Solar Optimization via Learning-driven Visual Intelligent Analytics

**Abstract**

Airborne particulate matter settling on photovoltaic (PV) modules significantly reduces energy output by inducing efficiency losses, necessitating an intelligent and cost-effective cleaning strategy. This paper introduces SOLVIA—**Solar Optimization via Learning-driven Visual Intelligent Analytics**—a comprehensive predictive cleaning framework originally designed for university campus PV installations but scalable to broader geographies. SOLVIA integrates real-time environmental monitoring, computational modeling, and economic analysis to determine optimal cleaning intervals. It leverages a robust sensor network comprising optical soiling sensors, PM2.5/PM10 particulate detectors, irradiance, temperature, and wind speed sensors to continuously monitor conditions contributing to dust accumulation. A regression-based model correlates this real-time data with historical performance to predict power losses from soiling. Unlike rigid, fixed-schedule cleaning routines, SOLVIA dynamically recommends cleaning actions based on a cost-benefit threshold, only initiating maintenance when expected energy losses surpass the associated cleaning costs. The system further includes a real-time visualization interface built with modern web technologies, displaying dust accumulation patterns, predicted energy and financial losses, and recommended cleaning schedules. Through intuitive dashboards and automated alerts, SOLVIA empowers stakeholders to make informed maintenance decisions, reduce operational costs, and maximize energy yield. Its modular architecture allows seamless integration with advanced machine learning models and autonomous robotic cleaning systems, making it a scalable and intelligent solution for smart solar energy management.

**Index Terms**—Photovoltaic Systems, Soiling Loss, Predictive Maintenance, Machine Learning, Dust Accumulation Forecasting, Web-Based Visualization, Smart Cleaning Systems.

1. **Introduction**

Photovoltaic (PV) systems have become fundamental to renewable energy solutions; however, their performance is significantly affected by soiling—the accumulation of dust and particulate matter on panel surfaces. Recent studies highlight the importance of predictive modeling, machine learning (ML) techniques, and optimized maintenance strategies in mitigating these effects and improving operational efficiency.

The issue of dust accumulation has been thoroughly examined in research that showcased specific soiling patterns based on location and seasonal variability. For instance, dust analysis in arid and semi-arid regions showed significant degradation due to the mineral composition of accumulated particles [[1](#one),[2](#two)]. The research by Mani and Pillai [[3]](#three) reported a 40% power loss caused by dust in desert environments, indicating a strong correlation between the tilt angle and the accumulation of dust. In Egypt, the analysis of particles showed that quartz, calcite, and gypsum were the primary components of soiling, which negatively affected PV efficiency due to optical degradation [[4]](#four). Likewise, research conducted at the Pasir Mas Solar Farm in Malaysia revealed heterogeneous particle sizes predominantly composed of SiO2 and Al2O3, with thermal imaging revealing that 19.91% of the strings were affected by hotspots [[5]](#five).

Optimal cleaning strategies have been a central concern. A study from Zimbabwe developed a soiling loss model and proposed a 15-day optimal cleaning cycle utilizing Particle Swarm Optimization (PSO) [[6]](#six). In Mesa, a cleaning frequency study found cleaning economically unjustified for residential PV modules at tilt angles above 20° [[7]](#seven). A Spain-based research integrated rainfall and environmental factors into a predictive model aimed at optimizing cleaning schedules for large-scale PV systems, resulting in a deviation of only 0.71% between the model predictions and actual sensor data [[8]](#eight). Another Spanish study presented a soiling forecasting model intended for PV plants with capacities exceeding 200 MW [[9]](#nine).

The integration of smart monitoring systems and real-time data acquisition has led to advanced decision-support frameworks. The Photovoltaic Soiling Monitoring System (PVSMS) was proposed as a low-cost, scalable system to evaluate soiling and suggest cleaning schedules. Field validation and a feasibility assessment at a mining site confirmed its practical utility [[10]](#ten10). Similarly, predictive modeling in India made use of machine learning techniques, with linear regression yielding an R2 score of 0.99994 in solar forecasting [[11]](#eleven).

A detailed breakdown of some of these impactful studies is provided in the following Table 1.

**Table 1 Comparison of ML model studies**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ref. No. | Study Location | Key Methodology | Model/Tool Used | Key Outcome |
| [9](#nine) | Spain | Soiling Forecasting Model | Environmental data integration | Informed optimal cleaning schedule for >200 MW PV plant |
| [11](#eleven) | India (Dhar) | ML & DL Model Comparison | 10 ML & 1 DL Algorithm | High accuracy forecasting across all models |
| [12](#twelve12) | UAE & Jordan | Predictive ML Models | SARIMAX, RF, GBM, Logistic Regression | R2 of 93.36%, cleaning classification accuracy of 91% |
| [13](#thirteen) | India | Ensemble ML & Time-Series Models | SARIMAX, RF, Voting Classifier | Accurate PR prediction and cleaning schedule optimization |
| [14](#fourteen14) | Egypt | Field study + ML forecasting | Regression models, feature selection | Identified rainfall, temperature as strong predictors |
| [15](#fifteen15) | India (Dhar) | Statistical & ML performance forecasting | Linear Regression, DL models | R2 of 0.99994, RMSE of 0.121 |
| [16](#sixteen16) | UAE & Jordan | Predictive PR-based cleaning schedule | SARIMAX, Ensemble classifiers | Performance-tailored maintenance models |

Material composition and environmental conditions significantly influence the impact of soiling on PV system performance. Laboratory tests simulating various dust types showed charcoal causing up to 98% degradation in short-circuit current, while salt caused only 7% [[17]](#seventeen17). Analysis in the UK correlated fine particulate accumulation with a 5–11% light transmittance drop impacted significantly by rain, humidity, and bird droppings [[18]](#eighteen18). Long-term research in Australia, spanning 18 years, concluded that dust accounted for 16–29% of performance degradation, with quartz and feldspar as the dominant elements [[19]](#nineteen19).

Additional predictive models have utilized hybrid approaches like (Particle Swarm Optimization paired with Random Forest) PSO-RF, achieving more than 94% accuracy in forecasting output power [[20]](#twenty20). These methods are crucial for minimizing uncertainty and supporting grid-level scheduling. The economic implications of soiling are substantial. In India, an ML model forecasted cleaning benefits by integrating irradiance forecasts and economic parameters. Cleaning once every ten days during high-dust seasons was proposed as a trade-off between the cost of cleaning and the energy gain.

Besides predictive modeling, effective system rollout is also based on integration of ML into back-end infrastructures. A recent publication emphasizes the challenge and opportunity with such integration, commenting on the requirement for strong data pipelines, scalable design, and model interpretability for effective real-time performance [[21]](#twenty1). This directly translates to the creation of smart PV monitoring frameworks, where integration of ML into back-end assists continuous analysis, forecasting, and operational decision-making.

In addition, AI and ML developments have revolutionized web applications from being passive interfaces to being interactive platforms that actively facilitate decision-making. Real-time vulnerability detection and system response is now possible thanks to the incorporation of ML into web applications [[22]](#twenty2). This feature also exists in modern PV monitoring dashboards, allowing for simple access to financial data, cleaning schedules, and forecasts, enhancing system usability and stakeholder engagement.

Several studies have highlighted the importance of ML and sensor-driven decision-making tools as essential facilitators. Techniques such as neural networks, fuzzy logic controllers, and decision trees have been commonly utilized to predict the need for cleaning and enhance energy output [[23-25]](#from23to25). Approaches like ensemble learners and deep learning techniques have shown superior performance compared to traditional statistical methods for cleaning prediction tasks [[12](#twelve12),[13](#thirteen),[25](#twenty5)].

