Received: 7 August 2022 | Revised: 20 January 2023 | Accepted: 27 January 2023

DOI: 10.1002/pip.3676

RESEARCH ARTICLE

PROGRESS IN PHOTOVOLTAICS WILEY

# Estimation of the impact of natural soiling on solar module operation through image analysis

Robinson Cavieres[1] | Jorge Salas[1] | Rodrigo Barraza[1] | Danilo Estay[1] | José Bilbao[2] | Patricio Valdivia[3]

[1]Department of Mechanical Engineering, Universidad Técnica Federico Santa Maria, Av. Vicuña Mackenna 3939, Santiago, Chile

[2]School of Photovoltaics and Renewable Energy Engineering, University of New South Wales, Sydney, NSW, 2052, Australia

[3]Department of Electrical Engineering, Universidad Técnica Federico Santa Maria, Av. Vicuña Mackenna 3939, Santiago, Chile

**Correspondence**
Rodrigo Barraza, Department of Mechanical Engineering, Universidad Técnica Federico Santa Maria, Av. Vicuña Mackenna 3939, Santiago, Chile.
Email: rodrigo.barraza@usm.cl

**Funding information**
ANID, Chile; Universidad Técnica Federico Santa María

## Abstract

This study presents a novel algorithm to evaluate the power losses in photovoltaic (PV) modules that are exposed to naturally accumulated soiling. The proposed algorithm uses red, green, and blue (RGB) images of the PV modules captured in the visible spectrum and measures features such as the average color intensity and standard deviation from every color channel. These features are then ranked based on the module performance by using statistical tools, and the most relevant set of features is selected. The selected features are used as an input, and the module power generation is used as the label to train an artificial neural network that outputs a continuous value that represents the instantaneous performance of the evaluated module when compared with a clean module operating under the same environmental conditions. The proposed method is effective for images captured within an irradiance ranging from 700 to 1000 ($W/m^2$), with the predictions achieving an $R^2$ score of 0.96 and RMSE of 0.74% for power losses ranging from 0% to 20% in natural soiling. This percentage corresponds to a clean condition, where irradiance is not used as an input, making it a cheap and reliable solution to monitor soiling conditions. This is the first work of its kind to demonstrate a correlation between the extracted color features of RGB images and module performance under outdoor conditions for the studied dataset. The proposed algorithm presents less accurate results when it is tested in modules exposed to non-homogenous soiling, suggesting that the proposed methodology might be effective only for naturally accumulated soiling.

**KEYWORDS**
artificial neural networks, computer vision, feature selection, image analysis, photovoltaic monitoring

## 1 | INTRODUCTION

The solar energy produced by photovoltaic (PV) modules, or solar modules, is affected by various factors. Technological factors such as the chemical composition of the cells and encapsulation determine performance under ideal conditions,[1] environmental factors such as irradiance, ambient temperature, and wind speed influence the performance of the module in the field,[2] and lastly, the installation and operational conditions affect the energy generated by the module.[3] The operators of solar installations generally aim to maximize the electrical output and performance of the modules. Therefore, it is crucial to understand the relevant variables affecting the performance of the

  

solar modules. Following the installation of the modules, the operators have no control over the module technology or the environmental conditions of the site. They can only control the operating conditions of the solar installation to improve the performance of the plant. Condition-based monitoring is one of the most useful methods to detect nonoptimal operational conditions at an early stage, which includes faults; thus, electricity generation can be maximized.[4] In an ideal operation, the operators use forecasting models to determine the amount of energy that a module must produce in a specific location under specific environmental conditions.[5] The prediction data can be compared with the actual energy production of the installation, and this information can help in taking maintenance measures when the on-site production values drift away from the predictions. Monitoring the operating conditions in a laboratory setting is relatively straightforward; however, it is difficult to scale these measurements to commercial facilities with hundreds of thousands of modules, particularly because of the increasing size of solar farms and the amount of equipment involved. A successful maintenance strategy can optimize energy generation and revenue while minimizing risk by increasing performance and reducing operation and management costs.[6]

Soiling is one of the operational conditions that deteriorate the energy output in solar power plants. Soiling primarily reduces the irradiance received by a PV module.[7] However, it can also produce other effects such as significant temperature gradients within the module based on the nature of the soiling.[8] It can also be a precursor to other causes of failure by promoting corrosion and humidity retention within the module. Soiling can severely deteriorate energy production if left unaddressed. A module can lose up to 10% of its power production within a month based on the location and tilt angle.[9] Such losses can be effectively mitigated by quantifying and addressing soiling, thus maximizing the energy yield and profits. For example, an installation that implements an optimal cleaning strategy can experience a reduction in power losses from 14.6% without cleaning to 4.6% with a monthly cleaning routine.[10] Therefore, understanding and quantifying the impact of soiling on a system are crucial to develop effective maintenance plans.

Soiling can be quantified by the surface density and particle size.[11,12] Different particle types with varying grades of chemical composition and sizes can be deposited on the surface of a module at different rates based on the plant locations and seasonal changes. Dust potency, which is the soiling loss per unit area of dust mass, can be used to quantify the total power loss in a soiled module.[13] Soiling losses can be extrapolated from environmental parameters. Particle matter and rainfall statistics demonstrate the best correlation with annualized soiling rates.[14] This approach is effective when on-site measurements are unavailable. Other approaches quantify the soiling impact by extrapolating losses from the parameters of the *I*–*V* curves.[15–17] Calculating the variations in the short circuit current ($I_{SC}$), open circuit voltage ($V_{OC}$), or maximum power point (MPP) for a given technology within a determinate timeframe can help in achieving accurate predictions of the soiling rates and power losses. Soiling affects the spectral distribution of solar radiation, primarily the blue and UV light.[18] This information can be used to formulate an equation that correlates the spectral distribution and the expected electrical losses. However, it is difficult to implement the experimental setup used to measure the spectral distribution in the operation of PV plants. The major drawback of this approach is the requirement for measuring stations with complex equipment to monitor a reference module. Because the soiling patterns in a PV plant may not be homogeneous, scaling these measurements to the entire system can produce inaccurate results. Conversely, optical solutions to soiling qualifications such as DUSST,[19] DUST IQ,[20] and MARS[21] are inexpensive and require low maintenance. These sensors can be deployed in multiple locations on an installation and help in obtaining an accurate estimation of the optical losses and soiling ratios, making it a commercially viable option. However, correlating optical losses to power losses can produce errors because the soiling patterns may not be homogeneous across the system or even in the module itself. Furthermore, the measures are considerably affected by ambient temperature, electrical activity, and mechanical vibrations. Other commercial solutions are based on measuring the *I*–*V* curve parameters from the reference modules located in the soiling stations.[22] An accurate estimate of the power loss can be obtained from a comparison of the electrical measurements of the clean and soiled modules. Similar to the commercial optical-based solutions, the *I*–*V* curve-based solutions deal with non-homogeneity in the soiling patterns across the system, which is a source of error when scaling measurements to all the modules.

