E3S Web of Conferences **391**, 01069 (2023) https://doi.org/10.1051/e3sconf/202339101069
ICMED-ICMPC 2023

Anomaly Detection in Solar Modules with Infrared Imagery

Ganapathi Raju N. V[1*], *Sai Narayana* G[2], *Raja Sai* A[3], *Vishnu Vardhan Rao* G[4], and *Yashwanth Reddy* Ch[5]

[1]Professor, Department of Information Technology, GRIET, India
[2,3,4,5]Student, Department of Information Technology, GRIET, India

Abstract. Image classification is a machine learning task that involves assigning a label or class to an input image. In the context of the Infrared Solar Modules dataset, image classification can be used to identify anomalies in solar panel imagery. To achieve this goal, A convolutional neural network (CNN) model trained from scratch and fine-tuned on the Infrared Solar Modules dataset from ai4earthscience. Model includes techniques such as dropout and image data generation to enhance its accuracy on this specific dataset. With these methods, Model can achieve high accuracy in identifying solar panel anomalies even with low-size images.

1 Introduction

Solar energy has become popular due to its cost-effectiveness, environmental benefits, and renewable nature. However, solar modules can experience efficiency decline due to factors such as environmental stress, material degradation, and manufacturing defects. To detect anomalies in solar modules, various techniques have been developed, including visual inspection, electrical measurements, and infrared imaging. Infrared imaging is a non-destructive testing method that is widely used to detect anomalies in solar modules. It enables the visualization of temperature variations within the solar module, helping to identify anomalies such as hotspots, shunt resistance, and cell mismatch. An algorithm can be developed to process infrared images much faster and more accurately than manual inspection. The development of an algorithm involves collecting infrared images of solar modules with known anomalies, selecting appropriate features, and evaluating the algorithm's performance. The selection of appropriate features is critical in the development of an algorithm, and the evaluation of the algorithm's performance involves measuring its accuracy, precision, and recall. Overall, infrared imaging and algorithms are effective techniques for detecting anomalies in solar modules, which can help identify and rectify issues that impact their performance and reduce their lifespan.

* Corresponding author: nvgraju@griet.ac.in

2 Literature Survey

The efficient detection and classification of operation anomalies in large-scale solar farms are crucial for enhancing their reliability and electricity generation. However, this task is challenging due to the high complexity and wide variety of anomalies, and the limited information provided by existing SCADA systems. Y. Zhao et.al., work on the paper [1] which proposes a data-driven solution for detecting and classifying photovoltaic system anomalies that can accurately identify different anomalies without requiring additional equipment or non-SCADA data collection. The proposed approach consists of two methods: a hierarchical context-aware anomaly detection method using unsupervised learning, and a multimodal anomaly classification method. The solution was deployed in two large-scale solar farms, and the results showed its effectiveness, robustness, and cost and computation efficiency. The proposed approach can provide better visibility into the health status of solar farms and help in making data-driven decisions for enhancing their operation reliability and electricity generation.

Gao, X et.al., [2] presents a solar panel defect detection system that automates the inspection process for large solar farms. The proposed system uses thermal infrared imaging to detect anomalies in solar panels without requiring expensive electrical detection circuitry. The system involves collecting infrared video sequences of each array of solar panels using an infrared camera mounted to a moving cart, which is driven from array to array in a solar farm. The image processing algorithm then segments the solar panels from the background in real time using only the height of the array as prior information to aid in the segmentation process. Frame-to-frame panel association is established using optical flow to "count" the number of panels within any given array. Local anomalies in a single panel, such as hotspots and cracks, are immediately detected and labeled as soon as the panel is recognized in the field of view. The proposed system also uses DBSCAN clustering to detect hot panels after the data from an entire array is collected. The results of the real-world test data containing over 12,000 solar panels show that the proposed system can recognize and count over 98% of all panels accurately, with 92% of all types of defects being identified by the system. This automated solar panel defect detection system could be a simple and reliable solution to achieving higher power generation efficiency and longer panel life.

