Title

Exploring Capability of Multimodal Foundation Model for Image-based Fault Detection of Photovoltaic Modules

Journal

2025 IEEE 53RD PHOTOVOLTAIC SPECIALISTS CONFERENCE, PVSC, 00

ISSN

0160-8371

ISBN

979-8-3315-3445-5

Authors

Li, Baojie

Chen, Xin

Jain, Anubhav

Publication Date

2025-06-13

DOI

10.1109/pvsc59419.2025.11133222

Exploring Capability of Multimodal Foundation Model for Image-based Fault Detection of Photovoltaic Modules

Baojie Li, Xin Chen, Anubhav Jain

Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

***Abstract* — Multimodal Foundation Model (MFM), like ChatGPT and Gemini, have emerged as powerful tools for their exceptional natural language processing capabilities and their emerging potential in image analysis. This paper investigates the application of MFMs for photovoltaic (PV) fault detection through image analysis, focusing on ChatGPT 4.0 and Gemini 1.5 Pro. Three types of PV images and the corresponding common PV faults are detected: bird droppings using visible images, cell cracks via electroluminescence (EL) images, and hotspots using infrared (IR) images. Among the two models, Gemini 1.5 Pro demonstrated superior performance, achieving near-perfect results with an average F1 score of 0.97, consistently outperforming ChatGPT 4.0 in accuracy and reliability. Unlike traditional machine learning (ML) models, MFMs can operate in a zero shot manner that does not require additional training by the user, and the input images are not limited by size, angle, scope, or PV technology. The strong adaptability and user-friendliness make MFM a promising tool for analyzing PV images and advancing health monitoring for PV modules.**

I. Introduction

Multimodal Foundation Models (MFM) [1], like ChatGPT [2] and Gemini [3], extended from Large Language Models (LLMs), have gained immense popularity in recent years for their outstanding capabilities not only in natural language processing tasks, but also in remarkable potential for image analysis [4], offering innovative solutions for tasks like classification, fault detection, and data interpretation [5].

With the wide deployment of photovoltaic (PV) imaging devices, such as drones and advanced cameras, an increasing number of images of PV modules have been captured, providing valuable insights into their health status [6]. However, analyzing these images remains a significant challenge due to variations in image quality, diverse module types, changing environmental conditions, and the complexity of developing models (*e.g.*, machine learning (ML) models), to process and interpret them effectively.

Some researchers have explored MFMs in the PV field. Sarahana et al. used ChatGPT to extract the solar policy [7]. Kalle et al. studied the public acceptance of solar energy via ChatGPT from media content [8]. Subir et al. reviewed the possible application of LLMs in the electric energy sector [9]. However, most of the existing work focuses on text-mining applications using MFMs. For the PV image analysis, most research focuses on custom-trained models [10], [11] rather than foundation models.

Seen in this light, this paper explores the capabilities of popular MFMs for analyzing the major types of PV image data, including visible, electroluminescence (EL), and infrared (IR) images. Their ability to detect corresponding common PV faults (bird droppings, cell cracks, and hot spots) will be evaluated, with an overview of the work depicted in Fig. 1. The major advantage of this framework is its ease of use and applicability to diverse image types and formats without user training.

The remainder of this paper is organized as follows: Section II presents the results of image-based fault detection performance using MFMs for visible, EL, and IR image data. Section III discusses the capabilities and limitations of MFMs, and Section IV concludes the paper.

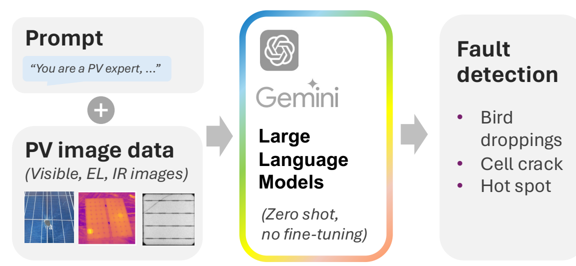

Fig. 1 Overview of methodology. Using PV image data and prompts, without fine-tuning, the LLMs are leveraged to perform the detection of various faults.

II. Fault Detection Using Image Data

Fault detection is a basic step of health monitoring of PV modules. Here, we study three types of PV images, visible, El, and IR images, and the corresponding fault detection performance.

