or scatter incident light, leading to power loss. Common contaminants include mineral dust, bird droppings, fungi, lichens, pollen, engine exhaust, and agricultural emissions. Contamination can affect conventional photovoltaic (PV) systems, concentrated photovoltaic (CPV) systems, and concentrated solar (thermal) power generation systems. However, contamination has a greater impact on concentrated systems than on non-concentrated systems. It is important to note that contamination refers both to the process of its accumulation and to the material itself. Various methods exist to mitigate the impact of contamination. Antifouling coatings are the most important solution in solar power projects. However, due to the lack of antifouling coatings in the past, water washing remains the most widely used technique. Power loss due to contamination varies by region and even within the same region. In areas with frequent rainfall, the average power loss due to contamination can be less than one percent..

There are losses that limit the power that PV modules can provide, i.e., array capture loss, degradation loss, shading, soiling, mismatch, and ohmic losses. If the solar panels aren't cleaned often, dust builds up on the PV surface and makes the power production drop a lot. If the solar PV surface isn't cleaned regularly, dust might build up and cut down on the amount of power it produces. Soiling loss is the name for this effect. Soiling is a major problem that lowers the performance of solar PV modules, which leads to power loss and a big change in the leveled cost of electricity. Dust tends to stick to surfaces that are slanted less than those that are flat. Shading, which is caused by dirt, has a big effect on the electrical properties of PV modules. The PV module cast a shadow because it had a lot of dust on it. For example, the study [11] showed that contaminants in the air and dust from the desert can cause energy losses of up to 20% every day if they are not cleaned up or dealt with in a timely manner. The researcher found that dust is a big influence on both polycrystalline and monocrystalline panels [12]. The efficiency losses are worse at angles that aren't suitable for cleaning. The authors [13] stressed that dirt not only affects the performance of individual panels, but also messes up planning at the grid level in hybrid PV-wind systems.

2.2 Existing Cleaning Strategies

In the early days of AI as an academic discipline, some researchers were interested in having machines learn from data. They attempted to approach the problem with various symbolic methods, as well as what were then termed "neural networks"; these were mostly perceptrons and other models that were later found to be reinventions of the generalised linear models of statistics. The surge in global energy demand has significantly depleted fossil fuel reserves. This is causing critical energy supply issues across various sectors. Integrating renewable energy into national energy mixes has become essential, especially for countries with abundant renewable resources. PV generation has grown exponentially, achieving an installed capacity of approximately 375 GW in 2023. The International Energy Agency anticipates continued growth in the PV sector driven by climate neutrality initiatives, projecting a capacity of around 540 GW by 2028. Photovoltaic (PV) panels can be cleaned in a number of methods, such as using dry cleaning, eco-friendly solutions, automated systems, and by hand. Mixing vinegar and water is a green way to clean, but cleaning by hand requires water and mild soap. Dry cleaning works well in situations where water is hard to get to. The other cleaning methods for solar panels can be divided into 5 groups. Manual tools, mechanized tools, installed hydraulic systems, installed robotic systems, and deployable robots are the methods for solar panel cleaning. The self-cleaning coatings, sprinkler systems, and robotic cleaners are the main types of Automated cleaning technologies. These technologies are used to get rid of dust with very little work. The quality of the water and how often it is cleaned are also very important. Professional cleaning services are available for large installations of PV plants. This is available for those systems that are hard to clean. It is important to turn off the PV system before cleaning to minimize electrical dangers and harsh chemicals or rough materials. The researchers Singh and Chandra [14] used the Helianthus Optimization method with a deep neural network to find the best cleaning times automatically. The authors [15] also came up with a scheduling framework based on reinforcement learning (RL). The framework can be used to adjust cleaning plans, with real-time performance as the basis. The model framework they used utilizes historical performance data and environmental inputs to find the best cleaning method with the lowest cost. The focus of this study is to use optimization algorithms to analyze cleaning costs and energy benefits [16]. The results show that seasonal dust trend weather forecasts can reduce cleaning frequency by

30% without affecting production. As it is a important step toward more sustainable operations.

2.3 ML Applications in Solar Energy Forecasting

Machine learning (ML) is a research area within artificial intelligence that focuses on developing and researching statistical algorithms that can learn from data and generalize to unseen data, thus enabling tasks to be performed without explicit programming language instructions. In a branch of machine learning, advances in deep learning have enabled neural networks (a class of statistical algorithms) to outperform many previous machine learning methods. Statistical and mathematical optimization (mathematical programming) methods form the foundation of machine learning. Data mining is a related research area that focuses on exploratory data analysis (EDA) through unsupervised learning. From a theoretical perspective, near-correct learning provides a mathematical and statistical framework for describing machine learning. Most traditional machine learning and deep learning algorithms can be described as minimizing empirical risk within this framework. As a scientific exploration, machine learning stems from the pursuit of artificial intelligence (AI). In the early days of AI as a discipline, some researchers focused on enabling machines to learn from data. They attempted to solve this problem using various symbolic methods and models then known as "neural networks"; these models were mostly perceptrons and other similar models, which were later found to be essentially reinventions of generalized linear models in statistics. Probabilistic reasoning has also been applied, particularly in the field of automated medical diagnosis.

