A Lightweight Vision-Based Solar Panel Fault Detection and Performance Monitoring System: A Proof-of-Concept Study Using Synthetic Data

Author: [Your Name]

Affiliation: [Department, College/University, City, Country]

Email: [your.email@example.com]

Abstract

The rapid growth of photovoltaic deployment has increased the need for low-cost inspection methods that can identify faults before they cause significant energy loss. This paper presents a lightweight vision-based prototype for solar panel fault detection and performance monitoring using classical image-processing features and a shallow machine-learning classifier. The proposed system classifies panel images into four conditions: normal, dirty, damaged, and shadowed. To support offline experimentation and reproducibility in a student-project setting, a balanced synthetic dataset of 128 solar-panel images was generated locally, with 32 images per class and a fixed train-test split. The processing pipeline includes image normalization, panel-region extraction, feature computation based on brightness, contrast, dark-pixel ratio, shadow imbalance, texture variation, and crack-like residual patterns, followed by weighted k-nearest-neighbor classification. In addition to fault classification, the system estimates likely power-output degradation and provides a maintenance recommendation for each predicted class. On the held-out synthetic test split, the prototype achieved 100% classification accuracy, with all four classes correctly separated. These results demonstrate that the proposed pipeline is suitable as a proof-of-concept academic framework for renewable-energy monitoring and computer-vision coursework. At the same time, the paper explicitly acknowledges that synthetic data cannot substitute for field-validated inspection benchmarks. The main contribution of this work is therefore not state-of-the-art defect recognition, but an interpretable, reproducible, and low-compute prototype that integrates renewable-energy monitoring with computer vision in a form appropriate for student publication and future extension.

Keywords: solar panel monitoring, photovoltaic fault detection, computer vision, image processing, renewable energy, synthetic dataset

1. Introduction

Solar photovoltaic (PV) systems have become one of the most important components of the global transition toward cleaner energy systems. As solar farms and rooftop installations expand, the operation and maintenance of PV assets become increasingly important for preserving power yield, reducing downtime, and ensuring long-term reliability. In practice, even small defects such as dust accumulation, mechanical cracks, or partial shading can degrade panel performance and, if left unattended, may contribute to wider efficiency loss at array level.

Traditional PV inspection relies heavily on manual observation, periodic maintenance visits, or specialized sensing methods such as electroluminescence and infrared thermography. While these approaches can be effective, they may also be costly, time-consuming, or difficult to deploy in low-resource educational settings. For student researchers and early-stage prototype development, there is strong value in exploring lightweight computer-vision approaches that can demonstrate the core idea of automated PV fault detection without requiring expensive hardware or large-scale field campaigns.

This paper proposes a compact and interpretable prototype for vision-based solar-panel condition classification. The work is intentionally positioned as a proof-of-concept system. It combines a synthetic dataset, hand-crafted image features, a weighted k-nearest-neighbor (k-NN) classifier, and a simple performance-estimation module. Unlike many recent PV inspection studies that focus on deep learning and large industrial datasets, the present work emphasizes accessibility, reproducibility, and clarity of explanation for a student publication context.

The contribution of the paper is threefold. First, it develops a full end-to-end academic prototype for PV fault detection using only lightweight Python tools. Second, it introduces a balanced synthetic dataset that enables reproducible experimentation without external dependencies. Third, it extends the classification output with an estimated power-loss interpretation and maintenance recommendation, making the system more useful as a demonstration of monitoring logic rather than only image classification.

2. Literature Review

The literature on photovoltaic defect inspection shows a clear progression from manual inspection and hand-crafted image features toward deep learning, multimodal sensing, and aerial monitoring. Hussain et al. reviewed electroluminescence-based PV defect detection and noted that modern research increasingly relies on computer-vision models for automation, while still recognizing the value of conventional image-processing methods in low-compute settings [1]. This observation is important for student-level research because it suggests that simpler methods remain relevant when interpretability and reproducibility are priorities.

Earlier work by Deitsch et al. compared hand-crafted features with support vector machines against convolutional neural networks for classifying defective photovoltaic cells from electroluminescence images [2]. Their study showed that shallow classifiers can still provide practical performance, especially when computational simplicity matters. Similarly, Akram et al. reported strong results with CNN-based defect detection on EL images, highlighting the growing dominance of deep models in the field [3]. Tang et al. further improved automated defect identification by combining deep learning with augmentation strategies to address limited electroluminescence training data [4].