A. Visible images

Visible PV images can reveal many faults, such as soiling, bird droppings, cracks, delamination, and discoloration. In this study, we focus on the bird-dropping "fault". Bird droppings on PV modules can cause localized shading, hotspot formation, and activation of bypass diodes, which not only leads to reduced energy output but also damages cells and encapsulation.

For the test data, 30 visible images of PV modules are collected from DuraMat Datahub [12] and open-source PV image datasets on Kaggle [13]. These images manually labeled as clean (15 images) or with bird dropping (15 images), with example images shown in Fig. 2. We input these images and a prompt to ChatGPT 4o [2] and Gemini 1.5 pro [3] (version of Jan 10, 2025) to detect the occurrence of dust. Note that the LLMs are without fine-tuning on PV images for this task; the models are used as provided. The prompt is "*This is a PV module image. Tell me whether it has bird dropping, yes or no, no verbose.*"

With the returned results, the confusion matrix for the detection performance using the two MFMs is calculated and shown in Fig. 2 and Fig. 3. We use three metrics to quantify the classification performance: precision, recall, and the F1 score, as defined in (1-3). They are calculated based on the following four key values: True Positives (TP), True Negatives (TN), False Positives (FP), False Negatives (FN), which corresponds to the confusion matrix shown in Fig. 2. The calculated values of these metrics for bird-dropping detection results are summarized in TABLE 1.

$$Precision = TP/(TP + FP) \quad (1)$$

$$Recall = TP / (TP + FN) \quad (2)$$

$$F1 = 2\,(Precision * Recall) / (Precision + Recall) \quad (3)$$

From the confusion matrix in Fig. 2 and the corresponding metrics in TABLE 1, Gemini achieves 100% accuracy in the classification of clean and dirty PV modules with the F1 score = 1.0. In comparison, the GPT model's performance is slightly lower, misclassifying 3 images as bird-dropping (as shown in Fig. 3), resulting in an F1 score of 0.9.

Notably, the PV visible images used for testing are diverse, with no restrictions on shooting angle, surroundings, or PV module type, as illustrated in Fig. 2. More interestingly, certain images even show a reflection of the environment (*e.g.*, Image 1 in Fig. 2), where the surrounding trees are reflected on the module surface. Despite these challenging conditions and the absence of fine-tuning, Gemini achieves 100% classification accuracy, highlighting its excellent generalization capability.

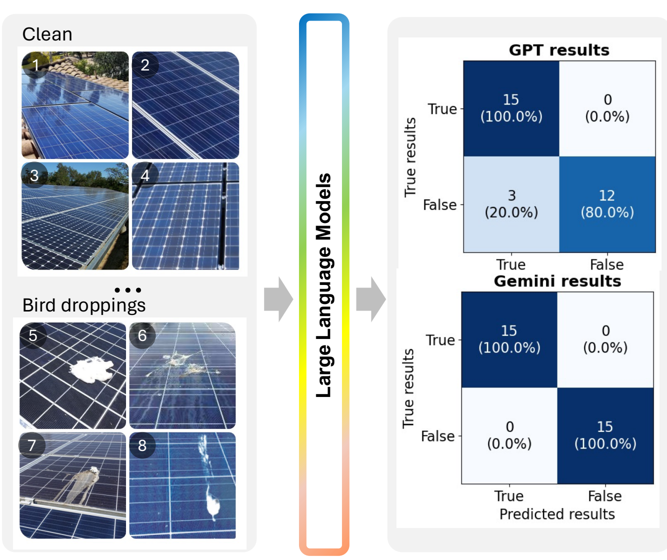

Fig. 2 Example of PV visible images (clean and with bird-dropping) and detection results (confusion matrix) using GPT 4o and Gemini 1.5 pro. Gemini 1.5 pro achieves 100% classification accuracy for all clean and dirty PV visible images.

Fig. 3 Example of misclassified PV visible images using GPT 4o

TABLE 1
SUMMARY OF METRICS FOR FAULT DETECTION USING TWO LLMS

Case	LLM	F1 score	Precision	Recall
Bird dropping	GPT	0.90	0.92	0.90
	Gemini	1.00	1.00	1.00
Cell crack	GPT	0.77	0.77	0.77
	Gemini	0.97	0.97	0.97
Hot spot	GPT	0.63	0.80	0.67
	Gemini	0.93	0.94	0.93
Average	**GPT**	**0.77**	**0.83**	**0.78**
	Gemini	**0.97**	**0.97**	**0.97**