The analysis of the operating conditions in the PV modules is performed using statistical methods based on images and computer vision, both in the visible and infrared spectra. Tsanakas *et al.*[23] proposed the usage of statistical processing techniques for images captured through infrared thermography, to automatically detect hot spots within a module during operation. This method involves the detection of edges along with the analysis of temperature histograms, enabling hotspot detection with high precision. However, this approach is binary in nature and only discriminates between healthy and faulty modules. Niazi *et al.*[24] also analyzed thermographs by extracting features such as contrast, energy, and entropy and constructed a Bayesian classifier to calculate an explicit probability for each state of the module (healthy and defective). This classification is binary similar to the previous study. Pivem *et al.*[25] analyzed images in the visible spectrum by using pixel frequency histograms in the red, green, and blue (RGB) channels. The level of soiling in the solar modules was located and determined, distinguishing between the punctual and distributed soiling on the surface of the PV module. Yap *et al.*[26] developed a toolbox to detect soiling presence in the modules by using features such as color histograms and statistical values. The results demonstrated that the developed tool identified varying degrees of soiling concentration on the surface of a module with high accuracy; however, these results are not correlated with the power losses. Soiling can be correlated to entropy[27] by establishing a relation between the power generated and the entropy in an image of a module. The results demonstrated that with an increase in the soiling accumulation, the power generation decreases along with the entropy of the image. Based on this information, a cleaning alarm can be set up, which activates when a predefined power threshold is crossed. However, the dust used in the test is obtained from a blowing machine in a laboratory setup and might not be representative of a real PV system.

Extensive research has been conducted on the evaluation of the operating conditions of a module using convolutional neural networks (CNNs). These methods use full images as inputs, and the feature extraction and classification stages run entirely within the layers of the neural network, contrary to the aforementioned algorithms. Consequently, more computational resources are required for both training and testing. CNNs have been employed to identify multiple fault modes in PV modules through electroluminescence imaging[28–30] and infrared imaging.[31,32] It has also been used to analyze visible spectrum images for soiling detection.[33–35] These techniques enable the extraction of more information from the image set and achieve greater precision in the classification. However, the hidden layer nature of these algorithms prevents access to the main feature being analyzed, and the method of processing the information remains unclear. Therefore, despite obtaining a good correlation, the causal link between the information provided and the results obtained is not assured. Furthermore, the classification effectiveness of the algorithm is highly dependent on the quality of the database. Consequently, the performance of the models outside the training and validation environment cannot be ensured.

This study presents an image analysis algorithm that uses statistical tools to extract the relevant features from an image to estimate the impact of naturally accumulated soiling on solar modules during operation and the quantification of power loss experienced through an artificial neural network (ANN). Most of the current approaches used to monitor soiling either require complex and expensive measuring stations that quantify the soiling impact on a reference module, or they cannot directly estimate the power losses. The proposed method does not require expensive computational equipment or complex measuring devices despite using neural networks, because it only uses a small set of features instead of using full RGB images and can run on standard computational hardware, making it a cheap and reliable tool to monitor soiling. Furthermore, this method can be used to independently evaluate multiple individual modules, by detecting different soiling levels across the entire PV system without using additional sensors or measurement stations. This study is the first to demonstrate a direct correlation between the features extracted from the images of PV modules and their performance to the best of the authors' knowledge. Contrary to the CNN algorithms that use full images as inputs, the proposed method only requires a small subset of the features extracted from the image, thereby minimizing the computational costs and ensuring a correlation between the inputs and outputs. The following sections describe the methodology implemented to perform the analysis, the experimental setup and the images used, the feature extraction and selection processes, the design and training of an ANN, the application of the proposed neural network on an external dataset, and lastly, the conclusions.

## 2 | METHODOLOGY DESCRIPTION

Extensive research is being conducted on the application of computer vision to PV installations. A condition-based maintenance tool that enables the monitoring of a large amount of modules with minimal financial investment can be established by using digital cameras and environmental monitoring stations. The monitoring range can be increased drastically by placing the digital camera on an unmanned flight unit. Figure 1 presents the images of a clean PV module (Figure 1A) and a PV module with soiling (Figure 1C), which are obtained in the laboratory. Although the soiled module can be easily distinguished, the differences between both modules must be explained. Figure 1B,D depicts the intensity frequency histograms of the pixels that decompose into the RGB channels of the images corresponding to the clean (Figure 1A) and soiled (Figure 1C) modules. This study aims to correlate the difference between the features extracted from Figure 1A,C and the energy loss of the soiled module when compared with the clean module.

Figure 2 depicts the proposed algorithm, which is based on four stages: image acquisition, segmentation and normalization, feature extraction and selection, and module performance prediction (regression). A digital camera captures the image of the module in the visible spectrum, which is then processed to extract the region of interest. The features of the image are extracted using various functions from the RGB; hue, saturation, and value (HSV); and grayscale channels, which are included in MATLAB's Balu toolbox,[36] as shown in Table 1. Mery *et al.*[37] presented a more detailed description of the extracted features and the used functions. The selection tool is then used as a statistical analyzer based on linear correlation, to identify and eliminate the redundant or irrelevant features in the analysis. Subsequently, the sequential forward selection (SFS) method is used to rank the features based on the correlation coefficient between the features and an energy loss label. The algorithm searches for the best subset of features that linearly correlates with the target label by first identifying the single best linearly correlated feature and then gradually adding more and more features until the $R^2$ score of a linear model constructed with the selected features reaches its maximum value. Last *et al.* provided an in-depth explanation of the SFS algorithm.[38] Lastly, the extracted information is processed by an ANN that estimates soiling impact and, by extension, the module performance. In this neural network, the selected features are used as the input layer. The number of neurons and hidden layers in a neural network are typically determined based on the research objective because there is no established procedure to determine these factors.[39] In this study, the number of hidden layers is determined through trial and error, optimizing the RMSE value, which is similar to the approach used in PV power forecasting.[40]

