



HAL
open science

Data-Driven Digital Inspection of Photovoltaic Panels Using a Portable Hybrid Model Combining Meteorological Data and Image Processing

Ayoub Oufadel, Alae Azouzoute, Hicham Ghennioui, Chaimae Soubai,
Ibrahim Taabane

► **To cite this version:**

Ayoub Oufadel, Alae Azouzoute, Hicham Ghennioui, Chaimae Soubai, Ibrahim Taabane. Data-Driven Digital Inspection of Photovoltaic Panels Using a Portable Hybrid Model Combining Meteorological Data and Image Processing. *Ieee Journal of Photovoltaics*, 2024, 14 (6), pp.937-950. 10.1109/jphotov.2024.3437736 . hal-04688823

HAL Id: hal-04688823

<https://hal.science/hal-04688823v1>

Submitted on 21 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Data-Driven Digital Inspection of Photovoltaic Panels Using a Portable Hybrid Model Combining Meteorological Data and Image Processing

Ayoub Oufadel , Alae Azouzoute , Hicham Ghennioui, Chaimae Soubai, and Ibrahim Taabane 

Abstract—This article proposes a novel approach to photovoltaic panel inspection through the integration of image classification and meteorological data analysis. Utilizing two convolutional neural network models with distinct architectures for classifying thermal and red, green, blue (RGB) images of photovoltaic installations, in addition to an support vector machines model for meteorological data classification, the results from these models are concatenated, allowing the fusion of visual and meteorological information for comprehensive defect detection. Data collection from photovoltaic panels is achieved using a portable device, followed by the application of advanced image processing techniques to identify faults rapidly and accurately with up to 96% accuracy. The inspection results are presented in a user-friendly format, facilitating straightforward interpretation and analysis. This new approach has the potential to significantly enhance the efficiency and durability of solar energy systems, enabling timely maintenance and repair for photovoltaic panel issues.

Index Terms—Convolutional neural network (CNN), image processing, innovative inspection, machine learning (ML), maintenance, photovoltaic.

I. INTRODUCTION

SOLAR energy is an important renewable energy source to meet the current and future energy needs of our planet [1]. Photovoltaic panels are a key component of this technology, are subject to defects and degradation that can affect their performance and lifespan [2]. Therefore, it is essential to regularly monitor the condition of these panels to ensure their optimal operation.

Solar field inspection generally includes two main components: manual inspection and automatic inspection. Manual inspection of solar panels presents several major

challenges. Difficulties in accessing panels installed at heights or in complex locations make inspection dangerous and time-consuming [3]. Moreover, human errors and inspector subjectivity can lead to gaps in defect detection. Advances in solar panel technology further complicate visual problem detection. Lastly, periodic inspections limit the implementation of preventive maintenance [4]. Since they typically involve scheduled visits at fixed intervals, such as once or twice a year, potential issues or defects in solar panels might not be detected or addressed before the next inspection.

For automatic inspection, this approach relies on the use of physical models or artificial intelligence (AI) models and data-driven techniques. Inspecting solar panels using physical models involves creating a mathematical representation of the panel, comparing model predictions with real data to identify anomalies, and making decisions based on simulation results. This can be a powerful tool to monitor and optimize solar installation performance.

By comparing simulation results with real data, it would be possible to detect anomalies or unexpected deviations. Generally, two physical parameters are used for the inspection of photovoltaic solar panels: power, measured by electrical panel parameters, and compared to the ideal power that should be produced by the installation (proportional to solar irradiation, solar panel surface area in square meters (m^2), and panel efficiency [5]). Another physical parameter that can be used to inspect solar panels is the solar panel temperature, referring to the temperature of the panel's surface itself. This temperature can differ from ambient temperature due to various factors such as the amount of received solar radiation, wind speed affecting convective cooling, humidity levels affecting evaporation, and the reflectivity of the surface on which the panel is installed [6]. By measuring the panel's temperature, emerging problems can be identified through trends and variations.

These physical models alone face difficulties in fully modeling the solar field inspection problem for several reasons. The complexity of the system, especially interactions among various parameters like material variations, environmental conditions, and other factors, makes it challenging to model. The natural variability of these factors can make accurate performance prediction using purely physical models difficult. This is why, to accurately address the solar field inspection problem, it is often necessary to use a combination of physical models, machine learning (ML, such as deep learning), and advanced

Manuscript received 24 April 2024; revised 13 July 2024; accepted 24 July 2024. This work was supported by the Mohammed VI Polytechnic University (UM6P). (Corresponding author: Alae Azouzoute.)

Ayoub Oufadel, Hicham Ghennioui, and Chaimae Soubai are with the Laboratory of Signals, Systems, and Components, Sidi Mohamed Ben Abdellah University, Fez 30000, Morocco (e-mail: ayouboufadel@gmail.com).

Alae Azouzoute is with the Fluid Mechanics Team, Laboratory of Mechanics and Energetics, Faculty of Sciences, University Mohammed First Oujda, Oujda 60000, Morocco (e-mail: azouzoute.alae@gmail.com).

Ibrahim Taabane is with the Institute of Electronics and Digital Technologies (IETR), University of Rennes, 35000 Rennes, France, and also with the Laboratory of Intelligent Systems, Geo-resources and Renewable Energies (SIGER), Sidi Mohamed Ben Abdellah University, Fez 30000, Morocco (e-mail: ibrahim.taabane@usmba.ac.ma).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JPHOTOV.2024.3437736>.

Digital Object Identifier 10.1109/JPHOTOV.2024.3437736

data processing techniques. These approaches can capture the complexity and variability of the system and integrate real-time data for better problem detection and maintenance that is more effective.

Modern solar installations typically feature monitoring systems that are commonly integrated into smart inverters. These inverters enable the detection of certain issues such as clutch problems, defective solar panels, wiring issues, grounding problems, and more. These defects are identified through the analysis of the electrical characteristics of the solar panel strings and the grounding cable. However, this technique has significant limitations as it cannot accurately detect various solar panel problems, especially those with minor impacts on the string. Moreover, it fails to precisely locate or identify the issue. Due to these constraints in inverter-based inspection and traditional manual or physics-based inspection methods, numerous studies in the literature have developed intelligent solutions for partial or comprehensive inspection based on data-driven approaches.

The current state of research in this field is summarized as follows: In a recent study [7] from 2023, the use of a thermal image database from billboards revealed that ResNet18, GoogleNet, SqueezeNet, and support vector machines (SVM) models exhibit high accuracy, but they are limited by the size of the database and interpretability. Another study [8] dated 2022 utilized an open-source dataset from the Solar Energy Laboratory with region-based convolutional neural network (RCNN) models, achieving acceptable accuracy of 73% for soiling and shading issues. However, the overall accuracy remains relatively low. In 2021, research [9] based on a dataset of 398 thermal images and CNN models provided detailed information with 11 classes, but the testing accuracy is low (0.57), suggesting potential overfitting. Another study in 2022 [10], employing real-time production data, indicates that CNN models offer high precision, but hardware requirements pose deployment challenges, with risks of invalidated overfitting. The use of you only look once version 3 (YOLOv3) on red, green, blue (RGB) and thermal images from two PV plants appears promising. The application of VGG16 and RCNN models on UAV and manual thermal images showed high accuracy (0.98) in 2022, although evaluation gaps persist due to the limited number of classes and the absence of new data application [11]. To summarize, the articles cover the years from 2020 to 2023. Various datasets are used, including thermal images, RGB images, as well as numerical data such as current, voltage, temperature, and irradiation level. Different classification models are employed, including ResNet18, GoogleNet, SqueezeNet, SVM, RCNN, and CNN with VGG16 architecture, saliency-constrained deep CNN (SC-DCNN), region of interest-DCNN (ROI-DCNN), and YOLOv3. Anomaly detection on solar panels using data-driven methods presents various substantial limitations. First, although these techniques allow for accurate detection of issues, they remain incapable of examining all potential anomalies. This restriction stems from the precise determination of the origin and location of problems remaining a major challenge. Data-driven approaches might lack precise categorizations to attribute the fundamental cause of identified problems, making effective resolution of these anomalies difficult. For instance, the works of [12] and

[13] successfully identified the region of the faulty PV module but did not specifically focus on the nature of the defect present in the image. The aspect of model accuracy also poses issues. While these models might perform well during their training, it remains crucial to verify their ability to generalize to new data to avoid overfitting and ensure real-world accuracy. This requires evaluations on novel datasets, a step that is often overlooked. Similar to paper [14], which used the difference between hot and cold spots to estimate panel condition, lacked adequate preprocessing to detect and eliminate outlines, preventing the algorithm from considering the outline as a cold spot on the panel, thereby reducing inspection accuracy. The reliability of results is also compromised due to data processing. Inaccuracies in input data, noise, and other factors can diminish the quality of predictions. Furthermore, the lack of testing on real installations limits the confidence that can be placed in these methods. For example, Zefri et al. [15] developed a highly accurate model without providing details about the used database or the specific solar fields involved, nor conducting tests on a real installation to validate data accuracy. Ultimately, integrating this solution into portable hardware for comprehensive solar installation inspection proves to be complex. Constraints in terms of computational power and memory can be significant obstacles to on-field implementation, thereby limiting solution portability and accessibility. Similar to the study conducted by Pratt et al. [16], where a photovoltaic panel failure detection system was integrated with significant accuracy using electroluminescence imaging analysis. However, this technique cannot be incorporated into a standalone device to inspect large-scale solar fields, primarily due to its high cost and the need to individually adjust each panel, hindering its large-scale implementation.