B. EL images

Electroluminescence (EL) imaging of a PV module visualizes variations in electrical activity and connections across the module. EL images are effective in diagnosing microcracks in PV modules, which are often undetectable through visible inspection. Using these images, we tasked the LLMs with detecting cell cracks. A dataset of 30 PV EL cell images was collected from open-source datasets on DuraMat Datahub [12] and Kaggle [14] and manually labeled as healthy (15 images) or containing cell cracks (15 images), with examples shown in Fig. 4. Note that, the test images include PV cells with 3 busbars (*e.g.*, Image 1 in Fig. 4) or 4 busbars (e.g., Image 3 in Fig. 4), featuring varying crack patterns and illumination intensities, as depicted in in Fig. 4.

The prompt is defined as "*This a PV cell EL image, tell me whether it presents cell cracks, yes or no, no verbose*". We input these images and the prompt to GPT 4o and Gemini 1.5 pro. The output results (confusion matrix) are shown in Fig. 4 and Fig. 5 with the metrics summarized in TABLE 1.

Without fine-tuning, GPT misclassified 7 images, including 3 false negatives (FN) and 4 false positives (FP), resulting in an F1 score of 0.77. In comparison, Gemini misclassified only a single image, likely due to its darkness (as shown in Fig. 5), achieving a significantly higher F1 score of 0.97. Despite variations in busbar configuration and illumination intensity, Gemini demonstrated near-perfect detection accuracy, which shows its exceptional capability to identify cell cracks under diverse conditions.

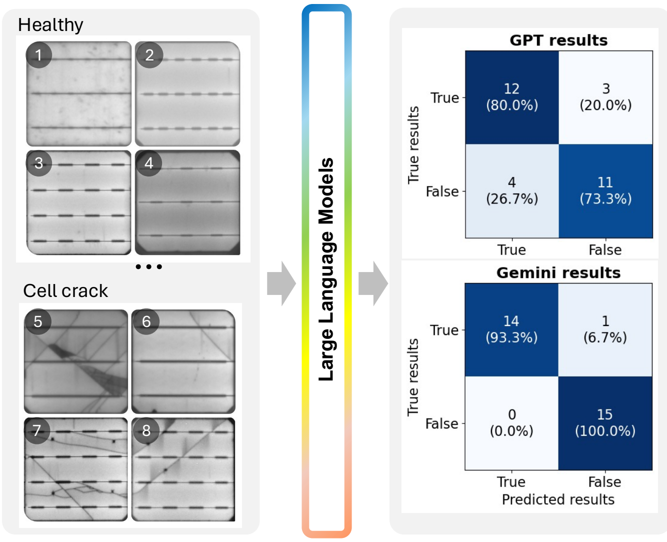

Fig. 4 Example of PV EL images (healthy and cracked) and detection results (confusion matrix) using GPT 4o and Gemini 1.5 pro. Gemini greatly outperforms GPT with only 1out of 30 images wrongly classified.

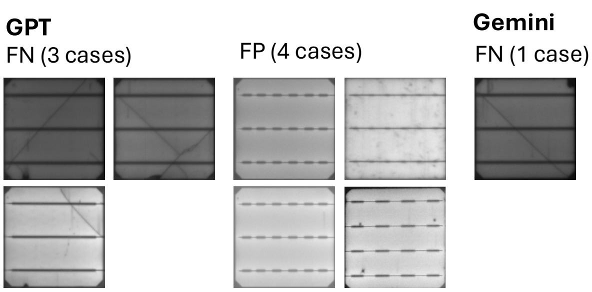

Fig. 5 Example of misclassified PV EL images using GPT 4o and Gemini 1.5 pro.

C. IR images

Infrared (IR) images of PV modules capture the thermal distribution across the module surface. The common fault IR images can visualize is the hot spot, which refers to a localized region with elevated temperature compared to the rest of the module. It typically occurs due to shading, cell damage, poor electrical connections, or bypass diode failure. Hotspots can reduce efficiency and potentially damage the module over time if not addressed.

We collected 30 PV IR images from open-source datasets [12], [15], [16] to evaluate the ability of MFMs to detect hotspots, with 15 images representing healthy conditions and 15 containing hotspots. The IR images were captured from various angles and include scenarios with one or multiple PV modules, as illustrated in Fig. 6.