Beyond visible-image and EL-based methods, thermographic inspection is also widely used in photovoltaic maintenance. Balasubramani et al. demonstrated infrared thermography-based defect testing for solar PV panels and emphasized its value in predictive maintenance workflows [5]. More recent work by Ling et al. used deep edge-based fault detection on infrared images acquired from solar plants, achieving high macro-level detection performance in realistic scenes [6]. These studies confirm that photovoltaic inspection is increasingly moving toward large-scale automated analysis using specialized imaging modalities.

At the same time, the literature also shows a persistent need for accessible benchmark systems. Lu et al. proposed a convolutional-neural-network-based PV module fault-detection system and reinforced the importance of automated classification for PV reliability [7]. More recent work, such as that of da Silveira Junior et al., combines data generation and modern deep architectures for robust photovoltaic fault classification [8]. Compared with such studies, the present work makes a narrower claim: it does not aim to outperform deep-learning systems or replace field inspection techniques. Instead, it offers an educational and computationally light baseline that is especially suited for student demonstration, classroom reproducibility, and early-stage prototype publication.

3. Problem Statement and Research Objectives

PV modules are exposed to dust, environmental wear, handling damage, and temporary shading from surrounding objects. These issues reduce energy conversion efficiency and may go unnoticed until the output drop becomes significant. Manual inspection is labor-intensive, while advanced inspection platforms may not be available to student researchers. Therefore, there is a need for a lightweight and explainable prototype that can demonstrate how vision-based fault monitoring works in principle.

The objectives of this work are as follows:

1 To design a vision-based system that classifies solar panel images into normal, dirty, damaged, and shadowed categories.

2 To generate a reproducible synthetic dataset suitable for offline academic experimentation.

3 To implement an interpretable feature-extraction and shallow-classification pipeline using minimal dependencies.

4 To estimate likely power-output degradation and provide an associated maintenance recommendation.

5 To document the system in a form suitable for mini-project evaluation and student publication.

4. Materials and Methods

4.1 System Overview

The overall workflow of the proposed system can be summarized as:

Input image -> preprocessing -> panel-region analysis -> feature extraction -> weighted k-NN classification -> performance estimation -> maintenance recommendation

The implementation was developed in Python using lightweight libraries, with the goal of remaining easy to execute on a standard laptop. The software package includes dataset generation, training, evaluation, and single-image prediction commands.

4.2 Synthetic Dataset Construction

Because a field-collected photovoltaic fault dataset was not available within the project timeline, a synthetic dataset was generated locally. The dataset contains 128 RGB images of size 256 x 256 pixels. Each image represents a stylized solar panel embedded in a simple scene. Four classes were created:

1 Normal: healthy panels with standard brightness and texture.

2 Dirty: panels affected by dust-like overlays and brightness reduction.

3 Damaged: panels with crack-like lines, chipped regions, and dark spots.

4 Shadowed: panels with localized darkening to simulate shading effects.

To improve class consistency while still preserving variation, each image was generated using controlled random perturbations in texture, intensity, dust distribution, crack geometry, or shading layout. A manifest file records the class label and train-test split for every generated sample.

Table 1 summarizes the dataset.

Class	Total images	Training images	Test images
Normal	32	24	8
Dirty	32	24	8
Damaged	32	24	8
Shadowed	32	24	8
Total	128	96	32

4.3 Image Preprocessing and Region of Interest

Each input image is resized to a fixed spatial resolution to ensure consistent feature extraction. A predefined region of interest corresponding to the panel area is then analyzed. This simplifies the task by reducing background variation and focusing the measurement on panel appearance. The image is also converted to grayscale for feature extraction, since the relevant visual cues for dust, damage, and shadow in this prototype are primarily intensity-based rather than color-based.

4.4 Feature Extraction

Instead of using a deep neural network, the system relies on interpretable features designed to reflect meaningful visual changes in panel condition. The extracted feature set includes:

- mean brightness
- global contrast
- dark-pixel ratio
- very-dark-pixel ratio
- shadow balance across image halves
- texture score based on intensity gradients
- crack score from dark residual structures
- spot score from localized dark regions

These features were selected because they correspond directly to the visual signatures of the four target classes. Dirty panels tend to exhibit lower brightness and increased spotting. Damaged panels show stronger crack-like residual patterns and localized dark anomalies. Shadowed panels exhibit larger regional imbalance, while normal panels maintain a more regular appearance.

