*9th North American Conference on Industrial Engineering and Operations Management*
*Washington D.C., USA, June 4-6, 2024*
Publisher: IEOM Society International, USA DOI: 10.46254/NA09.20240257
Published: June 4, 2024

# Detection of Soiling Events in Solar Photovoltaic Systems Using In-situ Optical Sensing and Machine Learning

**Nicholas Percarpio and Ahmed Aziz Ezzat**
Rutgers, The State University of New Jersey,
Piscataway, NJ, 08854, USA
nmp231@scarletmail.rutgers.edu, aziz.ezzat@rutgers.edu

## Abstract

The increasing worldwide adoption of solar energy motivates the need for innovative solutions to optimize the operations and maintenance (O&M) of solar energy systems. Soiling, i.e., the accumulation of material such as dust on solar photovoltaic (PV) arrays, poses a key challenge for solar operators as it can compromise the performance and reliability of solar energy systems. To address this challenge, we develop a machine-learning-(ML)-based in-situ monitoring solution in order to timely detect the occurrence of soiling events, thereby alerting operators to take necessary mitigation measures for the cost-effective management of their assets. First, an experimental testbed has been set up at the Energy Lab at Rutgers University–New Brunswick, wherein a new optical sensing technology is installed on an operational PV system in order to collect sensor measurements about the soiling ratio under normal and abnormal conditions. Then, a cluster of ML classifiers are trained on those sensor data in order to effectively disentangle the random variations in soiling ratio versus systematic drops due to ongoing soiling, and finally output a probabilistic decision on the soling status of the PV system. Our experiments on three artificially generated soiling events in 2023 suggest that kernel-based models such as support vector machines–when supplemented with a newly introduced feature that describes the rate of change in soiling ratio—can produce up to 99% accuracy in predicting soling events, while maintaining a sensible balance between false and missed alarms (up to 85.4% in F1-scores).

## Keywords
Machine Learning, Operations and Maintenance, Solar Energy, Soiling.

## 1. Introduction
In recent years, the penetration of solar photovoltaic (PV) power into modern-day electricity grids has significantly risen, as motivated by the global aspiration for cleaner and more sustainable energy sources. The global PV cumulative capacity grew to 1.6 TW as of the end of 2023, and as much as 446 GW of new PV system capacity has been commissioned, representing a 46% increase compared to 2022 (International Energy Agency, 2024). This rapid growth in solar PV system adoption motivates the need for technological innovations that can aid solar farm operators to ensure the cost-effective operations and maintenance (O&M) of their solar energy assets. O&M is a major contributor to the levelized cost of energy (LCOE) of renewable energy resources (including but not limited to solar PV systems), and as such, reductions in O&M expenditures can directly translate into significant enhancement in the economic outlook of solar energy systems.

Along those lines, this work is particularly focused on the challenges and impacts of soiling in PV systems. Soiling, or the process by which dust, dirt, and organic/inorganic matter accumulate on the surface of PV cells, is a major and often under-estimated factor that can negatively impact PV cell performance and reliability. In 2018, soiling reduced global solar power production by 3-4% leading to estimated annual losses of 3-5 billion euros that are only expected to increase as solar power becomes more prevalent (IIse et. al. 2019). Hence, finding innovative solutions to monitor the soiling of PV arrays in real time can prove to be economically viable especially in environments where visual inspection is not possible or too costly (e.g., large-scale solar farms). Modern solar power systems are actually equipped with sensors that monitor several parameters that can be potentially used for soiling monitoring. However, up until recently, there have not been “dedicated” sensing technologies for direct soiling sensing. With the emergence of optical soling sensors that primarily focus on measuring soling-related metrics (more on those in the literature review section), there is a need to design tailored alert systems and data-driven monitoring solutions that can act on those sensor measurements in order to timely detect soiling event occurrences and assist solar farm operators in optimizing their asset management strategies.