Ye Zhao et.al., [3] proposes a graph-based semi-supervised learning model for fault detection in solar photovoltaic (PV) arrays. Fault detection is crucial for increasing reliability and safety in PV systems. However, due to PV's nonlinear characteristics, conventional protection devices may not detect certain faults, which can lead to safety issues and fire hazards in PV fields. Although machine learning techniques have been proposed for fault detection based on measurements, such as PV array voltage, current, irradiance, and temperature, existing solutions typically use supervised learning models that require numerous labeled data for training. This paper proposes a graph-based semi-supervised learning model that only uses a few labeled training data, which are normalized for better visualization. The proposed model can detect faults and identify possible fault types, thus expediting system recovery. Moreover, the model can learn PV systems autonomously over time as weather changes. The proposed method is effective in fault detection and classification, as demonstrated by both simulation and experimental results. The proposed approach offers an alternative to traditional supervised learning methods, which require a large amount of labeled data and can be difficult to update and visualize.

Arena, Eleonora et.al., [4] explores a robust anomaly detection method applied to Enel Green Power's 3SUN solar cell production plant in Italy using a Monte Carlo pre-processing technique. Unlike traditional methods that only detect outliers, this approach includes outlier replacement and preserves temporal locality with respect to the training dataset. The authors then trained an anomaly detection model based on principal component analysis and defined key performance indicators for each sensor in the production line based on model errors. This allows anomalous conditions to be isolated by monitoring the indicators and triggering an alarm when exceeding a reference threshold. The proposed method was tested on standard operating conditions and an anomalous scenario, where it successfully anticipated a fault almost two weeks in advance while also demonstrating robustness to false alarms during normal conditions.

Fermín Mallor et.al., [5] their paper which presents a new method for automatically detecting outliers or faults in the solar energy production of identical sets of photovoltaic (PV) solar panels. The proposed method involves a two-stage unsupervised approach, which includes the creation of "in control" energy production data by removing global and local outliers from the dataset and the construction of control charts using both parametric and non-parametric methods. The control charts can be used to detect outliers or faults in the production data in real-time or at the end of the day. The method was applied to the real energy production data of six sets of identical PV solar panels over a period of three years. The results of the tests indicated that the proposed method successfully detected a reduction in efficiency in one of the solar panel sets by up to 5%. Both the parametric and non-parametric methods showed good performance results in constructing the control charts. The proposed method provides an effective and efficient solution for detecting anomalies or faults in solar energy production, which can help improve the reliability and efficiency of solar power generation.

3 System Architecture

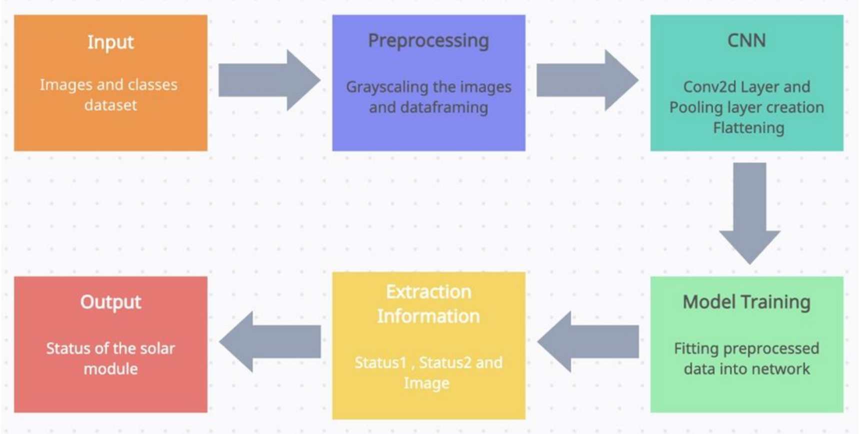

Fig. 1. System Architecture.

