2024 Tenth Indian Control Conference (ICC)
9 - 11 December, 2024
IISER Bhopal, India

Accurate Detection of Bright Spots in Solar Photovoltaic Panels

Rahul Nair
Reliance Industries Ltd.
Cochin, India
0009-0001-2706-1159

Anand Thirwani
Reliance Industries Ltd.
Delhi, India
0009-0008-0473-8471

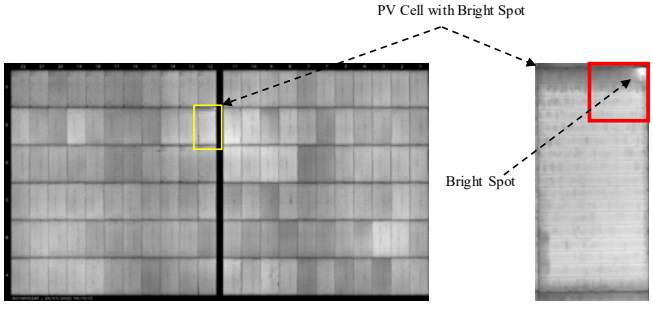
Kedar Kulkarni
Reliance Industries Ltd.
Bangalore, India
0009-0003-8517-7437

***Abstract*— For optimal performance and safety of photovoltaic (PV) solar panels, early detection of manufacturing defects is very critical. A major concern with PV solar panels is the presence of defects on its surface that show up as "bright spots" in their electro luminescence (EL) images. This defect can lead to malfunctions due to uncontrolled energy production, potentially resulting in short circuits, elevated temperatures, and even fire hazards. In this paper, we propose a robust machine learning (ML) based approach to accurately detect bright spots by optimally splitting the EL images of PV solar panels and engineering novel discriminative features. The proposed method's effectiveness was validated with an impressive F1 Score of 0.96 on real EL images captured directly from a functioning PV solar panel manufacturing unit. This signifies its real-world applicability for reliable bright spot detection in PV solar panels.**

Keywords*— *Artificial Intelligence, Machine Learning, Computer Vision, PV Solar Panel Defect Detection, Optimization

I. INTRODUCTION

Solar power is rapidly gaining traction as a key renewable energy source. Unlike fossil fuels, it doesn't contribute to greenhouse gas emissions, making it a weapon against climate change. Photovoltaic (PV) solar panels, a crucial technology for harnessing solar energy, have seen a surge in production and adoption due to their affordability, scalability, and environmental benefits. However, ensuring safe and efficient power generation from PV panels requires defect detection during manufacturing itself. Defects like cracks, voids, and impurities can significantly reduce a panel's lifespan and energy output.



(a) *(b)*

Fig. 1. a) EL image of a PV panel highlighting the PV cell containing a bright spot (yellow box), (b) Zoomed in PV cell with Subject Matter Expert (SME) bright spot annotation (red box).

The final step of PV panel manufacturing [1] involves lamination. Air is pumped out, and heat and pressure are applied to bind the layers contain PV cells together. This process can sometimes cause electrical wires to touch, creating a low-resistance area visible as a 'bright spot' in the panel's electro-luminescence (EL) image. These bright spots, if left undetected, can lead to short-circuits, overheating, and potential fires in deployed panels. Figure 1 shows a typical bright spot in an EL PV panel. Based on process knowledge, these bright spots are typically found on the corners and edges of PV cells.

While bright spots are infrequent defects in PV panels, achieving perfect detection (zero false negatives) presents a significant challenge. However, the consequences of missing even a single bright spot are severe. Undetected bright spots can lead to safety hazards, potentially causing fires and property damage. Additionally, they can significantly damage a manufacturer's reputation. Therefore, automating accurate bright spot detection during PV panel manufacturing is critical.

In this work, our aim is to improve existing bright spot detection models [2] by optimizing the size of the patches (annotating bounding box) that contains bright spots so that machine learning models can learn on more discriminative features of bright spot, thus improving the accuracy of bright spot detection in PV solar panels. The rest of the paper is organized as follows. Section II summarizes previous work on bright spot detection in PV panels. Section III provides a brief overview of the proposed solution formulation. Evaluation and discussion of results are presented in section IV, while section V concludes.