## 3 | EXPERIMENTAL SETUP

The operational conditions of the PV modules are analyzed at the solar energy laboratory at the Universidad Técnica Federico Santa María, located in Santiago. Two AstroEnergy, model ASM6612P, modules with a nominal power of 315 W are analyzed in this study. Each module is individually connected to a micro-inverter, which is in turn connected to the campus electrical grid. Zelada presented a more detailed description of the laboratory setup.[41] The voltage, current, and cell temperature of each module are measured individually. The environmental variables are continuously monitored by a Campbell

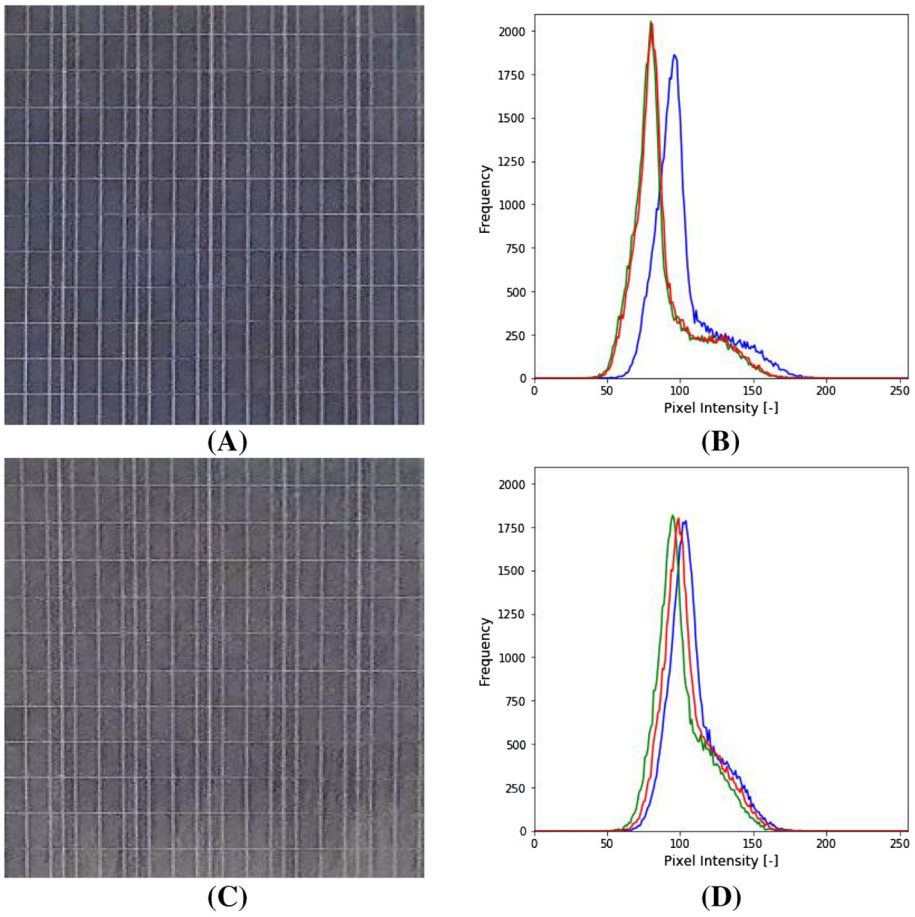


**FIGURE 1** Red, green, and blue (RGB) frequency histograms for (A and B) clean module and (C and D) soiled module.

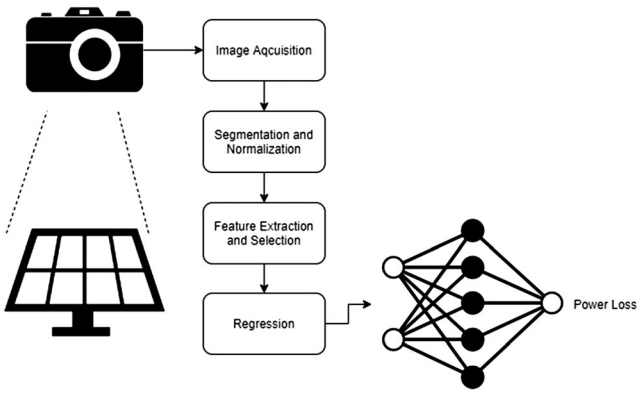


**FIGURE 2** Algorithm flowchart.

**TABLE 1** Features extracted for every channel.

| Channel | Features |
|---|---|
| • Grayscale<br>• Red<br>• Green<br>• Blue<br>• Hue<br>• Saturation<br>• Value | • Standard (mean, standard deviation, kurtosis, Laplacian, skewness)<br>• Local binary pattern<br>• Fourier and DCT<br>• Contrast |

Scientific Inc. RSR100 weather station, which can record the global horizontal, diffuse, and plane of array irradiance, along with the ambient temperature, wind speed, wind direction, and relative humidity. The data collected from the modules and stations are recorded every minute. The images are captured by a BOSCH IP9000 stationary camera, which is configured to photograph each module in the visible spectrum every 10 min. The images were captured between May and July 2019.

Figure 3 depicts the experiment setup, which comprises a reference module (left) that is cleaned daily. This module represents the optimal generation in the absence of soiling. The second module (right) is exposed to the accumulation of soiling because of natural causes at its specific location.

The electrical and environmental datasets are cleaned and filtered. The abnormalities of the experimental setup caused due to partial shading by the modules or failures in the recording equipment are eliminated. Similarly, the images that do not provide relevant information for the analysis because of either camera focus problems, sun reflection on the modules, or occlusion are not considered. After a preliminary analysis, a stable range of hours is selected (1:00 p.m. to 3:00 p.m., UTC-4), where there is no partial shading because of the camera or nearby structures, the sun does not reflect on the surface of modules (from the camera point of view), and there is an adequate level of irradiation. In the experiment, it was observed that greater noise was present in the

images captured under irradiances greater than 1100 W/m$^2$ because of the intensity of the sun. Therefore, the images captured under these conditions are discarded. To minimize variability in the scene lighting, the instances where irradiance lies within the range of 700 and 1000 W/m$^2$ are selected. The images are manually reviewed to eliminate any occlusion of the modules in case the maintenance staff are on site. Furthermore, the data with any irregularities in the current, voltage, and irradiance measurements are eliminated, and the final dataset contains 377 images.

## 4 | IMAGE PROCESSING

### 4.1 | Segmentation and normalization

After the set of valid images is selected, the images are segmented and normalized. The perspective is corrected to determine the region of interest in a homogeneous format by using an affine transformation, as shown in Figure 4. The perspective involves manually selecting the four corners of the modules being analyzed. The images are resized to 224 × 224 pixels.