As solar energy use increases, the integration of predictive analytics, environmental information, and adaptable cleaning methods becomes crucial for the sustainability and economic viability of PV operations. Future investigations should concentrate on developing region-specific AI models, integrating with autonomous cleaning technologies, and implementing real-time performance evaluations to ensure scalability in various regions.

Despite the breadth of existing research exploring the impact of soiling on PV performance and the application of machine learning for predictive modeling, several critical gaps remain unaddressed. While many studies have validated the effectiveness of AI and statistical models in forecasting soiling loss and optimizing cleaning schedules, there is a noticeable lack of an integrated, end-to-end, user-centric framework that is both technically robust and intuitively accessible. Existing models are often developed in siloed research environments with limited transition into deployable solutions, resulting in a clear disconnect between technological advancements and practical usability. No existing system currently offers a comprehensive, easy-to-use platform that integrates real-time inference, intelligent scheduling, economic evaluation, and interactive visualization into a seamless user experience. Moreover, most frameworks ignore seasonal adaptability, failing to account for the variability in soiling behavior across monsoon, summer, and winter cycles—thereby reducing the generalizability of their solutions. A particularly critical gap is the absence of region-aware solar analytics: current tools do not dynamically adapt predictions or visual outputs based on the user's geographic location, nor do they provide granular insights into solar irradiance, weather influence, or dust behavior specific to a region. Additionally, economic intelligence is often missing; few systems incorporate calculations for projected financial loss or provide a cost-benefit breakdown that justifies cleaning actions. Backend deployment and integration have also been largely overlooked, with existing systems lacking scalable, API-driven architectures to serve real-time predictions across platforms.

In this paper, the identified research gaps have been systematically addressed. The main contributions of our work can be summarized as follows:

* End-to-End Intelligent Framework- an integrated pipeline that seamlessly combines data ingestion, preprocessing, ML-based inference, and real-time frontend visualization, providing a holistic solution for soiling prediction and maintenance planning.
* Geolocation-Aware Analysis- it dynamically adapts predictions based on the user’s geographical input, offering region-specific solar insights including irradiance behavior, weather influence, and soiling trends—currently missing in other systems.
* Seasonal Modeling and Adaptability- Unlike conventional approaches, we implement seasonally segmented validation, ensuring that the model is adaptable across diverse climatic conditions and improves generalizability.
* Economic Decision Support- model generates not just technical predictions (power loss, optimal cleaning date), but also financial metrics, including estimated revenue loss and cleaning cost-benefit analysis—empowering users to make informed economic decisions.
* ML Model Interpretability and Optimization- It evaluate multiple machine learning algorithms and select the most optimal model based on validation metrics and interpretability, ensuring both performance and transparency.
* Modern, Interactive Web Dashboard- A user-friendly frontend built with Next.js, React, and Tailwind CSS renders key metrics, charts, and decisions using state-of-the-art dashboard—making the system accessible even to non-technical stakeholders.
* Real-Time Feedback Loop- the platform supports dynamic interaction and re-rendering based on user input ensuring responsiveness and continuous insight updates.

1. **Methodology**

The architecture (Fig. 1) of the Solar Optimization via Learning-driven Visual Intelligent Analytics (SOLVIA) framework- is a modular, data-driven system designed to enable predictive maintenance and optimize the operational efficiency of solar PV installations. It seamlessly integrates real-time environmental and electrical data acquisition, centralized storage, machine learning-based predictive modeling, and an intuitive visualization interface to provide actionable insights that reduce unnecessary maintenance efforts and enhance overall energy yield. Environmental parameters such as temperature, solar irradiance, PM2.5/PM10 particulate concentrations, nitrogen dioxide (NO₂), and dust levels are fetched in real-time through the OpenWeather API, while electrical metrics including AC/DC power output, voltage, and current are continuously gathered from on-site solar instrumentation.

These multi-source data streams are transmitted through a secure API gateway and stored in a centralized NoSQL database, built to handle high-frequency, scalable ingestion and retrieval. Once stored, the data is processed by a regression-based machine learning model, which has been trained to forecast dust accumulation and predict corresponding power losses. In addition to technical losses, the model integrates financial parameters such as revenue loss and cleaning costs to determine the optimal time for cleaning interventions. A feedback loop allows the system to adapt dynamically to changing environmental and seasonal conditions, ensuring improved predictive accuracy through continuous retraining. This predictive mechanism ensures that cleaning recommendations are economically justified and environmentally contextual.

The outputs generated by the model—including projected dust accumulation curves, financial losses, and cleaning date suggestions—are rendered in a robust frontend web interface developed using Next.js, React, and TypeScript. State management is implemented using Redux, while shadcn/ui components ensure a clean, responsive user interface. Secure user authentication is handled through Clerk, supporting OAuth logins from providers like Google and Apple. Although initially deployed at Netaji Subhas University of Technology (NSUT), Delhi, the SOLVIA framework has been developed with scalability and generalizability in mind. By abstracting core components such as data ingestion, predictive modeling, and visualization, the system can be easily adapted to solar deployments in different regions and climates—ranging from university campuses to industrial solar arrays and large-scale PV farms—simply by integrating relevant APIs and sensor systems.