4 Methodology

4.1 Dataset

The proposed model is convolutional neural networks (CNNs) for image classification on low-size image datasets and the dataset includes infrared solar modules. Dataset is taken from ai4earthscience.

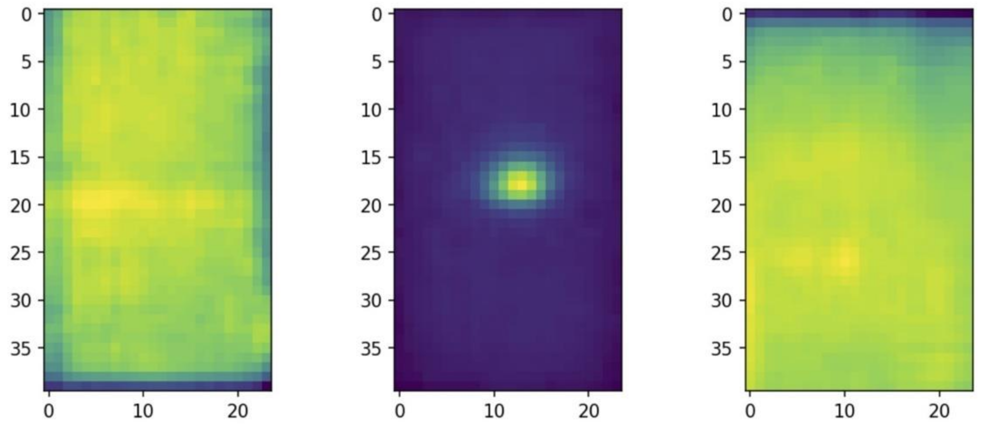

Fig. 2. Input Images.

Infrared Solar Modules is a machine-learning dataset that contains real-world imagery of different anomalies found in solar farms. This dataset can be used for machine learning research to gain efficiencies in the solar industry.

	image_filepath	anomaly_class	image_name
13357	images/13357.jpg	No-Anomaly	13357.jpg
5988	images/5988.jpg	Cell	5988.jpg
6796	images/6796.jpg	Hot-Spot	6796.jpg
270	images/270.jpg	Offline-Module	270.jpg
9528	images/9528.jpg	Vegetation	9528.jpg
2143	images/2143.jpg	Diode	2143.jpg
3519	images/3519.jpg	Shadowing	3519.jpg
7188	images/7188.jpg	Cracking	7188.jpg
876	images/876.jpg	Diode-Multi	876.jpg
8019	images/8019.jpg	Hot-Spot-Multi	8019.jpg
4592	images/4592.jpg	Cell-Multi	4592.jpg
8291	images/8291.jpg	Soiling	8291.jpg

Fig. 3. Dataframe of provided dataset json file.

The dataset consists of 20,000 infrared images that are 24 by 40 pixels each. There are 12 defined classes of solar modules presented with 11 classes of different anomalies and the remaining class being No-Anomaly (i.e., the null case).

4.2 Data loading and Preprocessing:

4.2.1 Function: split_data ():

Create training, validation, and test sets for a Convolutional Neural Network (CNN) model built from scratch to classify infrared images of solar modules as normal or anomalous. The initial step is to shuffle the data using a random seed value and split it into a training set, a validation set, and a test set, with the training set containing 80% of the data.

4.2.2 Function: generate_image_data ():

ImageDataGenerator objects are created to perform data augmentation by rescaling the pixel values of the images in the training and validation sets. The generator objects are then used to create three separate generator instances - one for each set - that allow for efficient loading of batches of images during training and testing.

4.2.3 Function: flow_from_dataframe ():

The generators are created using the flow_from_dataframe method, which takes the dataframes containing the image names and their corresponding labels as inputs. The method also specifies the image directory, image size, batch size, and color mode for the images. The color mode is set to grayscale since the infrared images are single-channel grayscale images. Finally, the flow_from_dataframe method is called separately for the train, validation, and test sets to create the corresponding generators.