The wealth of high-resolution soiling measurements from optical sensing technologies opens the door for innovative solutions that are rooted in machine learning (ML) to process those data, in real-time, and make informed predictions about the soling status of the solar PV system. In recent years, ML-based approaches have shown significant promise in a plethora of renewable energy applications (Rolnick et al., 2022). Some examples that show how ML have already had a positive impact on the renewable energy sector include applications in weather and energy forecasting (Ezzat, 2019, Ye et al., 2024), grid management and energy price prediction (Mhlanga, 2023), and renewable asset predictive maintenance (Shin et al., 2021, Papadopoulos et al., 2024).

We conclude that there are significant opportunities in leveraging emerging optimal sensing technologies in conjunction with ML capabilities to improve soiling detection and monitoring. Hence, *the goal of this work* is to develop and test an ML-based approach that acts on sensor data from optical measurement systems in order to timely detect soiling events and assist in alerting solar farm operators in optimally managing their solar energy assets.

### 1.1 Objectives

To achieve the aforementioned goal, the following three objectives have been carried out, from experimentation, modeling, to implementation: (*O1*) *Experimental*: To set up an experimental testbed wherein a specialized optical sensing technology is installed and used to monitor the soiling of a PV system, particularly during artificially generated soiling events; *(O2) Model Development:* Develop ML-based classification models that incorporate soiling sensor data, together with a newly introduced feature that is shown to have predictive power in detecting the presence of soiling on the solar assets; *(O3) Testing and Validation:* To evaluate the performance of the ML-based classification models using real-world data collected from the experimental testbed.

## 2. Literature Review

As solar energy operators strive to unlock more O&M cost reductions, the literature and practice on understanding, measuring, and monitoring soiling events has been gaining increased attention. A 2021 assessment of the landscape of PV soiling found that soiling research has a significant impact on solar energy operations because several of the places most conducive to solar power production are also areas at the highest risk of soiling due to elevated concentrations of air particles, minimal rainfall, and the prevalence of dust storms (Bessa et. al., 2021). This assessment also aligns with our survey of current soiling research, suggesting that existing soiling monitoring systems can be broadly characterized into three distinct clusters: soiling stations, soiling image analysis (SIA), and optical soiling measurement (OSM) sensors.

The first category is characterized by the use of a two-panel system: one panel that is regularly cleaned (used as a benchmark) and one that is allowed to become soiled. The power outputs of these panels are compared to calculate a soiling ratio, typically expressed as $r_s = \frac{Z_{soil}}{Z_{clean}}$, where $Z_{soil}$ and $Z_{clean}$ denote the power output of the clean and soiled panels, respectively. The soiling level can then be defined as $1 - r_s$ showing that a soiling ratio of 100% would equate to no soiling loss. These soiling stations are the most common commercially used soiling monitoring technique (Bessa et. al., 2021), but typically require manual or automatic cleanings using water as seen in (Gostein et al, 2014, 2015). Waterless soiling monitoring stations have also been proposed, using a rotating brush as an example (Toth et al., 2020). Major drawbacks of soiling stations include the significant uncertainty in measurements due to improper cleanings, as well scalability issues in covering large areas of a solar farm (Muller et al., 2018)

The second category corresponds to SIA sensors, which analyze aerial images using image segmentation or ML techniques to detect soiling and estimate the correspondent power loss by determining the areas of the modules covered by soiling. Drones are a popular choice for obtaining these images (Mehta et al., 2018, Miquela et al., 2023) but satellite images (Supe et al., 2020) have also been used. While these technologies have shown promise, constant monitoring of PV modules using images can be costly and/or impractical depending on the size of the solar farm.

Finally, OSM sensors, representing the third category, use the optical properties of the soiling that accumulates on PV glass to estimate the soiling ratio. They are designed to have little repair and maintenance requirements and have no moving parts which lowers the O&M costs, while ideally boosting measurement reliability (Bessa et. al., 2021). Our survey of the literature suggests that the first two commercial OSM sensors are the Mars Soiling Sensor (Gostein et al., 2018), which uses images taken by an internal camera to access the soiling impact, and the DustIQ sensor by Kipp and Zonen (Korevaar et al., 2017), which is used in this study. The DustIQ sensor works using an internal monochromatic light that is emitted on the back of a soiled glass. The reflection is then measured using a photodiode and is used to estimate the soiling ratio. Figure 1 shows a high-level depiction of the sensing methodology underpinning DustIQ. The two systems show great potential, but to the best of our knowledge, there are no detailed studies yet dedicated to analyzing and modeling the DustIQ data, further attesting to the importance of this study.