With this perspective, this article proposes a novel approach for inspecting photovoltaic panels by combining data and image processing. The developed portable hybrid model enables fast and accurate detection of defects in photovoltaic panels, facilitating timely maintenance and repair. This innovative approach significantly enhances the efficiency and effectiveness of photovoltaic panel inspection, resulting in improved performance and increased durability of solar energy systems through the digitization of the inspection process. This improvement is made possible through simple tools that do not require extensive expertise or special certifications while enhancing the customer experience to make it more accessible.

In this context, the use of advanced ML techniques and image analysis has emerged as a promising way to detect and classify anomalies in photovoltaic systems. The data-driven approach with AI, heavily relying on data to feed ML and AI algorithms [17], offers promising applications for solar panel inspection. By utilizing data analysis techniques, AI can process large amounts of information collected from sensors, drones, thermal and RGB images to precisely detect defects or anomalies in solar panels. These AI systems can identify issues such as hotspots, dirt, and abnormal temperature variations. In this study, by comparing collected data with previously established models to identify deviations [18], it enables accurate data interpretation, facilitating rapid problem detection and optimization of solar installation performance.

We propose an approach based on a combination of meteorological data for fouling rate modeling using machine-learning classification models like logistic regression and SVM, as well as image processing through ML and CNNs with different architectures for both RGB and thermal data, such as AlexNet and ResNet. Additionally, we incorporate environmental data such as temperature, solar irradiation, wind speed, and wind direction to enhance anomaly detection. This article combines satellite observations from an open-source database with real images collected from the photovoltaic installation at Mohammed VI Polytechnic University.

This innovative fusion enables a more comprehensive modeling of photovoltaic installation failures. By incorporating weather conditions and imagery, our models can capture complex relationships between environmental variables and the system. At the same time, using CNNs specifically tailored to RGB, and thermal data enhances prediction accuracy by leveraging the unique features of each data type. The outcome is a holistic approach with promising prospects for embedding the solution in a miniaturized system based on low-power and cost-effective microprocessors for widespread and rapid solar field inspection. The performance of this approach has been assessed using metrics such as precision, recall, and F-measure, and a comparison with the latest advancements in the field. This research has the potential to have a significant impact on the solar industry by enabling early anomaly detection, reducing maintenance costs, and enhancing photovoltaic installation performance. Furthermore, it contributes to the transition toward cleaner and more sustainable energy by optimizing solar energy utilization.

II. METHODOLOGY

A. Data Acquisition and Analysis

As we mentioned in the previous chapter, the use of thermal imaging, RGB images, and weather data is crucial for solar panel inspection. In this section, we will describe the collection and analysis of this data.

1) *Thermal and RGB Images*: Temperature significantly influences the efficiency of solar panels. Generally, as the temperature increases, the panel's efficiency decreases. This phenomenon is attributed to the "temperature coefficient," which affects the properties of the photovoltaic cells comprising the solar panels. Photovoltaic cells convert sunlight into electricity using a semiconductor material, typically silicon [19]. When the temperature rises, the electrical conductivity of the semiconductor material also increases, resulting in greater energy loss as heat.

Thermal images depict the temperature distribution on the surface of a solar panel. They are acquired using a thermal camera that measures the infrared radiation emitted by the solar panel. In contrast, RGB images are color images captured using a standard camera and can be presented as regular photos. In this study, FLIR E6xt camera is used to measure the temperature of objects using infrared thermal imaging technology. Unlike traditional cameras, thermal cameras not only capture visible images, but also infrared images emitted by objects based on


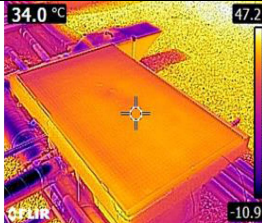

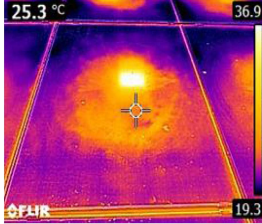

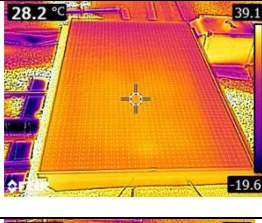

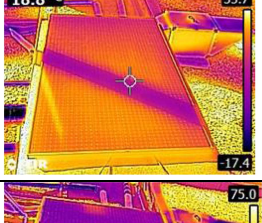

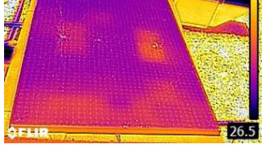
their temperature. The data were collected over the course of one year in 2022. We obtained real images from the campus installation of Mohammed 6 Polytechnic University in Rabat, Morocco, and organized them into folders based on their defect classes manually. The samples were collected from two types of solar panels. First, we used panels connected to the production facility to collect data on existing phenomena or those that we could simulate in this installation. Second, we used test panels that were disconnected from the installation but connected to an electrical load to simulate their actual operation. We utilized a dc projector as the load for these test samples.

We explore various phenomena that can influence the performance of solar panels. We particularly highlight five distinct classes, namely dirt, disconnection, shading, cracks, as well as a class of panels without any issues and well cleaned, the details of each class are shown in Table I. This information enhances our understanding of the data distribution in our study and aids in the evaluation of classification accuracy.

For the first phenomenon, soiling, a phenomenon degrades the efficiency and lifespan of panels [20], [21]. The dirt on solar panels acts as an insulating layer, increasing the panels' temperature and reducing their conversion efficiency. To optimize the efficiency of photovoltaic installations, maintaining a high level of cleanliness is crucial to prevent the deterioration and degradation of solar equipment. To ensure optimal performance, we aim to keep the soiling above 0.7. A threshold of 0.7 is often chosen in the literature as the minimum threshold to exceed before initiating cleaning, representing about 30% of performance loss [22]. This threshold varies depending on the meteorological conditions of the region and the cost of cleaning. However, for an installation in this region, the majority of operators generally take 0.7 as the cleaning threshold. If this threshold is exceeded, there is a risk of panel degradation, leading to a significant decrease in efficiency. Additionally, there is a risk of dust cementation, complicating the subsequent cleaning process. Thus, maintaining a high level of cleanliness is essential for ensuring optimal performance and facilitating the maintenance of photovoltaic installations [23], [24]. For the disconnection class, solar panels can be disconnected for various reasons such as extreme weather conditions, installation problems, or electrical issues. The third issue concerns shading, a significant challenge to solar panel efficiency. Several factors contribute to shading, including physical obstacles like trees, buildings, or surrounding walls, and environmental conditions such as clouds, snow, and the presence of dirt or debris. Additionally, installation problems, such as incorrect panel orientation, can exacerbate shading issues [25].

Finally, solar panels are susceptible to cracking and breaking, which can severely compromise their performance. Mechanical forces, such as wind, snow, hail, and physical impacts, pose significant risks to panel integrity. Additionally, temperature fluctuations and manufacturing defects can induce tension, leading to structural damage. These cracks and breaks render panels vulnerable to external elements like moisture, dust, rain, snow, and ice, exacerbating their deterioration. Ultimately, these damages can culminate in a complete loss of energy production [26].

TABLE I
CHARACTERISTICS FOR RGB AND THERMAL IMAGE DATASET

RGB image	Thermal image	Class and dataset size	Dataset Description
		No Anomaly, 1500	This database is a mixture of samples from panels in normal operation and others that were disconnected from the chain and connected to an electrical load to allow for proper cleaning of the panels and to resolve any other issues that may arise when connected in a chain.
		Soiling, 950	This database contains images showing significant degrees of dirt (soiling <math><0.7</math>) from panels installed in installations with an 80% rate based on 950 features, as well as images of panels connected to loads to simulate a level of dirt that is as low as (soiling <math><0.5</math>) [23] in real installations, which can be difficult to achieve.
		Offline, 600	This class also includes disconnected panels, both separate and adjacent to the installation. To simulate the panels in real installations with disconnected panels adjacent to other connected strings.
		Shaded, 1400	This class contains 1400 features devised in samples of connected and disconnected panels with a combination of multiple obstacles in different shapes. A large part of this class is elaborated with connected panels and weights to facilitate their movement in shading locations such as walls, trees, etc., without disrupting functioning installations.
		Cracked, 70	We use seven panels positioned in various locations to form the array, which are completely disconnected from the installations to avoid any impact on the string

2) *Meteorological Data*: Weather data, air pollution, and external environmental factors such as the presence of buildings or trees casting shadows on the solar panels, can have a combined effect. That is why our model will incorporate all the necessary data to detect the health of photovoltaic solar panels. This model will combine meteorological data from the NASA POWER Project database (Prediction Of Worldwide Energy Resource), which provides high-resolution spatial and temporal climate data to assess renewable energy potential and design sustainable energy systems [27]. The project includes a meteorological database called the POWER Data Access Viewer, which allows users to access and download weather data for various regions of the world.