4.3 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are deep neural networks commonly used for image classification and other computer vision tasks. They use convolutional layers to learn local patterns and features in the input image, which are combined by subsequent layers to form higher-level representations. This hierarchical feature extraction is followed by pooling layers and fully connected layers for classification or regression tasks. During training, the weights of the filters and fully connected layers are optimized using an optimization algorithm such as stochastic gradient descent (SGD). CNNs have achieved state-of-the-art performance in many computer vision tasks and are used in applications such as autonomous driving and medical image analysis. Their ability to learn local patterns and hierarchical representations makes them well-suited to tasks where the input is a grid-like structure, such as images. [6]

4.4 Dropout Technique

Dropout is a regularization technique used in convolutional neural networks (CNNs) to prevent overfitting. It works by randomly dropping out (i.e., setting to zero) a proportion of the neurons in a layer during each training iteration. The neurons that are dropped out are randomly selected, and their connections to the next layer are temporarily removed. This forces the remaining neurons to learn more robust and generalized representations of the input data. Dropout has been shown to be an effective method for improving the generalization performance of CNNs and reducing overfitting, particularly when applied to large datasets.

https://doi.org/10.1051/e3sconf/202339101069

4.5 Workflow and Algorithm

1. First, the necessary libraries are imported, including TensorFlow, NumPy, Pandas, and Matplotlib.
2. The dataset is loaded from a zip file and a Pandas dataframe is created from the metadata JSON file.
3. The images are displayed to get a sense of the dataset.
4. The dataset is split into training, validation, and test sets.
5. ImageDataGenerator is used to create generators for the training and validation sets. The images are resized to 40x24, normalized, and converted to grayscale.
6. The CNN is created using the Sequential model from Keras. The model contains two Conv2D layers with 32 and 64 filters, respectively, each followed by a MaxPooling2D layer. The output from these layers is flattened and passed through two Dense layers with 128 and 64 neurons, respectively, before being passed through a final output layer with 12 neurons (one for each class).
7. The model is trained using the fit() function and the training and validation generators. The Adam optimizer is used with a learning rate of 0.001, and the categorical cross-entropy loss function is used.
8. Finally, the model is evaluated on the test set using the evaluate() function.

5 Results Analysis

5.1 Input:

The input for the model is a grayscale image with a height of 40 pixels and a width of 24 pixels, represented as a 3D tensor of shape (image_height, image_width). In this case, the color_mode parameter is set to "grayscale", which means that the input images will be converted to grayscale before being fed into the model.

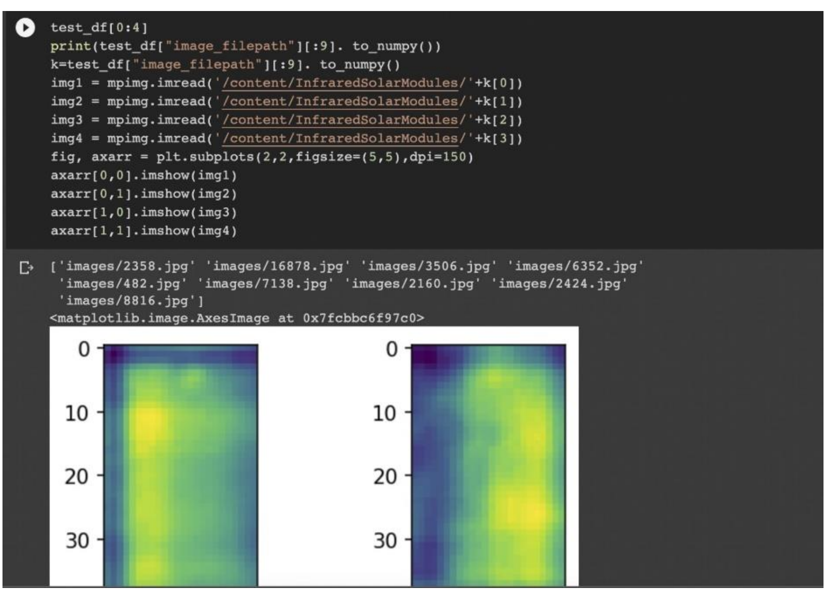

Fig. 4. Input Image

5.2 Output:

The Output for the model is the prediction the class input may belong to in class numbers.

```
[107] predicted_classes.shape
      predicted_classes[:4]

      array([0, 0, 1, 1])
```