### 4.2 | Labeling

The images are labeled based on the loss of power generation suffered by the module exposed to soiling, corresponding to the amount of power generated by the reference module at the same moment. The power loss is calculated as follows:

$$Loss = 100 \cdot \frac{P_{ref} - P_{mod}}{P_{ref}}, \quad (1)$$

**FIGURE 3** Reference module (left) next to a soiled module (right).

where $P_{ref}$ represents the power produced by the clean reference module and $P_{mod}$ represents the power generated by the module under evaluation; both the modules operate side by side under the same weather conditions. The resultant dataset contains 377 images, which are captured between 1:00 pm and 3:00 pm over a period of 2 months. Each image is labeled with its respective power loss value. In this experimental setup, the loss values range from 0% to 20%.

### 4.3 | Feature extraction and selection

The resultant database, which comprises the selected images and their respective labels, is processed with the help of the Balu toolbox,[36] in which the feature extraction tool and selection routines are used. Balu enables the automation of the feature extraction for a set of images. In this case, the loss generation parameter defined in Equation 1 is used as the target label. Table 2 presents the five selected features, where the skewness in the grayscale channel is the highest-ranking feature according to the SFS method. Figure 5 depicts the linear model constructed using the SFS method. This model has an $R^2$ score of 0.82 and an RMSE of 1.64%; these values will be used to compare the performance of the models designed in the next section. Irradiance was not selected as a feature by the SFS method; however, neural networks in the next step will be built using irradiance as a feature to analyze its impact on the accuracy of the model. Lastly, Figure 6 depicts a plot of loss against skewness in grayscale, which is the most important feature according to the SFS method, where a clear correlation is observed ($R^2$ score of 0.72).

## 5 | ANN CLASSIFIER

An ANN is built in MATLAB using the information extracted from the images, as shown in Figure 7. The proposed neural network requires a 5 × 1 dimension input layer, corresponding to the selected features of the images (a 6 × 1 dimension layer is used if irradiation is included as a feature). The output layer is a 1 × 1 dimension parameter that represents the power loss of the module. Various experiments are performed to determine the number of hidden layers and the neural network with the lowest RMSE. The activation functions of the hidden and output layers are defined as follows:

$$f(x_i) = \sum w_i x_i + b, \quad (2)$$

where $w_i$ represent the weights, $x_i$ represent the features, and $b$ represents the intercept coefficient for the linear approximation in every layer.

The dataset is divided into three random partitions, where 70% is used to train the network, 15% to validate each training stage, and the remaining 15% is used to test the trained network. These divisions on the dataset minimize the overfitting on the neural network. The training and validation sets are used to adjust the weights in every

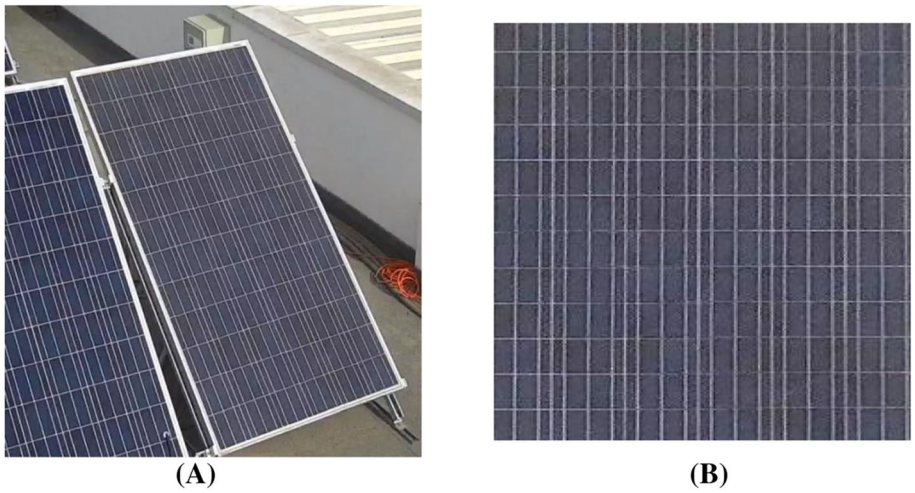


**FIGURE 4** (A) Original capture and (B) perspective corrected image.

**TABLE 2** Most relevant features ranked by Pearson correlation.

| Channel | Feature | Accumulated $R^2$ |
|---|---|---|
| Gray | Skewness | 0.72 |
| Saturation | Mean Laplacian | 0.76 |
| Hue | Standard deviation | 0.79 |
| Hue | Mean | 0.81 |
| Blue | Skewness | 0.82 |

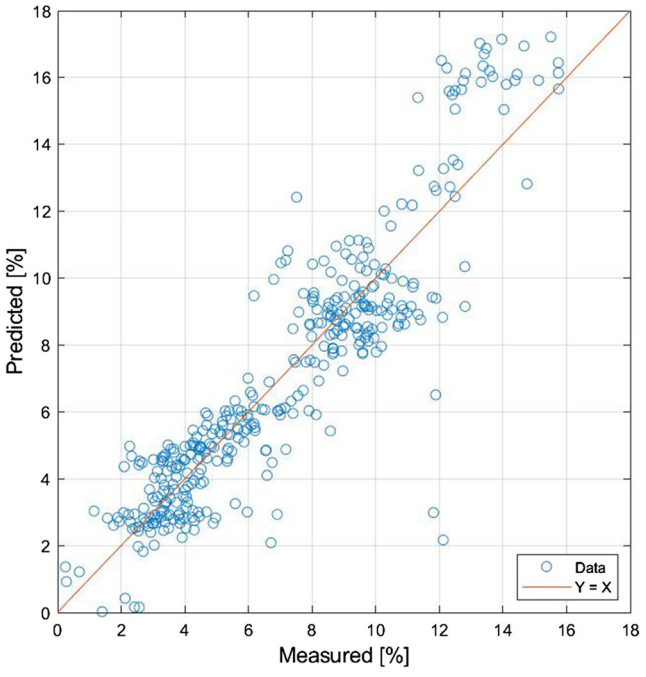


**FIGURE 5** Linear model constructed using selected features to predict loss power as a function of measured loss power.