Machine learning algorithms are increasingly being used to forecast solar energy. These models are more accurate and reliable than older methods. The older models were self-cleaning using brushes, squeegees, mops, and hoses. Another old cleaning model includes Semi-Automated Cleaning Systems, which are self-operating robots that clean panels without human intervention. The machine learning models, other than old models, anticipate how much solar energy will be generated in the future based on things like prior data, weather patterns, and more. This helps with managing the grid, making it work better, and making smart choices. Neural networks (NN), Long Short-Term Memory (LSTM), and Random Forest are examples of machine learning algorithms that learn from large collections of past weather and solar radiation data. They identify links in the facts that are intricate and not straightforward. This allows them to adjust how the system functions and use

energy in the greatest way by precisely anticipating solar power; power system operators can better deal with the fact that solar energy is unpredictable.

They require a lot of information about past solar radiation, weather, and other important factors in order to be able to forecast solar power. These kinds of data are necessary for machine learning models, such as ANN, LSTM, SVR, KNN, and other hybrid models, to make precise predictions. These datasets must be thoroughly cleaned and preprocessed in order to ensure accurate and high-quality predictions. Significant advancements in building energy scheduling and energy savings were discovered by the researchers of [17]. This implies that ML might be applied outside of utility-scale installations. In order to monitor and enhance solar energy systems in remote and arid regions, the authors of [18] proposed a hybrid ML-IoT framework that makes use of sensor data from the ground and cloud-based analytics. Utilized in Sub-Saharan Africa, the technology demonstrated that integrating environmental and soiling data in real time could help make maintenance decisions in advance and increase efficiency by over 18%.

The authors of [19] discussed how AI is becoming more important in connecting energy forecasts with systems for desalination, demand response, and predictive maintenance. The researches of [20] used numerical analysis methods improved by machine learning to help make Maximum Power Point Tracking (MPPT) algorithms better. Their work is mostly about control systems. But their works is mostly fits with the bigger trend of using machine learning. Recent studies have further explored ML-driven cleaning optimization and predictive maintenance by demonstrating improved PV output recovery and reduced operational costs [21-23]. Study [24] reported recent progress in transparent self-cleaning coatings for glass-panel applications, particularly in the photovoltaic (PV), automotive, and building-glass industries. The work highlighted how advances in surface wettability have significantly contributed to the development of self-cleaning properties, leading to growing industrial interest. It also argued that conventional cleaning approaches, such as mechanical and robotic cleaners, may not always be practical because they can damage glass surfaces, interrupt electricity generation during heavy rain, and involve high equipment costs.

According to [25], photovoltaic power generation has become one of the most important technologies for achieving energy-saving and carbon-reduction goals. Using the Web of Science database as its primary source, the study presented a comprehensive review of dust deposition mechanisms on PV panels, the factors

influencing soiling, power-loss prediction models, cleaning methods, and dirt-monitoring systems. The findings showed that dust accumulation can cause significant reductions in module power output, with losses ranging from 7% to 98.13% depending on the region. The study also found that automatic cleaning robots, as an emerging intelligent solution, can improve cleaning performance and raise PV efficiency by as much as 49.53%. In addition, [25] proposed an integrated dust-mitigation strategy that combines real-time monitoring of dust accumulation, predictive loss modeling, and optimized cleaning methods. The study further emphasized the importance of intelligent cleaning technologies, including robots and drone-based systems, and recommended wider use of artificial intelligence in PV monitoring and cleaning to support progress toward carbon neutrality.

Study [26] examined the effect of dust particles on the performance of photovoltaic solar panels through an experiment conducted at the University of Technology in Baghdad from March 15 to May 15, 2024. The results showed a substantial decline in PV conversion efficiency, dropping from 19.8% to 6.3% at a constant irradiance of 750 W/m². The study also found that, at a given particle density, conversion efficiency remained constant, while the overall relationship between particulate density and PV efficiency was nonlinear. More broadly, dust deposition was reported to reduce solar panel efficiency by 7.8% to 19.2%, while cleaning improved effectiveness by 6.8% to 7.9%. The study further noted that wind and rain can positively affect PV performance by removing accumulated dust from panel surfaces.

Similarly, [27] identified meteorological conditions as a major factor affecting solar photovoltaic systems, with dust being the most influential. The study investigated the impact of dust deposition on PV performance, particularly its role in reducing incoming radiation and operating output power, and it also sought to determine the optimal cleaning interval. The reported average mass concentration of dust in the samples was 187.63 ug/m³.

3. Methodology

3.1 Data Sources and Preprocessing

This study combines data from many different sources to create a full dataset. This dataset is used to look at how photovoltaic (PV) systems work in different weather conditions. Global Horizontal Irradiance (GHI), Ambient Temperature, Wind Speed, Humidity, Dust (AOD), Air Pollutants, PV output, and system losses are the main components of this dataset.