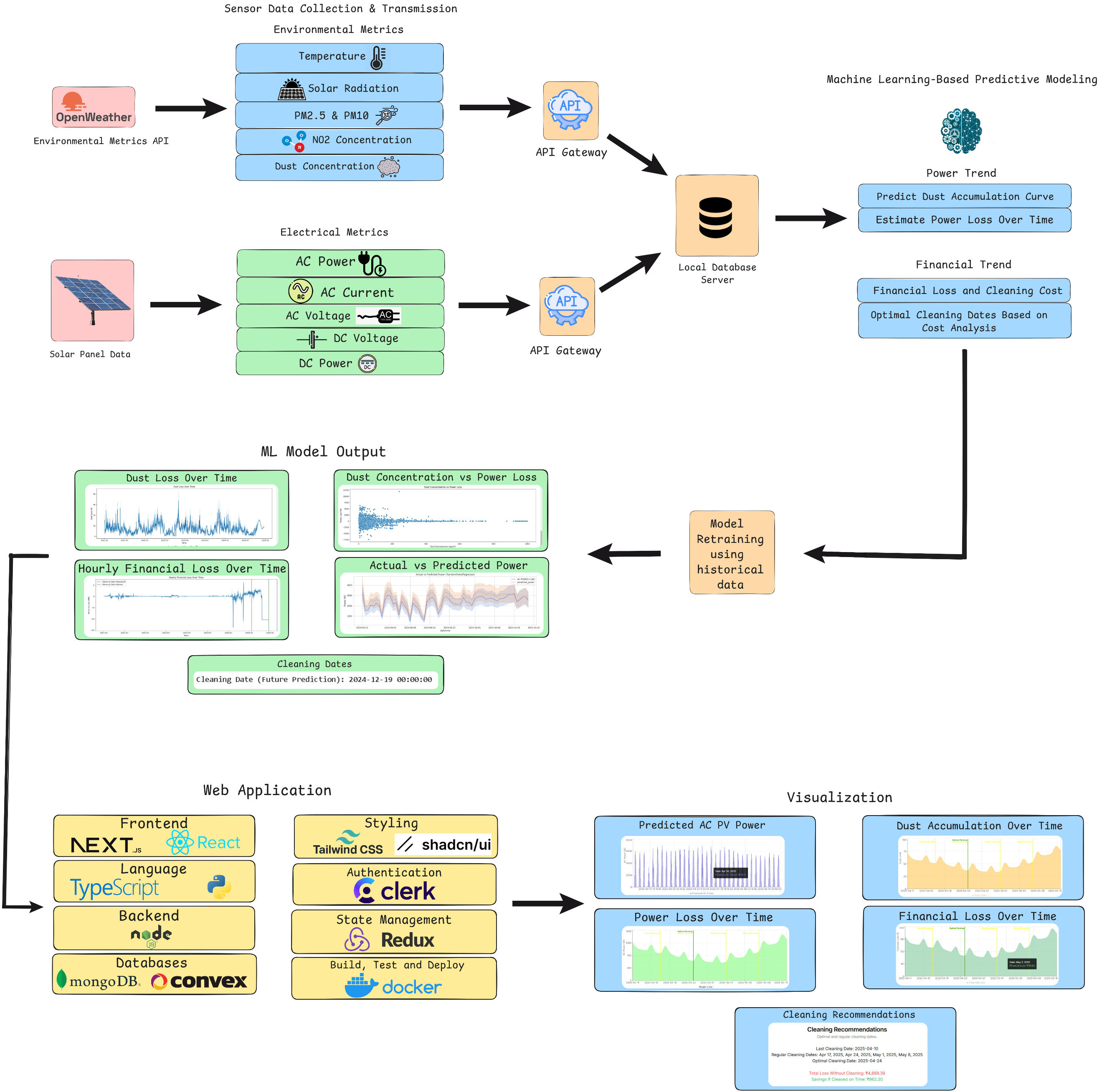


Fig. 1 Proposed Architecture

### **2.1 User Authentication and Interface Access**

### To ensure secure and seamless user access, SOLVIA integrates a robust authentication layer powered by Clerk, a modern identity and access management service. Clerk supports OAuth-based single sign-on (SSO) with major identity providers, such as Google, Apple, GitHub, and Microsoft, enabling users to log in using existing credentials without needing to create new ones. This integration streamlines the onboarding process and enhances security through industry-standard practices like multi-factor authentication (MFA), token-based sessions, and role-based access control (RBAC). Upon successful authentication, Clerk issues a secure session token, validated on both the frontend and backend components, ensuring restricted access to sensitive system resources.

### The frontend interface (Fig. 2), built using Next.js and React, provides a responsive, scalable user experience that adapts based on authentication status. Authenticated users gain access to dashboards and additional features only after a successful login, ensuring a secure user journey. By decoupling authentication from core application logic and leveraging Clerk for identity management, the system ensures scalability, maintainability, and compliance with security standards. This architecture is well-suited for institutional or commercial deployments where controlled access and secure user management are paramount.

A screenshot of a computer

AI-generated content may be incorrect.

Fig. 2 Home Screen

### **2.2 Sensor Data Collection and Transmission**

### The foundation of the SOLVIA framework lies in its robust sensor data acquisition pipeline, designed to provide real-time insights into both environmental conditions and solar panel performance. The system integrates two primary sources of data: the OpenWeather API and direct instrumentation from the PV modules. Environmental metrics are continuously fetched via the OpenWeather API, delivering high-resolution, location-specific atmospheric data, including ambient temperature, solar irradiance, PM2.5 and PM10 levels (particulate matter), NO₂ concentration, and dust concentration. These environmental parameters are critical for accurately modeling dust accumulation rates and predicting performance degradation over time. Simultaneously, the system gathers performance-related data directly from the PV modules and associated inverters, such as AC/DC power output along with voltage and current measurements. This dual-source data pipeline enables precise correlation of real-time power fluctuations with environmental stressors, thereby quantifying energy losses attributable to soiling and enhancing the predictive accuracy of the framework.

### **Machine Learning-Based Predictive Modeling**

### The ML framework (Fig. 3) within SOLVIA plays a pivotal role in forecasting the impact of environmental soiling on PV performance and recommending optimal cleaning schedules. The entire modeling lifecycle—from data ingestion to seasonal output generation—is constructed using a modular, interpretable pipeline.

A diagram of a computer

AI-generated content may be incorrect.

Fig. 3 ML Model Backend

### **2.3.1 Data Collection**

The modeling process in SOLVIA begins with the aggregation of multivariate data from three primary domains: weather, dust, and PV system performance. Weather data, sourced via the OpenWeather API, includes granular temporal variables such as solar irradiance, temperature, humidity, and wind speed—each of which significantly impacts both soiling behavior and overall power generation efficiency. Dust-related data is obtained from pollution monitoring datasets and dust index APIs, focusing on PM2.5 and PM10 levels to quantify atmospheric particulates that contribute to solar panel soiling. Complementing these, PV system data is harvested directly from on-site sensors, comprising AC/DC power output, voltage, current, and historical cleaning logs. This performance data serves as the dependent variable in regression modeling aimed at forecasting soiling impact. Data acquisition is implemented using Python, leveraging the requests and pandas libraries for seamless real-time API integration and structured storage.

**2.3.2 Data Processing**

Once collected, the dataset in SOLVIA undergoes a structured and methodical preprocessing pipeline to ensure consistency, reliability, and readiness for modeling. The first step involves addressing missing or null values using imputation techniques such as forward-fill and mean substitution, implemented through Python libraries like pandas and NumPy. These methods preserve temporal coherence while minimizing information loss. Following this, feature normalization is performed to bring input variables onto a common scale, thereby preventing skewed model behavior due to varying input ranges. This is achieved using scikit-learn’s StandardScaler or MinMaxScaler, depending on the nature of the data distribution. To maintain generalizability, the cleaned and normalized dataset is then divided into training and testing subsets using stratified sampling with an 80:20 split, ensuring that seasonal and environmental variations are well-represented in both segments. This preprocessing stage is critical for exposing the predictive models to a balanced set of conditions, thereby enhancing their robustness and accuracy.

**2.3.3 Model Development**

The modeling phase in SOLVIA involves a progressive application of machine learning algorithms to accurately capture the relationship between environmental factors, dust accumulation, and the resulting performance degradation of photovoltaic modules. Initially, linear regression is employed as a baseline to establish a fundamental correlation between the input features—such as temperature, irradiance, particulate concentration—and the output power. To account for non-linearities inherent in real-world soiling patterns, polynomial regression is introduced, enabling the model to capture higher-order interactions and curved trends in the data. To prevent overfitting, especially in scenarios involving multicollinearity among environmental variables, regularization techniques like Ridge and Lasso regression are integrated, helping to constrain the model complexity while preserving predictive performance. Beyond linear models, decision tree regression is utilized to generate interpretable, rule-based predictions by hierarchically partitioning the feature space into decision nodes based on environmental thresholds. To further enhance generalization and reduce variance, a random forest regression model—an ensemble of multiple decision trees trained via bootstrap aggregation—is deployed. Finally, for high-precision forecasting, SOLVIA incorporates gradient boosting using XGBoost, which sequentially optimizes an ensemble of weak learners by minimizing residual-based loss functions. This layered approach ensures a comprehensive understanding of soiling behavior and enables robust predictive performance under diverse environmental conditions.