## 3. Methods

We introduce an ML-based soiling detection system that directly acts on the DustIQ optical sensing measurements, using the following two step approach: *(S1)* Experimental setup and data collection; *(S2)* Feature engineering and classification. The details of each step are outlined in Sections 3.1 and 3.2, respectively.

### 3.1. Experimental Setup and Data Collection

The solar testbed used for the experiment is located on the balcony of the Energy lab in Richard Week Hall of Engineering at Rutgers, New Brunswick. The Energy Lab's precise location can be seen in Figure 2a. An off-grid 2.4 kW PV system consisting of eight solar panels were installed in the Summer of 2021 and have been operational ever since, powering several indoor loads. Figure 2b shows a profile view of the PV system used for this study. The DustIQ Sensor has been installed on one of the PV panels as can be seen in Figure 2c. It consists of two circles of PV glass that collect dust along with the solar panel array. Inside the sensor, there is a light emitting diode (LED) that emits a pulsed infrared light that gets scattered by the soiling that accumulates on the glass. This scattering is measured by a photodiode and fed to a calculation unit that uses a calibration constant to estimate transmittance loss ($TL$). Soiling ratio (SR) is then calculated using the formula $SR = 1 - TL$ (Wolfertstetter, et al, 2021, *Kipp and Zonen 2022*).

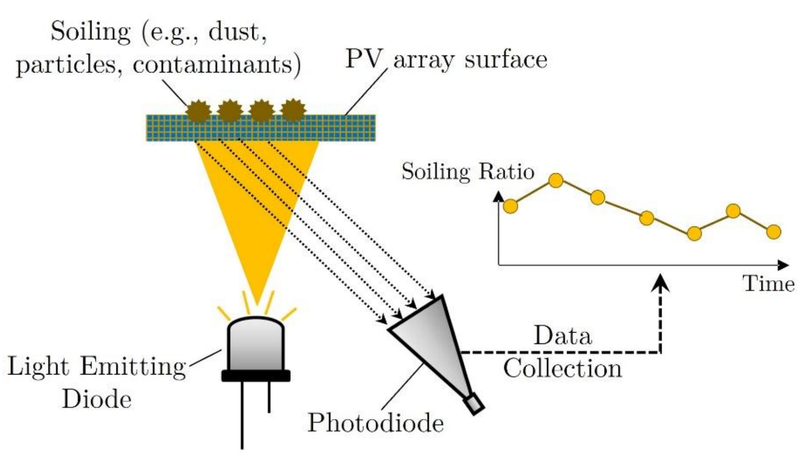


Figure 1. High-level depiction of DustIQ soiling measurement method


Figure 2. (a) Location of the Energy Lab at Rutgers University, New Brunswick; (b) Profile view of the PV system under study; (c) The DustIQ Sensor, mounted to the solar panel, showing artificially generated dust

Three separate soiling events were artificially generated in the month of October 2023. In those experiments, sand is sprayed on the PV panel on which the DustIQ sensor is installed. Those simulations, each lasting about 30 minutes-long, act as an "accelerated test" to produce positive instances of soiling. The soiling ratios are recorded by the DustIQ sensor in one minute resolution. The soiling ratio values recorded during the three experiments are shown in Figure 3 below. A time window of 1000 minutes before and after each of the on- and off-set points of the soiling events (lasting approximately 45 minutes each) is selected to create a dataset of soiling versus non-soiling conditions. The length of the time window was chosen to maintain class balance. Specifically, the data collected during the soiling event were labeled as soiling and assigned a numeric label of "1," whereas data before and after the soling events were labeled as unsoiled and assigned a numeric label of "0." Looking at Figure 3, small fluctuations in soiling ratio characterized by ~0.1% changes in soiling ratio per minute are present outside of the soiling regions. When soiling is present, however, the drop in soiling ratio is far more significant and the oscillations that occur consist of much larger minute-to-minute changes in soiling ratio than when normal conditions are present.