The data for this study were collected from NASA satellites, PVGIS, PM10, PM2.5, and AOD. The NASA satellite dataset for this study is used for solar radiation and meteorological parameters. The Photovoltaic Geographical Information System (PVGIS) is used for PV system performance benchmarks. Local Air Quality data is used for PM10 and PM2.5. The dataset spans from January 2025 to July 2025 with hourly measurements. The data have missing values. The missing values were handled using our proposed algorithm. Feature engineering is also included in this study, which includes lag features, rolling averages, and a derived soiling rate variable combining dust concentration and humidity.

Table 1

Summary of Collected Dataset Features and Units

Feature Name	Description	Unit	Source
Timestamp	Date and time of measurement	YYYY-MM-DD HH:MM	All sources
GHI	Global Horizontal Irradiance	W/m ²	NASA, PVGIS
DNI	Direct Normal Irradiance	W/m ²	NASA, PVGIS
Ambient Temperature	Surrounding air temperature	°C	NASA, PVGIS
Relative Humidity	Moisture level in air	%	NASA
Wind Speed	Wind velocity	m/s	NASA
PM10	Particulate matter concentration (dust proxy)	µg/m ³	AQI Data Portals
PV Output (Raw)	Actual energy output from the PV system	kWh	On-site Data Logs
Performance Ratio (PR)	Output normalized to theoretical potential	%	Calculated
Energy Loss	Reduction from ideal output	%	Calculated

From the dataset, Table 1 provides features, unit and data sources from which the data is taken. After collecting data across all the resources, the data cleaning is the second step in our framework. Duplicate entries and variations in time format are removed in this step during this cleaning procedure. This study's target variable, the system's energy loss percentage, allowed for the examination of how performance declined over time.

3.2 Machine Learning Methodology

Random Forest (RF), XGBoost, and Long Term Memory (LSTM) models were selected for analysis in this study. These models were chosen to combine the advantages of clustering tree models, including RF and XGBoost, with a deep learning model (LSTM). This model is capable of detecting temporal or sequential dependencies, which are often found in time-sensitive and event-driven data. The preprocessed dataset containing derived MM4 values was split using an 80 into 20 split for training and testing. This make sure that the model has sufficient data for learning and generalization.

Feature	AOD (Dust)	Temperature (°C)	Humidity (%)	PV Output (kWh)
AOD (Dust)	1.00	-0.02	0.05	-0.04
Temperature (°C)	-0.02	1.00	0.09	0.10
Humidity (%)	0.05	0.09	1.00	-0.15
PV Output (kWh)	-0.04	0.10	-0.15	1.00

Fig. 1. Heatmap of Feature Correlation (Dust, Temp, Humidity, Output)

Figure 1 shows a negative correlation between photovoltaic power output and dust at -0.04. This shows that higher dust concentrations reduce power output. The heat map shows a more negative correlation with photovoltaic power output at -0.15. This states that increased atmospheric humidity can significantly weaken solar performance. Temperature shows a weak positive correlation at 0.10. This states that the higher temperatures may improve power output. In real-world conditions, the excessive heat can also reduce panel efficiency. Random Forest and XGBoost models were chosen for their ability to handle nonlinear relationships. These models have the capacity to determine the significance of features based on clustering for future predictions. The LSTM model was chosen to capture temporal dependencies in time-series data.

4. Results

4.1 Model Accuracy

The classification accuracy of machine learning models can be validated using accuracy evaluation techniques. The holdout method splits the data into training and test sets (typically 2/3 training and 1/3 test), and evaluates the model's performance on the test set. In contrast, K-fold cross-validation randomly divides the data into K subsets, performs K trials, and uses one subset for evaluation in each trial, while the remaining K-1 subsets are used to train the model.

Besides holdout and cross-validation, bootstrap can also be used to evaluate model accuracy. Bootstrap draws n samples with replacement from the dataset. The model accuracy is an important element in any study. Here, in this case, predicting PV output loss is the most important element of this study. XGBoost, Random Forest, and LSTM are compared in this study. This comparison can be seen in both graphical and quantitative evaluation metrics in Table 2 and the figure given below. Model performance was evaluated using R^2 , which is resulting in superior accuracy ($R^2 = 0.93$) to assess the proportion of variance explained by predictions.

