

Article

PV Module Soiling Detection Using Visible Spectrum Imaging and Machine Learning

Boris I. Evstatiev ^{1,*}, Dimitar T. Trifonov ¹, Katerina G. Gabrovska-Evstatieva ², Nikolay P. Valov ¹
and Nicola P. Mihailov ¹

¹ Faculty of Electrical Engineering, Electronics, and Automation, University of Ruse “Angel Kanchev”, 7004 Ruse, Bulgaria; dtrifonov@uni-ruse.bg (D.T.T.); npvalov@uni-ruse.bg (N.P.V.); mihailov@uni-ruse.bg (N.P.M.)

² Faculty of Natural Science and Education, University of Ruse “Angel Kanchev”, 7004 Ruse, Bulgaria; kgg@ami.uni-ruse.bg

* Correspondence: bevstatiev@uni-ruse.bg

Abstract: During the last decades photovoltaic solar energy has continuously increased its share in the electricity mix and has already surpassed 5% globally. Even though photovoltaic (PV) installations are considered to require very little maintenance, their efficient exploitation relies on accounting for certain environmental factors that affect energy generation. One of these factors is the soiling of the PV surface, which could be observed in different forms, such as dust and bird droppings. In this study, visible spectrum data and machine learning algorithms were used for the identification of soiling. A methodology for preprocessing the images is proposed, which puts focus on any soiling of the PV surface. The performance of six classification machine learning algorithms is evaluated and compared—convolutional neural network (CNN), support vector machine (SVM), random forest (RF), k-nearest neighbor (kNN), naïve-Bayes, and decision tree. During the training and validation phase, RF proved to be the best-performing model with an F1 score of 0.935, closely followed by SVM, CNN, and kNN. However, during the testing phase, the trained CNN achieved the highest performance, reaching F1 = 0.913. SVM closely followed it with a score of 0.895, while the other two models returned worse results. Some results from the application of the optimal model after specific weather events are also presented in this study. They confirmed once again that the trained convolutional neural network can be successfully used to evaluate the soiling state of photovoltaic surfaces.

Keywords: soiling; photovoltaic; convolutional neural network (CNN); machine learning; imaging; classification



Citation: Evstatiev, B.I.; Trifonov, D.T.; Gabrovska-Evstatieva, K.G.; Valov, N.P.; Mihailov, N.P. PV Module Soiling Detection Using Visible Spectrum Imaging and Machine Learning. *Energies* **2024**, *17*, 5238. <https://doi.org/10.3390/en17205238>

Academic Editor: Philippe Leclère

Received: 10 September 2024

Revised: 12 October 2024

Accepted: 18 October 2024

Published: 21 October 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The installed photovoltaic (PV) power is constantly increasing worldwide and even though it is only available during the daytime, its share in the energy mix has become significant. One of the reasons for this is that in many cases PV installations require very little maintenance to operate, which is very appropriate for many small and medium-sized applications. Nevertheless, if the generated power, and therefore profit, are to be maximized, these facilities require regular maintenance.

The experience gained over the last decades showed that the long-term exploitation of a photovoltaic installation could face many potential problems, whose timely identification could save a lot of trouble and expenses for the plant operators. Such faults include module mismatch caused by uneven aging; micro-cracks, often caused by mechanical stress; degradation; and hotspots caused by shading, soiling, and uneven aging [1–3]. Other factors causing the abovementioned problems include manufacturing inconsistencies, temperature fluctuations, long UV exposure, extreme weather events, incorrect installation, and snow [1,4,5]. Some of the abovementioned factors can only be identified and acknowledged, while others, such as shading and soiling, can be mitigated. The most direct approach to

this is the application of different Maximum Power-Point Tracking (MPPT) algorithms, utilized by the contemporary inverters and solar controllers [6,7].

Soiling is known to be a problem for the operation of photovoltaic installations, which could significantly lower the power produced. Different studies have reported reductions ranging between 4% and 20% [8–12], which greatly depends on the geographic region and the available local impacts. This problem is especially important in desert climates, where the daily soiling rate could surpass 0.5% [10]. This rate is lower in regions with other climates and often has a seasonal character. For example, photovoltaic installations located near agricultural areas are known to receive a lot of soiling after the land has been plowed [13]. Similarly, pollutants created in urban and industrial areas and the presence of many birds or even snails can cause many problems for PV installation operators [14–16].

To handle this problem, PV surface cleaning is used in mid- and large-scale installations. Different cleaning methods exist, such as manual cleaning, tractors, robots, sprinklers, etc. [17–19], but all of them are relatively expensive. Therefore, the cleaning schedule should be carefully considered to ensure the maintenance costs do not surpass the potential energy gains. Different approaches exist for cleaning maintenance, based on power prediction and weather events, among others. A study in Palestine investigated the optimal frequency for cleaning PV panels [20]. The weekly cleaning of panels was compared with several other schedules—two-weekly, monthly, two-monthly, six-monthly, and annually. The results showed that the difference between weekly and two-weekly cleaning led to an output power difference varying between 0.51% during the winter months and 3.22% during the summer months. On the other hand, the difference between weekly and annual cleaning created a 13.1% increase in power production in favor of the weekly one.

In [21] the performance ratio of a PV installation was used for optimizing the scheduling of cleaning procedures. The Seasonal Autoregressive Integrated Moving Average with exogenous regressors (SARIMAX), Autoregressive integrated moving average (ARIMA), and Long Short-Term Memory (LSTM) models were investigated, out of which the first returned the highest R^2 metrics (92%). Other authors tried to optimize the PV cleaning schedule by comparing the actual and predicted PV power [22]. The proposed approach was based on the geographic location, the forecasted temperature, and the PV system specification. In [9], a model was developed that supports the decision-making process when large-scale PV installations should be cleaned based on the environmental conditions. Different soiling loss factors, rainfall characteristics, manual cleaning characteristics, and dust events were accounted for to create an empirical model. The validation was performed using different sensors, and, according to the authors, it achieved a 0.71% mean absolute error. However, the cleaning frequency requirements are different for different climate conditions and very often depend on certain instantaneous weather events such as rain storms, sand storms, and local pollution [23–25].

Several approaches exist for identifying problems with PV installations, which are based on different spectrum images, electrical, and meteorological data. In [26], a dataset of currents, voltages, temperatures, irradiation levels, and fault labels was used to predict five states of a PV installation—normal condition, degradation, short-circuit fault, open-circuit fault, and partial shading. The trained convolutional neural network (CNN) achieved a 95.20% overall accuracy and the accuracy for the different classes varied between 86.95% and 100%. Similarly, in [27], artificial intelligence (AI) for fault detection in a Saudi Arabian PV system was used. The input parameters used for the trained artificial neural network (ANN) included the maximum power, open-circuit voltage, short-circuit current, maximum power voltage, maximum power current, solar radiation, and ambient temperature. The classification included the following categories: no fault, partial shading, line-to-line fault, open-circuit fault, degradation fault, bridge fault, bypass diode fault, and hybrid fault. The optimal ANN achieved a 99.9% average performance, with the individual class success rates varying between 99.6% and 100%.

The diagnosis of PV installations can also be implemented using different spectrum images, such as infrared and electroluminescence. In [28], infrared images and machine

learning (ML) algorithms (quadratic discriminant analysis, naïve-Bayes, k-nearest neighbor, bagging ensemble, and support vector machine) were used to identify faulty and non-faulty hotspots over photovoltaic surfaces. The highest F1 score was achieved with the support vector machine (SVM), which reached 91%. Infrared imaging was also used in [29] to identify different PV surface faults, such as cracks, accumulated sand, soiling, covered modules, short-circuit modules, overheated bypass diode, and normally functioning modules. The authors developed a hybrid CNN–ML model, which achieved an overall accuracy of 88% and an F1 score varying between 74% and 97% for the different classes. UAV-obtained thermal imaging was used in [30] for fault detection and diagnosis of PV systems. The algorithms used to identify bypass diode faults and hotspots included neural networks, random forest (RF), k-nearest neighbor (kNN), and gradient boosting. All models achieved a similar F1 score, ranging between 0.930 and 0.941, with kNN having the highest performance. UAV-obtained infrared imaging was also used in [31] to investigate the automated identification and localization of defects on photovoltaic installations. The study used the You Only Look Once (YOLO) convolutional neural network to detect situations such as “disconnected substring”, “hot spot”, “disconnected string”, and “short circuit”. The obtained F1 varied between 0.42 and 0.899 for the different classes.

Another approach for fault identification is based on the electroluminescence phenomenon. In this case, the PV module is externally powered during night-time and its luminescence is observed. The main disadvantage of this method is that PV panels should be disconnected while examined. This approach was applied in [32] together with different types of CNN to identify several types of PV surface defects. The optimal model achieved an accuracy of 96.17%.