Figure 3. The values of soiling ratio, as measured by the DustIQ sensor, during the three soling events.

### 3.2. Feature Engineering and Classification

The observation about the behavior in the soling ratio during soling versus non-soiling event motivates us to define a new feature, which intentionally captures information about the rate of change of the soling ratio. In specific, we introduce a feature, which we dub as the squared soiling slope (SS) and denote as $s_t$, such that $s_t = (x_t - x_{t-1})^2$ where $x_t$ represents the soiling ratio at an arbitrary timestep $t$. It is used as an input in addition to the soiling ratio for the downstream ML models. Our goal is therefore to construct an ML-based classifier which translates the values of both features at time $t$, namely the soling ratio $x_t$ and the squared soiling slope $s_t$, into a binary decision of whether a soiling event is ongoing (or not). In ML parlance, our goal is to find the classification function $f(x_t, s_t|\theta)$, where $\theta$ are the classification parameters, such that $f(x_t, s_t|\theta) \rightarrow y_t \in \{0,1\}$, where $y_t$ is the final decision of the classifier, denoting the soiling status of the PV system being monitored.

For the choice of $f$, the following prevalent ML methods are evaluated for sensor-based soiling classification, comprising probability-based, neighborhood-based, and tree-based methods:

*Probability-based approaches:*

- **Logistic Regression (LR)** is a statistical method that uses a sigmoid function to predict the probability that an observation belongs to a particular class. A probability threshold is invoked to perform the final classification. In this work, we use a probability threshold of 0.5.

*Neighborhood-based approaches:*

- **K-nearest neighbors (kNN)** is a nonparametric neighborhood-based ML approach that makes a prediction for test data points based on the majority voting of its k-nearest neighbors. Grid search was used to find the best value of k, which in this work was k = 5.
- **Support vector machines (SVM)** is a kernel-based ML algorithm that seeks an optimal hyperplane that separates data from independent classes using the so-called "kernel trick." In this study, radial basis function (RBF) is chosen for the kernel due to superior training results.

*Tree-based approaches:*

- **Random Forests (RF)** is a tree-based ML algorithm that randomly subdivides the data in order to train multiple decision trees. The final prediction is assigned to the class that has the most tree "votes." In this work, grid search suggests that 50 trees are sufficient for achieving superior training results.
- **Extreme Gradient Boosting (XGBoost)** is an ensemble tree-based ML algorithm that iteratively builds decision trees that correct for the error of the previous trees using boosting and a regularized loss function (Chen et al., 2016)
- **Light Gradient Boosting Machine (LGBM)** is similar to XGBoost but is designed for speed and efficiency. Like XGBoost, it is an ensemble learning algorithm that builds multiple decision trees but uses gradient-based one-side sampling and exclusive feature bundling to handle larger datasets (Ke et. al. 2017)

## 4. Results and Discussion

### 4.1. Testing Procedure and Evaluation Metrics

The labeled data are randomly split according to an 80-20 train-test ratio to be used for training and testing the ML models described in Section 3.2. This procedure is repeated for a total of 20 times, 10 without the SS feature and 10 with the feature, in order to better assess the performance of each model and determine the impact of the newly introduced feature. To evaluate the performance of each model, four different evaluation metrics are used: accuracy, precision, recall, and F-score. If we denote $y_i$ and $\hat{y}_i$ as the true and predicted class for the *i*th soiling sample respectively, then accuracy is defined as $\frac{\sum_{i=1}^{n_{ts}} \mathbb{I}(y_i - \hat{y}_i)}{n_{ts}}$, where $n_{ts}$ is the number of testing samples and $\mathbb{I}(.)$ is the indicator function. Predictions is defined as $\frac{TP}{TP+FP}$ where *TP* and *FP* denote true and false positives, respectively. Recall, on the other hand, is defined as $\frac{TP}{TP+FN}$ where *FN* denotes false negatives. In this study, negative indicates the absence of soiling while a positive indicates that soiling is occurring, thus a false positive means that soiling was predicted when it was not occurring and a false negative means that soiling was not predicted when it was occurring. The F1 score is the harmonic mean of the precision and recall and is defined as $\frac{2 \times Precision \times Recall}{Precision + Recall}$ .