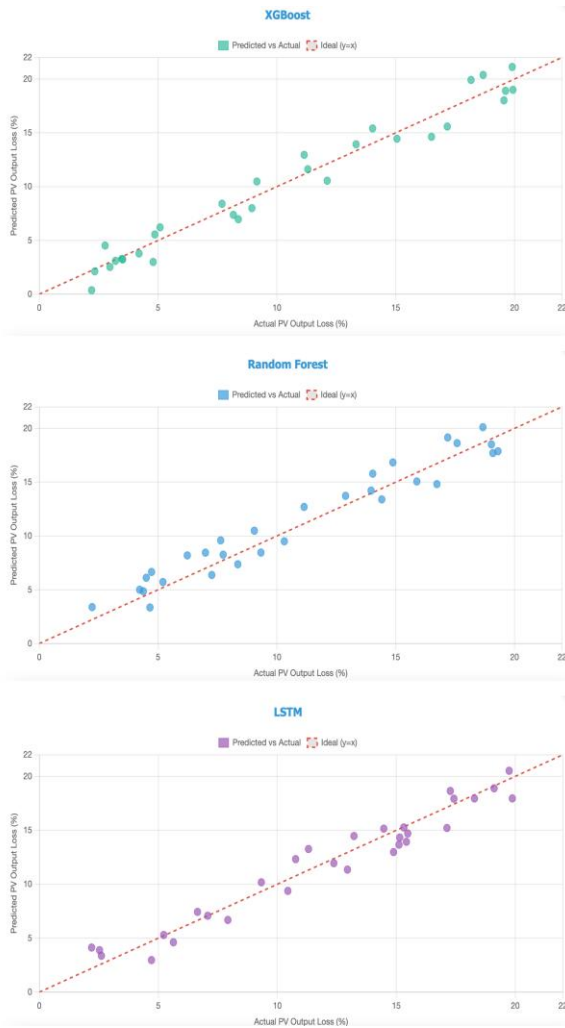


Fig 2. Predicted vs Actual PV Output Loss (Model Performance)

The scatter plots in Figure 2 show predicted losses against actual losses. This indicates perfect accuracy. The closer the data points lie to this line, the better the model predictions. The XGBoost model showed a tight clustering for actual losses near 10%. With predicted values ranging narrowly between 9% and 11%. This figure suggests a low prediction error and minimal bias. The predictions vary from 7% to 13% for a single true value. The spread of points around the diagonal is also indicative of the variance of the model predictions.

Table 2

Comparison of Model Performance (R^2 , RMSE, MAE)

Model	R^2	RMSE (%)	MAE (%)
XGBoost	0.93	1.32	0.84
Random Forest	0.91	1.48	0.96
LSTM	0.87	1.72	1.10

Table 2 showed quantitative metrics, root mean square error, and mean absolute error. XGBoost model perform better the others with an R^2 of 0.93. This indicates it explains 93% of the variance in actual energy losses. It also achieved the lowest RMS as 1.32% and the mean absolute error as 0.84%. This confirms its predictions are both accurate and consistent. The Random Forest model follows closely with an R^2 of 0.91, an RMSE of 1.48%, and an MAE of 0.96%. The LSTM model is capable of modeling time dependencies. This trails in accuracy with an R^2 of 0.87. RMSE of 1.72% and MAE of 1.10%.

4.2 Feature Importance Analysis

Feature importance is a technique that determines the contribution of each feature in a dataset toward a machine learning model's predictions. Figures 3 and 4 provide a comparative visualization of feature importance. This is determined by two prominent machine learning models used in predicting photovoltaic (PV) output loss.

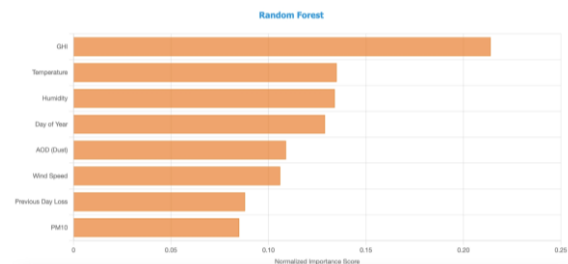


Fig. 3. Feature Importance Analysis results for Random Forest.

Each horizontal bar chart highlights the relative importance of input features. In both models, Figure 3-4, longer bars signify a higher contribution of that feature toward making accurate predictions.

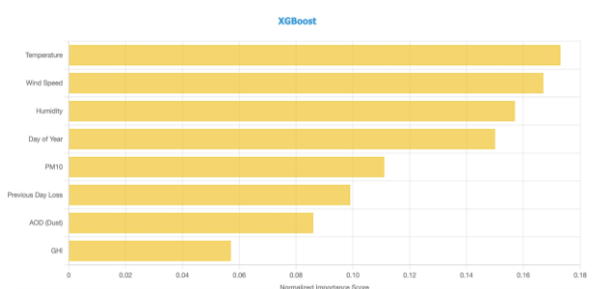


Fig. 4. Feature Importance Analysis results for XGBoost

XGBoost might emphasize fine-grained interactions and rank variables higher due to its gradient boosting mechanism. Random Forest may favor broader trends, which can be seen in the results. This highlights temperature, wind speed, and humidity.

4.3 Time Series Impact of Dust and Cleaning Intervals

The time series analysis highlights an inverse relationship between dust concentration (AOD) and PV energy output. The time series plot (Figure 5) shows that higher dust levels lead to reduced solar efficiency. Figure 5 also shows a time-series visual representation of daily PV energy output alongside dust concentration levels. This visual representation shows Aerosol Optical Depth (AOD) over a defined period ranging from January 2022 to January 2023.