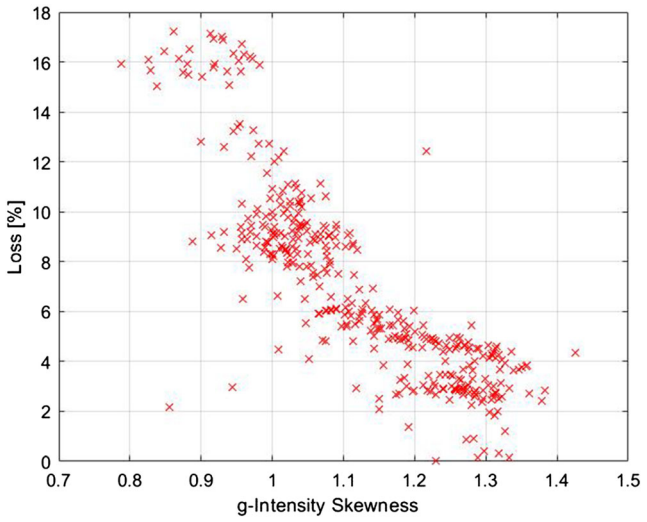


**FIGURE 6** Loss power as a function of skewness in grayscale.

hidden layer of the network, and the test set is used to evaluate the performance of the trained network. This process of randomly dividing the dataset into three groups, followed by training and testing, is repeated 10 times to ensure that the model performs adequately for unseen data. The scores obtained from every training and testing iteration are averaged to represent the performance of all the configurations that were analyzed. Table 3 presents the average performance of the network in the test sets for different numbers of hidden layers. The smallest RMSE of the test set is obtained when the network has 20 hidden layers; therefore, this configuration is selected for the proposed neural network. The correlation coefficient, $R^2$, indicates the dependence of the variation of the target label on the selected features. The maximum correlation coefficient is achieved with the

lowest RMSE, when the network has 20 layers. It can be observed that the optimal ANN presents better overall scores when compared with the accuracy of the linear model constructed in the previous section. There is no significant change in the RMSE when irradiance is included as a feature.

Figure 8 depicts the performance of the proposed network. The continuous line represents the ideal case where the predictions perfectly concur with the test set data. It is observed that the samples are close to the ideal case throughout the entire set, within an RMSE of 0.74% in a dataset with power loss values ranging from 0% to 20%. Therefore, it is concluded that the classifier performs accurately. Figure 9 displays the absolute error distribution for the entire dataset, as the difference between predicted and measured power loss. In most of the sets, the error is low, except for isolated anomalies. Furthermore, there is no skewness in the distribution, implying that the proposed classifier does not underestimate or overestimate soiling impact. Hence, the classifier developed in this study is a competent tool for evaluating the natural soiling of PV modules studied in this dataset, predicting with high accuracy soiling energy losses on PV modules. It is important to note that the results shown in this section only apply to the studied dataset. Future work will focus on validating the algorithm in multiple datasets with naturally accumulated soiling.

## 6 | APPLICATION IN EXTERNAL DATASETS

Deep Solar Eye[34] is an experiment conducted by researchers at IBM to estimate the impact of soiling on the power generation of PV modules using CNNs. This method presents some similarities to the experimental setup used in our research, such as the use of a stationary camera, a reference module that is constantly cleaned, and the relative power loss production metric. Because of these similarities, we have applied Deep Solar Eye's image dataset to the proposed methodology. The key difference between the two experiments is in the use of the images. In the Deep Solar Eye, full images are used as inputs in the CNN architecture, where features are extracted and processed entirely within the neural network to predict power losses in a categorical manner. Conversely, our algorithm takes an image as an input, extracts the target features, and uses these data as input in a separate ANN to predict the power losses as a continuous function. Because our method uses only a small set of features obtained from the input image, it requires less computational resources than the Deep Solar Eye method.

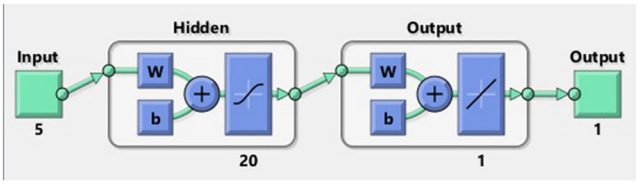


**FIGURE 7** Artificial neural network designed to evaluate soiling.

The set of images obtained from the Deep Solar Eye experiment illustrates different soiling patterns produced by various particles, primarily because of artificial soiling in cases where dust or other substances are manually deposited on the module surface. This type of soiling is not common in power plants. The dataset contains more than 45 000 different images, where each one is labeled with the plane of array irradiance observed at the moment of image capture. The images also present the relative power loss of the module by comparing the power output to a clean healthy module (without failures), with values ranging from 0% to 100%. The proposed methodology is applied to the Deep Solar Eye image dataset using the Balu toolbox to extract the image features. The images that fit the power loss between 0% and 20% are selected, similar to the datasets used in this study.

Table 4 presents the features extracted from the Deep Solar Eye dataset based on the methodology presented in Section 4.3. The accumulated $R^2$ is significantly lower than the value obtained for our dataset. The SFS method does not yield accurate results for the Deep Solar Eye dataset. Figure 10 depicts the saturation skewness, which is the most relevant feature based on the $R^2$ score, where no correlation can be observed. A new neural network is trained on the Deep Solar Eye data, as shown in Figure 11. This network is constructed by using the five most relevant features, which are taken as the input, and considering 20 hidden layers, similar to the previous experiment. It can be observed that the results of the proposed network vary considerably from the target values. Figure 12 presents the samples from the images in (a) our dataset and (b) the Deep Solar Eye dataset, which can explain this variation. The soiling patterns observed in the two experiments are completely different, both in

| Hidden layers | RMSE–test set (%) | RMSE–test set (with irradiance) (%) | $R^2$ |
|---|---|---|---|
| 1 | 2.02 | 2.21 | 0.88 |
| 5 | 1.65 | 1.72 | 0.91 |
| 10 | 1.43 | 1.35 | 0.93 |
| 15 | 0.88 | 0.89 | 0.94 |
| 20 | 0.74 | 0.76 | 0.96 |
| 30 | 0.86 | 0.92 | 0.95 |