Different approaches also exist for the identification of soiling. The first one is based on the prediction of the photovoltaic energy yield. In [33], a regression model was created that predicted the energy yield based on the so-called soiling ratio, which is the ratio between the performance of a soiled PV panel and the performance of the same panel without soiling. The performance was evaluated using the Nash–Sutcliffe efficiency (R^2) and Mean Absolute Error measures. Optimal results were obtained with the Gaussian process regression, reaching $R^2 = 0.98$ during the training phase and 0.86 during the testing phase. In [34], clustering analysis and artificial neural networks were used to analyze the yield of photovoltaic installations. The goal was to detect and identify defects and degradation in the PV modules using a wide range of features such as irradiance, relative humidity, wind velocity, ambient temperature, string current, string voltage, string power, module temperature, and yield. In another study, a hybrid LSTM-KNN algorithm was proposed to predict the PV power output and power losses based on the day number, sunshine duration, humidity, temperature, solar radiation, and power output [35]. The trained model achieved a coefficient of determination of 0.9963 during the validation stage and 0.9822 during the testing stage.

Another approach in soiling analysis is the evaluation of power losses. In [11], artificial intelligence was applied to estimate the soiling losses of a PV installation. The following input parameters were used: temperature of the panel, short-circuit current, global irradiation, relative humidity, ambient temperature, atmospheric pressure, and solar altitude. The trained artificial neural network achieved a correlation coefficient of $R = 0.91$. Another study tried to predict the soiling losses of PV modules using UAV-obtained RGB images [8]. The proposed method was based on the increased brightness of soiled PV surfaces, which was used to calculate the transmission loss of the soiling layer. The study also considered the irradiation at the moment of measurement and the viewing point. Similarly, in [23] the soiling losses were estimated based on 479 RGB images and a machine-learning regression. The trained model achieved an R^2 equal to 0.98, which indicates it can be used to predict soiling loss using images made under similar conditions. In [36], the soiling losses were evaluated using a combination of images of the PV surface and time series with solar radiation as input. The power loss was analyzed by splitting it into 16 and 21 categories

and the performance of different CNN models was evaluated. The optimal ones achieved accuracies of 77% and 66%, respectively, and F1 scores of 71% and 64%, respectively.

Two main approaches exist for the identification of soiling on the PV surface. The first one is based on different electric and environmental data. Several machine learning algorithms were used in [37] to predict the necessity for cleaning the photovoltaic panels based on the arrays' voltage and current. Two categories were used: "Cleaning" and "No cleaning". Furthermore, four models were implemented based on logistic regression (LR), SVM, ANN, and RF. The highest classification accuracy (>90%) was obtained for the RF model. Similarly, in [38], raw photovoltaic solar energy data and machine learning techniques were used for identifying soiling in Cyprus. The study utilized unsupervised k-means clustering to perform a daily soiled/non-soiled classification.

Soiling can also be identified using RGB imaging. In [39], different methods for semantic segmentation based on supervised and unsupervised machine learning algorithms were investigated. The achieved F1 score for recognizing PV panel soiling varied between 83% and 85%, and the achieved accuracy was up to 98%. A total of 2231 RGB images was used in [40] to train a convolutional neural network for assessing the dusting of PV panels in Bangladesh. All randomly shaped original images were resampled to 227×227 px and were classified into "clean" and "dirty" classes. The average precision obtained was 98.2%, and the obtained precision for clean and dirty panels was 97.6% and 95.7%, respectively. In [41], a soiling recognition algorithm for application in automated PV surface cleaning robots was presented. Different approaches were used, such as segmentation and thresholding, to estimate masks of panels with different levels and distribution of dirtiness (uniform and non-uniform). The study reported a precision of up to 90%. Other authors trained a neural network for recognizing PV panels soiling based on RGB images and the MobileNetV2 pre-trained network [42]. They reported an accuracy of up to 97% with a validation loss of 9.7%. The study concluded that the output of the CNN was highly dependent on the quality of the images. In [43], satellite- and UAV-obtained images from 12 countries around the world were used. They were preprocessed using rotation and were classified into several categories: bird droppings, cement, cracks, soiling, and clean panels. Several neural networks were trained with different backbones (VGG19, MobileNet, InceptionV3, ResNet50, and EfficientNetB0). The highest precision, recall, F1 score, and accuracy were achieved by the VGG19-based model.

Another approach is the simultaneous identification of shading and soiling. In [44], a dataset of images with panels was used that had been exposed to various problems including soiling and shading. VGG-16 and VGG-19 architectures were applied for training CNNs. The obtained accuracies were 97% and 99%, and after considering the precision and recall, the obtained F1 scores were 0.85 and 0.89, respectively. Similarly, in [45], RGB images were analyzed to assess soiling and partial shading problems. A CNN was trained using several images, which allowed the authors to achieve an accuracy of 73%.

Other potentially related approaches to the investigated problem include the recognition of the PV installation extents, which might be required as a preliminary step for creating analysis masks. In [46], several machine learning algorithms (SVM, RF, and NB) for the detection of photovoltaic installations were used, relying on satellite data (Sentinel 1, Sentinel 2, PlanetScope) and different indices (Normalized Difference Vegetation Index, Normalized Difference Water Index, and Photovoltaic Spectral Index). All algorithms returned a satisfactory accuracy, although SVM and RF were the best-performing, with accuracies of above 95%. A similar approach was demonstrated in [47], where aerial images were analyzed using different tools of the GRASS GIS v.8.2 software.

The performed analysis shows that soiling can take different forms and its influence on photovoltaic installations greatly depends on its properties and chemical composition. In other words, the soiling rate, color, and impact depend on the specific geographic location and local dusting factors. To the best of our knowledge, no such studies exist for Bulgaria and the region of Ruse in particular. Many authors identify soiling based on energy yield reduction, infrared spectrum, and visible spectrum images, and when image processing is

used, segmentation and deep learning are the common approaches. The literature review showed that few authors tried to solve this problem using the classification approach in combination with image processing. Moreover, in most cases, the performance of a single machine-learning algorithm or of different versions of the same algorithm was investigated. Our analysis also showed that limited comparative results exist on the performance of different classification algorithms for the identification of clean and dirty PV panels. The above-mentioned indicates that a knowledge gap can be identified in this area.

Therefore, this paper aimed to propose and test a methodology based on the classification approach, which categorizes PV modules as either clean or dirty. It should rely on ground-based or UAV-based RGB images and different machine-learning algorithms. This study also aimed to compare their performance and identify the optimal algorithm for the proposed methodology.

2. Materials and Methods

2.1. Location and Means of the Study

The experimental photovoltaic park used in this study is located in the city of Ruse, Bulgaria, on the territory of the University of Ruse “Angel Kanchev”, geographic coordinates 43.853388283689526, 25.969217923215588 (Figure 1). Ruse is located on the boundary with Romania and is 60 km away from its capital Bucharest. It is characterized by relatively cold winters and hot summers. The average monthly temperatures vary between $-2\text{ }^{\circ}\text{C}$ in January and $24.1\text{ }^{\circ}\text{C}$ in July, and the average maximum monthly temperatures between $1\text{ }^{\circ}\text{C}$ in January and $30\text{ }^{\circ}\text{C}$ in July. The average monthly precipitation is between 36 mm and 80 mm, though during recent years the rainfall has decreased dramatically in the summer and early autumn months of July to September. The maximum solar irradiance ranges between $522\text{ W}\cdot\text{m}^{-2}$ in the winter and $1162\text{ W}\cdot\text{m}^{-2}$ in the summer. The lowest number of sunny days is in January (7.6) and the highest in August (10.9), and the number of days with partial cloudiness varies between 19.5 and 25 in the different months. The number of days with precipitation varies between 9.2 in the autumn and 15.4 in June. The average annual relative humidity ranges between 60% and 80%, and the average wind speed at the site of the PV installation is $0.38\text{ m}\cdot\text{s}^{-1}$. The presented meteorological data were obtained from [48] and from a Vantage Pro2 meteorological station by Davis Instruments (Hayward, Charlotte, CA, USA).

Two main soiling factors exist in the region. On one hand, the Ruse region is an agricultural area, which, combined with the climate, also creates a soiling on the photovoltaic modules, especially during the summer and early autumn. On the other hand, there are many industries in the city such as chemical plants, factories for aluminum components, granite tiles, and faience, among others.

The PV Park Kanev is an experimental facility of the University of Ruse with a cumulative installed power of 12.6 kWp. It includes three strings with 12 BSM350P-72 polycrystalline panels by Bluesun Solar Co. Ltd. (Shushan District, Hefei, China). They were installed in 2022 and have never been cleaned since then.

In this study, we relied on the Orange Data Mining v.3.36 software [49], developed and maintained by the University of Ljubljana (Ljubljana, Slovenia). It is a free data processing tool available for Microsoft Windows, Mac OS, and other operating systems. Orange provides a wide range of instruments for classification, clustering, testing, validation, visualization, and others. Furthermore, its functionality can be extended with Python scripts, which makes it an excellent choice for implementing machine learning experiments. For the abovementioned reasons, it was chosen as the main instrument of our research, as it fully met the requirements of this study.