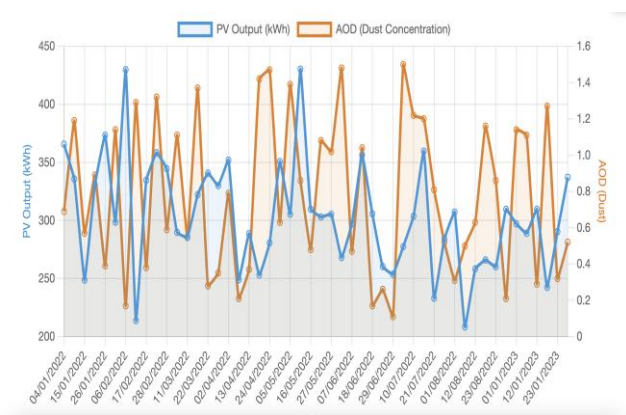


Fig. 5. Time-Series Plot: Daily Energy Output vs Dust Concentration

Figure 5 shows a notable lag post-January 10. This lag shows AOD spiked to 1.19. PV output declined gradually from 335.7 kWh to 248.7 kWh over five days. PV output showed a quick recovery following periods of low AOD. These findings from the above figure underscore the critical importance of regular maintenance. This also optimized cleaning intervals to counteract dust-related efficiency losses in solar energy systems.

4.4 Visualization of Cleaning Optimization

Figure 6 shows a cleaning optimization curve. This curve offers a visual depiction of the theoretical relationship between the frequency of photovoltaic (PV) panel cleaning. The figure shows that monthly cleaning jumps significantly from 0% to 10%. This also shows a reach of 14% with biweekly intervals. Decreasing returns are indicated by the curve. The increase to 16.5% is due to the weekly cleaning. A gain of 17.5% is achieved with every three-day cleaning. The maximum gain that can be achieved with daily cleaning is 18%.

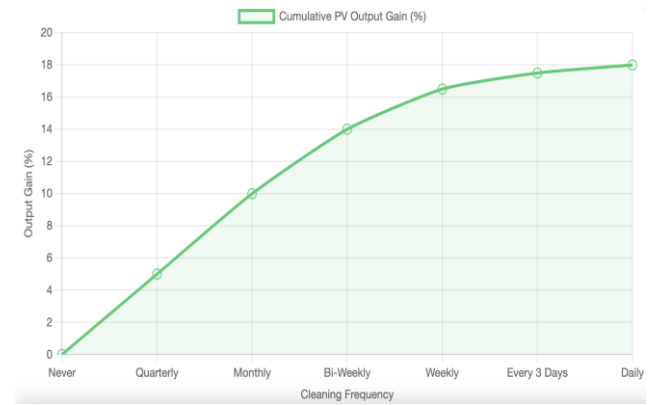


Fig. 6. Cumulative PV output gain

Table 3 provides a estimate of the impact of pre-cleaning and post-cleaning on the soiling rate. Table 3 reinforces this in a synthetic scenario. It shows that the pre-cleaning soiling may reduce energy output by up to 20%. Post-cleaning intervals significantly help recover the lost energy.

Table 3

Soiling Rate Estimation Pre- and Post-Cleaning (Synthetic and Real Cases)

Cleaning Frequency	Cleaning Cost (Units)	Energy Revenue Loss (Units)	Total Effective Cost (Units)
Never	0	20	20
Quarterly	3	16	19
Monthly	6	12	18
Bi-weekly	8	5	13
			(Optimal)
Weekly	12	3	15
Every 3 Days	16	2.5	18.5
Daily	20	2	22

4.5 Cost-Energy Trade-Off Simulation

Figure 7 provides a visual representation of the cost-benefit balance associated with different solar panel cleaning frequencies. The simulation showed a clear inverse relationship between cleaning frequency and energy revenue loss. The lower the losses due to dirt build, the significantly higher the cleaning frequency. This improvement in energy production comes at the cost of higher operating expenses.



Fig. 7. Cost-Benefit Graph: Cleaning Cost vs Energy Revenue Loss

Table 4 shows a comparative breakdown of costs and energy revenue losses at different cleaning frequencies. The daily cleaning minimizes energy loss to just 2%. This Table 4 shows that the biweekly interval achieves a favorable balance and moderate cleaning cost. This substantially reduced energy loss. This resulting in the lowest total effective cost.

Table 4

Cost vs Benefit Simulation

Cleaning Frequency	Cleaning Cost (Units)	Energy Revenue Loss (Units)	Total Effective Cost (Units)
Never	0	20	20
Quarterly	3	16	19
Monthly	6	12	18
Bi-weekly	8	5	13
			(Optimal)
Weekly	12	3	15
Every 3 Days	16	2.5	18.5
Daily	20	2	22

5. Proposed Optimization Framework

5.1 Cleaning Trigger Threshold Logic

This paper proposes an optimization framework based on a cleaning threshold logic. This framework enables real-time scheduling of photovoltaic panel cleaning using historical and future data, achieving dynamic response. Location, season, and historical performance data can all be used to adjust the threshold. Thus, cleaning is only initiated when it is economically feasible.