### 4.2. Numerical Results and Discussion:

The evaluation metrics described in Section 4.1 are used to evaluate each of the ML models during each of the 20 experiments. The metrics are then averaged, and the standard deviations are calculated for the group of experiments without and with the SS feature, as can be seen in Tables 1 and 2, respectively.

Table 1. Evaluation Metrics for the six ML models without the SS feature. The numbers in parentheses denote the standard deviation across the ten randomized experiments. Bold-faced values indicate the best performance across each metric.

| Classifier | Precision | Recall | F1 Score | Accuracy |
|---|---|---|---|---|
| LR | 90.2% (6.45%) | 65.5% (7.80%) | 75.7% (6.77%) | 98.5% (0.48%) |
| kNN | **93.4**% (6.77%) | 73.3% (8.61%) | 81.6% (4.95%) | 98.8% (0.41%) |
| RF | 92.9% (4.18%) | 75.6% (8.78%) | 83.0% (4.79%) | **98.9**% (0.32%) |
| SVM | 92.9% (4.18%) | **76.4**% (7.95%) | **83.5**% (4.40%) | **98.9**% (0.34%) |
| XGBoost | 92.9% (4.18%) | **76.4**% (7.95%) | **83.5**% (4.40%) | **98.9**% (0.34%) |
| LGBM | 93.7% (4.15%) | 75.4% (8.10%) | 83.2% (4.33%) | **98.9**% (0.32%) |

Table 2. Evaluation Metrics for the six ML models with the SS feature. The numbers in parentheses denote numbers in parentheses denote the standard deviation across the ten randomized experiments. Bold-faced values indicate best performance across each metric.

| Classifier | Precision | Recall | Avg. F1 Score | Accuracy |
|---|---|---|---|---|
| LR | **93.5**% (6.72%) | 65.5% (7.74%) | 76.8% (6.1%) | 98.5% (0.39%) |
| kNN | 92.4% (7.11%) | 76.5% (8.85%) | 83.2% (5.28%) | 98.9% (0.45%) |
| RF | 91.0% (4.29%) | 75.2% (10.3%) | 82.5% (6.08%) | 98.8% (0.40%) |
| SVM | 91.4% (4.46%) | **80.6**% (7.71%) | **85.4**% (4.59%) | **99.0**% (0.34%) |
| XGBoost | 91.5% (3.64%) | 79.0% (7.67%) | 84.5% (4.41%) | **99.0**% (0.31%) |
| LGBM | 92.8% (4.14%) | 77.9% (10.24%) | 84.2% (5.46%) | **99.0**% (0.31%) |

By looking at the results in Tables 1 and 2, few key observations can be drawn. Firstly, LR appears to significantly under-perform relative to the other five models, likely due to the simplicity of logistic regression which seeks a simple linear-regression-based fit to the data. This is particularly evident in terms of the F1-score, which is around 76% due to low recall values. In contrast, more advanced ML models have significant improvements in F1-scores (of up to 10%) relative to LR. Secondly, the new SS feature appears to have a noticeable impact on the performance of each of the models (with the exception of RF). This improvement is most noticeable in terms of improving the recall of the model (in the order of ~4 to 5% improvement), i.e. its ability to detect positive soiling instances. While the accuracy of the models remains relatively comparable across most models, the F1 scores of the three highest performing models, SVM, XGB and LGBM, increased ~1-2% demonstrating the models' improved ability to accurately predict true positives. This improvement comes from the improved recall scores balancing out slightly worse precision scores. The improved recall indicates that the newly introduced SS feature helped classify true positive values that were previously classified as false negative values. With that said, we can conclude that the newly proposed feature had an overall positive effect on the classifiers' performance.