Figure 1. Geographic location of the experimental photovoltaic park.

2.2. Methodology of the Study

The experimental methodology developed and implemented in this study is summarized in Figure 2 and included the following steps:

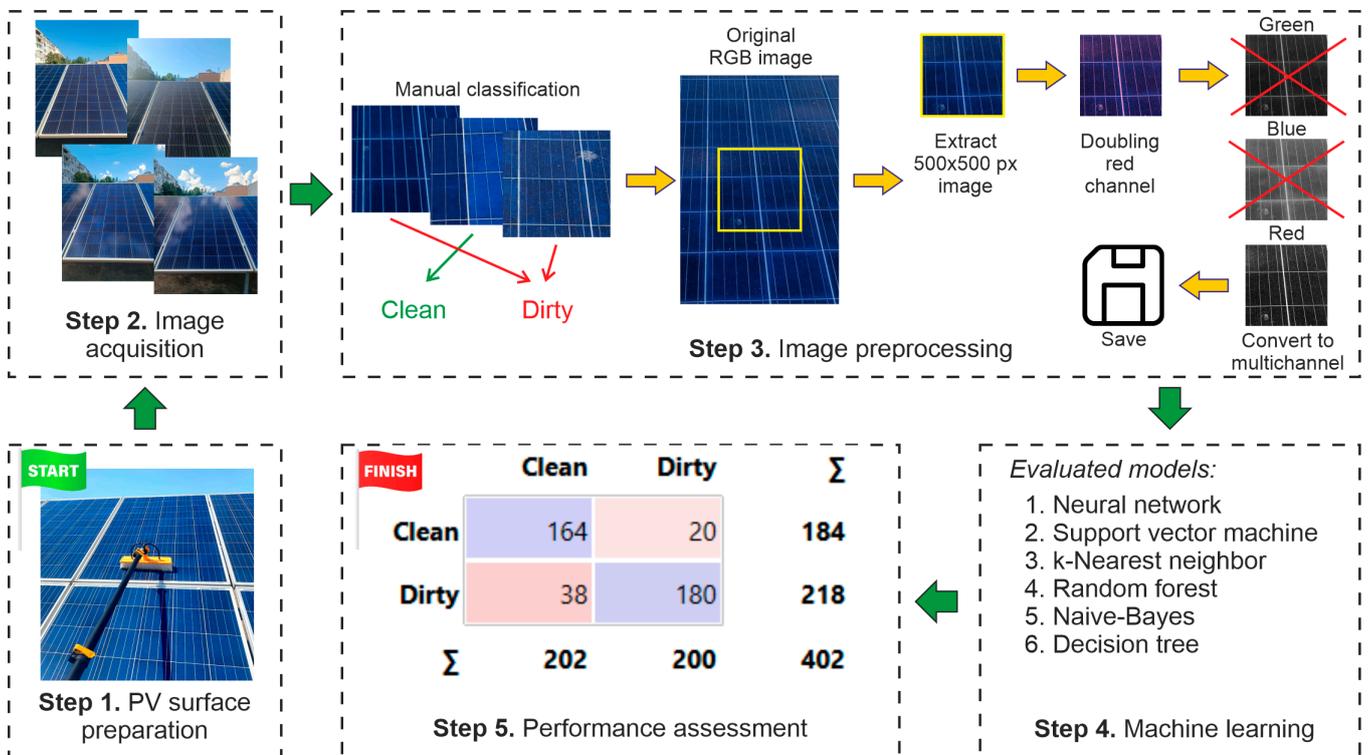


Figure 2. Summary of the study methodology.

Step 1. The surface of the selected photovoltaic modules is cleaned.

Half of the PV panels (18) are cleaned early in the morning (between 7 and 8 AM) according to the scheme presented in Figure 3. This task is implemented manually and it is assumed the corresponding photovoltaic surfaces will remain clean for the next couple of days.

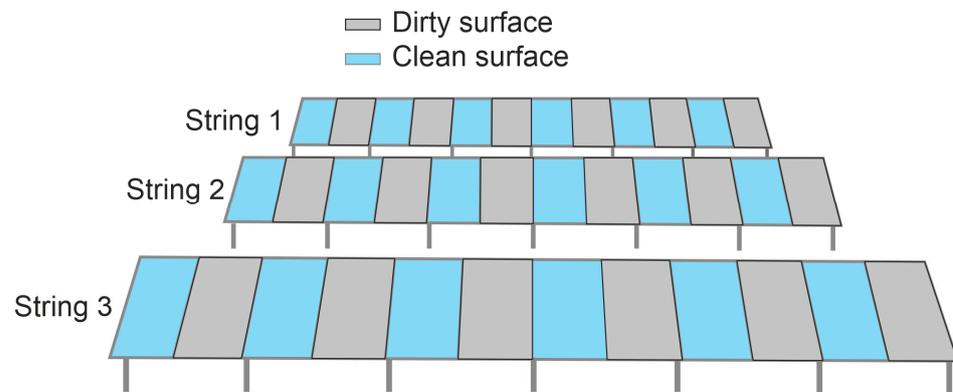


Figure 3. General schematic of the experimental setup.

Step 2. Multitemporal photos of each photovoltaic surface are made.

After the corresponding PV panels are cleaned, photos of all PV panels are taken using a camera. This scenario is repeated numerous times during the day at different astronomic hours, from early morning until late evening. Previous studies have shown that the angle of view and the sun's position could influence the ability to quantify soiling [50]. Therefore, the idea is to ensure numerous images under various lighting conditions.

In the current study, this step was implemented with on-the-ground means; however, in general cases this could also be achieved using cameras mounted on unmanned aerial vehicles or satellites.

Step 3. Image preprocessing.

The image preprocessing is aimed at preparing training and validation datasets and includes the following subtasks:

1. Preliminary filtering of the images is performed and those with inappropriate quality are removed;
2. The original PV module images are manually classified as either dirty or clean.
3. Furthermore, they are divided into training and validation datasets.
4. From the created datasets, 500×500 px images of the photovoltaic surfaces are cropped. When extracting the image fragments, the following requirements are defined:
 - a. They should contain the area of at least one PV cell.
 - b. They should not contain areas not part of the PV surface.

Previous studies have shown that color is an important feature for improving the performance of machine learning algorithms [51]. During the preliminary investigations and analysis, it was noticed that soiling contrasts better on the red channel of the photos. Therefore, the following subtasks were performed to focus on the dirty fragments of the PV panels:

5. The value of the image red channel is doubled.
6. The obtained image is converted to a multichannel format, where each channel (red, green, and blue) is represented with 256 shades of gray.
7. Finally, only the red channel (represented as gray with 256 shades) is saved for further analysis.

Step 4. Training machine learning models.

Several machine learning algorithms were selected for comparison: convolutional neural network, support vector machine, k-nearest neighbor, random forest, decision tree (DT), and naïve-Bayes (NB). The reason for choosing them was that from previous studies they are known to provide good results when dealing with photovoltaic surface analysis and image recognition [28,29,37,52]. Next, the prepared image datasets were represented with a vector of numbers using Google's Interception v3 deep neural network. Furthermore, the hour of the day (taking values from 0 to 23) was added as an additional feature, representing the hour in which the photo was taken. The reason for this is that the reflectance of the sun over the photovoltaic surfaces might significantly change the way they look. Finally, the selected features were fed to the 6 machine-learning models for training and validation.

Step 5. Performance assessment of the trained models.

The performance of each model is assessed to obtain the optimal one. This is achieved using several metric indicators, whose meaning is explained below:

- Accuracy—measures the overall correctness of the model:

$$\text{Accuracy} = \frac{\text{True positives} + \text{True negatives}}{\text{Total number of samples}} \quad (1)$$

- Precision—measures the quality of positive predictions:

$$\text{Precision} = \frac{\text{True positives}}{\text{True positives} + \text{False positives}} \quad (2)$$

- Recall—measures the quality of false negative predictions:

$$\text{Recall} = \frac{\text{True positives}}{\text{True positives} + \text{False negatives}} \quad (3)$$

- F1 score—balances between the precision and recall:

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

3. Results and Discussion**3.1. Conduction of the Experimental Study and Preprocessing of the Obtained Images**

Two experimental studies were conducted at the PV Park Kanev on 7 June 2024 and 4/5 July 2024, according to the schedule, presented in Table 1. Additional photos of the PV panels were taken on 6 July and 8 July. A total of 828 pictures were taken, and after filtering out those with low quality, 402 images were selected for training purposes, out of which 184 were of clean and 218 were of dirty PV modules from the 7 June, 4 July, and 5 July datasets.

Two additional datasets were created after different weather events:

- On 9 June 2024 a weather event occurred, creating dust soiling. Therefore, photos of all PV panels were taken on 10 June 2024.
- On 21 July 2024 a storm occurred with 9 mm rainfall, measured with a Vantage Pro2 meteorological station by Davis Instruments (Hayward, Charlotte, CA, USA). Therefore, photos of all PV panels were taken on 22 July 2024.