II. PREVIOUS WORK

Researchers have been actively exploring machine learning (ML) based computer vision approaches for defect detection in EL images of PV solar panels. For instance, one study [3] trained traditional computer vision, machine learning, and deep learning models (e.g., KAZE + SVM, ResNet50, VGG19) on high-quality EL images for defect prediction. Another study [4] employed a bi-directional attention feature pyramid network for defect detection in EL images of solar panels. While a method [5] combining thermal imaging and GoogleNet captured "hot spots," it relied on *thermal* imaging data, not the desired EL data.

Recent advancements offer promising solutions for more targeted detection. A Faster-RCNN based scheme [6] tackled

complex backgrounds and small defects in near-infrared EL images, laying the groundwork for automated solutions. Additionally, a U-Net based approach [7] focused specifically on bright spots caused by shielding, effectively segmenting and classifying defects using infrared images. Furthermore, YOLO object detection is being explored for surface defect detection in EL images [8]. However, all the approaches quoted here did not directly address the bright spot detection problem in EL images of solar PV panels.

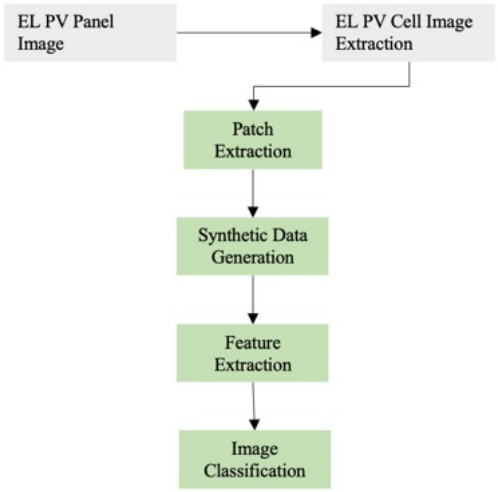


Fig. 2. Bright spot detection approach proposed in [2]

Our previous work [2] identified this gap in existing methodologies and established a novel methodology as shown in Figure 2 to directly address bright spot detection within EL PV images. We also benchmarked our proposed approach with the existing state of the art and found our approach to be superior. This proposed approach has the following key stages: patch extraction, synthetic bright spot generation, novel feature extraction, and finally image classification. This comprehensive approach enabled reliable detection of bright spots within EL PV images.

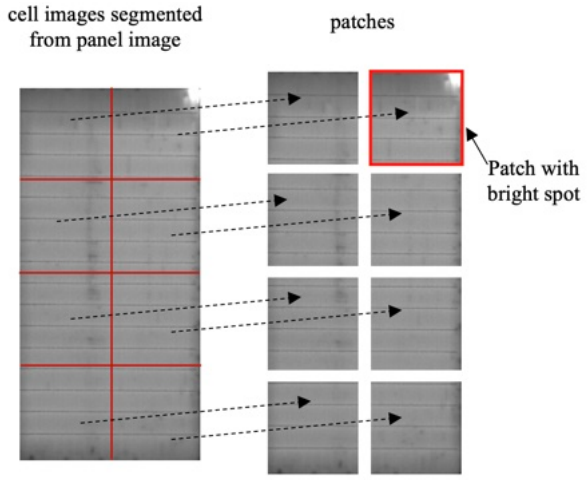


Fig. 3. Patches extracted from 4 x 2 grid of EL PV cell image.

Our previous work started by splitting EL PV cell images from EL PV panel images. Then with patch extraction, we split PV cell images into equal sized non-overlapping patch images. The patch size in the EL PV panel data was adjusted to be approximately the same as that of the bright spot *bounding box* provided by subject matter experts (SMEs). This adjusted patch-size corresponded to a 4 x 2 (4 rows, 2 columns) grid on each PV cell, that yields 8 patches as shown in Figure 3.

The limited number of bright spot instances within the dataset presented the challenge of class imbalance. In other words, the number of images containing bright-spots was much smaller than the number of images without bright-spots (i.e., *normal* images). To mitigate this issue, we generated synthetic data with realistic representations of bright spots, leading to a more balanced data distribution. For feature extraction, we engineered *regular* pixel-intensity based as well as *novel*, discriminative features for bright spots by observing the Inverse Cumulative Density Function (ICDF) of PV cell patches with and without bright spots. Figure 4 shows the difference in ICDF plots between patch images with and without bright spots. As can be observed, there is a steep increase in pixel intensity after 95th percentile for the patch with bright spot. Utilizing this characteristic of bright spots, we built custom features shown in Table I to enable machine learning models to learn discriminative features for bright spot detection.