5.2 Integration of Dust Index Forecasting

To make the framework more proactive, in addition to the cleaning trigger logic, a module for predicting dust indexes has been integrated. This module can predict dust storms and persistent droughts. To save time and effort, cleaning operations can be postponed until after an event occurs. Cleaning operations can be performed in advance when panel performance is expected to improve. This predictive cleaning mechanism minimizes operating costs. Weather is unpredictable, but this predictive cleaning mechanism can handle dust accumulation patterns in arid and semi-arid regions throughout the year.

5.3 Decision-Making Workflow

The proposed intelligent photovoltaic cleaning optimization system operates on a decision-making

workflow. This workflow first collects and processes input data in real time. The proposed optimization framework utilizes data from our designed dataset, which we have previously introduced; this dataset contains detailed information about weather and other variables. Duplicates in the data are then removed. After cleaning, machine learning algorithms are applied to analyze the data. This is the model's fouling loss prediction algorithm. An improved version of the framework is then proposed, capable of predicting energy output with and without cleaning intervention. With its predictive capabilities, the system can simulate potential outcomes under various conditions. Guided by this model, we can make better decisions.

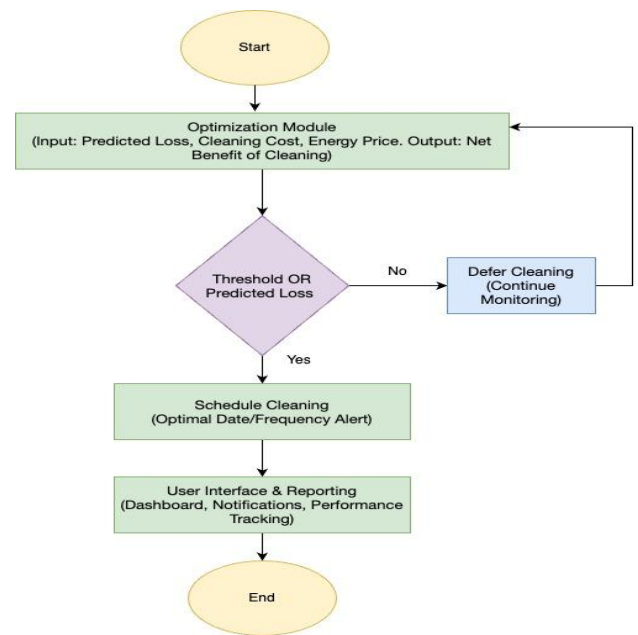


Fig. 8. Proposed ML-Based Cleaning Scheduler Flowchart

Figure 8 flowchart shows that the data is passed to the optimization module. This optimization module then calculates the total economic benefit of performing cleaning actions. The factors used in this module are estimated energy recovery, cleaning costs, and current energy prices. The proposed system algorithm will then trigger a cleaning alert to the monitoring system. This will continue monitoring based on this logic. A user interface and reporting dashboard visualize ongoing performance metrics. This provides alerts and logs historical cleaning efficiency. This is all done as a part of the feedback loop from the proposed framework and flowchart.

5.4 Regional Scheduling Adaptation

Regional adjustments to photovoltaic module cleaning schedules are important for maximizing energy output under varying environmental conditions. This study further includes two other cities in Saudi Arabia: Dammam and New City (NEOM).

Dammam is the capital of Saudi Arabia's Eastern Province and an industrial port city. While New City is an eco-city under development in Saudi Arabia's Tabuk Province. Figure 9 compares the optimal cleaning intervals across multiple regions. The results show that the recommended cleaning interval decreases with increasing dust concentration. Based on the results, cleaning every 28 days is recommended under low aerosol optical thickness (AOD) conditions. However, when dust indices rise to moderate and high levels, the optimal cleaning intervals shorten to 14 days and 7 days, respectively. This indicates that more frequent maintenance is required during dusty periods. This also helps prevent a significant decrease in photovoltaic module output power due to dust accumulation.

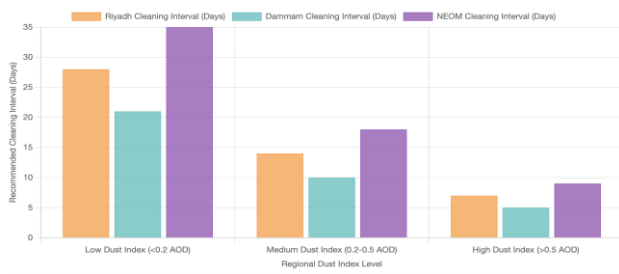


Fig. 9. Regional Cleaning Intervals by Dust Index (Illustrative)

Figure 9 illustrates regional differences. These differences suggest that cleanup periods should not be uniformly applied. Even at similar dust index levels, Riyadh may require a 14-day cleanup period at the average aerosol optical thickness level. The cleanup period in the NEOM area may extend to 18 days due to potentially lower pollution levels. This is influenced by local dust particle composition, weathering factors, and surface reflectivity. These conclusions are based on data provided in the dataset. Site-specific recommendations should be incorporated into the cleanup planning logic.