The photos of the PV modules were taken using different mobile phone cameras from approximately 2 m distance, as shown in Figure 4. It was decided to photograph each module individually, as this would make it easier to pre-process and analyze the datasets. Next, according to the developed methodology, all photos were manually classified as either clean or dirty and 500×500 px fragments were cropped from them. Considering the significant number of images, the extraction process was automated with a Corel

Photopaint batch process (script). To make sure the extracted fragments corresponded to the defined requirements (they should contain the area of at least one PV cell and should not contain areas that are not part of the PV surface), they were observed by an operator. Those that did not meet these requirements were processed manually.

Table 1. Schedule of the performed experiment.

No.	Date	Star Time	Action
1	7 June 2024	7:00	PV panels cleaning
2		8:00	Make photos of all PV panels
3		9:00	Make photos of all PV panels
4		10:00	Make photos of all PV panels
5		11:00	Make photos of all PV panels
6		13:00	Make photos of all PV panels
7		14:00	Make photos of all PV panels
8		15:00	Make photos of all PV panels
9		16:00	Make photos of all PV panels
10		17:00	Make photos of all PV panels
11		18:00	Make photos of all PV panels
12		19:00	Make photos of all PV panels
13	10 June 2024	08:00	Make photos of all PV panels
14	4 July 2024	18:00	PV panels cleaning
15		19:00	Make photos of all PV panels
16	5 July 2024	8:00	Make photos of all PV panels
17		12:00	Make photos of all PV panels
18		18:00	Make photos of all PV panels
19	6 July 2024	8:00	Make photos of all PV panels
20		12:00	Make photos of all PV panels
21		18:00	Make photos of all PV panels
22	8 July 2024	8:00	Make photos of all PV panels
23		12:00	Make photos of all PV panels
24		18:00	Make photos of the cleaned PV panels
25	22 July 2024	16:00	Make photos of all PV panels

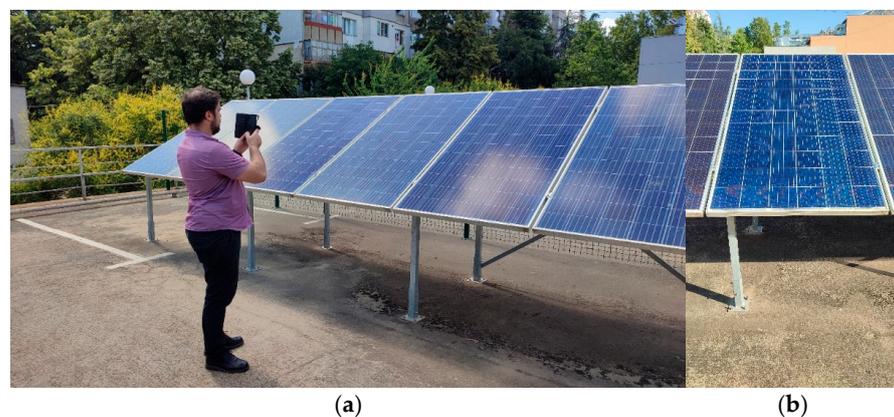


Figure 4. Examples of taking individual photos of each PV module (a) and approximate point of view of each photo (b).

All images were further preprocessed according to the proposed methodology by doubling their red channel, converting them to multichannel, and exporting only the red greyscale channel. Examples from the preprocessing of dirty and clean PV panel fragments are presented in Figure 5.

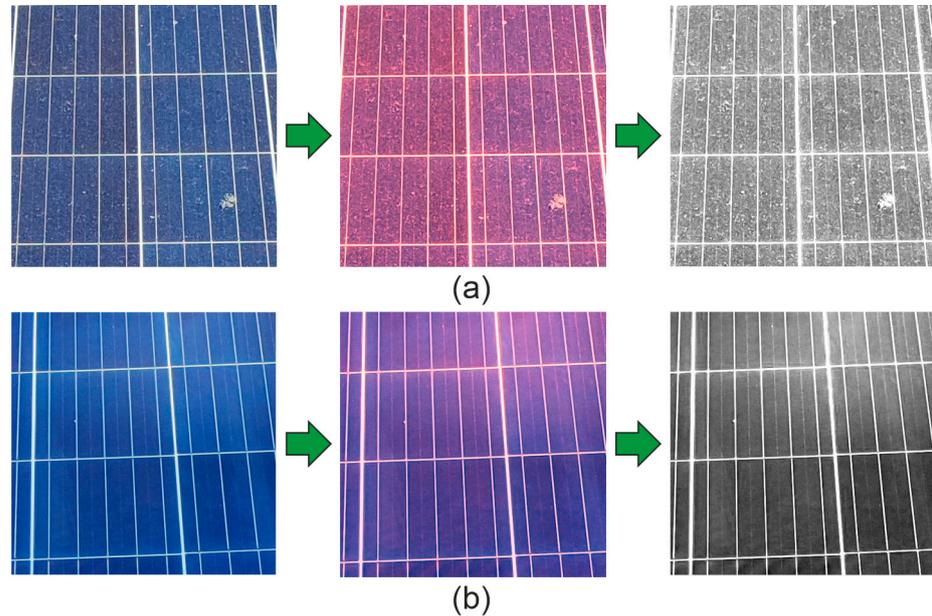


Figure 5. Example of preprocessing of 500×500 px fragments of dirty (a) and clean (b) PV panels.

3.2. Training and Validation of the Machine Learning Models

The machine learning and validation steps were implemented in the Orange Data Mining v.3.36 tool, as shown in Figure 6. The optimal parameters of the six ML algorithms were obtained experimentally and are summarized in Table 2.

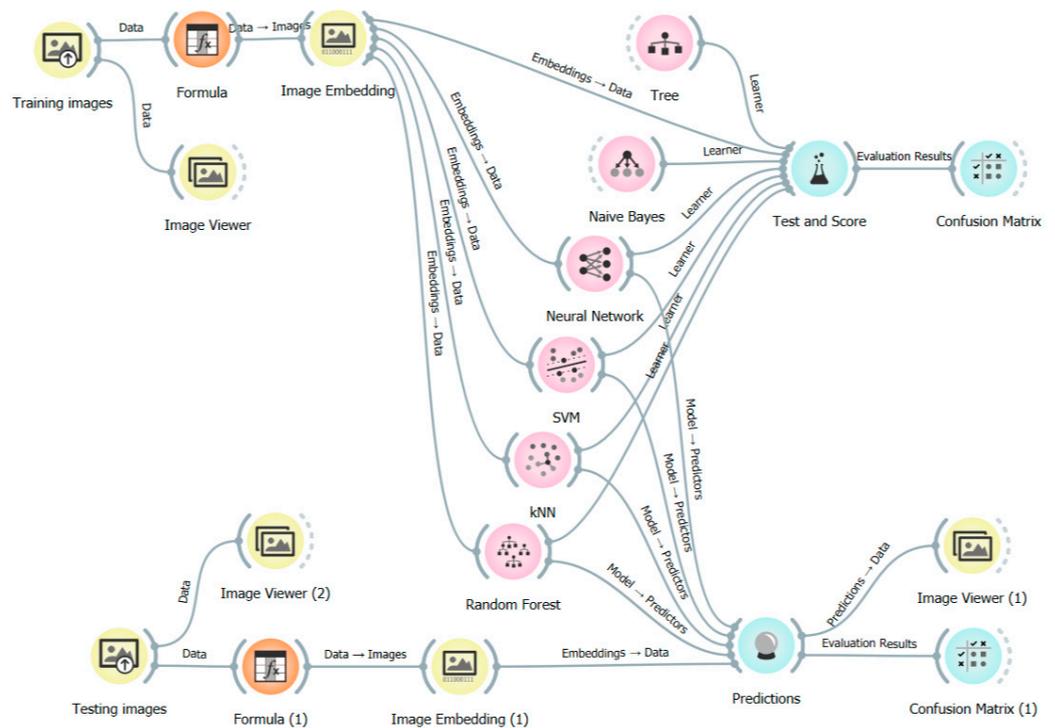


Figure 6. Implementing the training, validation, and testing stages in Orange Data Mining.

Table 2. Optimal parameters of the machine learning algorithms.