**2.3.4 Model Validation**

To ensure the reliability and robustness of predictive performance, each machine learning model developed within the SOLVIA framework undergoes a comprehensive validation process using standard statistical regression metrics. These include Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Squared Error (MSE), and the Coefficient of Determination (R²), all computed using the scikit-learn metrics module. These metrics are critical in evaluating the model's accuracy, bias, and variance, ensuring it performs well not only on the training dataset but also on unseen data. Furthermore, SOLVIA incorporates a unique seasonal validation strategy by partitioning the dataset into distinct environmental segments—monsoon, summer, and winter. This seasonal cross-validation helps in assessing the model's adaptability to varying climatic conditions, which is crucial for maintaining prediction accuracy across real-world deployment scenarios where soiling rates and atmospheric dynamics fluctuate significantly.

#### **2.3.5 Model Selection and Seasonal Output Generation**

Following rigorous validation, the most suitable model within the SOLVIA pipeline is selected based on a balance of predictive accuracy, generalization performance, and interpretability. The chosen model is then serialized and deployed via FastAPI, transforming it into a scalable RESTful API endpoint capable of real-time inference. Once deployed, the model is used to generate season-specific predictions by leveraging the segmented environmental data for monsoon, summer, and winter conditions. These predictions include critical insights such as the estimated percentage loss in power output due to soiling, the optimal date for the next cleaning cycle, the projected financial loss incurred from delayed maintenance, and a cost-benefit analysis to guide cleaning decisions. All outputs are formatted as structured JSON responses, which are seamlessly consumed by the frontend interface for intuitive visualization and actionable decision-making. This integration enables SOLVIA to deliver precise, data-driven guidance tailored to fluctuating environmental conditions across seasons.

**2.4 Frontend Integration and Visualization**

### The integration between the backend machine learning models and the frontend dashboard in SOLVIA is designed for real-time responsiveness, reliability, and scalability. The system ensures that the predictive insights generated by Python-based models are seamlessly delivered and displayed to the user through a modern web interface.

The integration of a ML backend, powered by FastAPI, with a frontend interface built using Next.js, React, and TypeScript, enables real-time delivery of predictive insights to users. This seamless integration facilitates the presentation of key metrics, including power loss due to soiling, financial implications, and cleaning recommendations, through an interactive web interface.

**2.4.1 API Communication Layer**

The **API Communication Layer** in SOLVIA acts as the critical bridge between the machine learning backend and the frontend interface, enabling real-time interaction and data-driven decision-making. Built on FastAPI, the backend exposes inference endpoints like /api/predict, which are optimized to receive POST requests with structured JSON payloads. These payloads typically include real-time environmental parameters such as temperature, solar irradiance, particulate matter (PM2.5/PM10), timestamps, dust index, and metadata like location identifiers. Upon receiving a request, the backend parses the input, processes it through the trained ML model, and returns predictions that detail expected power loss, associated financial implications, and recommended cleaning intervals.

On the frontend, the integration is handled using asynchronous HTTP requests through tools like fetch() or Axios, which are embedded within React’s useEffect hooks. This ensures that API calls are triggered on component mount or when specific state variables—such as location selection or time range—change. The responses from the backend are then deserialized and seamlessly integrated into the dashboard, where users can visualize insights in an intuitive, responsive manner. This two-way communication layer ensures SOLVIA maintains a fluid and scalable experience, delivering predictive insights in near real-time with minimal latency.

**2.4.2 Backend Model Inference and Response**

When an API request is received by the SOLVIA backend, the first step involves validating the incoming data to ensure consistency and correctness. FastAPI uses built-in Pydantic models to perform this validation, ensuring that all required input parameters, such as environmental data and solar panel performance metrics, are correctly formatted and free from errors. This validation process is crucial for maintaining the integrity of the system and ensuring that the model receives accurate inputs for inference. Following validation, the backend proceeds to preprocess the data using tools from the scikit-learn library, such as the StandardScaler. This step ensures that the input features are appropriately normalized, transforming the data into a consistent format suitable for model inference.

Once the data is validated and preprocessed, the backend loads the pre-trained machine learning models, such as Random Forest, XGBoost, or Ridge Regression, using serialization tools like joblib or pickle. These models, trained on historical data, are then used to generate predictions related to the solar panel’s performance under soiled conditions. The backend computes key outputs, including the percentage loss in power caused by soiling, the optimal cleaning date, the estimated financial loss from reduced energy output, and a cost-effectiveness analysis to decide whether cleaning is necessary. The results are returned as a well-structured JSON response, which includes both numerical outputs (such as the percentage loss) and categorical outputs (such as the binary decision for cleaning). This structured response is then sent to the frontend for visualization, enabling users to take actionable steps based on the insights provided.

**2.4.3 Frontend State Management and User Interface Rendering**

Upon receiving the JSON response from the backend, the SOLVIA frontend system updates the user interface by parsing the data and storing it within the component-level state. This is achieved using React’s useState hook or, for more complex scenarios, through global state management libraries such as Redux or Zustand. By storing the prediction results in the state, the frontend ensures that the UI remains responsive and dynamic, reflecting the latest insights in real time. The key display elements include KPI cards, which are visually styled using React and Tailwind CSS to present critical metrics such as power loss, estimated financial loss, and cleaning recommendations in an easy-to-read format. In addition to these cards, the system leverages time-series visualizations generated with libraries like Chart.js or Recharts to depict historical performance trends, environmental conditions, and scheduled cleaning cycles. These charts provide users with a detailed, interactive view of system performance over time, helping them make informed decisions based on real-time data and predictive insights.

**2.4.4 Re-fetching and Re-rendering Logic**

User interactions, such as selecting a new date range, choosing a different region, or updating the weather data source, automatically trigger a series of processes in the SOLVIA system to ensure that the displayed information is always relevant and up-to-date. When a user makes a change, the frontend system sends an updated API request to the backend with the new parameters. This request prompts the backend to re-execute the machine learning inference model, using the newly provided data to generate updated predictions. These predictions may include metrics such as soiling-induced power loss, optimal cleaning dates, and financial loss estimates, all of which are recalculated based on the revised input.

Once the backend processes the request and returns the updated results, the frontend system uses React’s virtual DOM to re-render the user interface, ensuring that the most recent data is reflected dynamically. React’s efficient re-rendering process ensures that only the components that have changed are updated, keeping the UI responsive and minimizing unnecessary re-renders. This feedback loop between the backend and frontend guarantees that the system remains highly responsive, offering users the most accurate and contextually relevant insights based on real-time inputs. By continually refreshing the information displayed, SOLVIA provides an intuitive and seamless user experience, ensuring that decision-making is always backed by the most current data available.

1. **Real Field-Case Study**

The *Predictive Cleaning Framework* was practically implemented at NSUT, Dwarka (Fig. 4), located in Delhi, India. This site was selected owing to its strategic geographical location and its susceptibility to environmental stressors that significantly influence PV system performance. Positioned at 28.56° N latitude and 77.03° E longitude, NSUT experiences a continental climate, characterized by high seasonal variability, elevated particulate matter levels, and frequent atmospheric changes. These factors collectively contribute to substantial soiling losses on solar modules due to dust accumulation.

A map of a city

AI-generated content may be incorrect.

Fig. 4 NSUT Dwarka Campus

To adapt the framework to real-world conditions, region-specific machine learning models were developed by integrating historical environmental data with advanced regression-based algorithms. These models were trained to recognize the influence of climatic variables—such as temperature, humidity, wind speed, and solar irradiance—on dust deposition rates and corresponding PV efficiency degradation.