Dust accumulation on photovoltaic panels can indeed have a negative impact on their performance and energy production. This is due to the dust's ability to reduce the amount of solar light reaching the photovoltaic cells. Our model will be capable

of predicting the condition of photovoltaic panels by combining both meteorological data and thermal and RGB images of each panel. The advantage of this combination is to gather a maximum amount of digital data, which is either available as open-source or easy to collect. These data include temperature, relative humidity, wind speed, wind direction, precipitation, and cleanliness. These parameters are correlated with each other, and we can observe the data used in our recent studies [28], [29], [30].

Data are generated by Xlsat (as it can be seen in Table II), which uses meteorological data as input, cleans it, and then calculates the correlation between these various parameters, with values on the diagonal representing the correlation between each variable and itself (always equal to 1), and values outside the diagonal representing the correlation between two different variables. To calculate a regression matrix with one output variable and multiple input variables, multiple linear regression is used. The steps include preparing the data by adding a bias column

TABLE II
CORRELATION MATRIX OF METEOROLOGICAL DATA AND CLEANLINESS

Variables	Cleanliness	Temperature	Relative humidity	Wind speed	Wind direction	Precipitation
Cleanliness	1	0283	-0294	0179	0 243	0288
Temperature	0283	1	-0912	-0250	0 026	0199
Relative humidity	-0294	-0912	1	0365	-0030	-0142
Wind speed	0179	-0250	0365	1	0 032	0001
Wind direction	0243	0026	-0030	0032	1	-0084
Precipitation	0288	0199	-0142	0 001	-0084	1

to the input variable matrix, then calculating the regression coefficients using the following [31]:

$$\beta = (X^T X)^{-1} X^T y \quad (1)$$

where X is the matrix of input variables with the added bias column and y is the output vector.

For example, the correlation between temperature and relative humidity is -0.9122 , indicating a moderate negative correlation between these two variables. We can observe that temperature is strongly correlated with relative humidity (0.912), suggesting a significant impact of temperature on relative humidity.

Other interactions between variables cannot be captured by a simple linear correlation, which is why it is necessary to use nonlinear models like the latest multiple linear regression (MLR) model we developed, yielding a correlation factor of 0.85 . This phenomenon can be detected using normal RGB imaging or thermal imaging, which will increase the accuracy of dust detection and related phenomena such as soiling or snail trails.

For the parameter ‘‘cleanliness,’’ it represents the inverse of the soiling ratio (SR_i), which is a parameter used to evaluate the level of soiling or dirt accumulation on the solar panels of a photovoltaic installation. It is determined by the ratio of the maximum current of the soiled string to that of the clean string, as indicated in the following:

$$SR_{i,PV} = I_{max_s} / I_{max_c} \quad (2)$$

where $SR_{i,PV}$ is the soiling ratio of the solar panels;

I_{max_s} represents the daily average of the maximum current of the soiled string;

I_{max_c} represents the daily average of the maximum current of the clean string.

The soiling ratio (SR_{i,PV}) is typically expressed as a percentage. The higher the soiling ratio, the greater the level of dirt or soiling on the solar panels, resulting in reduced electricity production from the photovoltaic installation. It is therefore important to maintain proper cleanliness of the solar panels to maximize electricity production and ensure the efficient operation of the photovoltaic system.

According to our latest study, there is a strong correlation between the soiling ratio and meteorological data. Since measuring and integrating this parameter into the prediction model can be

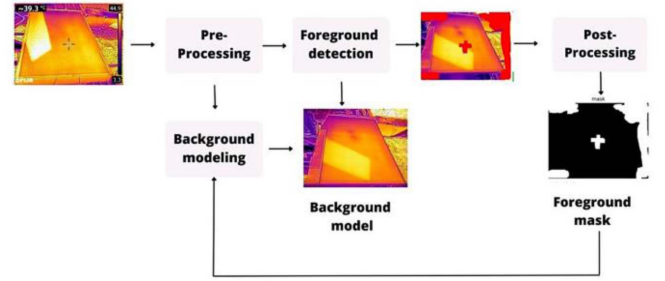


Fig. 1. Thermal image processing diagram.

challenging, we rely solely on meteorological data as inputs for the model. Additionally, we use the minimum, maximum, and daily average values for each meteorological parameter, as well as images that can describe the state of dust on the panels, through either normal RGB imaging or thermal imaging.

To ensure a more reliable inspection, it is recommended to conduct it during peak production hours, typically between 12 and 2 P.M. when solar irradiation is high [15]. This means a higher production current, and the heat effect will be more visible automatically.

B. Data Processing

1) *Image Processing*: When preprocessing RGB and thermal images (see Fig. 1), several essential steps are followed to ensure accurate and consistent results. In this section, we describe these three key steps: The first step is to eliminate any unwanted areas present in the images. Next, we select the focal point as the center of the thermal image. The third step is to resize the images:

a) *Eliminating unwanted areas*: To address the specific issue, we have adapted the techniques proposed in the article [32].

Our approach involves analyzing the image features to identify the blemishes and their surrounding context. We then apply advanced image processing algorithms to consistently fill the affected areas with visually realistic data.

First, we use blemish detection techniques to locate and identify the affected regions in the image. These techniques can be based on color analysis, image segmentation, or CNN. This

provides accurate information about the location and size of the blemishes.

Once the blemishes are detected, we analyze the surrounding context to understand the patterns and structures within the image. We apply a mask to the image to either hide or remove the unwanted object using pixel-by-pixel operations. If the image and mask are in grayscale, you can simply multiply each pixel of the image by the corresponding pixel in the mask. If the mask pixel is white, retain the pixel value from the image. If the mask pixel is black, replace the pixel value in the image with a chosen value (e.g., 0 for grayscale images). If the image and mask are in color (e.g., represented using the RGB model), you need to apply the operations on each color channel (red, green, blue) independently. You can multiply each color channel of the image by the corresponding pixel of the mask in the same way as for grayscale images. The result will be a new image in which the unwanted object specified by the mask is either removed or hidden.

b) Resizing images before training may be necessary for several reasons: First, popular models have been pretrained on 224x224-sized images [33]. This requires resizing the images to the same size for optimal use of these models. Additionally, resizing can reduce the complexity of the images. High-resolution images can be complex and require significant computational resources during model training, especially for RGB images with high resolution. Moreover, resizing helps normalize the input data size. By using the “resize” function as a data preprocessing step, it is possible to adjust the image size to fit the model requirements or other considerations. In addition to the size of 224x224 pixels, other dimensions can be used depending on specific requirements.

c) Data augmentation: For data augmentation, two essential libraries were used: Keras and scikit-image. Keras provides functionality for generating augmented image data, while scikit-image allows for image reading. The ImageDataGenerator function from Keras was used to augment the image data [34]. This function is highly convenient as it allows for defining various transformations to apply to the original image, generating additional variations in the dataset.

Data augmentation offers several advantages. First, it helps increase the size of the dataset, which is particularly beneficial when data are limited. This provides more diversity to the model and can improve its ability to generalize and recognize new images. By subjecting the model to variations such as rotations, deformations, zooms, horizontal flips, and changes in brightness, data augmentation can also enhance the model’s robustness. This diversity helps the model adapt to real-world scenarios where images may exhibit such variations.

Lastly, data augmentation helps reduce the risk of overfitting. When the model becomes too specific to the training data, it may struggle to generalize to new data. By introducing variations in the dataset through augmentation, we assist the model in developing better generalization capabilities [35].

The program utilizes the libraries Keras, skimage, TensorFlow, NumPy, Matplotlib, and cv2 to perform data augmentation on images. It traverses the specified directory, and for each file,

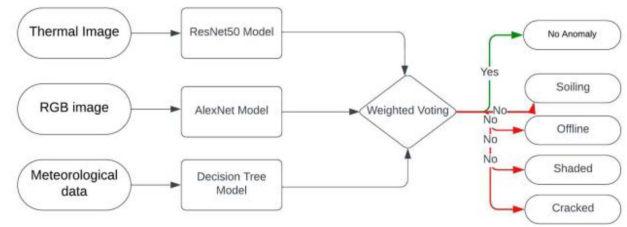


Fig. 2. Multimodal anomaly detection in photovoltaic systems using thermal and RGB images with meteorological data integration.

it loads the image, applies specific transformations using ImageDataGenerator, generates batches of augmented images, and saves them to specific directories. This allows for the creation of an augmented dataset for each type of anomaly.