Fig. 5. Model Output

6 Conclusion

The detection of anomalies in solar modules is crucial for maintaining their performance and ensuring their longevity. The proposed project aims to develop an algorithm that can detect anomalies in solar modules using IR imagery. The algorithm is expected to be more efficient and reliable than the current manual inspection method. The project is expected to deliver an algorithm that can aid in the identification and rectification of anomalies in solar modules, ultimately contributing to the wider adoption of solar energy.

7 Future Enhancements

There are several potential areas for future expansion and improvement for the project of finding anomalies in solar modules using infrared imagery. Increase the size and scope of the dataset: Currently, the model is trained on a relatively small dataset of infrared images of solar modules. Expanding the dataset could improve the accuracy of the model and enable it to detect a wider range of anomalies. Develop a real-time monitoring system: The current model is designed to analyse static images of solar modules. However, a real-time monitoring system that can continuously monitor solar modules could provide more valuable information about the performance of the modules and help detect anomalies in real-time. Incorporate other types of data: In addition to infrared imagery, there are other types of data that could be used to improve anomaly detection in solar modules. For example, data on temperature, weather conditions, and energy production could be incorporated into the model to provide more context and enable more accurate predictions.

Expand the types of anomalies detected: The current model is designed to detect anomalies related to hotspots and cracks in solar modules. However, there are other types of anomalies that can occur in solar modules, such as shading, soiling, and delamination. Expanding the types of anomalies detected could make the model more versatile and useful in a wider range of applications. Develop a more user-friendly interface: While the current model is effective at detecting anomalies in solar modules, it requires some technical expertise to operate. Developing a more user-friendly interface that makes it easier to input data, visualize results, and interpret the findings could make the model more accessible to a wider range of users.

Overall, the potential for expansion and improvement for the project of finding anomalies in solar modules using infrared imagery is significant. With continued development and refinement, this technology could become an increasingly valuable tool for improving the efficiency and performance of solar energy systems.

References

1. Y. Zhao, Q. Liu, D. Li, D. Kang, Q. Lv and L. Shang, "Hierarchical Anomaly Detection and Multimodal Classification in Large-Scale Photovoltaic Systems," in IEEE Transactions on Sustainable Energy, vol. 10, no. 3, pp. 1351-1361, (July 2019)
2. Xiang Gao, Eric Munson, Glen P. Abousleman, and Jennie Si "Automatic solar panel recognition and defect detection using infrared imaging", Proc. SPIE 9476, Automatic Target Recognition XXV, 94760O (22 May 2015)
3. Y. Zhao, R. Ball, J. Mosesian, J. -F. de Palma and B. Lehman, "Graph-Based Semi-supervised Learning for Fault Detection and Classification in Solar Photovoltaic Arrays," in IEEE Transactions on Power Electronics, vol. 30, no. 5, pp. 2848-2858, (May 2015)
4. Arena, E., Corsini, A., Ferulano, R., Iuvara, D. A., Miele, E. S., Ricciardi Celsi, L., Sulieman, N. A., & Villari, M. Anomaly Detection in Photovoltaic Production Factories via Monte Carlo Pre-Processed Principal Component Analysis. Energies, 14(13), 3951.(2021)
5. Mallor, Fermín, et al. "A method for detecting malfunctions in PV solar panels based on electricity production monitoring." Solar Energy 153: 51-63 (2017).
6. L. O. Chua, "CNN: A Vision of Complexity," International Journal of Bifurcation and Chaos, pp. 2219-2425, (1997).
7. A. F. Agarap, "Deep learning using rectified linear units (relu).," arXiv preprint, (2018).