Environmental metrics were sourced via the OpenWeather API, supplemented with physical sensors deployed on-site (e.g., temperature sensors, irradiance sensors). In parallel, electrical performance data, including power output, voltage, and AC current, was obtained directly from NSUT’s operational solar PV systems. Fusion of environmental and electrical data within the machine learning pipeline enabled accurate forecasting of efficiency loss, financial implications, and recommended cleaning intervals. This facilitated the formulation of a data-driven and economically optimized cleaning schedule, significantly reducing operational losses while maximizing solar energy yield.

Moreover, integration with the frontend web application ensured that the predicted outcomes were presented in an intuitive and user-friendly manner. Stakeholders—including non-technical personnel—could conveniently visualize key performance indicators, interpret financial trends, and act upon automated cleaning recommendations generated by the backend models. This successful deployment demonstrates the framework’s robustness and scalability. With minor adjustments to environmental input parameters, the Predictive Cleaning Framework can be readily extended to other regions facing similar dust-related challenges, thereby enhancing the performance and longevity of PV infrastructure across diverse climatic zones.

1. **Results and Findings**

Based on the seasonal and temperature changes, the levels of dust accumulation, particulate matter, and many other environmental factors vary, and this affects the soiling patterns in the PV system. Change in seasons also causes a change in the outcomes of the predictive analysis performed by the machine learning model. This, in return, affects the optimal clean times and the generated power output. The conventional cleaning practices thus prove inefficient, and ML model based predictive cleaning is adopted as an alternative [[Table 2]](#table2). The graphical trends of these outputs are displayed in the web application for seamless and convenient viewing of energy generation, potential financial and power losses and suggested cleaning times.

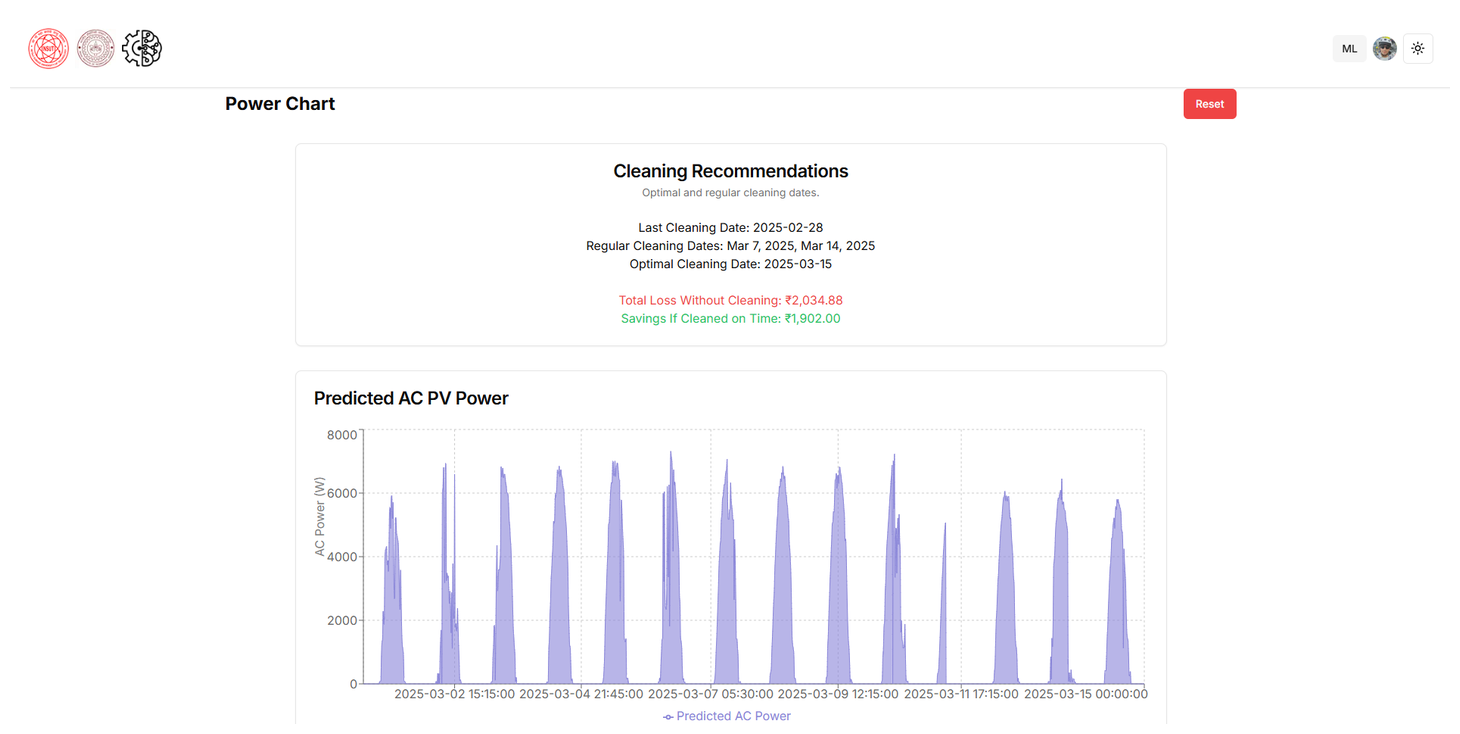


Fig. 5 Output Screen

#### **4.1. Machine Learning Model Outputs**

The machine learning regression model was trained and evaluated in a Google Colab environment using real-world environmental and electrical datasets collected from the NSUT PV installation. The model was developed using **Python** with libraries such as **Pandas, NumPy, Scikit-learn, and Matplotlib**, and employed **multi-variable linear regression** to predict power loss due to dust accumulation.

A blue line graph with numbers

AI-generated content may be incorrect.

Fig. 6a Dust loss over time

A graph with blue lines

AI-generated content may be incorrect.

Fig. 6b Dust concentration vs power loss

A graph showing a number of data

AI-generated content may be incorrect.

Fig. 6c Hourly Financial loss overtime

A graph of a graph

AI-generated content may be incorrect.

Fig. 6d Actual vs Predicted Power

These findings served as a validation phase prior to deploying the model in the production-ready backend.

#### **2. Application Output Visualizations**

The integrated web application, developed using Next.js, React, and TypeScript, delivers real-time insights from the ML model through an intuitive dashboard. Predictive outputs—processed from live sensor and weather data—are visualized using dynamic charts powered by Chart.js and Recharts. This translation of raw numerical data into actionable insights enables technicians and facility managers to make informed, efficient decisions.

**Case-1 (1st March 2025 to 15th March 2025)**

During the early summer season, a detailed analysis of PV performance and environmental data revealed critical insights into maintenance efficiency. The last manual cleaning of the solar panels was performed on February 28, 2025, followed by routine cleanings scheduled for March 7 and March 14, 2025. However, the machine learning model identified March 15, 2025, as the optimal cleaning date based on real-time environmental stressors and predicted soiling impact. Had the panels remained uncleaned until this optimal point, the projected energy loss would have amounted to ₹2,034.88. By adhering to the model’s recommendation and cleaning on time, the system would realize a substantial cost saving of ₹1,902.00 (Fig. 7e). This insight underscores the financial and operational value of predictive maintenance, especially during high-soiling periods like early summer.