2) Weather Data Processing: Preprocessing meteorological data are crucial steps to ensure accurate and reliable results in subsequent analysis or modeling. In this article, we describe two main steps in preprocessing CSV-format meteorological data: normalization and handling missing values. These steps are essential in preparing meteorological data for tasks such as weather prediction or climate trend analysis.

a) Normalization of data: Data normalization is a crucial step in preprocessing meteorological data. It aims to bring all variables to the same scale, preventing certain variables from dominating others in terms of magnitude. This ensures that all variables contribute to the analysis in a balanced manner. Various techniques can be used for normalization, including scaling values between 0 and 1. This step standardizes the data and facilitates comparisons between different variables.

b) Handling-missing values: Missing values in meteorological data can pose challenges during analysis. They can introduce biases or errors in the obtained results. There are several approaches to handling missing values, such as removing rows or columns with missing values, imputing missing values using techniques like mean, median, or regression, or utilizing more advanced methods based on ML models. In our case, we chose to handle missing values by removing the affected rows or columns. However, it is important to ensure the overall integrity of the dataset is maintained during this step.

C. Model Description

The proposed model (see Fig. 2) has a specific objective: to detect and classify thermal leaks in solar panels using thermal and RGB images, while also making anomaly predictions based on weather data. It is divided into four sections, each playing a distinct role. The first section focuses on loading a pretrained model that specializes in classifying thermal images. To accomplish this task, it utilizes the powerful ResNet architecture, which has been proven effective in image classification. Moving on to the second section, it handles the loading of a pretrained model designed for classifying RGB images. Here, the widely recognized AlexNet architecture is employed, which is known for its exceptional performance in image classification tasks.

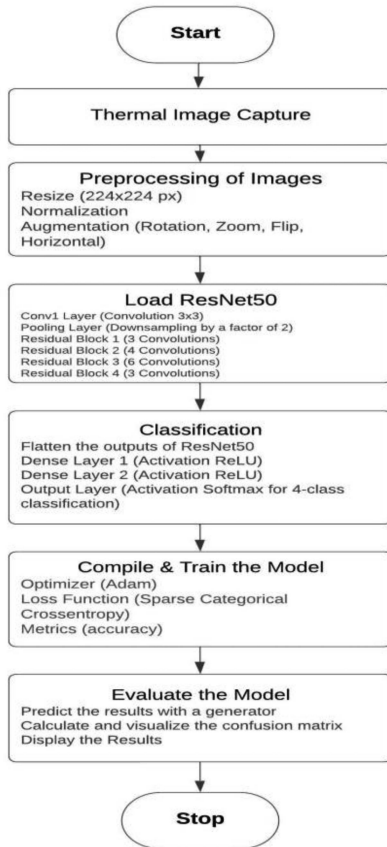


Fig. 3. Workflow for thermal image-based anomaly detection in photovoltaic systems using ResNet50.

The third section is dedicated to anomaly prediction, utilizing weather data. In this case, the program reads the relevant data from a CSV or Excel file.

Preprocessing steps are applied to the weather data, and predictions are made using a specialized model that corresponds to this type of input. Finally, the fourth section combines the predictions generated by the three aforementioned models. To achieve this, predefined weights are assigned to each model. The thermal and RGB images are preprocessed, and predictions are made using their respective models. Similarly, the weather data are also preprocessed, and predictions are made using the corresponding model. These probabilities are then weighted according to the predefined weights, and the class with the highest weighted probability determines the final predicted outcome. The ultimate goal of this program is to consolidate the predictions made by different models, including those for thermal images, RGB images, and weather data. By leveraging multiple sources of information, the program enhances the accuracy of anomaly predictions, providing a comprehensive assessment of potential anomalies present in the images.

1) *ResNet50 for Thermal Images*: The ResNet50 model (see Fig. 3) is a deep CNN architecture that was introduced by Microsoft Research in 2015. ResNet50 belongs to the family of ResNet models (residual networks), which were designed to address the problem of performance degradation with increasing

network depth. One of the key features of ResNet models, including ResNet50, is the use of residual blocks. These blocks introduce skip connections to bypass certain layers of the network, allowing information to propagate more easily through the network. This helps alleviate the issue of vanishing or exploding gradients during the training of deep networks [36], [37].

The ResNet50 model consists of 50 convolutional layers, pooling operations, and fully connected layers. It utilizes 3×3 convolutions and pooling layers with a subsampling factor of 2 to reduce the spatial dimension of features. The ResNet50 architecture has been pretrained on a large dataset of images called ImageNet. ResNet50 can be finetuned for specific tasks by adding additional layers for classification or regression, depending on the problem at hand. It has been widely used in areas such as object recognition, object detection, image segmentation, and other computer vision tasks.

The model uses libraries such as TensorFlow and Keras, which are popular libraries for building and training deep learning models. Next, the training and testing data directories are defined. Data preprocessing is performed using the `ImageDataGenerator` class from Keras. This class allows generating batches of augmented and preprocessed images from the data directories. The images are resized, their value scale is normalized, and transformations such as rotation, zoom, and horizontal flipping are applied to increase the diversity of the training data. The pretrained ResNet50 model is then loaded using the `ResNet50` function from Keras, with weights pretrained on the ImageNet dataset. The fully connected layers are excluded, and the weights of the layers in the pretrained model are frozen to avoid gradient backpropagation during training. The classification model is built by adding additional layers on top of the ResNet50 model. A Flatten layer is used to flatten the outputs of the ResNet50 model, followed by two Dense (fully connected) layers with ReLU and softmax activation functions. The output layer has a softmax activation to perform classification into four classes. The model is then compiled by specifying the optimizer (Adam), the loss function (`sparse_categorical_crossentropy`), and the evaluation metrics (accuracy). The model is trained using the fit method with the previously defined training and testing data generators and the specified number of epochs. Finally, the trained model can be evaluated using a separate test dataset. The model predictions are obtained using the `predict_generator` method, and the confusion matrix is computed from the predictions and the true labels. The confusion matrix can be visualized using the `matplotlib` library to evaluate the model's performance.

2) *AlexNet for RGB Images*: AlexNet model (see Fig. 4) is one of the most popular architectures for CNNs. It is known for introducing several innovative concepts when it was presented in 2012. The AlexNet architecture is characterized by the direct stacking of multiple convolutional layers, which allows the network to capture complex features and use them for accurate classifications [38].

The first convolutional layer in AlexNet is an input layer that processes input images of size $227 \times 227 \times 3$. Here, " 227×227 " refers to the spatial size of the image, and " 3 " indicates the three-color channels (red, green, blue) in the RGB image. The first convolutional layer uses 96 filters of size $11 \times 11 \times 3$. Each

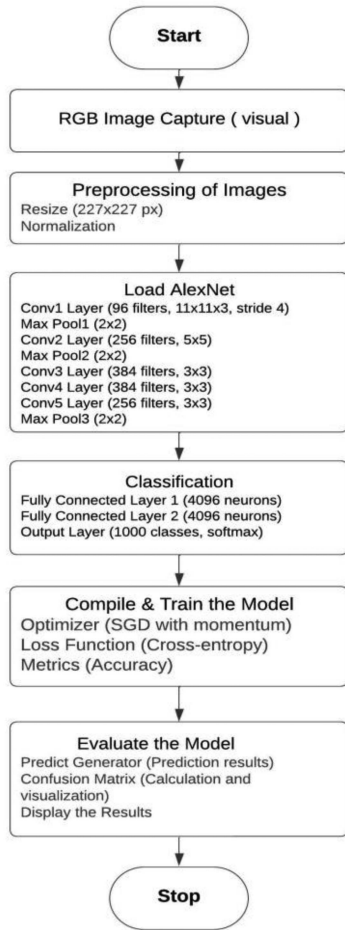


Fig. 4. Workflow for RGB image-based anomaly detection in photovoltaic systems using AlexNet.

filter is a small weight matrix that is applied to the input image to extract specific features. The filter size, in this case, is 11×11 , and it is applied to all three-color channels of the image. The result of this convolution produces 96 filter maps. The stride of 4 pixels indicates that the filter moves 4 pixels at a time during convolution. This means that the extracted features are sampled at intervals of 4 pixels, which reduces the spatial resolution of the filter maps produced by this layer.

The filters in the second, fourth, and fifth convolutional layers in AlexNet are connected only to the filter maps from the previous layer residing on the same graphics processing unit (GPU). This means that these layers efficiently exploit parallelization using multiple GPUs to accelerate computations. The third convolutional layer in AlexNet is connected to all the filter maps from the second convolutional layer. This means that the features extracted in this layer consider information from all the filter maps of the previous layer.

Finally, the neurons in the fully connected layers in AlexNet are connected to all the neurons in the previous layer. This allows for global communication between neurons, which is important for classification tasks.

In summary, AlexNet is a CNN architecture that introduced innovative concepts in the field of computer vision. It uses

stacked convolutional layers, filters of different sizes, and specific connections between layers to extract features and perform accurate classifications on RGB images.

3) *Decision Tree*: In the CSV dataset containing weather data, the decision tree ML algorithm is used. It is a nonparametric supervised learning algorithm used for both classification and regression tasks. It has a hierarchical structure, and tree-like structure, consisting of a root node, branches, internal nodes, and leaf nodes [39].