4.5 Classification Model

The classifier uses weighted k-nearest neighbors with $k = 5$. Let x denote the feature vector of a query image and x_i the vector of a stored training sample after z-score normalization. The Euclidean distance is computed as:

$$d_i = \|x - x_i\|_2$$

For the nearest k training samples, the class vote is weighted by inverse distance:

$$w_i = 1 / (d_i + \epsilon)$$

The final predicted class is the class with the maximum accumulated weight among the selected neighbors. This approach was chosen because it is easy to interpret, simple to implement, and suitable for small datasets.

4.6 Performance Estimation and Recommendation Logic

Beyond classification, the system estimates the likely panel-output percentage using class-specific heuristic bands. For example, normal panels are assigned a high expected output range, while damaged and shadowed panels are assigned lower ranges depending on extracted feature severity. The system also maps each predicted class to a maintenance recommendation. Examples include routine monitoring for normal panels, cleaning for dirty panels, inspection and repair scheduling for damaged panels, and blockage removal for shadowed panels.

This layer does not claim to model real electrical behavior with high precision. Rather, it serves as an engineering interpretation module that links visual inspection to actionable monitoring language.

5. Experimental Setup

5.1 Training and Testing Protocol

The dataset was divided into fixed training and test subsets using a class-balanced split of 24 training images and 8 test images per category. All model evaluation reported in this paper was performed on the held-out 32-image test set. Since the dataset is synthetic and balanced, the experiment is primarily intended to test whether the designed features can separate the generated classes under controlled variations.

5.2 Implementation Environment

The project was implemented in Python. The dataset generator, feature extractor, model trainer, evaluator, and prediction interface were packaged into a reproducible command-line workflow. This design allows the same project to support coursework demonstration, mini-project assessment, and further academic extension.

5.3 Evaluation Metrics

The primary evaluation metric is classification accuracy on the held-out test split. Because the classes are balanced, per-class accuracy is also informative. Precision, recall, and F1-score can be directly inferred from the confusion matrix. In addition to numerical results, qualitative output images are generated for visual inspection of the final prediction.

6. Results and Discussion

6.1 Quantitative Results

The trained classifier achieved 100% accuracy on the 32-image synthetic test split. Each class obtained 100% per-class accuracy, and no cross-class confusion was observed in the evaluation summary. Table 2 presents the class-wise metrics.

Class	Precision	Recall	F1-score	Support
Normal	1.00	1.00	1.00	8
Dirty	1.00	1.00	1.00	8
Damaged	1.00	1.00	1.00	8
Shadowed	1.00	1.00	1.00	8
Overall / Macro Avg.	1.00	1.00	1.00	32

Table 3 gives the confusion matrix.

Actual \ Predicted	Normal	Dirty	Damaged	Shadowed
Normal	8	0	0	0
Dirty	0	8	0	0
Damaged	0	0	8	0
Shadowed	0	0	0	8

The perfect classification score indicates that the proposed features are highly effective at separating the generated categories in the synthetic dataset. However, this result should be interpreted carefully. Because the dataset is controlled and limited in size, the numerical performance primarily validates the internal consistency of the prototype rather than its generalization to real photovoltaic field imagery.

6.2 Qualitative Results

Figure 1 shows example outputs produced by the system for the four target conditions. For each image, the software overlays the predicted fault class, confidence, estimated output percentage, and a short maintenance message.