**TABLE 3** Mean net performance for various depth levels.

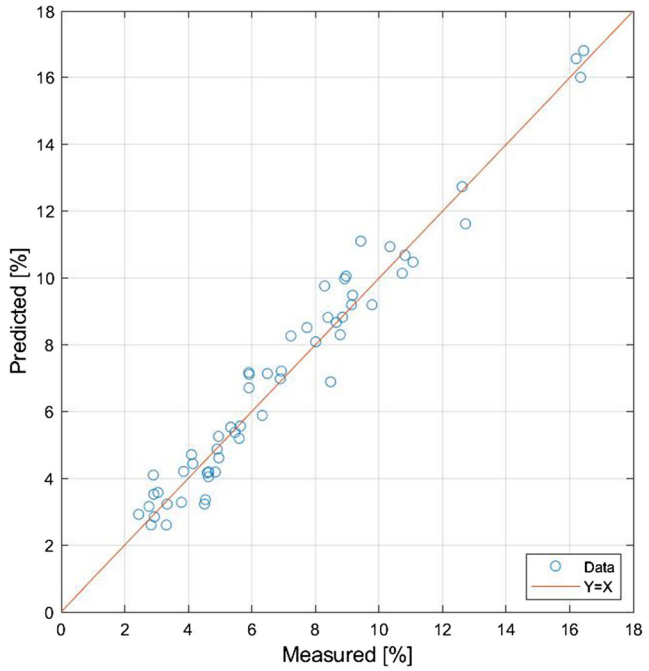


**FIGURE 8** Predicted power loss values as a function of the measured values for the test set.

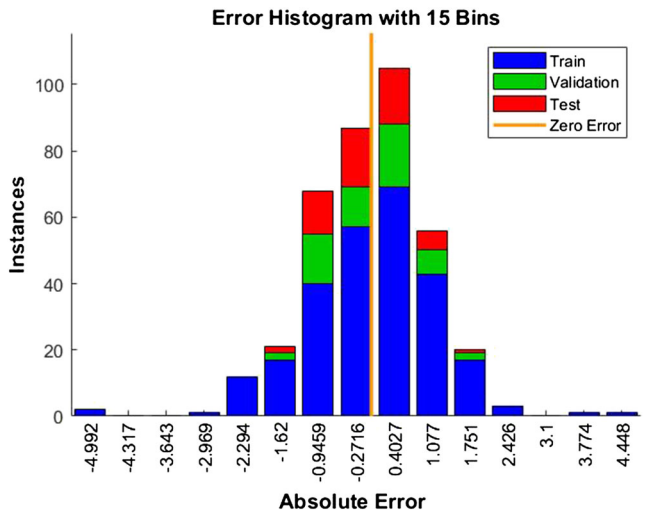


**FIGURE 9** Absolute error distribution for the complete dataset with 15 bins.

terms of particle types and distribution over the module surface. The results of the accumulated $R^2$ and the new neural network indicate that this methodology cannot be used to create a competent classifier for Deep Solar Eye's dataset, which suggests that it might be ineffective for datasets containing artificial soiling. Future work will be focused on evaluating the algorithm on different artificial soiling patterns and particle types.

**TABLE 4** Most relevant features ranked by Pearson correlation in the Deep Solar Eye Dataset.

| Channel | Feature | Accumulated $R^2$ |
|---|---|---|
| Saturation | Skewness | 0.081 |
| Green | Local binary pattern | 0.092 |
| Hue | Contrast *K* | 0.112 |
| Saturation | Mean Laplacian | 0.125 |
| Hue | Standard deviation | 0.130 |