The program imports necessary libraries such as pandas and matplotlib.pyplot to work with the data and generate visualizations. Then, it loads the data from an Excel file using the “pd.read_excel()” function. The data are stored in a DataFrame object called “data.” Next, the program splits the data into independent variables “X” and dependent variable “Y.” The columns in “X” contain all the features or explanatory variables, except for the last column which contains the target labels or classes in “Y.” The program displays the first few rows of “X” and “Y” to provide an overview of the data. It also shows information about “X” using the “info()” method, showing the total number of entries, data types, and memory consumption. Additionally, it checks for any null values in “X” using “isnull().sum().sort_values(ascending = False).” Then, the program performs feature scaling using “MinMaxScaler” from “sklearn.preprocessing.” The values in each column of “X” are scaled to the range [0,1] using the “fit_transform()” method. The scaled data are stored in a new DataFrame called “min_max_Scalar_df.” In summary, the program loads the data, splits it into independent and dependent variables, performs feature scaling, and displays information about the data before and after scaling. This prepares the data to be used in ML models or for further analysis.

4) *Weighted Voting*: We are exploring a weighted voting approach for multiclass classification using three different models. The objective is to combine the predictions of these models by assigning them weights based on their performance and reliability. To begin, we evaluate the performance of each model on a validation set [40]. This allows us to determine the weights to assign to each model. Models that achieve better results are assigned higher weights, reflecting their confidence and accuracy. Once the weights are assigned, we weigh the predictions of each model by multiplying each prediction by its corresponding weight. For example, if Model 1 has a weight of 0.4, Model 2 has a weight of 0.3, and Model 3 has a weight of 0.3, we adjust the predictions of Model 1 by multiplying them by 0.4, Model 2 predictions by 0.3, and Model 3 predictions by 0.3. By combining the weighted predictions from each model, we obtain a final prediction for each data instance. This weighted voting approach allows us to leverage the individual strengths of each model while compensating for any potential weaknesses they may have.

III. RESULTS

This section describes the results for kernel fragmentation and overlength recognition using the Faster R-CNN variants. We compare our model variants against a

TABLE III
THERMAL AND RGB IMAGE CLASSIFICATION RESULTS

Model	Accuracy training	Accuracy test	Loss training	Loss test
ResNet for Thermal images	0.9793	0.97934	0.0552	0.05524
AlexNet for RGB images	0.9287	0.9287	0.033706	0.033707

baseline naive Faster R-CNN trained with standard parameters as defined in Section II-C. This way we can evaluate and compare against a standard practice of deep learning development of simply training on a large dataset. The naive models are denoted Baseline for kernel fragmentation and all for overlenghts.

A. Image Processing

Table III presents the evaluation results of two machine-learning models applied to different types of images: AlexNet for RGB images and ResNet for thermal images. The models' performances are measured in terms of accuracy on the training and test sets, as well as the loss during training and testing.

For the AlexNet model, the accuracy is 0.9287 on both the training and test datasets. This indicates that the model is able to predict approximately 92.87% of the samples in both sets. The loss values during training and testing are 0.033706 and 0.033707 respectively, suggesting that the model has successfully minimized errors during training.

Regarding the ResNet model applied to thermal images, the accuracy is 0.9793 on the training dataset and 0.97934 on the test dataset. This indicates that the model is able to predict approximately 97.93% of the samples in both sets. The loss values during training and testing are 0.0552 and 0.05524, respectively, suggesting low error during both the training and testing phases.

This suggests that both models are capable of generalizing well and providing accurate predictions for the specific types of images on which they were trained.

For the ResNet model, the training and validation performances of a neural network model are provided for 500 epochs. The architecture and objective of the model are not specified, but we can provide a general analysis based on the given information.

The model starts with a training loss of 1.3929 and an accuracy of 0.3349 (see Fig. 5). Over the initial epochs, the loss and accuracy values fluctuate, but there is an overall improvement in both metrics. The validation loss and accuracy follow a similar pattern.

As the subsequent epochs progress, the model appears to further improve. The training and validation losses decrease, indicating that the model is learning and better fitting the data. The accuracy values also increase, suggesting that the model is making more accurate predictions.

Similarly, for the AlexNet model illustrated in Fig. 6, it seems that the model's performance improves with each epoch, as both the training loss and validation loss decrease, while the accuracy

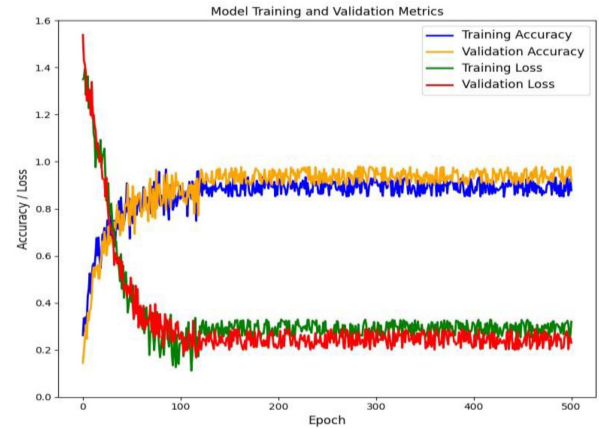


Fig. 5. Accuracy convergence graph images for ResNet model.

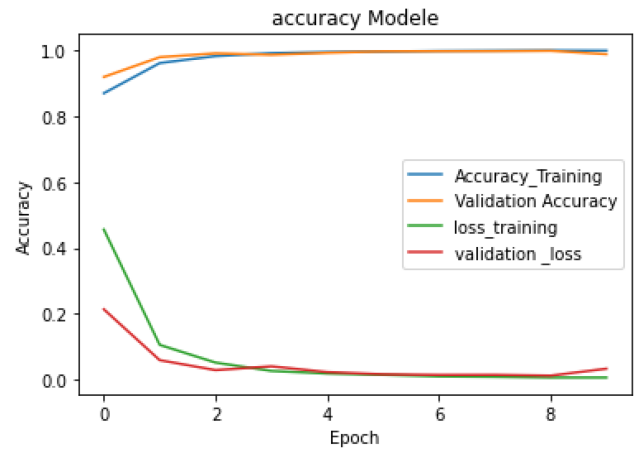


Fig. 6. Accuracy convergence graph images for AlexNet model.

TABLE IV
METEOROLOGICAL DATA CLASSIFICATION RESULTS

Model	Accuracy	Precision	Recall
KNN	0.4547	0.4547	0.4547
ANN	0.5433	0.5433	0.5433
SVM	0.6422	0.6422	0.6422

increases. This suggests that the model is learning and enhancing its ability to classify the provided data.

B. Machine Learning

After running tests on four ML models, the results are as follows in Table IV.

The accuracy, precision, and recall values are all identical, indicating that the model's predictions are balanced between positive and negative classes. In other words, the model makes an equal number of correct predictions for both positive and negative cases. Among the four models, the SVM performs the best; therefore, we choose to model this problem using the SVM model.

```

1/1 [=====] - 2s 2s/step
1/1 [=====] - 0s 309ms/step
The predicted image is: shadowing
No Anomaly: 2.912408858537674%
offline Module: 16.404107365323075%
shadowing: 80.6829412866092%
soiling: 0.0005394216259446694%

```

Fig. 7. Example of sample prediction results.

C. Model Exploitation

The program loads pretrained models, performs predictions on thermal images, and combines weighted predictions to obtain a final prediction. It uses three different models (ResNet, AlexNet, and a classification model), loads the weights of the models, and preprocesses the images before feeding them into the models for predictions. The accuracies of the models are used as weights to weigh the predictions, and the predictions from each model are combined using these weights. Finally, the predicted final class and the probabilities associated with each class are displayed.

Example in Fig. 7, we tested an image from the shadowing class. The program searched for the class with the highest probability among the weighted predictions. The results are displayed, showing the predicted final class and the probabilities associated with each class.

The prediction speed is 0.3 s, demonstrating the swiftness of our prediction program. When we have high percentages for two or more classes, this may be due to issues in the panel, especially regarding soiling problems. The model can detect both dust and another issue, such as offline status, and assign them significant percentages compared to others. This implies a potential coexistence of multiple problems in the panel, significantly influencing the results.

IV. DISCUSSION

The article presents an innovative approach for inspecting photovoltaic panels using a portable hybrid model that combines digital data processing and image analysis. The goal of this approach is to detect defects quickly and accurately in photovoltaic panels, enabling timely maintenance and repair. The approach involves collecting data from the photovoltaic panels using a portable device, which is then processed using image processing techniques to identify any defects. The inspection results are presented in a user-friendly format, allowing for easy interpretation and analysis. This approach has the potential to significantly improve the efficiency and effectiveness of photovoltaic panel inspections, leading to enhanced performance and increased durability of solar energy systems. This technology is important as photovoltaic panels are a key component of solar energy infrastructure, and defects in these panels can reduce their performance and lifespan. Regular inspection of photovoltaic panels is therefore crucial to ensure efficient and reliable energy production. However, there are challenges to overcome with

this approach. First, the quality of the data collected from the photovoltaic panels can be affected by various environmental factors, such as sunlight, weather conditions, and dirt. Additionally, the accuracy of defect identification relies on the quality of the image processing algorithm used. Finally, it is important to note that this approach requires technical skills and training for the use of the portable device and image processing algorithm. Therefore, adequate training and awareness among operators will be necessary to ensure effective and safe utilization of this technology.