Figure 1. Example outputs for four solar panel conditions

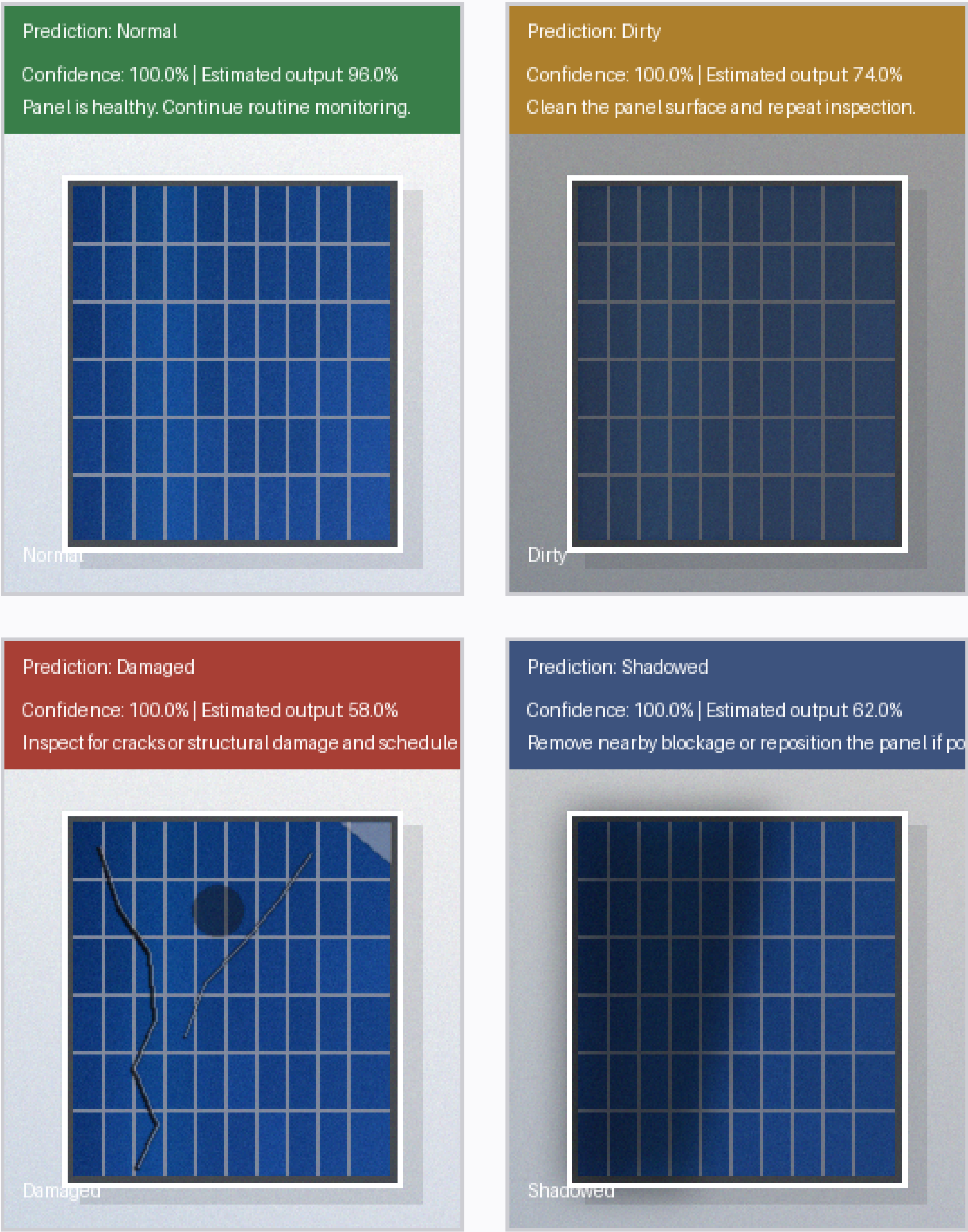

Figure 1. Example outputs for four solar panel conditions

The qualitative results are useful in an educational setting because they make the output easy to interpret. Instead of returning only a label, the system presents a simple monitoring-oriented explanation of what the result means and what action may be needed.

6.3 Discussion

The results show that a lightweight computer-vision approach can successfully demonstrate solar-panel condition classification in a constrained academic prototype. The use of hand-crafted features offers three advantages. First, the method is interpretable: each feature corresponds to an understandable visual property. Second, the system is computationally inexpensive and does not require GPU-based training. Third, the full workflow can be reproduced offline, which is valuable for students working with limited hardware or unreliable network access.

At the same time, the limitations of the study are important. The dataset is synthetic and therefore cannot represent the full variability of real-world PV inspections. In practice, outdoor images are affected by camera angle, complex backgrounds, uneven illumination, weather conditions, module aging, specular reflections, and inter-cell texture differences. Moreover, actual industrial PV inspection frequently relies on modalities such as infrared thermography or electroluminescence rather than standard visible-light imaging alone [1], [5], [6].