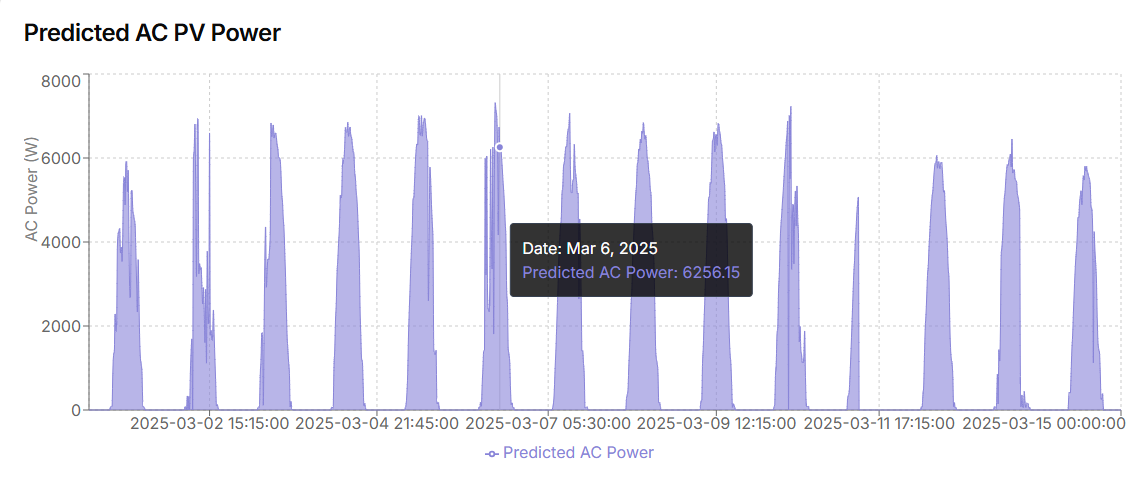


Fig. 7a AC Power

A graph with numbers and a black rectangle

AI-generated content may be incorrect.

Fig. 7b Dust accumulation

A green graph with yellow text

AI-generated content may be incorrect.

Fig. 7c Power loss over time

A graph with green lines and black text

AI-generated content may be incorrect.

Fig. 7d Financial loss over time

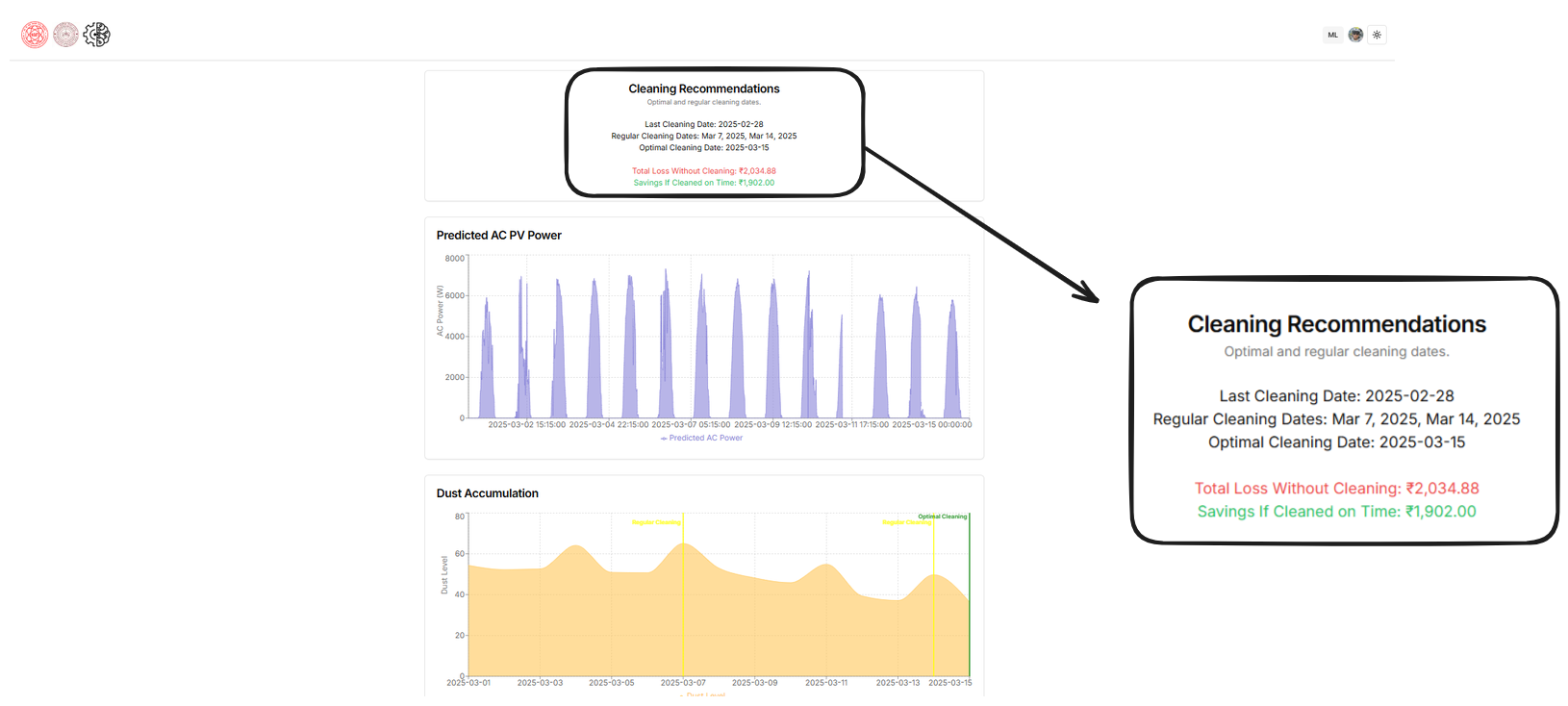


Fig. 7e Cleaning Recommendations

**Case-2 (1st June 2025 to 15th June 2025)**

During the early monsoon period of June, marked by increased humidity and the onset of rainfall-driven particulate deposition, the last cleaning was performed on May 28, 2025. Routine maintenance was scheduled for June 4 and June 11; however, predictive analytics highlighted June 3, 2025, as the optimal date for intervention. Despite regular cleanings falling close to the recommended date, slight misalignment led to a projected energy loss of ₹1,962.23 if cleaning were delayed. Following the model’s guidance could have minimized this loss, resulting in savings of ₹347.36 (Fig. 8e)—underscoring the significance of precision scheduling even during rain-induced natural cleaning phases.

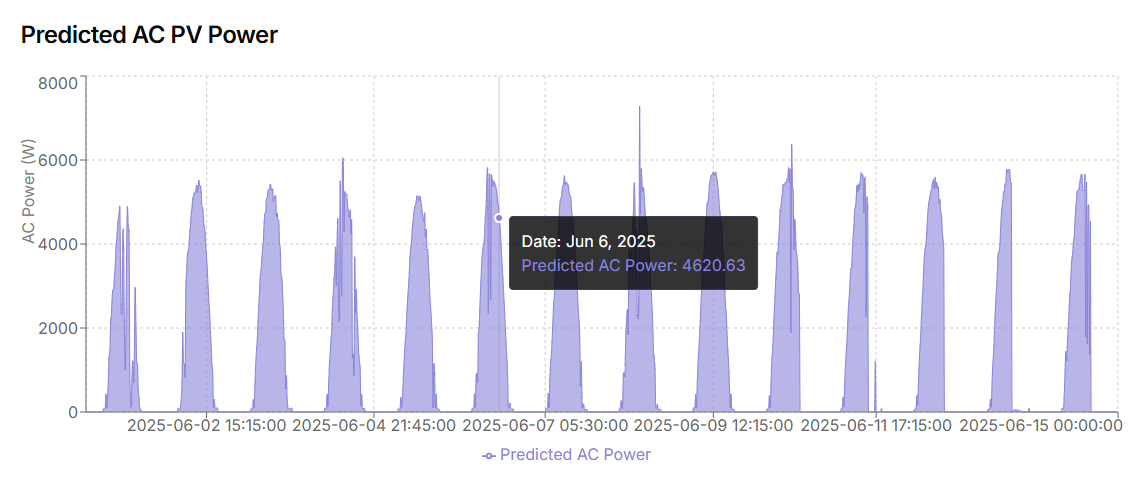


Fig. 8a AC Power

A graph with numbers and a black rectangle

AI-generated content may be incorrect.