A. Benefits of Combining Data and Images

Temperature and its distribution are crucial parameters for diagnosing the condition of solar panels and for classifying faults based on temperature distribution. Other phenomena can also be detected through RGB imaging. However, the connection to the external environment of the panel facilitates the detectability of faults. That is why we have integrated a ML model, which can detect phenomena that are only detectable through this data.

The combination of imaging and data processing helps reveal the relationships between different types of data in order to better classify problems. Therefore, the use of a SVM model helps to better understand how different parameters affect the temperature and performance of the solar panel. These relationships can then be used to classify different problems with an acceptable accuracy of 0.64 when combined with the two classification models for thermal and RGB imaging, which achieve a high accuracy of up to 0.98.

In the existing literature, various studies utilize image processing and modeling techniques for anomaly classification in photovoltaic systems. However, the distinctiveness of our study lies in the flexibility to integrate three types of data with varying accuracies depending on data quality and model adaptability. For instance, the performance of ML models applied to thermal images has been explored in several studies like [41] tested Faster R-CNN, SSD, and SSD FPN on three PV plants, revealing that Faster R-CNN achieved test accuracies ranging from 95.81% to 98.97%, while SSD and SSD FPN showed significant performance variations across different sites (e.g., 73.54% to 99.23% for SSD), indicating potential instability of the models in varying environments. Additionally, a study in Procedia Computer Science [42] applied models such as ResNet-50 and Faster R-CNN for hotspot detection in thermal images, where ResNet-50 achieved an F1-Score of 85.37%, outperforming other models like DenseNet-169 (F1-Score of 0.72), and Faster R-CNN achieved a mean average precision of 0.67. Conversely, for RGB images, a study [43] demonstrated the use of a simple DCNN architecture and VGG-16 model for classifying healthy and faulty classes, achieving accuracies of 98.39% and 99.91%, respectively, with VGG-16 maintaining similar accuracy for five-class segmentation (99.80%). Moreover, another study published [44] utilized WEKA for feature selection and fault classification in photovoltaic modules using k-Nearest Neighbor (kNN), J48 decision tree, and SVM, where ensemble models combining these classifiers improved classification accuracy, with the two-class ensemble of kNN

and SVM achieving 98.30%. These findings emphasize the importance of selecting models tailored to specific data types and conditions, highlighting the innovative aspect of our study in combining different data types and models to enhance accuracy and robustness.

B. Integrating Data-Driven Solutions

A data-driven system based on a Raspberry Pi board is used for data acquisition, preprocessing, and prediction using the generated model. This data-driven solution relies on specific hardware, including a Flir camera capable of simultaneously capturing thermal and RGB images. Additionally, downloading satellite observations is essential for this embedded system, which is frequently used in similar studies [45].

A human-machine interface facilitates the management of this system by allowing the triggering of data acquisition, preprocessing, and applying the model to generate predictions corresponding to the sample's class. The compatibility of the Raspberry Pi board with the Python programming language is a major advantage of this approach. The model, developed in Python, can be deployed, and executed directly on the Raspberry Pi board thanks to support for popular libraries such as TensorFlow, PyTorch, or Keras. This compatibility and the accessibility of Python libraries fully leverage the computational power of the Raspberry Pi board [46], ensuring efficient model execution and accurate results. This integration also simplifies the development, deployment, and execution process of the model on the Raspberry Pi board, providing an integrated and convenient solution for our data-driven system.

The Raspberry Pi board and the FLIR camera are connected to the local WiFi network, facilitating the solution's portability. The camera can be moved during inspections, powered by an integrated lithium battery with a 4-h autonomy. As for the Raspberry Pi board, it can be powered by either a wall outlet or a 5 V battery. A notable advantage of this board is its low power consumption, which is around 6 W.

This approach comes with a total cost of US\$ 2000, resulting in a significant reduction in operational expenses compared to the traditional manual inspection method. The latter takes more than five times the inspection time compared to the method described here, which allows for an average of 5 s per panel inspection. This efficiency translates to an 80% reduction in labor costs.

C. Time Integration (Time Series and Real-Time Inverter Data)

As a perspective for our study, we propose integrating models based on time series analysis, which can be used to detect solar panel issues by analyzing energy production data over time, as the majority of problems are cumulative phenomena that change over time, such as dust accumulation.

Trend detection, seasonality analysis, prediction, and event impact evaluation: The time dimension allows for the analysis of event impacts on data. Volatility analysis in the data: On the other hand, time series models are specifically designed for analyzing data that varies over time, such as annual data. These models are capable of capturing trends, seasonal patterns,

and recurring cycles in the data, enabling accurate forecasting and understanding of temporal fluctuations. Time series models can be used to analyze various features of annual data, such as the overall trend, seasonal effects, volatility, and correlations between different years. This provides real-time information on the installation's state using past and current measurements. This method allows for the rapid detection of variations and anomalies, facilitating the diagnosis of potential problems and enabling necessary actions to improve installation efficiency.

We can also include in our inputs the data from inverters, which provide real-time production information, such as input voltage and current for each string, cumulative production, inverter output power, and overall equipment consumption. This provides a better understanding of the installation's state, particularly by comparing real-time predicted production with meteorological data and the estimated installation state, which considers factors such as tilt angle, dust, and shading, using real-time energy measurements from the inverter. This technique also expands the scope of installation inspection to other unaddressed issues, such as electrical and connectivity problems, overvoltage, and overloading. Furthermore, it enables better production management by leveraging the installation's state to optimize energy efficiency, as demonstrated in our recent study [47], where we used IoT to manage the issue of excess production in the same installation at Mohammed 6 Polytechnic University.

V. CONCLUSION

In conclusion, this article presents a novel and innovative approach for the inspection of photovoltaic panels by combining data and image processing. This approach enables the rapid and accurate detection of panel defects, facilitating timely maintenance and repair. Regular inspection of photovoltaic panels is crucial to ensure their proper functioning, enhance their performance, and prolong their lifespan.

The proposed approach in this article offers numerous advantages for improving the efficiency and effectiveness of photovoltaic panel inspection, as well as enhancing the performance and durability of solar energy systems. It can be applied to various types of solar installations, including residential, commercial, or industrial, based on new AI approaches.

We have employed two CNN models, AlexNet and ResNet, to classify thermal and RGB images of photovoltaic systems. The performance of our models has been evaluated using accuracy, loss, and a weighted voting scheme. We have utilized a large dataset collected from the solar installation at Mohammed 6 Polytechnic University in Rabat, Morocco, along with meteorological observations from NASA POWER database. Our results have been compared with other state-of-the-art methods, demonstrating comparable or better performance.

Furthermore, we have discussed the challenges associated with training and evaluating CNN models for fault detection in photovoltaic systems. These challenges include acquiring a diverse and extensive dataset of images, avoiding overfitting the models to the training data, and deploying the models in real-world scenarios.

We believe that our methodology shows promise in fault detection for photovoltaic systems. Our future work involves further improving the performance of our models, addressing the challenges, and deploying them in real-world scenarios.

APPENDIX

A. Classification by Machine Learning

ML is a branch of AI that focuses on the development of algorithms and models capable of learning from data and making decisions or predictions without being explicitly programmed. Classification is one of the most common tasks in ML, where the goal is to predict the class or label of a data sample based on its features. Classification involves training a model on a training dataset where each sample is labeled with a known class. The model learns from these examples to find patterns or relationships between the features and the corresponding classes. Once the model is trained, it can be used to predict the classes of new, unlabeled samples.

There are different types of classification algorithms in ML, such as logistic regression, decision trees, SVM, neural networks, etc. Each algorithm has its own strengths, weaknesses, and underlying assumptions, and it is important to choose the appropriate model based on the data characteristics and the objectives of the classification problem.

The performance evaluation of a classification model is typically done using metrics such as accuracy, recall, F-measure, and the confusion matrix. These metrics measure the accuracy of the model's predictions and its ability to correctly discriminate between different classes.

B. CNN Model

Convolutional layers are capable of learning visual features from raw images by applying convolution filters. These filters are weight matrices that are applied to adjacent areas of the image to detect specific patterns and features.

The CNN model is composed of multiple layers, each of which is responsible for a specific step in the processing of the input image. The first layer is typically a convolutional layer that applies filters to the input image [45]. This layer is followed by an activation layer that adds nonlinearity to the model by applying an activation function to the output of the convolutional layer. Several CNN architectures have been developed and used for computer vision tasks. Here are some of the most popular architectures: LeNet-5, AlexNet, VGGNet, GoogleNet, and ResNet. For our study, we have chosen two models that are most suitable for this type of problem, namely AlexNet for RGB images and ResNet for thermal images. The main difference between AlexNet and ResNet lies in their architecture. AlexNet is simpler and less deep, while ResNet is deeper and uses residual connections to facilitate learning in deep networks. Since thermal images contain more details and it can be challenging to distinguish between different anomalies, we have opted for a more complex model that is better suited for classifying the various anomalies.