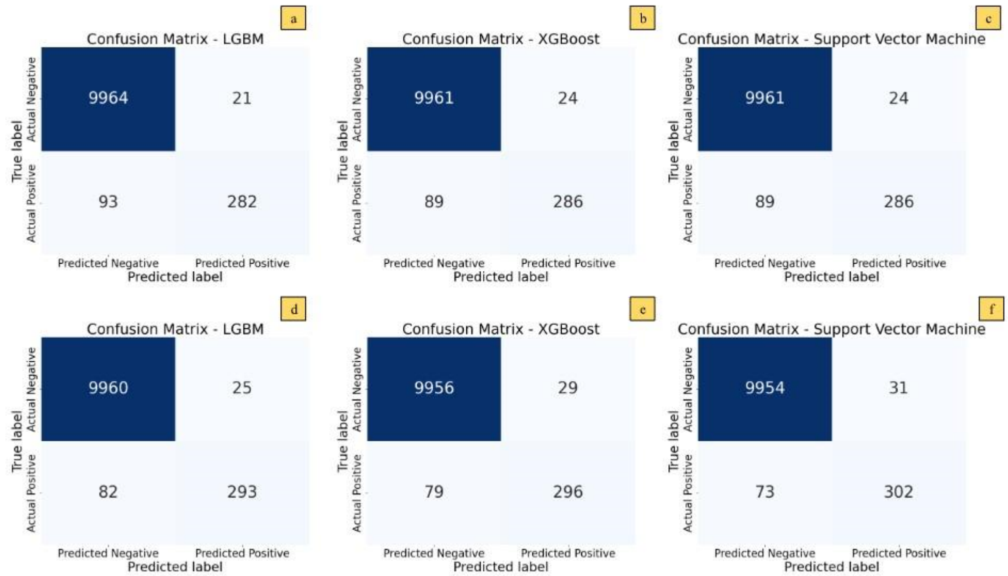


Figure 4. Confusion matrices of the three highest performing models before (a-c) and after (d-f) feature addition

These results are further highlighted in the confusion matrices shown in Figure 4, which are created by summing the results from all ten trials in both experiment types and can provide further insight into model performance. The matrices make it obvious that the models disproportionately struggle with false negatives compared to false positives both before and after the SS feature was introduced. They also confirm what the numerical evaluation metrics suggested in that the feature reduced the instances of false negatives but slightly increased the instances of false positives all while increasing the total number of true positive predictions. We believe that those false negatives could be further reduced by introducing additional features that could capture historical data beyond one previous time step. Furthermore, collecting more soiling data could potentially improve the models' performance. With that said, F1 scores approaching 85% on our highest performing models demonstrate the potential of ML to effectively detect the occurrence of soiling events in solar PV systems.

## 5. Conclusion and Future Research

The rapid expansion of solar PV systems worldwide has created an emerging need for cost-effective ways to reduce the O&M expenses of solar farms. In this work, we present a ML-powered soiling detection system that uses data from a state-of-the-art optical soiling sensor, DustIQ, that has been mounted on a PV solar array at the Energy Lab at Rutgers University in New Jersey, USA. Data from three different soiling experiments were analyzed and used to train ML classifiers that have been shown to effectively detect the occurrence of soiling, especially when supplemented with a newly introduced feature that captures the rate of change in soiling ratios. This work demonstrates an example of how new sensor technology can be successfully combined with ML to help optimize the O&M operations of a solar farm, thus contributing towards enhancing the economic outlook of solar energy.

This work paves the way for exploring a number of research avenues for further development. Firstly, our models were trained on a relatively small dataset that utilized one data source: soiling ratio. Future works could incorporate additional soiling experiments, as well as exogenous variables such as temperature or humidity to finetune the soiling detection system. Likewise, new features, that are defined in a similar manner to the square soiling slope (SS), can be created and introduced to further bolster the training data, perhaps through a semi-automated feature engineering and selection approach. Finally, although real time detection of soiling is indeed valuable, anticipating future soiling events is of equal importance, but is understandably a more challenging problem to address.