No.	Algorithm	Parameters
1	CNN	Neurons in hidden layers—200 Activation—ReLU Solver—L-BFGS-B Regularization—0 Maximal number of iterations—200 Replicable training—Checked
2	SVM	SVM type—SVM Cost (C)—0.50 Regression loss epsilon (ϵ)—0.20 Kernel—Polynomial g—auto c—3.00 d—3.0 Numerical tolerance—0.0010 Iteration limit—200
3	kNN	Number of neighbors—8 Metric—Euclidean Weight—Uniform
4	RF	Number of trees—23 Replicable training—checked Do not split subsets smaller than—5
5	DT	Induce binary tree—checked Min. number of instances in leaves—12 Do not split subsets smaller than—10 Limit the maximal tree depth to—100 Step when majority reaches—95%
6	NB	N/A

The “hour of the day” feature was extracted from the images’ filenames using the “Formula” component in the Orange Data Mining tool. It was implemented as a number variable, whose value was estimated using the following Python v.3.11 script:

$$\text{int}(\text{image_name}[13 : 15]) \quad (5)$$

The six algorithms were evaluated using the “Test and Score” component. It was set up to use cross-validation with the number of folds set to 5. This means that 4/5 of the dataset would be used for training and 1/5 for validation. The obtained evaluation metrics, with and without the hour of the day included as an additional feature, are presented in Table 3. It can be seen that when the hour of the day was not included, the best-performing algorithm was random forest with $F1 = 0.935$, followed by SVM with 0.933 and CNN with 0.928. On the other hand, if the hour of the day was added as an additional feature, the best-performing algorithm was CNN, with $F1 = 0.938$, followed by SVM with 0.933 and random forest with 0.915. In other words, this additional feature did not have a significant impact on the accuracy of the models, even though it led to small changes and reordered their rank in terms of performance. It can also be seen that the SVM model was practically not influenced by this additional feature.

Table 3. Evaluation metrics from the training of the six models with and without the “hour of the day” feature.

Model	Classification Accuracy	Average Precision	Average Recall	F1 Score
Without “hour of the day” as an additional feature				
RF	0.935	0.936	0.935	0.935
SVM	0.933	0.933	0.933	0.933
CNN	0.928	0.928	0.928	0.928
kNN	0.920	0.922	0.920	0.921
Tree	0.871	0.872	0.871	0.871
NB	0.856	0.860	0.856	0.856
With “hour of the day” as an additional feature				
CNN	0.938	0.938	0.938	0.938
SVM	0.933	0.933	0.933	0.933
RF	0.915	0.916	0.915	0.915
kNN	0.903	0.905	0.903	0.903
Tree	0.871	0.872	0.871	0.871
NB	0.856	0.860	0.856	0.856

It should be noted that if an “hour of the day” feature is added, it might also be necessary to add a “month of the year” feature because of the different lighting conditions of the PV panels throughout the year. This, as well as the lack of significant difference in the accuracy of the top performing models (CNN and SVM) with and without the “hour of the day” feature, allows us to conclude that it was better not to use it, unless the training dataset was significantly larger and included numerous images for each hour of the day, for each month of the year, and under different environmental conditions. Therefore, from now on, this study only considers the trained models without the additional “hour of the day” feature.

Furthermore, from the obtained results it can be seen that all of the tested algorithms performed reasonably well; nevertheless, the DT- and NB-based ones performed slightly worse with an F1 score below 0.9, and were therefore not considered in the further analysis.

If compared with previous studies, the obtained results position themselves quite well. In [39], different deep learning algorithms for identifying random-form soiling on the PV surface were compared. The best-performing ones were based on U-net and achieved accuracies of 95.81% and 98.10%, and F1 scores of 85.48% and 84.05%, respectively. Similarly, image segmentation was used in [42]; although this was with a MobileNetV2-based model for identifying different types of soiling. The authors claimed a 97% accuracy, though they agreed that their 50-image dataset was quite small. Furthermore, they also stated that the F1 metric should be used to bring more insightful information, though their study did not provide this.

Similarly, in [40], a SolNet CNN for distinguishing between clean and dirty panels was trained, which achieved an accuracy of 98.2%. The precision for identification of clean and dirty surfaces was 97.6% and 95.7%, respectively, though no recall or F1 score was reported. In [43], different CNNs were trained to detect soiling, cracks, bird drops, and other factors influencing the PV surface. The VGG19 backbone model provided the highest accuracy and F1 score of approximately 99%. In [44], panels under the influence of shadows, soiling, and bird droppings were also identified with an accuracy of 99%. The detection of normal panels achieved a precision, recall, and F1 of 100%, 75%, and 86%, respectively. On the other hand, the identification of faulty panels reached 80%, 100%, and 89%, respectively. Slightly

poorer results were obtained in [45], where the identification of soiling, grass breakage, and cracks achieved a 73% accuracy.

A summary of the comparison performed is presented in Table 4. It can be seen that most of the previous studies achieved a higher accuracy, though this is known to be misleading, especially when imbalanced datasets are used. Therefore, the F1 score is a more reliable metric, and, out of the reviewed articles, only [43] achieved a higher performance. It can be noted though, that in [43] a larger dataset was used, which might explain the better results. Furthermore, in [40,42,45], no information was provided about precision, recall, or the F1 score.

Table 4. Comparison of the obtained results with previous studies.

Study	Model	Accuracy	F1 Score
Cruz-Rojas et al. [39]	U-net-based CNN	95.81%	85.48%
		98.10%	84.05%
Selvi et al. [42]	MobileNetV2 CNN	97%	N/A
Onim et al. [40]	SolNet CNN	98.2%	N/A
Shaik et al. [43]	VGG19-based CNN	98.6%	99%
Yanboiy et al. [44]	VGG19 based CNN	99%	89%
Cavieres et al. [45]	Custom CNN	73%	N/A
Ours	Interception v3 + RF	93.5%	93.5%
	Interception v3 + SVM	93.3%	93.3%
	Interception v3 + CNN	92.8%	92.8%
	Interception v3 + kNN	92.0%	92.1%

3.3. Testing of the Machine Learning Models with Previously Unused Images

Next, the four best-performing models were tested with a different dataset, which was not used during the training phase. It contained 61 images of clean PV panels and 151 images of dirty PV panels, extracted from the 7 June, 10 June, 6 July, and 8 July datasets. The results from the testing process are presented with confusion matrices for the CNN, SVM, RF, and kNN models in Table 5, Table 6, Table 7, and Table 8, respectively.

Table 5. Confusion matrix from the testing of the trained CNN model.

		Predicted			Metrics		
		Clean	Dirty	Σ	Precision	Recall	F1
Actual	Clean	59	2	61	0.776	0.967	0.861
	Dirty	17	134	151	0.985	0.887	0.934
	Σ	76	136	212	0.925	0.910	0.913

Table 6. Confusion matrix from the testing of the trained SVM model.

		Predicted			Metrics		
		Clean	Dirty	Σ	Precision	Recall	F1
Actual	Clean	59	2	61	0.738	0.967	0.837
	Dirty	21	130	151	0.985	0.861	0.919
	Σ	80	132	212	0.914	0.892	0.895

Table 7. Confusion matrix from the testing of the trained RF model.

		Predicted			Metrics		
		Clean	Dirty	Σ	Precision	Recall	F1
Actual	Clean	56	5	61	0.659	0.918	0.767
	Dirty	29	122	151	0.961	0.808	0.878
	Σ	85	127	212	0.874	0.840	0.846

Table 8. Confusion matrix from the testing of the trained kNN model.

		Predicted			Metrics		
		Clean	Dirty	Σ	Precision	Recall	F1
Actual	Clean	59	2	61	0.615	0.967	0.752
	Dirty	37	114	151	0.983	0.755	0.854
	Σ	96	116	212	0.877	0.816	0.824

It can be noticed that the precision metrics for clean panels and the recall metrics for dirty panels were relatively lower in all confusion matrices. This behavior was caused by two factors:

- The testing dataset was imbalanced, i.e., the ratio between clean and dirty panels was approximately 1 to 3.
- The successful identification rate for dirty panels varied between 11% and 24% for the different models.

It can be seen that the highest average precision, recall, and F1 were obtained for the CNN model, with 0.913, 0.910, and 0.913, respectively. It was closely followed by the SVM, whose metrics were 0.914, 0.892, and 0.895, respectively. A closer look into the confusion matrices shows that CNN performed slightly better when identifying dirty panels and had equal performance with the SVM when identifying clean panels.

The other two models showed noticeably lower performance during the testing phase. The RF model achieved an average F1 score of 0.911 and showed a worse identification performance for both classes. The kNN model achieved the worst F1 score (0.824). It had the same performance for clean panels as CNN and SVM; however, its identification rate for dirty panels was more than twice as bad as CNN's. The obtained results show the following:

- All four models deal well with the identification of clean panels (56–59 out of 61 correct identifications).
- Their performance with dirty panels was different, though identifying dirty panels was their primary goal.

The obtained performance indicators show that the number of images used during the training phase (184 of clean and 218 of dirty panels) was sufficient for training high-performing CNN and SVM models with the proposed methodology. This was confirmed by the high F1 scores obtained with both the training and testing datasets. On the other hand, the significant decrease in the F1 score for RF and kNN during the testing phase (from 0.935 to 0.846 and from 0.921 to 0.824, respectively) indicates that these algorithms require a larger training dataset to improve their performance. Nevertheless, considering that the F1 score decreased between the training and testing phases for all four evaluated models, all of them could benefit from a slight increase in the volume of the training dataset.