C. Model Performance

To calculate the metrics of a CNN classification model, we use the following formulas. They will be the same for ML models, except for ML, where we include three metrics: Accuracy, Precision, and Recall, which are equal in our case because there are no false positives, false negatives, or misclassifications in the model's outputs [48].

1) *Accuracy (Precision)*: Accuracy measures the proportion of correct predictions made by the model compared to the total number of predictions.

$$\text{Accuracy} = \text{TP} / (\text{TP} + \text{FP}) \quad (3)$$

TP: Number of correct predictions on the training set for the 5 classes (shaded, no anomaly, dust, etc.)

FP: Number of incorrect predictions on the training set for the 5 classes (shaded, no anomaly, dust, etc.).

It is applied to both the training and test datasets, which are split before training. These datasets allow for a realistic estimation of the model's performance on new data. This evaluation helps determine whether the model is good enough to be deployed in a production environment or if it still requires improvements.

2) *Loss*: Loss is a measure of the model's error when making predictions. It quantifies the quality of the model's predictions and provides necessary feedback to adjust the model's weights during training. In our models, we have used the `sparse_categorical_crossentropy` function, which offers advantages in terms of memory efficiency, simplicity of data flow, and similar performance to `categorical_crossentropy`. The following formula clarifies this calculation, where S refers to samples, C refers to classes, and $(S\&C)$ represents the samples belonging to class C [49]:

$$\text{Loss} = -\log p(S\&C). \quad (4)$$

These metrics are calculated using TensorFlow, a numerical computation library.

3) *Final Model Performance*: The final model is a combination of three models using weighted voting, which assigns more weight to optimal predictions. The final prediction is a combination of classes based on the assigned weights. Therefore, the final performance of the model will fall between that of the least performing model and the optimal model, with convergence toward the optimal model in most cases. Weighted voting allows the final model to leverage the strengths of each individual model, improving the overall performance by considering their expertise. However, it is important to note that the final performance of the combined model will depend on the individual performances of the underlying models and the quality of the assigned weights

$$\hat{Y} = \text{mode} \{C1(x), C2(x), \dots, Cm(x)\}. \quad (5)$$

We predict the label \hat{Y} of the case by voting of each classifier Ci using (1), we compute a weighted majority vote by associating a weight Wj with classifier Cj : where Xi is the characteristic function [$Cj(X) = I \in A$], and A is the set of unique class labels [21].

While the final model typically outperforms the least performing model, it may not reach the same level of performance as the optimal model. The final performance will depend on the quality of predictions from individual models and the weights assigned to each mode.

ACKNOWLEDGMENT

The authors appreciate the opportunity provided by UM6P to collaborate on the solar installation project at the Rabat campus in Morocco. This work is made possible by their invaluable partnership.

REFERENCES

- [1] A. Azouzoute, A. A. Merrouni, and S. Touili, "Overview of the integration of CSP as an alternative energy source in the MENA region," *Energy Strategy Rev.*, vol. 29, 2020, Art. no. 100493.
- [2] M. B. Haines, S. Moore, and T. Adornetto, "Suspending democratic (dis)belief: Nonliberal energy politics of solar power in Morocco and Tanzania," *Energy Res. Soc. Sci.*, vol. 96, Feb. 2023, Art. no. 102942, doi: [10.1016/j.erss.2023.102942](https://doi.org/10.1016/j.erss.2023.102942).
- [3] P. Nooralishahi et al., "Drone-based non-destructive inspection of industrial sites: A review and case studies," *Drones*, vol. 5, no. 4, Sep. 2021, Art. no. 106, doi: [10.3390/drones5040106](https://doi.org/10.3390/drones5040106).
- [4] N. Sakib and T. Wuest, "Challenges and opportunities of condition-based predictive maintenance: A review," *Procedia CIRP*, vol. 78, pp. 267–272, 2018, doi: [10.1016/j.procir.2018.08.318](https://doi.org/10.1016/j.procir.2018.08.318).
- [5] A. Aslam et al., "Advances in solar PV systems; A comprehensive review of PV performance, influencing factors, and mitigation techniques," *Energies*, vol. 15, no. 20, Oct. 2022, Art. no. 7595, doi: [10.3390/en15207595](https://doi.org/10.3390/en15207595).
- [6] K. Hasan et al., "Effects of different environmental and operational factors on the PV performance: A comprehensive review," *Energy Sci. Eng.*, vol. 10, no. 2, pp. 656–675, Feb. 2022, doi: [10.1002/ese3.1043](https://doi.org/10.1002/ese3.1043).
- [7] R. Cavieres et al., "Automatic soiling and partial shading assessment on PV modules through RGB images analysis," *Appl. Energy*, vol. 306, Jan. 2022, Art. no. 117964, doi: [10.1016/j.apenergy.2021.117964](https://doi.org/10.1016/j.apenergy.2021.117964).
- [8] R. H. F. Alves, G. A. de Deus Júnior, E. G. Marra, and R. P. Lemos, "Automatic fault classification in photovoltaic modules using convolutional neural networks," *Renewable Energy*, vol. 179, pp. 502–516, Dec. 2021, doi: [10.1016/j.renene.2021.07.070](https://doi.org/10.1016/j.renene.2021.07.070).
- [9] S. A. Memon et al., "A machine-learning-based robust classification method for PV panel faults," *Sensors*, vol. 22, no. 21, Nov. 2022, Art. no. 8515, doi: [10.3390/s22218515](https://doi.org/10.3390/s22218515).
- [10] P. Haidari, A. HajiAhmad, A. Jafari, and A. Nasiri, "Deep learning-based model for fault classification in solar modules using infrared images," *Sustain. Energy Technol. Assessments*, vol. 52, Aug. 2022, Art. no. 102110, doi: [10.1016/j.seta.2022.102110](https://doi.org/10.1016/j.seta.2022.102110).
- [11] Y. Su, F. Tao, J. Jin, and C. Zhang, "Automated overheated region object detection of photovoltaic module with thermography image," *IEEE J. Photovolt.*, vol. 11, no. 2, pp. 535–544, Mar. 2021, doi: [10.1109/JPHOTOV.2020.3045680](https://doi.org/10.1109/JPHOTOV.2020.3045680).
- [12] P. Masita, A. Hasan, and T. Shongwe, "75MW AC PV module field anomaly detection using drone-based IR orthogonal images with res-CNN3 detector," *IEEE Access*, vol. 10, pp. 83711–83722, 2022, doi: [10.1109/ACCESS.2022.3194547](https://doi.org/10.1109/ACCESS.2022.3194547).
- [13] M. W. Akram et al., "Automatic detection of photovoltaic module defects in infrared images with isolated and develop-model transfer deep learning," *Sol. Energy*, vol. 198, pp. 175–186, Mar. 2020, doi: [10.1016/j.solener.2020.01.055](https://doi.org/10.1016/j.solener.2020.01.055).
- [14] A. Di Tommaso, A. Betti, G. Fontanelli, and B. Michelozzi, "A multi-stage model based on YOLOv3 for defect detection in PV panels based on IR and visible imaging by unmanned aerial vehicle," *Renewable Energy*, vol. 193, pp. 941–962, Jun. 2022, doi: [10.1016/j.renene.2022.04.046](https://doi.org/10.1016/j.renene.2022.04.046).
- [15] Y. Zefri, I. Sebari, H. Hajji, and G. Aniba, "Developing a deep learning-based layer-3 solution for thermal infrared large-scale photovoltaic module inspection from orthorectified big UAV imagery data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 106, Feb. 2022, Art. no. 102652, doi: [10.1016/j.jag.2021.102652](https://doi.org/10.1016/j.jag.2021.102652).
- [16] L. Pratt, D. Govender, and R. Klein, "Defect detection and quantification in electroluminescence images of solar PV modules using U-Net semantic segmentation," *Renewable Energy*, vol. 178, pp. 1211–1222, Nov. 2021, doi: [10.1016/j.renene.2021.06.086](https://doi.org/10.1016/j.renene.2021.06.086).
- [17] T. Ahmad, H. Chen, Y. Guo, and J. Wang, "A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review," *Energy Buildings*, vol. 165, pp. 301–320, Apr. 2018, doi: [10.1016/j.enbuild.2018.01.017](https://doi.org/10.1016/j.enbuild.2018.01.017).
- [18] G. B. Goh, N. O. Hodas, and A. Vishnu, "Deep learning for computational chemistry," *J. Comput. Chem.*, vol. 38, no. 16, pp. 1291–1307, Jun. 2017, doi: [10.1002/jcc.24764](https://doi.org/10.1002/jcc.24764).
- [19] A. A. E. Amin and M. A. Al-Maghrabi, "The analysis of temperature effect for mc-Si photovoltaic cells performance," *Silicon*, vol. 10, no. 4, pp. 1551–1555, Jul. 2018, doi: [10.1007/s12633-017-9639-5](https://doi.org/10.1007/s12633-017-9639-5).
- [20] A. Azouzoute, A. A. Merrouni, and M. Garoum, "Soiling loss of solar glass and mirror samples in the region with arid climate," *Energy Rep.*, vol. 6, pp. 693–698, 2020.
- [21] A. Azouzoute, M. Garoum, F. Jeffali, and A. Ghennioui, "Experimental study of dust effect on the transmission of a glass PV panel for a fixed and tracking system," *Mater. Today Proc.*, vol. 27, pp. 3091–3094, 2020.
- [22] A. Azouzoute et al., "Developing a cleaning strategy for hybrid solar plants PV/CSP: Case study for semi-arid climate," *Energy*, vol. 228, 2021, Art. no. 120565.
- [23] M. Abraim et al., "Techno-economic assessment of soiling losses in CSP and PV solar power plants: A case study for the semi-arid climate of Morocco," *Energy Convers. Manage.*, vol. 270, 2022, Art. no. 116285.
- [24] M. Abraim et al., "PVSMS: A system for quantifying soiling effects and optimizing cleaning schedule in PV solar plants," *Energy Convers. Manage.*, vol. 284, 2023, Art. no. 116978.
- [25] G. Lobaccaro et al., "A methodological analysis approach to assess solar energy potential at the neighborhood scale," *Energies*, vol. 12, no. 18, Sep. 2019, Art. no. 3554, doi: [10.3390/en12183554](https://doi.org/10.3390/en12183554).
- [26] C. Buerhop et al., "Evolution of cell cracks in PV-modules under field and laboratory conditions," *Prog. Photovolt. Res. Appl.*, vol. 26, no. 4, pp. 261–272, Apr. 2018, doi: [10.1002/pip.2975](https://doi.org/10.1002/pip.2975).
- [27] J. W. White, G. Hoogenboom, P. W. Stackhouse, and J. M. Hoell, "Evaluation of NASA satellite- and assimilation model-derived long-term daily temperature data over the continental US," *Agricultural Forest Meteorol.*, vol. 148, no. 10, pp. 1574–1584, Sep. 2008, doi: [10.1016/j.agrformet.2008.05.017](https://doi.org/10.1016/j.agrformet.2008.05.017).
- [28] A. Maftah et al., "Soiling investigation for PV and CSP system: Experimental and ANN modelling analysis in two sites with different climate," *Int. J. Sustain. Energy*, vol. 41, no. 6, pp. 629–645, Aug. 2021, doi: [10.1080/14786451.2021.1965605](https://doi.org/10.1080/14786451.2021.1965605).
- [29] A. Oufadel et al., "CSP mirror soiling modeling from measured weather factors and forecasting using openweathermap server," in *Proc. 3rd Int. Conf. Electron. Eng. Renewable Energy Syst.*, 2023, vol. 954, pp. 805–813, doi: [10.1007/978-981-19-6223-3_83](https://doi.org/10.1007/978-981-19-6223-3_83).
- [30] H. Zitouni et al., "Experimental investigation and modeling of photovoltaic soiling loss as a function of environmental variables: A case study of semi-arid climate," *Sol. Energy Mater. Sol. Cells*, vol. 221, 2021, Art. no. 110874.
- [31] L. Breiman and J. H. Friedman, "Predicting multivariate responses in multiple linear regression," *J. Roy. Statist. Soc. Ser. B, Statist. Methodol.*, vol. 59, no. 1, pp. 3–54, Jan. 1997, doi: [10.1111/1467-9868.00054](https://doi.org/10.1111/1467-9868.00054).
- [32] P. Kulshreshtha, B. Pugh, and S. Jiddi, "Feature refinement to improve high resolution image inpainting," [cs.CV], 2022, [arXiv:2206.13644](https://arxiv.org/abs/2206.13644).
- [33] Z. Dai, H. Liu, Q. V. Le, and M. Tan, "CoAtNet: Marrying convolution and attention for all data sizes," *Adv. Neural Inf. Process. Syst.*, vol. 34, pp. 3965–3977, 2021.
- [34] G. Dimauro, F. Girardi, M. Gelardi, V. Bevilacqua, and D. Caivano, "Rhino-Cyt: A system for supporting the rhinologist in the analysis of nasal cytology," in *Intelligent Computing Theories and Application*, vol. 10955, D.-S. Huang, K.-H. Jo, and X.-L. Zhang, Eds. Berlin, Germany: Springer-Verlag, 2018, pp. 619–630, doi: [10.1007/978-3-319-95933-7_71](https://doi.org/10.1007/978-3-319-95933-7_71).
- [35] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding data augmentation for classification: When to warp?," in *Proc. Int. Conf. Digit. Image Comput.: Techn. Appl.*, 2016, pp. 1–6, doi: [10.1109/DICTA.2016.7797091](https://doi.org/10.1109/DICTA.2016.7797091).
- [36] G. S. Nijaguna et al., "Quantum fruit fly algorithm and ResNet50-VGG16 for medical diagnosis," *Appl. Soft Comput.*, vol. 136, Mar. 2023, Art. no. 110055, doi: [10.1016/j.asoc.2023.110055](https://doi.org/10.1016/j.asoc.2023.110055).