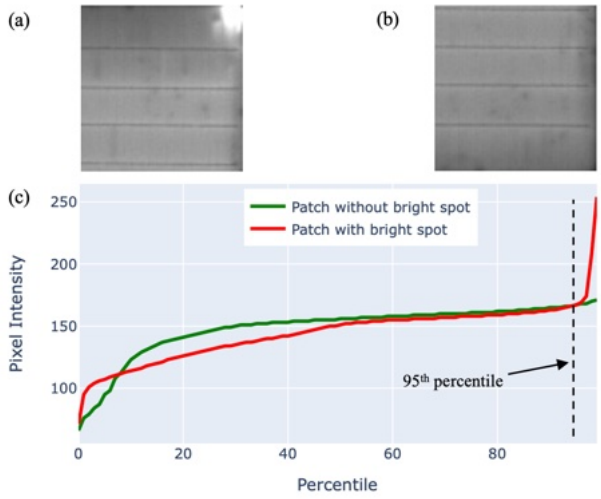


Fig. 4. (a) Patch with bright spot, (b) Patch without bright spot, (c) ICDF plots of a PV patch with bright spot and a PV patch with no bright spot.

Furthermore, in our previous work, we had assumed P and Q (refer Table I) to equal 0.95 and 95 respectively to generate custom features for subsequent image classification model development. The best model was XGBoost [9] which yielded a precision of 0.93, recall of 0.93 and F1 score of 0.93.

In this paper, our objective is to develop a superior model to detect bright spots in EL PV panel images by implementing a novel strategy to *optimize the patch size* (i.e., the size of bounding box containing the bright spot) used for feature extraction, thus improving prediction accuracy. Rather than using bounding boxes provided by SMEs as the basis to choose a patch size, an *optimized* patch size should assist in extracting more discriminative features that will capture the unique characteristics of bright spots within the EL images.

This, in turn, is expected to lead to a more accurate and robust model for bright spot detection.

TABLE I. FEATURES EXTRACTED FOR EACH PATCH

Regular Features	
Feature Name	**Feature Description**
Maximum intensity	Maximum value of the pixel intensities in a PV patch
High intensity flag with threshold **P**	1 if maximum intensity >= P, 0 otherwise (Pixel intensity is scaled between 0 and 1)
Novel Features	
Feature Name	**Feature Description**
Slope of ICDF, Slope between maximum and **Qth** percentile pixel intensities	$\frac{ICDF(100) - ICDF(Q)}{100 - Q}, where\ Q \in [0,100]$
Ratio of ICDF, Ratio between maximum and **Qth** ICDF pixel intensities	$\frac{ICDF(100)}{ICDF(Q)}, where\ Q \in [0,100]$

III. METHODOLOGY

A. Motivation Behind Patch Size Optimization

Selecting the right patch size is extremely important in extracting discriminative features for bright spot detection. Dividing the original cell image into patches of varying sizes can potentially influence the resulting performance of the detection model. To investigate this hypothesis, we compared the ICDF plots of images with different patch sizes generated using two grids: 4 x 2 and 8 x 4. Analyzing these plots allows us to assess the impact of patch size on the distribution of pixel intensities within the patches, potentially affecting the model's ability to learn discriminative features for accurate bright spot detection.

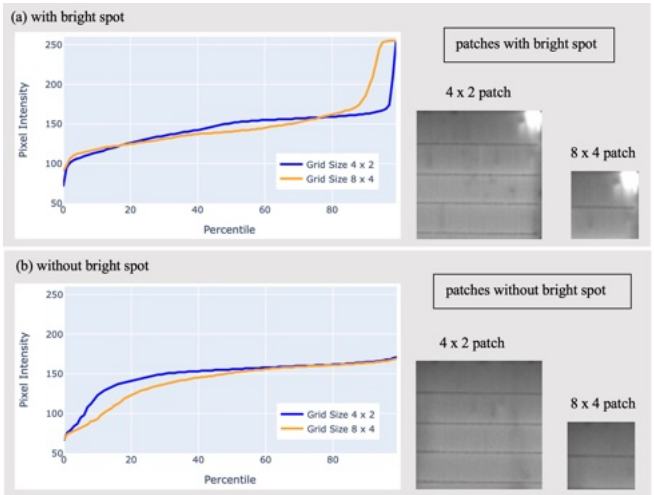


Fig. 5. (a) ICDF plots for patches with bright spots for 4 x 2 and 8 x 4 grids, (b) ICDF plots for patches without bright spots from 4 x 2 and 8 x 4 grids