6. Conclusion

This study demonstrates that machine learning can optimize the cleaning scheme for photovoltaic (PV) panels in the arid climate of Riyadh, Saudi Arabia. To address the major issue of efficiency loss due to dust accumulation, the proposed framework utilizes environmental data and predictive models. The study found that XGBoost and Random Forest algorithms can adjust the maintenance schedule based on actual and anticipated conditions. The finding that PV power generation decreases with increasing dust concentration is significant, as system efficiency decreases at higher dust concentrations. Time series analysis indicates that timely cleaning intervention

can restore system performance to normal. However, over-cleaning reduces power generation. Cost-energy tradeoff simulations suggest an optimal cleaning cycle of once every two weeks. This reduces operating costs and energy waste. Due to varying dust accumulation patterns in different regions, unique cleaning plans are required for each area. The ideal cleaning intervals differ between Riyadh, Dammam, and New City (NEOM). This paper proposes a cleanroom cleaning scheduling system using machine learning techniques. To aid decision-making before actual maintenance is needed, dust prediction is combined with performance prediction and cost-benefit analysis. This approach not only increases energy output but also reduces unnecessary cleaning costs. Therefore, it is well-suited for large-scale solar PV projects. Large-scale solar power plants can benefit from our recommended cleanroom scheduling system. To achieve this, cleaning intervals are adjusted based on dust forecasts, reducing energy waste, labor costs, and water consumption. These results confirm the applicability of smart maintenance strategies in arid regions globally. With future advancements in Internet of Things (IoT) integration and real-time analytics, these models can be further refined, enabling photovoltaic systems to self-regulate and self-optimize. By demonstrating how data science and artificial intelligence can reduce the environmental impact of solar power generation, this research contributes to the broader goal of sustainable energy management and helps achieve a more efficient and cleaner energy transition globally.

7. References

- [1]. Kumar Dora, Bimal, Sunil Bath, Arghya Mitra, Damien Ernst, Adrian Halinka, Daria Zychma, and Pawel Sowa. "The Global Electricity Grid: A Comprehensive Review." *Energies* 18, no. 5 (2025).
- [2]. Saavedra, Alexander, Nataly Alejandra Galvis, Fredy Mesa, Edison Banguero, Mónica Castaneda, Sebastian Zapata y Andrés Julián Aristizábal. "Current state of the worldwide renewable energy generation: A review." *International Journal on Engineering Applications* 9, no. 3 (2021): 115-127.
- [3]. Perea-Moreno, Alberto-Jesús. "Renewable energy and energy saving: Worldwide research trends." *Sustainability* 13, no. 23 (2021): 13261.
- [4]. Bigerna, Simona, Carlo Andrea Bollino e Paolo Polinori. "Convergence in renewable energy sources diffusion worldwide." *Journal*