A closer analysis of the incorrectly identified dirty panels by the CNN model showed that in many cases (8 out of 17) there was a reflection of the sun on their surface, as shown in Figure 7. A reflection of the sun can also be observed in one of the two incorrectly identified clean PV panels. This indicates that in theory the model performance could be

further improved if the angle of images made is carefully chosen, depending on the hour of the day and the meteorological conditions.

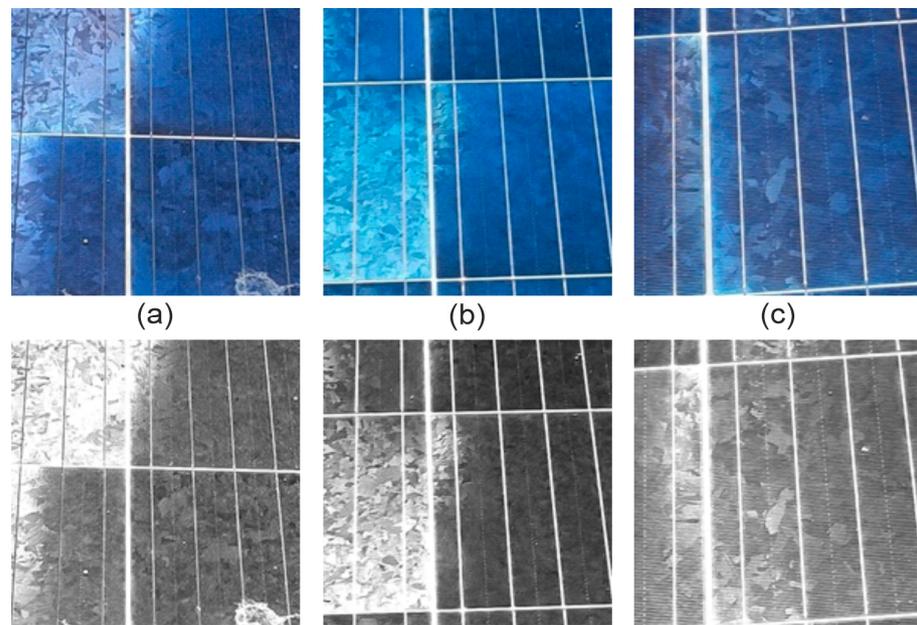


Figure 7. Examples of incorrectly identified dirty panels: sample panel 1 (a); sample panel 2 (b); sample panel 3 (c).

3.4. Case Studies after Specific Weather Events

The optimal CNN model was tested on two datasets, which were obtained after two specific weather events:

- Case study 1: A dust soiling occurred on 9 June 2024, which made all panels dirty (Figure 8).
- Case study 2: a 9 mm rain storm occurred on 21 July 2024, which had a positive effect on the soiling state of all PV panels (Figure 9).

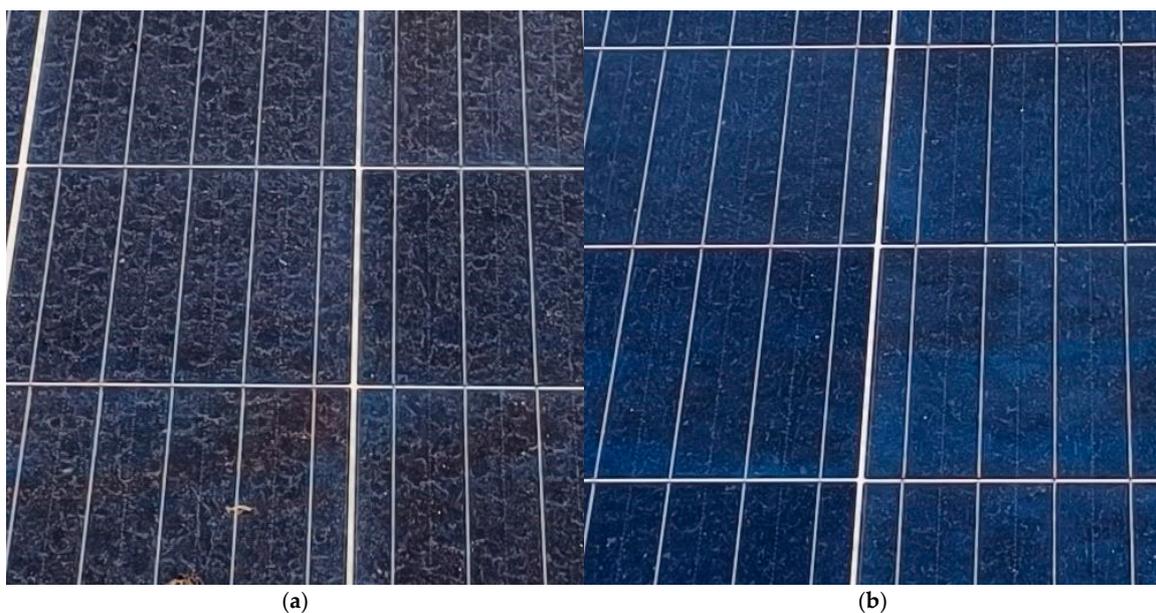


Figure 8. Examples of the PV surface condition on 10 June 2024 after a soiling weather event: (a) a previously dirty panel; (b) a previously clean panel.

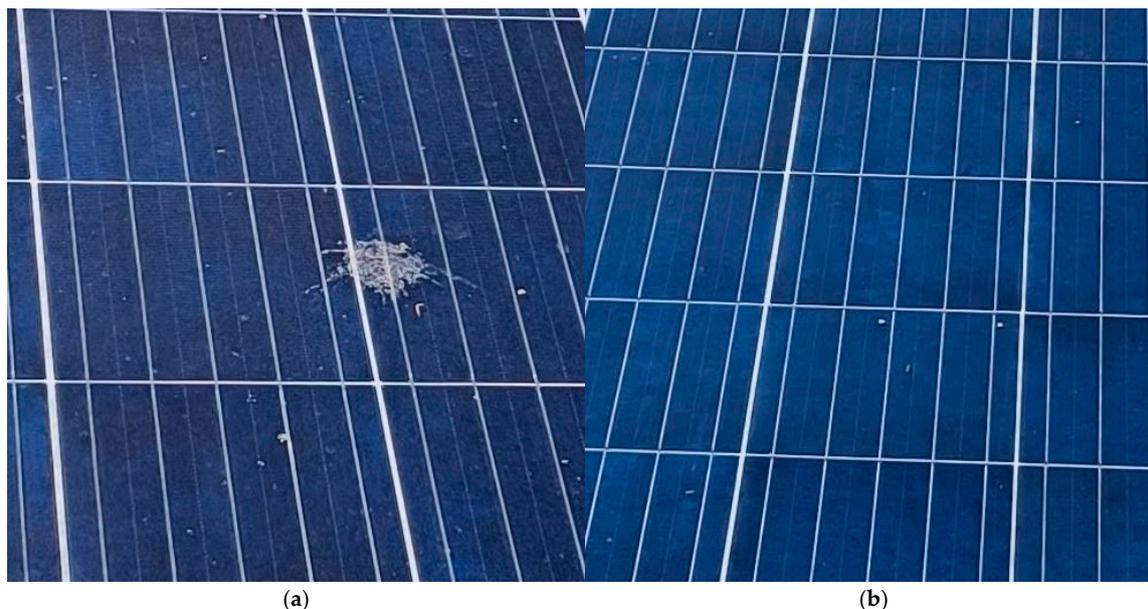


Figure 9. Examples of the PV surface condition on 22 July 2024 after a 9 mm rain: (a) a relatively dirty panel; (b) a relatively clean panel.

Therefore, on the next day, pictures of all 36 PV panels were taken. The two datasets were processed according to the proposed methodology and were fed to the optimal CNN model. The obtained results are summarized in Table 9. In the first scenario, after a soiling weather event, the trained neural network identified 31 out of 33 PV panels as dirty. This means that only 6% of them were classified as clean. Such results indicate that a cleaning procedure was likely required by the time the photos were taken and could lead to an increase in the PV installation's energy yield.

Table 9. Summary of the results from the two case studies.

Case Study	Number of Classified Images			Clean PV Panels Ratio, %
	As Clean	As Dirty	Total	
1	2	31	33	6.1%
2	32	4	36	89%

On the other hand, the results for the second scenario showed that 32 out of 36 PV panels (including those identified as dirty on 8 July 2024) were identified as clean, i.e., the share of the clean panels was approximately 89%. Such metrics allow us to make the following conclusions:

- By the time the photos were taken it was most likely not necessary to implement a cleaning procedure for the photovoltaic installation.
- The relatively strong rainstorm reduced the soiling of the PV modules.