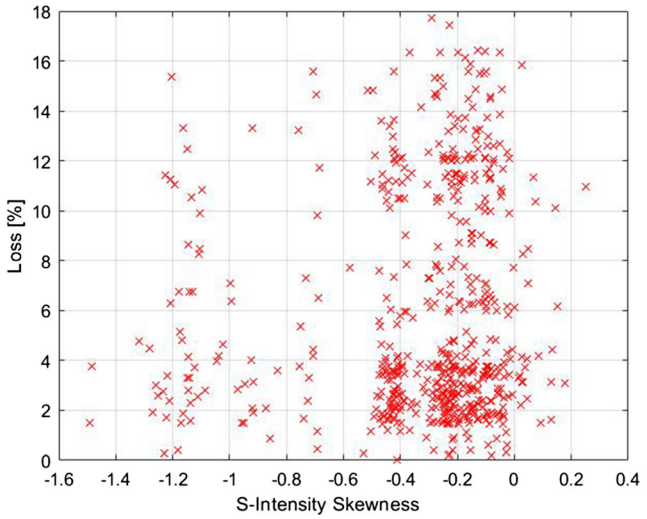


**FIGURE 10** Loss power as a function of skewness of the saturation from the Deep Solar Eye dataset.

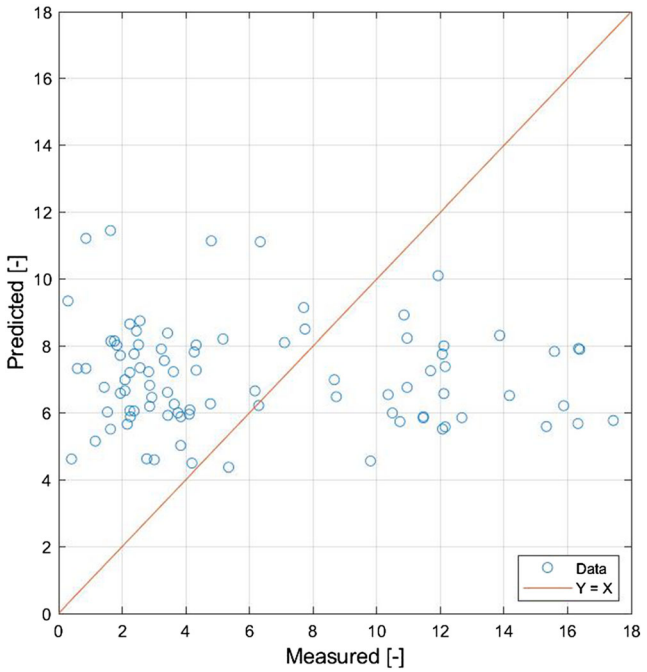


**FIGURE 11** Predicted power loss values as a function of the measured ones for the Deep Solar Eye dataset.

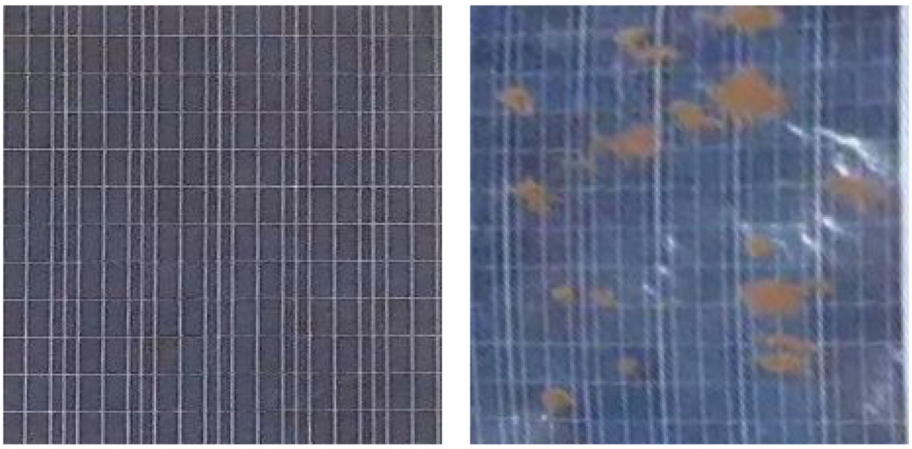


**FIGURE 12** Sample images from (A) own dataset and (B) Deep Solar Eye.

## 7 | CONCLUSIONS

This study presents a novel methodology that quantifies the impact of naturally accumulated soiling on a solar PV module performance through RGB image analysis. In the proposed solution, the images of a PV module are captured in the visible spectrum. These images are segmented and normalized to extract the region of interest in a homogeneous format. The relevant features that describe the performance of the module are then obtained through statistical analysis. Lastly, the selected features are fed to the ANN, and the performance is analyzed. The proposed method does not depend on the irradiance data as an input, as long as the values lie in the range of 700–1000 W/m$^2$, achieving an RMSE of 0.7% and $R^2$ of 0.96 for power loss. The dataset included power loss values ranging from 0% to 20%, demonstrating the effectiveness of the classifier in the defined intervals of time and irradiance for the studied dataset. This methodology can be implemented in any installation because of its low complexity because it only requires a camera and a clear sky, thereby reducing the requirement for expensive sensors or high-maintenance monitoring techniques. Further testing is required to validate the algorithm on datasets with different technologies and locations.

Furthermore, it is observed that the skewness in the gray-scale channel correlates to the module performance within the studied dataset. When the methodology was tested on an external dataset based on artificial soiling, the classifier was unable to effectively diagnose the images using the previously trained neural network or by training a new neural network. However, this type of soiling is not common in PV plants. Therefore, results suggest that the proposed methodology might only be suitable for the accurate prediction of PV power loss in modules subject to the natural accumulation of soiling. It is important to note that the results shown in this work only apply to the datasets under study. With additional datasets with both natural and accumulated soiling, the proposed methodology may be used to generalize these results.

### ACKNOWLEDGEMENTS

The authors appreciate the financial support from projects ANID/FONDECYT/1221878, ANID/FONDEF/ID21I10424, ANID/FONDEQUIP/EQM200183, and ANID/FONDAP/15110019 "Solar Energy Research Center" (SERC), Chile. In addition, Robinson Cavieres would like to acknowledge the Universidad Técnica Federico Santa María for the financial support through a Scientific Initiation and Innovation Program PIIC.

### DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### ORCID

*Rodrigo Barraza* https://orcid.org/0000-0002-6737-9145
*Patricio Valdivia* https://orcid.org/0000-0002-7457-1346

### REFERENCES

1. Green MA, Dunlop ED, Hohl-Ebinger J, et al. Solar cell efficiency tables (version 60). *Prog. Photovoltaics Res. Appl.* 2022;30(7):687-701. doi:10.1002/pip.3595
2. Besharat F, Dehghan AA, Faghih AR. Empirical models for estimating global solar radiation: a review and case study. *Renew Sustain Energy Rev*. 2013;21:798-821. doi:10.1016/j.rser.2012.12.043
3. Köntges M, Kurtz S, Packard CE, et al. Review of failures of photovoltaic modules. Report IEA-PVPS T13-01; 2014. ISBN 978-3-906042-16-9.
4. Kala P, Joshi P, Agrawal S, Yadav LK, J.M. Introduction to condition monitoring of PV system. In: *Soft Computing in Condition Monitoring and Diagnostics of Electrical and Mechanical Systems*. Singapore: Springer; 2020.
5. Ahmed R, Sreeram V, Mishra Y, Arif MD. A review and evaluation of the state-of-the-art in PV solar power forecasting: techniques and optimization. *Renew Sustain Energy Rev*. 2020;124:109792. doi:10.1016/j.rser.2020.109792
6. National Renewable Energy Laboratory (NREL). (2018). Best Practices for Operation and Maintenance of Photovoltaic and Energy Storage Systems; 3rd Edition. *Nrel/Tp-7a40-73822*, (December), 153.
7. Zorrilla-Casanova J, Piliougine M, Carretero J, et al. Losses produced by soiling in the incoming radiation to photovoltaic modules.