Figure 5 visually demonstrates the influence of patch size on ICDF plots. As the patch size changes from 4 x 2 to 8 x 4, the ICDF distribution for patches containing bright spots (Figure 5.a) exhibits a shift. Notably, the sharp change observed at the 95th percentile for 4 x 2 patch changed to a gentler gradient shift near the 90th percentile with 8 x 4 patch. In case of patches without bright spot (Figure 5.b), the area under the ICDF curve shows a slight decrease for 8 x 4 patch compared to that of 4 x 2 patch. To quantify the effect of patch size, we calculated features described in Table I for the 4 x 2 and 8 x 4 patches.

TABLE II. FEATURES EXTRACTED FOR 4 X 2 AND 8 X 4 PATCH

Patch Size	**Bright Spot**	**Regular Features**		**Novel Features**		**Euclidean Distance**
		Maximum Intensity	**High Intensity Flag**	**Slope of ICDF**	**Ratio of ICDF**	
4 x 2	Yes	255	1	**17.60**	**1.52**	**15.21**
	No	179	0	**2.40**	**1.07**	
8 x 4	Yes	255	1	*0.40*	*1.00*	*1.00*
	No	179	0	*1.40*	*1.04*	

Note: refer Table I for formula, P and Q are taken as 0.95 and 95 respectively

While regular features remained unaffected by patch size (Table II), novel features exhibited significant changes. To measure this effect, we calculated the Euclidean distance between features from bright spot and non-bright spot patches for each patch size. Larger Euclidean distances indicate more discriminative features, meaning they better distinguish between bright spot and non-bright spot patches. In this case, the Euclidean distance for patch size of 4 x 2 was significantly higher compared to that of 8 x 4. This suggests that a patch size of 4 x 2 is more discriminative and has a higher potential to improve the machine learning model's ability to detect bright spots.

B. Rigorous HPO-Based Patch Size Optimization

To find the optimal patch size which gives the best bright spot detection accuracy, hyper-parameter optimization (HPO) was conducted on the best ML model from previous work (mentioned in Section II), to evaluate model performance variations across different patch-sizes. We limited the search space to determine the optimal patch-size to 16 pair-wise combinations of the following 4 row and column sizes: 2, 4, 6, and 8. This was done due to the significant increase in computational cost associated with training and tuning models for larger number of patches resulted by increasing row and column size. Even with this limited search space, the HPO took several hours, thus inspiring the search for a greedy heuristic that would potentially reduce the computation time and still yield equally good results, if not better. Additionally, along with the patch sizes, the variables P and Q, mentioned in Section II, were also optimized by HPO.

In contrast to our previous work [2], in this work we opted against synthetic data generation through (GANs) [10] to address the class imbalance due to a low number of bright-spot images. This decision was motivated by the substantial computational overhead associated with training a GAN for each of the 16 patch sizes considered for HPO. However, to mitigate class imbalance, class weights were adjusted to facilitate a balanced training on each class.

C. Entropy-Based Patch Size Optimization

In this sub-section, we explain the concept behind the proposed entropy-based patch size optimization as a greedy heuristic to reach optimal patch-size. This greedy heuristic systematically evaluates different patch sizes by calculating a *neighborhood entropy score* as explained in Algorithm I. In terms of data used for optimization, EL PV cell images from the validation set which contains bright spots were used. As shown in the pseudocode (Algorithm I), the calculation for a neighborhood entropy score (E_{ij}) for a given patch size corresponding to the grid *i* x *j* involves the following steps:

1) *Patch Generation*: PV cell images containing bright spots are divided into grids of size *i x j* to create patches.
2) *Patch Entropy Calculation*: For each patch P_{mn}^{b} in cell *b* (where *b* denotes the index of a cell with bright spot, and *m* and *n* are subsets of *[1, 2, …, i]* and *[1, 2, …, jj]*), the Shannon entropy [11] is calculated to quantify the randomness of its pixel intensities. Similarly, Shannon entropy is also calculated for all neighboring patches surrounding P_{mn}^{b}, including side and corner neighbors (as illustrated in Figure 6). This captures the local distribution of intensity around the bright spot patch.
3) *Patch-Level Neighborhood Entropy Calculation*: A patch-level neighborhood entropy score, denoted as X_{mn}^{b} (shown in Algorithm I.8), is computed for each patch P_{mn}^{b} containing a bright spot. This is the ratio of the difference in entropy of P_{mn}^{b} and median of neighboring patch entropies to the sum of *i* and *j*.
4) *Cell-Level Neighborhood Entropy Calculation*: For each cell *b*, the median value of its X_{mn}^{b} scores are taken as the cell-level neighborhood entropy score, denoted as K^{b}.
5) *Final Neighborhood Entropy Calculation*: Finally, after calculating K^{b} for all cells with bright spots in the validation set, the median of all these K^{b} values are taken as the final neighborhood entropy score, E_{ij} for patch size *i x j*.

Once the final neighborhood entropy score is calculated for each patch size, it is used to rank them, allowing us to identify the most promising optimal patch-size candidates. While Shannon entropy serves as the foundation to our approach for patch size ranking, other metrics like Euclidean distance, Jensen-Shannon divergence [12], slope of ICDF and ratio of ICDF (Table I) can be also used to systematically evaluate different patch sizes.

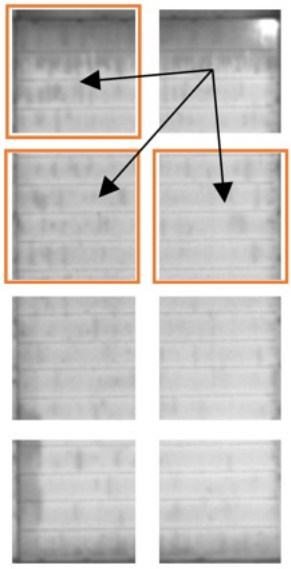


Fig. 6. Neighborhood patches (marked with orange box) of the patch containing bright spot on an EL PV cell split with 4 x 2 grids.

ALGORITHM I : ALGORITHM TO CALCULATE NEIGHBORHOOD ENTROPY SCORE

Initialize S: $S = \{2, 4, 6, 8\}$, *where S is the predefined search space for patch / grid sizes*
Initialize B: *B = number of cells containing brightspots in validation data*
for $i \in S$
  for $j \in S$
    for *b = 1,…, B*
      Split cell b into i x j patches
      for *each patch* $\boldsymbol{P_{mn}^{b}}$, *that contains bright spot, where* $m \in \{1, 2, \ldots, i\}$ *and* $n \in \{1, 2, \ldots, j\}$
        Compute $\boldsymbol{X_{mn}^{b}} = \dfrac{\left(\boldsymbol{Pixel\ Intensity\ Entropy\ of\ P_{mn}^{b}} - \textit{median}\ (\boldsymbol{Pixel\ Intensity\ Entropy\ of\ Neighborhood\ Patches})\right)}{\boldsymbol{i + j}}$
      end
      Compute $\boldsymbol{K^{b}} = median(\{\boldsymbol{X_{mn}^{b}}\})$
    end
    Compute the final neighborhood entropy score for patch i x j, $\boldsymbol{E_{ij}} = median\ (\{\boldsymbol{K^{b}}\})$
  end
end

IV. Results and Discussions

This section explains the results obtained with the HPO approach and the proposed entropy-based greedy heuristic optimization of the patch size. We utilized a dataset consisting of 13 EL panel images containing bright spots and 193 images without them. To ensure consistency and facilitate comparison with our previous work [2], the train, validation, and test splits for this dataset were maintained. Details regarding the specific split ratios can be found in Table III.