3.5. Applicability of the Obtained Results

While the obtained results with the trained CNN are quite promising, it is important to discuss its applicability for remote diagnostics on a larger scale. Making on-the-ground photos of a large PV park would not be appropriate as this would require a amount of time and workforce. As previous studies have shown large-scale monitoring can be implemented with UAV-obtained images. However, in this situation, depending on the flying height and the lens angle of the camera, each image will contain numerous PV modules. Therefore, to apply the proposed methodology, the following options exist:

1. The UAV could be operated at a low height. This would allow appropriate image quality even with a lower-resolution camera.
2. The UAV could be operated at medium height. This way more PV panels could be captured at once; however, this implies the need to use more expensive cameras with a higher resolution.

In both situations, the obtained photos will likely contain numerous PV panels, so additional preprocessing would be required. The extraction of images with an appropriate size could be implemented using the following methodology:

1. The extent of the available PV modules on each photo is recognized. This could be implemented by training an object-based deep learning model for PV panel identification.
2. A square fragment of the image is cropped from each identified PV module.
3. The dimensions of each image fragment are reduced to 500×500 px.

Another important aspect is the number of analyzed PV modules. When it comes to large-scale PV installations, taking pictures of all available modules might be difficult. An appropriate approach would be to divide the facility into zones and to take a limited number of photos in each zone. This way, the decision making on whether a cleaning procedure is required could be made for each zone independently using the available images.

4. Conclusions

In this study ground-based visible spectrum imaging and machine learning were used to identify soiling on photovoltaic installations. All images were preprocessed following the proposed methodology to focus on the available soiling. This included doubling the red channel, converting the image to multichannel, and exporting the modified grayscale red channel for further analysis. Thereafter, the images were manually classified into clean or dirty categories and were divided into training and testing datasets. The performance of six machine learning algorithms was evaluated.

During the training and cross-validation phase, a total of 402 images was used. The best-performing model was RF, followed by SVM, CNN, kNN, DT, and NB. They achieved an F1 score of 93.5%, 93.3%, 92.8%, 92.1%, 87.1%, and 85.6%, respectively. The first four models showed similar evaluation metrics and were further investigated using 212 previously unused testing images. During this phase, the models were reordered and the trained convolutional neural network achieved the best performance, with an average F1 score of 91.3%. It was closely followed by the SVM model with 89.5%. Even though the RF model was the best-performing one during the validation phase, it achieved a significantly lower F1 score (84.6%) using the testing dataset. The F1 metric of the kNN model was even lower (82.4%), though it kept its ranking. These results allow us to conclude that deep learning performs slightly better when identifying PV surface soiling in comparison to other machine learning algorithms.

Even though in the current study ground-based images were used, the application of UAV-obtained images for medium- and large-scale facilities is fully applicable. The proposed methodology could be applied to support the decision-making process for operators of PV installations. It could help them decide when the solar panels require cleaning so that their profit is maximized.

No threshold value was proposed in this study regarding when cleaning is required because, as previous studies have observed, this strongly depends on the specifics and characteristics of soiling in the corresponding region. To do this, the results of this paper should be correlated with the local soiling losses, the prices of energy, and the cleaning procedure. Such investigations are important for the region of Ruse and Bulgaria and are an object for future studies.

Author Contributions: Conceptualization, B.I.E. and N.P.M.; methodology, B.I.E.; software, K.G.G.-E. and B.I.E.; validation, B.I.E., N.P.V. and N.P.M.; formal analysis, B.I.E.; investigation, D.T.T., and B.I.E.; resources, N.P.M., N.P.V. and B.I.E.; data curation, B.I.E. and N.P.M.; writing—original draft preparation, B.I.E.; writing—review and editing, B.I.E.; visualization, B.I.E.; supervision, B.I.E., and N.P.M.; project administration, B.I.E. and N.P.M.; funding acquisition, B.I.E. and N.P.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research is financed by the European Union-NextGenerationEU, through the National Recovery and Resilience Plan of the Republic of Bulgaria under project no. BG-RRP-2.013-0001-C01.

Data Availability Statement: The datasets used in this study are published under the CC BY 4.0 license and can be found at <https://doi.org/10.6084/m9.figshare.26977123>.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Islam, M.; Rashel, M.R.; Ahmed, M.T.; Islam, A.K.M.K.; Tlemçani, M. Artificial Intelligence in Photovoltaic Fault Identification and Diagnosis: A Systematic Review. *Energies* **2023**, *16*, 7417. [CrossRef]
- Bassi, H.; Salam, Z.; Ramli, M.Z.; Sindi, H.; Rawa, M. Hardware Approach to Mitigate the Effects of Module Mismatch in a Grid-connected Photovoltaic System: A Review. *Energies* **2019**, *12*, 4321. [CrossRef]
- Mariani, V.; Adinolfi, G.; Buonanno, A.; Ciavarella, R.; Ricca, A.; Sorrentino, V.; Graditi, G.; Valenti, M. A Survey on Anomalies and Faults That May Impact the Reliability of Renewable-Based Power Systems. *Sustainability* **2024**, *16*, 6042. [CrossRef]
- Bošnjaković, M.; Stojkov, M.; Katinić, M.; Lacković, I. Effects of Extreme Weather Conditions on PV Systems. *Sustainability* **2023**, *15*, 16044. [CrossRef]
- Hepp, D.; Hempelmann, S.; Behrens, G.; Friedrich, W. Detection of snow-coverage on PV-modules with images based on CNN-techniques. In *EnvirolInfo 2022*; Gesellschaft für Informatik e.V.: Bonn, Germany, 2022.
- Yilmaz, M. Comparative Analysis of Hybrid Maximum Power Point Tracking Algorithms Using Voltage Scanning and Perturb and Observe Methods for Photovoltaic Systems under Partial Shading Conditions. *Sustainability* **2024**, *16*, 4199. [CrossRef]
- Celikel, R.; Yilmaz, M.; Gundogdu, A. A voltage scanning-based MPPT method for PV power systems under complex partial shading conditions. *Renew. Energy* **2022**, *184*, 361–373. [CrossRef]
- Winkel, P.; Wilbert, S.; Röger, M.; Krauth, J.J.; Algner, N.; Nouri, B.; Wolfertstetter, F.; Carballo, J.A.; Alonso-Garcia, M.C.; Polo, J.; et al. Cell-Resolved PV Soiling Measurement Using Drone Images. *Remote Sens.* **2024**, *16*, 2617. [CrossRef]
- Redondo, M.; Platero, C.A.; Moset, A.; Rodríguez, F.; Donate, V. Soiling Modelling in Large Grid-Connected PV Plants for Cleaning Optimization. *Energies* **2023**, *16*, 904. [CrossRef]
- Kam-Lum, E.; Meyers, B.E.; Cosme, D.; Aissa, B.; Scabbia, G. Soiling Rate Determination from Referenced Systems in Desert Climate using PVInsight Soiling Algorithm. In Proceedings of the 2021 IEEE 48th Photovoltaic Specialists Conference (PVSC), Fort Lauderdale, FL, USA, 20–25 June 2021. [CrossRef]
- Simal Pérez, N.; Alonso-Montesinos, J.; Battles, F.J. Estimation of Soiling Losses from an Experimental Photovoltaic Plant Using Artificial Intelligence Techniques. *Appl. Sci.* **2021**, *11*, 1516. [CrossRef]
- Al-Ibrahim, E. Impact of Dust and Shade on Solar Panel Efficiency and Development of a Simple Method for Measuring the Impact of Dust in any Location. *J. Sustain. Dev. Energy Water Environ. Syst.* **2023**, *30*, 1–4. [CrossRef]
- Jung, D.; Gareis, G.H.; Staiger, A.; Salmon, A. Effects of soiling on agrivoltaic systems: Results of a case study in Chile. *AIP Conf. Proc.* **2022**, *2635*, 1. [CrossRef]
- Pareek, A.; Gupta, R. Analysis and insights into snail trail degradation in photovoltaic modules. *Sol. Energy* **2024**, *275*, 112613. [CrossRef]
- Liu, L.; Li, Q.; Liao, X.; Wu, W. Bird Droppings Defects Detection in Photovoltaic Modules Based on CA-YOLOv5. *Processes* **2024**, *12*, 1248. [CrossRef]
- Toth, S.; Hannigan, M.; Vance, M.; Deceglie, M. Enhanced photovoltaic soiling in an urban environment. In Proceedings of the 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC), Chicago, IL, USA, 16–21 June 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 2904–2907. [CrossRef]
- Jothi Venkatesh, K.; Sankaranarayanan, S.; Kannan, K.; Arjun, P. AI Based Solar Panel Cleaning Robot. *Int. J. Eng. Technol. Manag. Sci.* **2023**, *7*, 313–318. [CrossRef]
- Zahedi, R.; Ranjbaran, P.; Gharehpetian, G.B.; Mohammadi, F.; Ahmadihangar, R. Cleaning of Floating Photovoltaic Systems: A Critical Review on Approaches from Technical and Economic Perspectives. *Energies* **2021**, *14*, 2018. [CrossRef]
- Najmi, N.; Rachid, A. A Review on Solar Panel Cleaning Systems and Techniques. *Energies* **2023**, *16*, 7960. [CrossRef]
- Abdallah, R.; Juaidi, A.; Abdel-Fattah, S.; Qadi, M.; Shadid, M.; Albatayneh, A.; Çamur, H.; García-Cruz, A.; Manzano-Agugliaro, F. The Effects of Soiling and Frequency of Optimal Cleaning of PV Panels in Palestine. *Energies* **2022**, *15*, 4232. [CrossRef]
- Abuzaid, H.; Awad, M.; Shamayleh, A. Enhancing Photovoltaic System Cleaning Using Machine Learning Algorithms. *J. Propuls. Technol.* **2024**, *45*, 2119–2124.