*Prog Photovoltaics Res Appl*. 2013;21(4):790-796. doi:10.1002/pip.1258

8. Xu L, Li S, Jiang J, et al. The influence of dust deposition on the temperature of soiling photovoltaic glass under lighting and windy conditions. *Sol. Energy*. 2020;199:491-496. doi:10.1016/j.solener.2020.02.036
9. Ullah A, Amin A, Haider T, Saleem M, Zafar N. Investigation of soiling effects, dust chemistry and optimum cleaning schedule for PV modules in Lahore, Pakistan. *Renew Energy*. 2020;150:456-468. doi:10.1016/j.renene.2019.12.090
10. Azouzoute A, Zitouni H, El Ydrissi M, et al. Developing a cleaning strategy for hybrid solar plants PV/CSP: case study for semi-arid climate. *Energy*. 2021;228:120565. doi:10.1016/j.energy.2021.120565
11. Sayyah A, Horenstein MN, Mazumder MK. Energy yield loss caused by dust deposition on photovoltaic panels. *Sol. Energy*. 2014;107:576-604. doi:10.1016/j.solener.2014.05.030
12. Figgis B, Ennaoui A, Ahzi S, Rémond Y. Review of PV soiling particle mechanics in desert environments. *Renew Sustain Energy Rev*. 2017;76(February):872-881. doi:10.1016/j.rser.2017.03.100
13. Javed W, Guo B, Figgis B, Aïssa B. Dust potency in the context of solar photovoltaic (PV) soiling loss. *Sol Energy*. 2021;220:1040-1052. doi:10.1016/j.solener.2021.04.015
14. Micheli L, Deceglie MG, Muller M. Predicting photovoltaic soiling losses using environmental parameters: an update. *Prog. Photovoltaics Res. Appl.* 2019;27(3):210-219. doi:10.1002/pip.3079
15. Al-Addous M, Dalala Z, Alawneh F, Class CB. Modeling and quantifying dust accumulation impact on PV module performance. *Sol. Energy*. 2019;194:86-102. doi:10.1016/j.solener.2019.09.086
16. Olivares D, Ferrada P, Bijman J, et al. Determination of the soiling impact on photovoltaic modules at the coastal area of the Atacama Desert. *Energies*. 2020;13(15):3819. doi:10.3390/en13153819
17. Haddad AG, Dhaouadi R. Modeling and analysis of PV soiling and its effect on the transmittance of solar radiation. *Adv Sci Eng Technol Int Conf ASET*. 2018;2018:1-5.
18. Smestad GP, Germer TA, Alrashidi H, et al. Modelling photovoltaic soiling losses through optical characterization. *Sci Rep*. 2020;10(1):58. doi:10.1038/s41598-019-56868-z
19. Muller M, Micheli L, Solas AF, et al. An in-depth field validation of "DUSST": a novel low-maintenance soiling measurement device. *Prog. Photovoltaics Res. Appl.* 2021;(8):1-15. doi:10.1002/pip.3415
20. Alonso-Montesinos J, Martínez FR, Polo J, Martín-Chivelet N, Batlles FJ. Economic effect of dust particles on photovoltaic plant production. *Energies*. 2020;13(23):6376. doi:10.3390/en13236376
21. Gostein M, Bourne B, Farina F, Stueve B. (2019) Field evaluation of Mars™ optical soiling sensor. *36th Eur*. Photovolt Sol Energy. Conf *Exhib*. 1471-1473.
22. FRACSUN. Photovoltaic vs. Optical Soiling Measurement. 1–7.
23. Tsanakas JA, Chrysostomou D, Botsaris PN, Gasteratos A. Fault diagnosis of photovoltaic modules through image processing and Canny edge detection on field thermographic measurements. *Int J Sustain Energy*. 2015;34(6):351-372. doi:10.1080/14786451.2013.826223
24. Niazi K, Akhtar W, Khan HA, Sohaib S, Nasir A.K. (2018) Binary Classification of Defective Solar PV Modules Using Thermography. *2018 IEEE 7th World Conf. Photovolt. Energy Conversion, WCPEC 2018 - A Jt. Conf. 45th IEEE PVSC, 28th PVSEC 34th EU PVSEC*, 753–757.
25. Pivem T, Oliveira de Araujo FD, Oliveira de Araujo LD, de Oliveira GS. Application of a computer vision method for soiling recognition in photovoltaic modules for autonomous cleaning robots. *Signal Image Process an Int J*. 2019;10(3):43-59. doi:10.5121/sipij.2019.10305
26. Yap WK, Galet R, Yeo KC. (2015) Quantitative Analysis of Dust and Soiling on Solar PV Panels in the Tropics Utilizing Image-Processing Methods. Asia - Pacific Sol. Res. Congr., (2014).
27. Tribak H, Zaz Y. (2019) Dust Soiling Concentration Measurement on Solar Panels based on Image Entropy. *Proc. 2019* 7th Int. Renew. Sustain. Energy Conf. IRSEC 2019.
28. Deitsch S, Christlein V, Berger S, et al. Automatic classification of defective photovoltaic module cells in electroluminescence images. *Solar Energy*. 2019;185:455-468.
29. Karimi AM, Fada JS, Liu J, Braid JL, Koyuturk M, French RH. (2018) Feature extraction, supervised and unsupervised machine learning classification of PV cell electroluminescence images. *2018 IEEE 7th World Conf. Photovolt. Energy Conversion, WCPEC 2018 - A Jt. Conf. 45th IEEE PVSC, 28th PVSEC 34th EU PVSEC*, 418–424.
30. Rey G, Kunz O, Green M, Trupke T. (2022) Luminescence imaging of solar modules in full sunlight using ultranarrow bandpass filters. *Prog. Photovoltaics Res. Appl.*, (March), 1–7.
31. Bommes L, Hoffmann M, Buerhop-Lutz C, et al. Anomaly detection in IR images of PV modules using supervised contrastive learning. *Prog. Photovoltaics Res. Appl.* 2022;30(6):597-614. doi:10.1002/pip.3518
32. Dunderdale C, Brettenny W, Clohessy C, van Dyk EE. Photovoltaic defect classification through thermal infrared imaging using a machine learning approach. *Prog. Photovoltaics Res. Appl.* 2020;28(3):177-188. doi:10.1002/pip.3191
33. Wenjie Z, Liu S, Member OG, Member G, Member HQ, Srinivasan D. *Deep-Learning-Based Probabilistic Estimation of Solar PV Soiling Loss*. 2021;3029(c):1-9.
34. Mehta S, Azad AP, Chemmengath SA, Raykar V, Kalyanaraman S. (2018) DeepSolarEye: Power Loss Prediction and Weakly Supervised Soiling Localization via Fully Convolutional Networks for Solar Panels. *Proc. - 2018 IEEE Winter Conf. Appl. Comput. Vision, WACV 2018*, **2018-Janua**, 333–342.
35. Cavieres R, Barraza R, Estay D, Bilbao J, Valdivia-Lefort P. Automatic soiling and partial shading assessment on PV modules through RGB images analysis. *Appl Energy*. 2022;306:117964. doi:10.1016/j.apenergy.2021.117964
36. Mery, D. (2011) BALU: A Matlab toolbox for computer vision, pattern recognition and image processing.
37. Mery D, Pedreschi F, Soto A. Automated design of a computer vision system for visual food quality evaluation. *Food Bioproc Tech*. 2013;6(8):2093-2108. doi:10.1007/s11947-012-0934-2
38. Last M, Kandel A, Maimon O. Information-theoretic algorithm for feature selection. *Pattern Recognit Lett*. 2001;22(6–7):799-811. doi:10.1016/S0167-8655(01)00019-8
39. Grimaccia F, Leva S, Mussetta M, Ogliari E. ANN sizing procedure for the day-ahead output power forecast of a PV plant. *Appl Sci*. 2017;7(6):622. doi:10.3390/app7060622
40. Netsanet S, Zheng D, Zhang L, Hui M. (2017) Input parameters selection and accuracy enhancement techniques in PV forecasting using artificial neural network. *2016 IEEE Int. Conf. Power Renew*. Energy, ICPRE 2016, 565–569.
41. Zelada, W. (2019) "DISEÑO E IMPLEMENTACIÓN DE UN LABORATORIO PARA ESTUDIO DE FALLAS EN PANELES FOTOVOLTAICO."

**How to cite this article:** Cavieres R, Salas J, Barraza R, Estay D, Bilbao J, Valdivia P. Estimation of the impact of natural soiling on solar module operation through image analysis. *Prog Photovolt Res Appl*. 2023;31(7):690-699. doi:10.1002/pip.3676