Fig. 8b Dust accumulation

A green graph with yellow text

AI-generated content may be incorrect.

Fig. 8c Power loss over time

A graph with green and yellow lines

AI-generated content may be incorrect.

Fig. 8d Financial loss over time



Fig. 8e Cleaning Recommendation

**Case-3 (1st September 2025 to 15th September 2025)**

In the post-monsoon transition of September—characterized by fluctuating humidity and residual airborne particulates—the last manual cleaning was recorded on August 28, 2025. Scheduled cleanings were carried out on September 4 and September 11 as part of the regular maintenance cycle. However, the predictive model, informed by real-time weather and dust data, identified September 15, 2025, as the most effective cleaning date. Missing this optimal window would have led to an estimated energy loss of ₹2,028.21 (Fig. 9e). Timely cleaning based on the model’s recommendation could have yielded savings of ₹1,887.52, showcasing the advantage of data-driven scheduling over static routines during seasonal shifts.

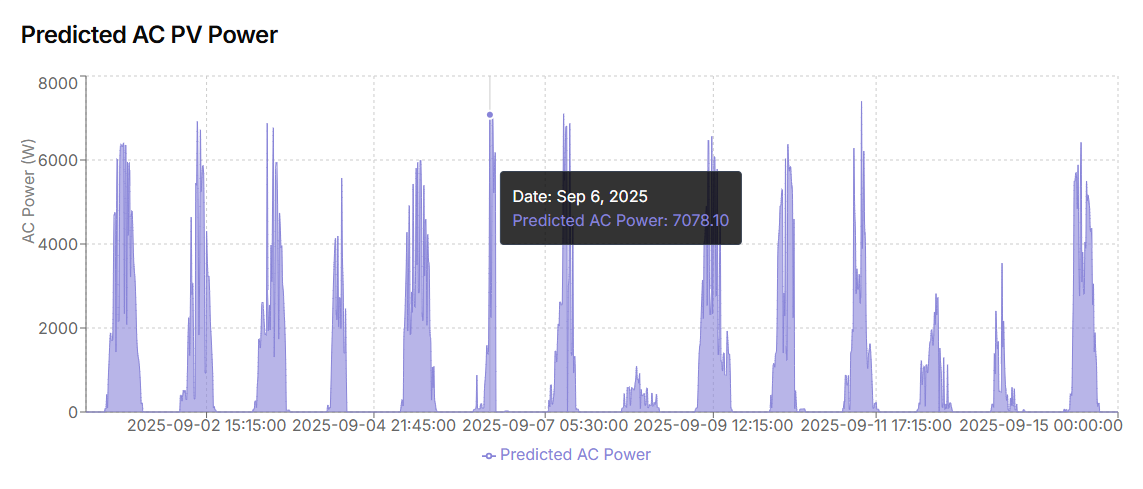


Fig. 9a AC Power

A graph with numbers and a black rectangle

AI-generated content may be incorrect.

Fig. 9b Dust accumulation

A green graph with numbers and a green line

AI-generated content may be incorrect.

Fig. 9c Power loss over time

A graph with green and black text

AI-generated content may be incorrect.

Fig. 9d Financial loss over time



Fig. 9e Cleaning Recommendation

**Case-4 (1st December 2025 to 15th December 2025)**

In the cooler and drier conditions of December—when atmospheric dust tends to linger due to lower humidity and reduced rainfall—the last cleaning was conducted on November 28, 2025. Scheduled maintenance dates were set for December 5 and December 12, with the predictive model identifying December 12, 2025, as the optimal cleaning date. Despite alignment with the second scheduled cleaning, the delay in addressing accumulating soiling could have resulted in a substantial energy loss of ₹2,710.0 (Fig. 10e). By adhering to the model's optimal timing, an estimated ₹1,091.72 could have been saved, highlighting the importance of data-driven scheduling even during winter months when soiling progression may appear less aggressive but remains economically significant.

A graph of a graph

AI-generated content may be incorrect.

Fig. 10a AC Power

A graph with numbers and a black rectangle

AI-generated content may be incorrect.

Fig. 10b Dust accumulation

A green graph with yellow text

AI-generated content may be incorrect.

Fig. 10c Power loss overtime

A green graph with numbers and a line

AI-generated content may be incorrect.

Fig. 10d Financial loss overtime



Fig. 10e Cleaning Recommendation

Table 2 Conventional vs. ML model based Predictive Cleaning

|  |  |  |
| --- | --- | --- |
| *Aspect* | *Conventional Cleaning (Rule-Based/Manual)* | *ML Model-Based Predictive Cleaning* |
| Cleaning Schedule | Fixed intervals (e.g., weekly or biweekly) regardless of environmental conditions | Dynamically computed based on real-time weather, pollution, and performance data |
| Early Summer Example (March) | Cleanings on Mar 7 & Mar 14—missed the actual critical point | Predicted optimal cleaning on Mar 15 minimized soiling loss with potential savings of ₹1,902.00 |
| Monsoon Example (June) | Cleanings on Jun 4 & Jun 11 did not align with actual soiling trends | Model predicted Jun 3 as optimal, but conventional cleaning missed early onset, reducing potential savings to ₹347.36 |
| September Example | Scheduled on Sep 4 & Sep 11—cleanings occurred too early | Optimal cleaning identified as Sep 15, enabling savings of ₹1,887.52 if followed |
| Winter Example (December) | Cleanings on Dec 5 & Dec 12 matched optimal date but were not guided by environmental severity | Model confirmed Dec 12 as optimal, showing cleaning was effective; however, savings of ₹1,091.72 justified model use |
| Adaptability | Not adaptive—fails to account for weather variation, pollution surges, or dust storm effects | Highly adaptive—responsive to real-time environmental data, seasonal variations, and performance metrics |
| Energy Loss Without Cleaning | Often higher due to delayed or premature cleaning decisions | Quantified in real-time; loss estimated and minimized proactively |
| Decision Intelligence | Based on fixed assumptions or manual heuristics | Backed by trained ML models (Random Forest, Gradient Boost, etc.) analyzing multivariate relationships |
| Financial Optimization | Cannot accurately calculate ROI of cleaning or predict financial implications | Provides real-time estimates of energy loss, cleaning cost, and savings |
| Operational Efficiency | May lead to unnecessary maintenance cycles or missed cleaning opportunities | Optimizes cleaning intervals, reducing both excessive cleaning and power degradation |
| User Accessibility | Lacks digital accessibility and predictive insights | Integrated into a modern web interface for visualization and easy decision-making |

1. **Conclusion and Future Works**

The Opti-Clean system presents a robust and intelligent approach to mitigating power losses in PV modules caused by environmental soiling and dust accumulation. By integrating real-time environmental sensing with machine learning-based predictive modeling, it delivers actionable insights through a streamlined web interface built using modern frameworks like Next.js, React, and TypeScript. The system dynamically determines optimal cleaning intervals, quantifies power degradation and financial losses, and aids in efficient maintenance planning. Its deployment at NSUT, Delhi, demonstrated high accuracy across seasonal variations—summer, monsoon, and winter—highlighting the adaptability and precision of the model. Visualization of key performance indicators and real-time alerts significantly improved user interaction, enabling both technical and non-technical stakeholders to make faster, data-backed decisions.