- of Environmental Management 292 (2021): 112784.
- [5]. Wikipedia Contributors "Solar power in Saudi Arabia" Available on https://en.wikipedia.org/wiki/Solar_power_in_Saudi_Arabia Accessed on 29 May 2025
 - [6]. Sumayli, Hisham, Abdelrahman El-Leathy, Syed Noman Danish, Hany Al-Ansary, Zeyad Almutairi, Zeyad Al-Suhaibani, Nader S. Saleh et al. "Integrated CSP-PV hybrid solar power plant for two cities in Saudi Arabia." *Case Studies in Thermal Engineering* 44 (2023): 102835.
 - [7]. Alqahtani, Nasser, and Nazmiye Balta-Ozkan. "Assessment of rooftop solar power generation to meet residential loads in the city of Neom, Saudi Arabia." *Energies* 14, no. 13 (2021): 3805.
 - [8]. Olorunfemi, Benjamin Oluwamuyiwa, Omolola A. Ogbolumani, and Nnamdi Nwulu. "Solar panels dirt monitoring and cleaning for performance improvement: a systematic review on smart systems." *Sustainability* 14, no. 17 (2022): 10920.
 - [9]. Olorunfemi, Benjamin O., Nnamdi I. Nwulu, and Omolola A. Ogbolumani. "Solar panel surface dirt detection and removal based on Arduino color recognition." *MethodsX* 10 (2023): 101967.
 - [10]. Vedula, Gowtham, Anbazhagan Geetha, and Ramalingam Senthil. "Review of strategies to mitigate dust deposition on solar photovoltaic systems." *Energies* 16, no. 1 (2022): 109.
 - [11]. Rashid, Haseeb, Liu Ming Hua, Lyu Guanghua, Rija Hasan, Abdulrahman AlKaseem, Aqsa Ali, Syed Hadi Hussain Shah, Shoaib Shaikh, Arsalan Muhammad Soomar, and Piotr Musznicki. "A comprehensive review on economic, environmental impacts, and future challenges for photovoltaic-based electric vehicle charging infrastructures." *Frontiers in Energy Research* 12 (2024): 1411440.a
 - [12]. Hua, Lyu-Guang, Qasir Ali Memon, Muhammad Fawad Shaikh, Shoaib Ahmed Shaikh, Rehan Ali Rahimoon, Syed Hadi Hussain Shah, and Abdul Qadir. "Comparative analysis of power output, fill factor, and efficiency at fixed and variable tilt angles for polycrystalline and monocrystalline photovoltaic panels The case of Sukkur IBA University." *Energies* 15, no. 11 (2022): 3917.
 - [13]. Soomar, Arsalan Muhammad, Lyu Guanghua, Shoaib Shaikh, Syed Hadi Hussain Shah, and Piotr Musznicki. "Scrutiny of power grids by penetrating PV energy in wind farms: a case study of the wind corridor of Jhampir, Pakistan." *Frontiers in Energy Research* 11 (2023): 1164892.
 - [14]. Singh, Upendra Pal, and Subhash Chandra. "Optimal cleaning schedule in solar PV using the biology-based helianthus optimization enabled coupled deep network." *Solar Energy* 266 (2023): 112138.
 - [15]. An, Heungjo. "Reinforcement learning-based dynamic cleaning scheduling framework for solar energy system." *한국지능시스템학회 논문지* 35, no. 1 (2025): 84-97.
 - [16]. Shboul, Bashar, Mohamed E. Zayed, Waqar Muhammad Ashraf, Muhammad Usman, Dibyendu Roy, Kashif Irshad, and Shafiqur Rehman. "Energy and economic analysis of building integrated photovoltaic thermal system: Seasonal dynamic modeling assisted with machine learning-aided method and multi-objective genetic optimization." *Alexandria Engineering Journal* 94 (2024): 131-148.
 - [17]. Houchati, Mahdi, AbdelMonem H. Beitelmal, and Marwan Khraisheh. "Predictive modeling for rooftop solar energy throughput: a machine learning-based optimization for building energy demand scheduling." *Journal of Energy Resources Technology* 144, no. 1 (2022): 011302.
 - [18]. Kerama, Benjamin Nyabera. "A Novel Hybrid Machine Learning-IoT Framework for Optimizing Solar Energy Efficiency in Arid Regions: A Case Study of Sub-Saharan Africa." *International Journal of Innovative Science and Research Technology* 10, no. 4 (2025): 3730-3736.
 - [19]. Jathar, Laxmikant D., Keval Nikam, Umesh V. Awasarmol, Raviraj Gurav, Jitendra D. Patil, Kiran Shahapurkar, Manzoore Elahi M. Soudagar et al. "A comprehensive analysis of the emerging modern trends in research on photovoltaic systems and desalination in the era of artificial intelligence and machine learning." *Heliyon* 10, no. 3 (2024).

- [20]. Guanghua, Lyu, Farah Andleeb Siddiqui, Muhammad Mohsin Aman, Syed Hadi Hussain Shah, Aqsa Ali, Arsalan Muhammad Soomar, and Shoaib Shaikh. "Improved maximum power point tracking algorithms by using numerical analysis techniques for photovoltaic systems." *Results in engineering* 21 (2024): 101740.
- [21]. Howlader, Abdus Salam. "Machine learning applications in renewable energy: Predictive analytics for solar cell performance optimization and energy yield forecasting." *American Journal of Scholarly Research and Innovation* 4, no. 01 (2025): 392-427.
- [22]. Behl, Devansh, Anuradha Tomar, Hritik Kumar, and Satvika Sharma. "Solar Optimization via Learning-driven Visual Intelligent Analytics." *Renewable Energy* (2026): 125171.
- [23]. Al-Humairi, Ali, Enmar Khalis, Zuhair A. Al Hemyari and Peter Jung. "The Impact of Data Augmentation on AI-Driven Predictive Algorithms for Enhanced Solar Panel Cleaning Efficiency." *Processes* 13, no. 4 (2025): 1195.
- [24]. Syafiq, Amirul, Vengadaesvaran Balakrishnan, Mohd Syukri Ali, Sanjay J. Dhoble, Nasrudin Abd Rahim, Azimah Omar, and Ab Halim Abu Bakar. "Application of transparent self-cleaning coating for photovoltaic panel: a review." *Current Opinion in Chemical Engineering* 36 (2022): 100801.
- [25]. Wan, Letao, Liqian Zhao, Wensheng Xu, Feihong Guo, and Xiaoxiang Jiang. "Dust deposition on the photovoltaic panel: A comprehensive survey on mechanisms, effects, mathematical modeling, cleaning methods, and monitoring systems." *Solar Energy* 268 (2024): 112300.
- [26]. Abd, Hussain Saad, Hadia Kadhim Judran, Salah H. Abid Aun, Ameer Abed Jaddoa, Karrar A. Hammoodi, Saif Ali Kadhim, and Jafaar Mohammed Daif. "Dust deposition and cleaning effect on PV panel: Experimental approach." *Results in Engineering* 27 (2025): 10604.
- [27]. Fatima, Kulsoom, Ahmad Faiz Minai, Hasmat Malik y Fausto Pedro García Márquez. "Experimental analysis of dust composition impact on Photovoltaic panel Performance: A case study." *Solar Energy* 267 (2024): 112206