## Acknowledgment
We would like to acknowledge the senior design team consisting of Stephanie Lopez-Perez, Carlos Arias, Meriem Mrabbaj, and Viviana Flower for their tremendous help in setting up the experimental testbed and data collection needed for this work. Nickolas Percarpio has been supported through The Research Experiences for Undergraduates (REU) program by the U.S. National Science Foundation (NSF Grant #: ECCS-2114422).

## References
Bessa, João Gabriel, Monitoring Photovoltaic Soiling: Assessment, challenges, and perspectives of current and potential strategies." iScience, vol. 24, no. 3, p. 102165, 2021. https://doi.org/10.1016/j.isci.2021.102165.

Chen, Lin, et al. "Artificial Intelligence-based solutions for climate change: A Review." Environmental Chemistry Letters, vol. 21, no. 5, pp. 2525–2557, 2023. https://doi.org/10.1007/s10311-023-01617-y.

Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." In Proceedings of the 22nd ACM KDD international conference on knowledge discovery and data mining, pp. 785-794. 2016.

Ezzat, Ahmed Aziz, Jun, M., and Ding, Y. "Spatio-temporal short-term wind forecast: A calibrated regime-switching method." The Annals of Applied Statistics, vol. 13, no. 3, 1, 2019, https://doi.org/10.1214/19-aoas1243.

Gostein, Michael, J. Riley Caron, Measuring soiling losses at utility-scale PV power plants." 2014 IEEE 40th Photovoltaic Specialist Conference (PVSC), 2014, https://doi.org/10.1109/pvsc.2014.6925056.

Gostein, Michael, Kendra Passow, Local variability in PV soiling rate." 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC), 2018, https://doi.org/10.1109/pvsc.2018.8548049.

Gostein, Michael, Stan Faullin, et al. "Mars soiling sensorTM." 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC), 2018, https://doi.org/10.1109/pvsc.2018.8547767.

Gostein, Michael, Tim Duster, Accurately measuring PV soiling losses with soiling station employing module power measurements." 2015 IEEE 42nd Photovoltaic Specialist Conference (PVSC), 2015, https://doi.org/10.1109/pvsc.2015.7355993.

Herweijer, C., Combes, B. and Gillham, J., How AI Can Enable a Sustainable Future? Technical Report, 2020, URL: https://www.pwc.co.uk/sustainability-climate-change/assets/pdf/how-ai-can-enable-a-sustainablefuture.pdf (Last accessed: May 2024).

Ilse, Klemens, Techno-economic assessment of soiling losses and mitigation strategies for solar power generation." Joule, vol. 3, no. 10, pp. 2303–2321, 2019. https://doi.org/10.1016/j.joule.2019.08.019.

International Energy Agency (IEA), PVPS Programme, Snapshot of Global PV Markets 2024, Technical Report, IEA-PVPS T1-44:2024, 2024, URL: https://iea-pvps.org/wp-content/uploads/2024/04/Snapshot-of-Global-PV-Markets_20241.pdf (last accessed: May 2024).

Ke, Guolin, LightGBM: A Highly Efficient Gradient Boosting Decision Tree." Advances in Neural Information Processing Systems 30, 2017.

Kipp und Zonen, "DustIQ Soiling Measurement Instruction Manuel," https://www.kippzonen.com/Product/419/DustIQ-Soiling-Monitoring-System (2022), accessed: May 2024

Korevaar, Marc, et al. Novel Soiling Detection System for Solar Panels, 2017, https://doi.org/10.4229/EUPVSEC20172017-6BV.2.11.

Laarabi, Bouchra, Deep analysis of soiling effect on glass transmittance of PV modules in seven sites in Morocco." Energy, vol. 213, p. 118811, 2020. https://doi.org/10.1016/j.energy.2020.118811.

Mehta, Sachin, DeepSolarEye: Power loss prediction and weakly supervised soiling localization via fully convolutional networks for solar panels." 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), 2018, https://doi.org/10.1109/wacv.2018.00043.

Mhlanga, David. Artificial Intelligence and machine learning for energy consumption and production in emerging markets: A Review." Energies, vol. 16, no. 2, 9 , p. 745, 2023. https://doi.org/10.3390/en16020745.