Looking ahead, the SOLVIA framework holds potential for further advancement in multiple directions. Enhancing its generalizability through inclusion of geographically diverse environmental datasets would improve performance across different climatic zones. Integration with autonomous robotic cleaning systems can automate responses based on model recommendations, creating a closed-loop system. Incorporating advanced machine learning algorithms such as LSTM networks or hybrid ensembles can offer deeper insights into the non-linear and temporal patterns of soiling behavior. A dedicated mobile application could extend usability to field technicians, supporting remote access and real-time decision-making in distributed solar installations. Finally, introducing a long-term energy forecasting module, built on historical pollution and seasonal trends, would further strengthen the platform’s role in strategic energy and financial planning.

**References**

1. Herrmann, J., Klimm, E., Koehl, M., Slamova, K., Weiss, K. A., Farah, A. A., & Jaber, A. A. (2015, June). Dust mitigation on PV modules in western saudi arabia. In *2015 IEEE 42nd Photovoltaic Specialist Conference (PVSC)* (pp. 1-5). IEEE.
2. Alshareef, M. J. (2023). A comprehensive review of the soiling effects on PV module performance. *IEEE Access*, *11*, 134623-134651.
3. Sayyah, A., Horenstein, M. N., & Mazumder, M. K. (2013, June). Mitigation of soiling losses in concentrating solar collectors. In *2013 IEEE 39th Photovoltaic Specialists Conference (PVSC)* (pp. 0480-0485). IEEE.
4. Javed, W., Guo, B., Figgis, B., & Aïssa, B. (2021). Dust potency in the context of solar photovoltaic (PV) soiling loss. *Solar Energy*, *220*, 1040-1052.
5. Islam, M. I., Jadin, M. S., & Al Mansur, A. (2025). Characterization of accumulated unknown and soil dust on the PV system: An experimental study on the impact and performance analysis. *Renewable Energy*, *240*, 122200.
6. Chiteka, K., Arora, R., Sridhara, S. N., & Enweremadu, C. C. (2020). A novel approach to Solar PV cleaning frequency optimization for soiling mitigation. *Scientific African*, *8*, e00459.
7. Naeem, M., & TamizhMani, G. (2015, June). Cleaning frequency optimization for soiled photovoltaic modules. In *2015 IEEE 42nd Photovoltaic Specialist Conference (PVSC)* (pp. 1-5). IEEE.
8. Redondo, M., Platero, C. A., Moset, A., Rodríguez, F., & Donate, V. (2023). Soiling modelling in large grid-connected PV plants for cleaning optimization. *Energies*, *16*(2), 904.
9. Redondo, M., Platero, C. A., Moset, A., Rodríguez, F., & Donate, V. (2021, September). Soiling forecasting in large grid-connected PV plants and experience in Spain. In *2021 IEEE International Conference on Environment and Electrical Engineering and 2021 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe)* (pp. 1-4). IEEE.
10. Abraim, M., El Ydrissi, M., Ghennioui, H., Ghennioui, A., Hanrieder, N., Wilbert, S., & Azouzoute, A. (2023). PVSMS: A system for quantifying soiling effects and optimizing cleaning schedule in PV solar plants. *Energy Conversion and Management*, *284*, 116978.
11. Sharma, P., Mishra, R. K., Bhola, P., Sharma, S., Sharma, G., & Bansal, R. C. (2024). Enhancing and optimising solar power forecasting in Dhar district of India using machine learning. *Smart Grids and Sustainable Energy*, *9*(1), 16.
12. Sanganaboina, A. B., Ruttala, S., Mandadapu, H., Kanigiri, S. V. U. M., Kumar, S. D., & Kumar, S. V. P. (2024, June). Prediction of Solar Panel Maintenance using BiLSTM. In *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (pp. 1880-1886). IEEE.
13. Abuzaid, H., Awad, M., Shamayleh, A., & Alshraideh, H. (2025). Predictive modeling of photovoltaic system cleaning schedules using machine learning techniques. *Renewable Energy*, *239*, 122149.
14. Cherif, M., Bouaziz, B., & Bacha, F. (2024). Analysis of the weather impact on the photovoltaic panel efficiency based on machine learning techniques. *In 2024 IEEE International Multi-Conference on Smart Systems & Green Process (IMC-SSGP)* (pp. 1–6). IEEE. <https://doi.org/10.1109/IMC-SSGP63352.2024.10919865>
15. Garg, U., Chohan, D. K., & Dobhal, D. C. (2021). The prediction of power in solar panel using machine learning. *In 2021 International Conference on Computational Performance Evaluation (ComPE)* (pp. 354–358). IEEE. <https://doi.org/10.1109/ComPE53109.2021.9751901>
16. Abuzaid, H., Awad, M., Shamayleh, A., & Alshraideh, H. (2024). Predictive modeling of photovoltaic system cleaning schedules using machine learning techniques. *Renewable Energy, 239*, 122149. <https://doi.org/10.1016/j.renene.2023.122149>
17. Chanchangi, Y. N., Ghosh, A., Sundaram, S., & Mallick, T. K. (2020). An analytical indoor experimental study on the effect of soiling on PV, focusing on dust properties and PV surface material. *Solar Energy, 203*, 46–68. <https://doi.org/10.1016/j.solener.2020.03.089>
18. Ghazi, S., & Ip, K. (2014). The effect of weather conditions on the efficiency of PV panels in the southeast of UK. *Renewable Energy, 69*, 50–59. <https://doi.org/10.1016/j.renene.2014.03.018>
19. Tanesab, J., Parlevliet, D., Whale, J., Urmee, T., & Pryor, T. (2015). The contribution of dust to performance degradation of PV modules in a temperate climate zone. *Solar Energy, 120*, 147–157. <https://doi.org/10.1016/j.solener.2015.06.052>
20. Pan, J., & Tian, H. (2024, November). Prediction of Output Power of PV Power Stations Based on Particle Swarm Optimization-Random Forest. In *2024 IEEE 4th International Conference on Data Science and Computer Application (ICDSCA)* (pp. 773-779). IEEE.
21. Paguthaniya, S. A., Patel, F. Y., Aminu, E., & Adeleye, A. Integration Of Machine Learning Models Into Backend Systems: Challenges And Opportunities.
22. Tiwari, C., Pillai, S., Obaid, A. J., Saear, A. R., & Sabri, A. K. (2023, September). Integration of artificial intelligence/machine learning in developing and defending web applications. In *AIP Conference Proceedings* (Vol. 2736, No. 1). AIP Publishing.
23. Mirkar, S., Juvekar, S., Popat, A., & Jain, H. (2025, February). Smart Solar Panel Maintenance: A Data-Driven Approach Using IoT and Machine Learning. In *2025 International Conference on Sustainable Energy Technologies and Computational Intelligence (SETCOM)* (pp. 1-6). IEEE.
24. Ding, C., Wu, Z., Wang, L., & Gong, M. (2024, July). Automatic Cleaning and Maintenance System for Photovoltaic Power Generation Based on Integrated Machine Learning Algorithm. In *2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS)* (pp. 1255-1259). IEEE.
25. Shen, Y., Fouladirad, M., & Grall, A. (2024). Impact of dust and temperature on photovoltaic panel performance: A model-based approach to determine optimal cleaning frequency. *Heliyon, 10*(16), e35390. <https://doi.org/10.1016/j.heliyon.2024.e35390>