We feed these images and a prompt to GPT 4o and Gemini 1.5 pro, where the prompt is set as "*This a PV IR image, tell me whether it presents hot spot, yes or no, no verbose*". The detection results are presented in Fig. 6 and Fig. 7 with the metrics listed in TABLE 1.

From the results, GPT's performance is lower than Gemini's, with 10 hot spot images misidentified as healthy (as shown in Fig. 7). Its F1 score is only 0.63. Comparatively, Gemini only has 2 cases of FP, with the F1 score reaching 0.93.

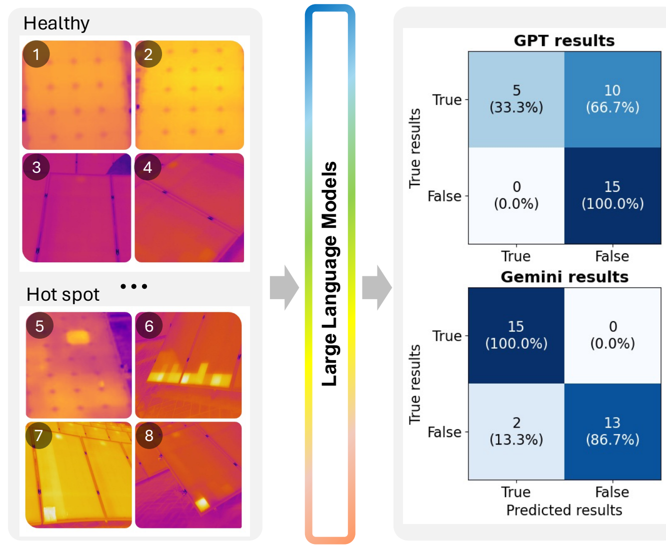

Fig. 6 Example of PV ER images (health and with hot spot) and detection results (confusion matrix) using GPT 4o and Gemini 1.5 pro.

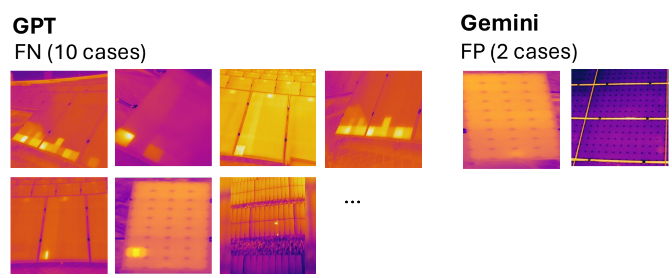

Fig. 7 Example of misclassified PV IR images using GPT 4o and Gemini 1.5 pro.

IV. Discussion

Globally, Gemini 1.5 Pro delivers better and more consistent results, achieving an average F1 score of 0.97 compared to ChatGPT 4o's F1 score of 0.77. Among the three image types, detection performance using visible images is notably better than with EL or IR images. This is reasonable, as visible images are more common in real life and are therefore more likely to be included in the training datasets of the LLMs.

Compared to traditional machine learning (ML) models for PV fault detection, like the ones reviewed in [11], the primary advantage of MFMs is their simplicity and ease of use. Users do not need to undergo the complex processes of data collection, preprocessing, model selection, construction, training, validation, and testing. All results presented in this paper are direct outputs from existing MFMs, allowing users to obtain immediate results with just an image and a simple prompt. Future work will quantitatively compare MFM results with ML models on the same datasets.

Furthermore, as shown in the example images, the inputs are not constrained by size, angle, scope, or module type, showcasing the strong adaptability and user-friendliness of MFMs.

For the PV application, MFMs show great potential for pre-screening images for fault detection or as a complementary tool for double-validating results alongside specialized PV image-based fault detection tools.

It is worth noting that MFMs are updated frequently, and the results presented here are based on ChatGPT 4o and Gemini 1.5 Pro, both are free upon registration as of Jan 2025. Future versions may achieve even better outcomes. However, it is important to consider that companies like OpenAI and Google offer different model tiers at varying price points.

IV. Conclusion

In this paper, we evaluated the performance of image-based fault detection using both ChatGPT 4.0 and Gemini 1.5 Pro (without fine-tuning). Three fault types were addressed: bird droppings detected via visible images, cell cracks detected via electroluminescence (EL) images, and hot spots detected via infrared (IR) images. Gemini 1.5 Pro consistently delivered near-perfect results, achieving an average F1 score of 0.97, outperforming ChatGPT 4.0 in both accuracy and consistency.