Miquela, A., Bagul, D., and Aziz Ezzat, A., , Defect Detection in Solar Photovoltaic Systems Using Unmanned Aerial Vehicles and Machine Learning. In 8th North America Conference on Industrial Engineering and Operations Management, 2023. https://doi.org/10.46254/NA8.20230375.

Muller, Matthew, A method to extract soiling loss data from soiling stations with imperfect cleaning schedules." 2017 IEEE 44th Photovoltaic Specialist Conference (PVSC), 2017, https://doi.org/10.1109/pvsc.2017.8366214.

Papadopoulos, Petros, David W. Coit, and Ahmed Aziz Ezzat. "STOCHOS: Stochastic opportunistic maintenance scheduling for offshore wind farms." IISE Transactions 56, no. 1, 1-15. 2024

Rolnick, David, Tackling Climate Change with Machine Learning." ACM Computing Surveys, vol. 55, no. 2, 7, pp. 1–96, 2022. https://doi.org/10.1145/3485128.
Shin, Won, AI-assistance for predictive maintenance of Renewable Energy Systems. Energy, vol. 221, p. 119775, 2021. https://doi.org/10.1016/j.energy.2021.119775.
Supe, Hitesh, Google Earth engine for the detection of soiling on photovoltaic solar panels in Arid Environments." Remote Sensing, vol. 12, no. 9, 5, p. 1466, 2020, https://doi.org/10.3390/rs12091466.
Toth, Sarah, Predicting photovoltaic soiling from air quality measurements." IEEE Journal of Photovoltaics, vol. 10, no. 4, pp. 1142–1147, 2020. https://doi.org/10.1109/jphotov.2020.2983990.
Wolfertstetter, F., Incidence angle and diffuse radiation adaptation of soiling ratio measurements of indirect optical soiling sensors." Journal of Renewable and Sustainable Energy, vol. 13, no. 3, 1, 2021, https://doi.org/10.1063/5.0048001.
Ye, Feng, AIRU-WRF: A physics-guided spatio-temporal wind forecasting model and its application to the U.S. Mid Atlantic Offshore Wind Energy Areas." Renewable Energy, vol. 223, p. 119934, 2020. https://doi.org/10.1016/j.renene.2023.119934.

## Biographies

**Nicholas Percarpio** is a rising undergraduate senior majoring in Industrial and Systems Engineering at Rutgers University, New Brunswick, NJ, USA. He has worked as an undergraduate research assistant with the Renewables and Industrial Analytics (RIA) Research Group under Prof. Aziz Ezzat since the summer of 2022. His research interests include machine learning and artificial intelligence applied to renewable energy systems. He will be interning as part of Colgate's Global Engineering Team on an energy reduction project during the upcoming summer of 2024.

**Ahmed Aziz Ezzat** is an Assistant Professor of Industrial & Systems Engineering at Rutgers University, NJ, where he leads the Renewables & Industrial Analytics (RIA) research group. Before joining Rutgers, Dr. Aziz Ezzat received his Ph.D. from Texas A&M University in 2019, and his B.Sc. degree from Alexandria, Egypt, in 2013, both in Industrial Engineering. His broad research interests are in the areas of data and decision sciences, forecasting, machine learning, quality and reliability engineering, with focus on renewable energy applications and industrial informatics. His awards include the Walter A. Tyson Assistant Professorship Award from Rutgers University, 2023 Best Track Paper Award from the IISE Energy Systems Division (jointly with co-authors), the 2024 and 2022 Teaching Awards from the IISE Operations Research (OR) and Data Analytics (DAIS) divisions, respectively, and the 2020 IIF-SAS® research award from the Institute of International Forecasters and SAS corporation. He serves as the 2023-2024 president of the IISE Energy Systems (ES) Division. His research has been supported by the National Science Foundation (NSF), The National Offshore Wind Consortium (NOWRDC), The NJ Economic Development Authority, The Rutgers Energy Institute, as well as industry. He is a member of INFORMS, IISE, and IEEE-PES.