- [37] G. Singh and K. K. Yogi, "Comparison of RSNET model with existing models for potato leaf disease detection," *Biocatalysis Agricultural Biotechnol.*, vol. 50, Jul. 2023, Art. no. 102726, doi: [10.1016/j.bcab.2023.102726](https://doi.org/10.1016/j.bcab.2023.102726).
- [38] T. Shanthi and R. S. Sabeenian, "Modified alexnet architecture for classification of diabetic retinopathy images," *Comput. Elect. Eng.*, vol. 76, pp. 56–64, Jun. 2019, doi: [10.1016/j.compeleceng.2019.03.004](https://doi.org/10.1016/j.compeleceng.2019.03.004).
- [39] A. Priyam, R. Gupta, A. Rathee, and S. Srivastava, "Comparative analysis of decision tree classification algorithms," *Int. J. Curr. Eng. Technol.*, vol. 3, no. 2, pp. 334–337, 2013.
- [40] A. Rojarath and W. Songpan, "Cost-sensitive probability for weighted voting in an ensemble model for multi-class classification problems," *Appl. Intell.*, vol. 51, no. 7, pp. 4908–4932, Jul. 2021, doi: [10.1007/s10489-020-02106-3](https://doi.org/10.1007/s10489-020-02106-3).
- [41] I. Segovia Ramírez, F. P. García Márquez, and J. Parra Chaparro, "Convolutional neural networks and Internet of Things for fault detection by aerial monitoring of photovoltaic solar plants," *Measurement*, vol. 234, Jul. 2024, Art. no. 114861, doi: [10.1016/j.measurement.2024.114861](https://doi.org/10.1016/j.measurement.2024.114861).
- [42] S. P. Pathak, S. Patil, and S. Patel, "Solar panel hotspot localization and fault classification using deep learning approach," *Procedia Comput. Sci.*, vol. 204, pp. 698–705, 2022, doi: [10.1016/j.procs.2022.08.084](https://doi.org/10.1016/j.procs.2022.08.084).
- [43] N. Kellil, A. Aissat, and A. Mellit, "Fault diagnosis of photovoltaic modules using deep neural networks and infrared images under algerian climatic conditions," *Energy*, vol. 263, Jan. 2023, Art. no. 125902, doi: [10.1016/j.energy.2022.125902](https://doi.org/10.1016/j.energy.2022.125902).
- [44] N. V. Sridharan, S. Vaithyanathan, and M. Aghaei, "Voting based ensemble for detecting visual faults in photovoltaic modules using AlexNet features," *Energy Rep.*, vol. 11, pp. 3889–3901, Jun. 2024, doi: [10.1016/j.egy.2024.03.044](https://doi.org/10.1016/j.egy.2024.03.044).
- [45] A. Oufadel et al., "In-situ heat losses measurements of parabolic trough receiver tubes based on infrared camera and artificial intelligence," *Environ. Challenges*, vol. 10, Jan. 2023, Art. no. 100679.
- [46] A. A. Abed and S. A. Rahman, "Python-based raspberry Pi for hand gesture recognition," *Int. J. Comput. Appl.*, vol. 173, no. 4, pp. 18–24, Sep. 2017, doi: [10.5120/ijca2017915285](https://doi.org/10.5120/ijca2017915285).
- [47] A. Oufadel et al., "Optimization of the energy lost in the current limitation mode of the In grid-connected photovoltaic systems: A case study of the PV installation of the Mohammed 6 polytechnic campus of Rabat," in *Proc. IEEE 9th Int. Conf. Wireless Netw. Mobile Commun.*, 2022, pp. 1–5.
- [48] A. A. Mubarak, H. Cao, and I. M. Hezam, "Deep analytic model for student dropout prediction in massive open online courses," *Comput. Elect. Eng.*, vol. 93, Jul. 2021, Art. no. 107271, doi: [10.1016/j.compeleceng.2021.107271](https://doi.org/10.1016/j.compeleceng.2021.107271).
- [49] N. Rahimi, F. Eassa, and L. Elrefaei, "One- and two-phase software requirement classification using ensemble deep learning," *Entropy*, vol. 23, no. 10, Sep. 2021, Art. no. 1264, doi: [10.3390/e23101264](https://doi.org/10.3390/e23101264).