Unlike the complex process of building a traditional machine learning (ML) model, MFMs enable researchers to achieve high-accuracy results instantly with a simple prompt. Notably, the images used for testing are not constrained by size, angle, scope, or module type, demonstrating the robust adaptability and user-friendliness of MFMs. While they may not fully replace specialized fault detection models, LLMs can serve as effective tools for the automatic pre-screening of PV images or as complementary tools to validate results alongside dedicated PV image-based fault detection ML models or tools.

References

[1] N. Fei *et al.*, "Towards artificial general intelligence via a multimodal foundation model," *Nature Communications 2022 13:1*, vol. 13, no. 1, pp. 1–13, Jun. 2022, doi: 10.1038/s41467-022-30761-2.

[2] OPENAI, "ChatGPT 4o." Accessed: Jan. 15, 2025. [Online]. Available: https://chatgpt.com/

[3] Google AI., "Gemini." Accessed: Jan. 15, 2025. [Online]. Available: https://ai.google/gemini/

[4] J. Wu, X. Tang, Z. Yang, K. Hao, L. Lai, and Y. Liu, "An Experimental Evaluation of LLM on Image Classification," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* , vol. 15449 LNCS, pp. 506–518, 2025, doi: 10.1007/978-981-96-1242-0_37.

[5] J. Wang, E. Zurich, and L. Ke, "LLM-Seg: Bridging Image Segmentation and Large Language Model Reasoning," 2024. Accessed: Jan. 11, 2025. [Online]. Available: https://github.com/wangjunchi/LLMSeg.

[6] A. Mellit, G. M. Tina, and S. A. Kalogirou, "Fault detection and diagnosis methods for photovoltaic systems: A review," *Renewable and Sustainable Energy Reviews*, vol. 91, pp. 1–17, Aug. 2018, doi: 10.1016/j.rser.2018.03.062.

[7] S. Shrestha, I. Bittencourt, A. S. Varde, and P. Lal, "AI-Based Modeling for Textual Data on Solar Policies in Smart Energy Applications," *2024 15th International Conference on Information, Intelligence, Systems & Applications (IISA)*, pp. 1–8, Jul. 2024, doi: 10.1109/IISA62523.2024.10786713.

[8] K. Nuortimo, J. Harkonen, and K. Breznik, "Global, regional, and local acceptance of solar power," *Renewable and Sustainable Energy Reviews*, vol. 193, p. 114296, Apr. 2024, doi: 10.1016/J.RSER.2024.114296.

[9] S. Majumder *et al.*, "Exploring the capabilities and limitations of large language models in the electric energy sector," *Joule*, vol. 8, no. 6, pp. 1544–1549, Jun. 2024, doi: 10.1016/j.joule.2024.05.009.

[10] X. Chen, T. Karin, and A. Jain, "Automated defect identification in electroluminescence images of solar modules," *Solar Energy*, vol. 242, pp. 20–29, Aug. 2022, doi: 10.1016/J.SOLENER.2022.06.031.

[11] B. Li, C. Delpha, D. Diallo, and A. Migan-Dubois, "Application of Artificial Neural Networks to photovoltaic fault detection and diagnosis: A review," *Renewable and Sustainable Energy Reviews*, p. 110512, Nov. 2020, doi: 10.1016/j.rser.2020.110512.

[12] DuraMat, "DuraMat Data Hub." Accessed: Jan. 21, 2025. [Online]. Available: https://datahub.duramat.org/

[13] Afroz, "Solar Panel Images Clean and Faulty Images." Accessed: Jan. 21, 2025. [Online]. Available: https://www.kaggle.com/datasets/pythonafroz/solar-panel-images

[14] Y. ZHANG, "Dataset of solar cells defect segmentation." Accessed: Jan. 21, 2025. [Online]. Available: https://www.kaggle.com/datasets/yaozhang01182010/dataset-of-solar-cells-defect-segmentation

[15] E. L. C. H. F.-M. É. R.-G. A.-D. N. S. Alfaro Mejia, "Photovoltaic System Thermal Images." Accessed: Jan. 21, 2025. [Online]. Available: https://www.kaggle.com/datasets/saurabhshahane/photo-voltaic-cell-images

[16] UITM, "Hotspot Computer Vision Project." Accessed: Jan. 21, 2025. [Online]. Available: https://universe.roboflow.com/uitm-eyqpa/hotspot-ahbdm/dataset/1