22. Alfaris, F.E. A Sensorless Intelligent System to Detect Dust on PV Panels for Optimized Cleaning Units. *Energies* **2023**, *16*, 1287. [[CrossRef](#)]
23. Yang, M.; Javed, W.; Guo, B.; Ji, J. Estimating PV Soiling Loss Using Panel Images and a Feature-Based Regression Model. *IEEE J. Photovolt.* **2024**, *2*, 1–8. [[CrossRef](#)]
24. Song, Z.; Liu, J.; Yang, H. Air pollution and soiling implications for solar photovoltaic power generation: A comprehensive review. *Appl. Energy* **2021**, *15*, 117247. [[CrossRef](#)]
25. Valerino, M.; Ratnaparkhi, A.; Ghoroi, C.; Bergin, M. Seasonal photovoltaic soiling: Analysis of size and composition of deposited particulate matter. *Sol. Energy* **2021**, *227*, 44–55. [[CrossRef](#)]
26. Memon, S.A.; Javed, Q.; Kim, W.-G.; Mahmood, Z.; Khan, U.; Shahzad, M. A Machine-Learning-Based Robust Classification Method for PV Panel Faults. *Sensors* **2022**, *22*, 8515. [[CrossRef](#)] [[PubMed](#)]
27. Al-Katheri, A.A.; Al-Ammar, E.A.; Alotaibi, M.A.; Ko, W.; Park, S.; Choi, H.-J. Application of Artificial Intelligence in PV Fault Detection. *Sustainability* **2022**, *14*, 13815. [[CrossRef](#)]
28. Ali, M.U.; Khan, H.F.; Masud, M.; Kallu, K.D.; Zafar, A. A machine learning framework to identify the hotspot in photovoltaic module using infrared thermography. *Sol. Energy* **2020**, *208*, 643–651. [[CrossRef](#)]
29. Benghanem, M.; Mellit, A.; Moussaoui, C. Embedded Hybrid Model (CNN–ML) for Fault Diagnosis of Photovoltaic Modules Using Thermographic Images. *Sustainability* **2023**, *15*, 7811. [[CrossRef](#)]
30. Baltacı, Ö.; Kırıl, Z.; Dalkılıç, K.; Karaman, O. Thermal Image and Inverter Data Analysis for Fault Detection and Diagnosis of PV Systems. *Appl. Sci.* **2024**, *14*, 3671. [[CrossRef](#)]
31. Starzyński, J.; Zawadzki, P.; Harańczyk, D. Machine Learning in Solar Plants Inspection Automation. *Energies* **2022**, *15*, 5966. [[CrossRef](#)]
32. Wang, J.; Bi, L.; Sun, P.; Jiao, X.; Ma, X.; Lei, X.; Luo, Y. Deep-Learning-Based Automatic Detection of Photovoltaic Cell Defects in Electroluminescence Images. *Sensors* **2023**, *23*, 297. [[CrossRef](#)]
33. Faskari, S.A.; Ojim, G.; Falope, T.; Abdullahi, Y.B.; Abba, S.I. A novel machine learning based computing algorithm in modeling of soiled photovoltaic module. *Knowl. Based Eng. Sci.* **2022**, *3*, 28–36. [[CrossRef](#)]
34. Dassler, D.; Malik, S.; Kuppanna, S.B.; Jaeckel, B.; Ebert, M. Innovative Approach for Yield Evaluation of PV Systems Utilizing Machine Learning Methods. In Proceedings of the 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC), Chicago, IL, USA, 16–21 June 2019. [[CrossRef](#)]
35. Tanyıldızı Ağır, T. Prediction of Losses Due to Dust in PV Using Hybrid LSTM-KNN Algorithm: The Case of Saruhanlı. *Sustainability* **2024**, *16*, 3581. [[CrossRef](#)]
36. Fang, M.; Qian, W.; Qian, T.; Bao, Q.; Zhang, H.; Qiu, X. DGImlNet: A deep learning model for photovoltaic soiling loss estimation. *Appl. Energy* **2024**, *376*, 124335. [[CrossRef](#)]
37. Heinrich, M.; Meunier, S.; Samé, A.; Quéval, L.; Darga, A.; Oukhellou, L.; Multon, B. Detection of cleaning interventions on photovoltaic modules with machine learning. *Appl. Energy* **2020**, *263*, 114642. [[CrossRef](#)]
38. Martin, J.; Jaskie, K.; Tofis, Y.; Spanias, A. PV array soiling detection using machine learning. In Proceedings of the 2021 12th International Conference on Information, Intelligence, Systems & Applications (IISA), Chania Crete, Greece, 12–14 July 2021; IEEE: New York, NY, USA; pp. 1–4. [[CrossRef](#)]
39. Cruz-Rojas, T.; Franco, J.A.; Hernandez-Escobedo, Q.; Ruiz-Robles, D.; Juarez-Lopez, J.M. A novel comparison of image semantic segmentation techniques for detecting dust in photovoltaic panels using machine learning and deep learning. *Renew. Energy* **2023**, *217*, 119126. [[CrossRef](#)]
40. Onim, M.S.H.; Sakif, Z.M.M.; Ahnaf, A.; Kabir, A.; Azad, A.K.; Oo, A.M.T.; Afreen, R.; Hridy, S.T.; Hossain, M.; Jabid, T.; et al. SolNet: A Convolutional Neural Network for Detecting Dust on Solar Panels. *Energies* **2023**, *16*, 155. [[CrossRef](#)]
41. Pivem, T.; de Araujo, F.D.O.; de Araujo, L.D.O.; de Oliveira, G.S. Application of a computer vision method for soiling recognition in photovoltaic modules for autonomous cleaning robots. *Signal Image Process. Int. J.* **2019**, *10*, 43–59. [[CrossRef](#)]
42. Selvi, S.; Devaraj, V.; PS, R.P.; Subramani, K. Detection of soiling on PV module using deep learning. *Int. J. Electr. Electron. Eng.* **2023**, *10*, 93–101. [[CrossRef](#)]
43. Shaik, A.; Balasundaram, A.; Kakarla, L.S.; Murugan, N. Deep Learning-Based Detection and Segmentation of Damage in Solar Panels. *Automation* **2024**, *5*, 128–150. [[CrossRef](#)]
44. El Yanboiy, N.; Khala, M.; Elabbassi, I.; Eloutassi, O.; El Hassouani, Y.; Messaoudi, C. Enhancing Surface Defect Detection in Solar Panels with AI-Driven VGG Models. *Data Metadata* **2023**, *2*, 81. [[CrossRef](#)]
45. 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**, *15*, 117964. [[CrossRef](#)]
46. Dalagan, A.G.; Principe, J.A. Spatial Inventory of Solar Photovoltaic (PV) Installations Using Remote Sensing and Machine Learning: Case of Pampanga Province, Philippines. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2024**, *48*, 115–123. [[CrossRef](#)]
47. Giussani, F.; Wilczynski, E.; Zandonella Callegher, C.; Dalle Nogare, G.; Pozza, C.; Novelli, A.; Pezzutto, S. Use of Machine Learning Techniques on Aerial Imagery for the Extraction of Photovoltaic Data within the Urban Morphology. *Sustainability* **2024**, *16*, 2020. [[CrossRef](#)]
48. Meteoblue. Available online: <https://www.meteoblue.com> (accessed on 2 October 2024).

49. Demsar, J.; Curk, T.; Erjavec, A.; Gorup, C.; Hocevar, T.; Milutinovic, M.; Mozina, M.; Polajnar, M.; Toplak, M.; Staric, A.; et al. Orange: Data Mining Toolbox in Python. *J. Mach. Learn. Res.* **2013**, *14*, 2349–2353.
50. Yang, M.; Ji, J.; Guo, B. Soiling Quantification Using an Image-Based Method: Effects of Imaging Conditions. *IEEE J. Photovolt.* **2020**, *10*, 1780–1787. [[CrossRef](#)]
51. Marinov, M.; Kalmukov, Y.; Valova, I. Applying Object Recognition to Improve Image Retrieval by Color Features. In Proceedings of the 2024 23rd International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina, 20–22 March 2024. [[CrossRef](#)]
52. Mladenova, T.; Valova, I. Comparative analysis between the traditional K-Nearest Neighbor and Modifications with Weight-Calculation. In Proceedings of the 2022 International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, Turkey, 20–22 October 2022; pp. 961–965. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.