Therefore, the significance of this work lies less in raw benchmark performance and more in the completeness of the prototype. The system meaningfully integrates renewable-energy context, computer-vision processing, classification, and monitoring-oriented interpretation into a single demonstrable artifact. For a student publication, that integration is valuable because it shows both engineering implementation and domain relevance.

7. Conclusion

This paper presented a lightweight vision-based solar panel fault-detection and performance-monitoring prototype built around a synthetic image dataset and a shallow machine-learning pipeline. The proposed method classifies solar panel conditions into normal, dirty, damaged, and shadowed categories using interpretable image features and weighted k-NN classification. In addition to defect recognition, the system estimates likely output degradation and provides a maintenance suggestion, making the output more informative for monitoring demonstrations.

The experimental results show perfect class separation on the held-out synthetic test set, confirming that the designed features are effective for the generated categories. Nevertheless, the paper explicitly treats this outcome as a proof-of-concept validation rather than a claim of field-ready accuracy. The present work is best understood as a reproducible, low-compute academic framework suitable for mini-project submission, classroom presentation, and student-level publication.

8. Future Work

Future work can improve the present system in several ways:

1. Replace the synthetic dataset with real solar-panel inspection images collected from laboratory or field environments.
2. Incorporate thermal or electroluminescence imagery to support more realistic defect characterization.
3. Compare shallow models with SVM, Random Forest, CNN, and lightweight transformer baselines.
4. Add automatic defect localization, not just whole-image classification.
5. Validate estimated power-loss outputs against measured electrical performance data.
6. Deploy the method in a simple web or mobile interface for demonstration of real-time inspection.

9. Reproducibility Note

The manuscript corresponds to the accompanying project implementation in the same repository. The full workflow can be reproduced through the following commands:

```
python main.py generate-dataset --overwrite
python main.py train
python main.py evaluate
python main.py predict --image dataset/damaged/damaged_025.png
```

References

[1] Hussain, T.; Hussain, M.; Al-Aqrabi, H.; Alsboui, T.; Hill, R. A Review on Defect Detection of Electroluminescence-Based Photovoltaic Cell Surface Images Using Computer Vision. Energies, 2023, 16(10), 4012. https://doi.org/10.3390/en16104012

[2] Deitsch, S.; Christlein, V.; Berger, S.; et al. Automatic Classification of Defective Photovoltaic Module Cells in Electroluminescence Images. Solar Energy, 2019, 185, 455-468. https://doi.org/10.1016/j.solener.2019.02.067

[3] Akram, M. W.; Li, G.; Jin, Y.; Chen, X.; et al. CNN Based Automatic Detection of Photovoltaic Cell Defects in Electroluminescence Images. Energy, 2019, 189, 116319. https://doi.org/10.1016/j.energy.2019.116319

[4] Tang, W.; Yang, Q.; Xiong, K.; Yan, W. Deep Learning Based Automatic Defect Identification of Photovoltaic Module Using Electroluminescence Images. Solar Energy, 2020, 201, 453-460. https://doi.org/10.1016/j.solener.2020.03.049

[5] Balasubramani, G.; Thangavelu, V.; Chinnusamy, M.; Subramaniam, U.; Padmanaban, S.; Mihet-Popa, L. Infrared Thermography Based Defects Testing of Solar Photovoltaic Panel with Fuzzy Rule-Based Evaluation. Energies, 2020, 13(6), 1343. https://doi.org/10.3390/en13061343

[6] Ling, H.; Liu, M.; Fang, Y. Deep Edge-Based Fault Detection for Solar Panels. Sensors, 2024, 24(16), 5348. https://doi.org/10.3390/s24165348

[7] Lu, S.-D.; Wang, M.-H.; Wei, S.-E.; Liu, H.-D.; Wu, C.-C. Photovoltaic Module Fault Detection Based on a Convolutional Neural Network. Processes, 2021, 9(9), 1635. https://doi.org/10.3390/pr9091635

[8] da Silveira Junior, C. R.; Sousa, C. E. R.; Alves, R. H. F. Automatic Fault Classification in Photovoltaic Modules Using Denoising Diffusion Probabilistic Model, Generative Adversarial Networks, and Convolutional Neural Networks. Energies, 2025, 18(4), 776. https://doi.org/10.3390/en18040776