TABLE III. Data Distribution of PV Panels

Data Split	Number of bright spot panels	Number of non-bright spot panels
Training	16	96
Validation	3	19
Test	13	78

As mentioned in section III, we conducted HPO over a predefined search space of patch sizes corresponding to 16 pair-wise combinations of 4 row and column sizes each of 2, 4, 6, and 8. For each patch size, data re-labeling was necessary due to the impact of patch size on the position of bright spots after patch extraction. Table IV shows the results of the HPO by summarizing the performance metrics achieved by the optimal models trained for each patch size. All the model hyperparameters have been tuned using the Hyperopt [13] library. Surprisingly, the best values for parameters P and Q, associated with the features mentioned in Section II, remained the same regardless of the patch size. Notably, all HPO optimized models had P and Q of values 0.7 and 85, respectively.

TABLE IV. Performance of Different Patch Sizes Based on Rigorous HPO

*Patch size	Model	Precision	Recall	F1 Score	Run Time
2 x 2	XGBoost	0.22	1.0	0.36	~ 3 min
2 x 4		0.75	0.92	0.83	~ 5 min
2 x 6		0.67	0.92	0.77	~ 8 min
2 x 8		0.71	0.92	0.8	~ 13 min
4 x 2		0.8	0.92	0.86	~ 6 min
4 x 4		0.85	0.85	0.85	~ 14 min
4 x 6		0.67	0.77	0.71	~ 26 min
4 x 8		0.5	0.77	0.61	~ 41 min
6 x 2		1	0.92	0.96	~ 8 min
6 x 4		0.56	0.77	0.65	~ 25 min
6 x 6		0.53	0.62	0.57	~ 50 min
6 x 8		0.39	0.54	0.45	~ 1 hr. 34 min
8 x 2		0.8	0.92	0.86	~ 14 min
8 x 4		0.62	0.62	0.62	~ 58 min
8 x 6		0.28	0.62	0.38	~ 1 hr. 38 min
8 x 8		0.4	0.46	0.43	~ 2 hr. 43 min

*Note: patch size is represented by the grid size using which the EL PV cell images were split

Patch-level performance metrics would be non-comparable due to the variation in sample size caused by differing patch sizes. Thus, for a fair performance comparison, we chose to calculate performance metrics aggregated at the panel level, because the number of bright spot and non-bright spot images remains the same regardless of the patch size. As can be seen in Table IV, the 6 x 2 patch size emerged as the best performer, achieving an F1 score of 0.96. Patch sizes of 4 x 2 and 6 x 2 were close contenders, both achieving F1 scores of 0.86. The total runtime for HPO was approximately 9 hours on a late-2021 16-inch MacBook Pro with M1 Pro chip. However, including the manual data re-labeling effort, the total time taken was about 12 hours.

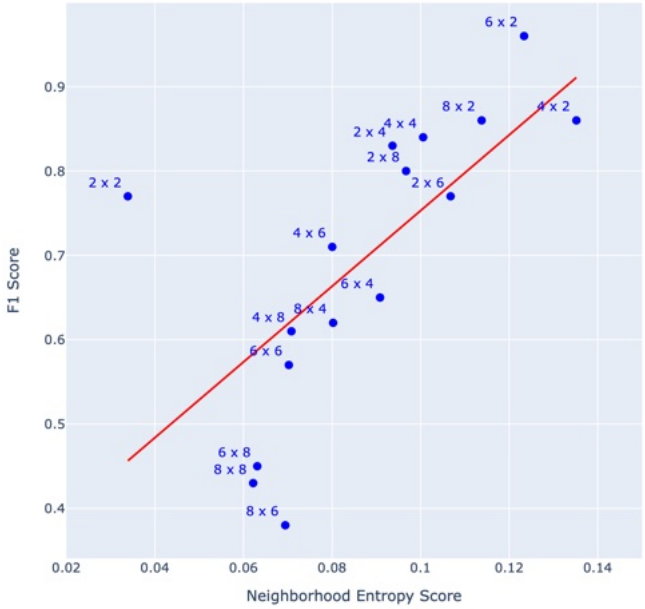


Fig. 7. Positive correlation between neighborhood entropy scores and F1 scores of all patch sizes investigated.

To evaluate the effectiveness of the proposed greedy heuristic-based optimization method (Section III.C), we considered the same patch size combinations used in rigorous HPO. Notably, the optimization exercise only took less than 5 seconds to complete. As can be seen in Figure 7, the F1 score increases with an increase in the neighborhood entropy score, thus demonstrating the strong positive correlation between the final neighborhood entropy score generated by the greedy heuristic and the F1-score generated by rigorous HPO.

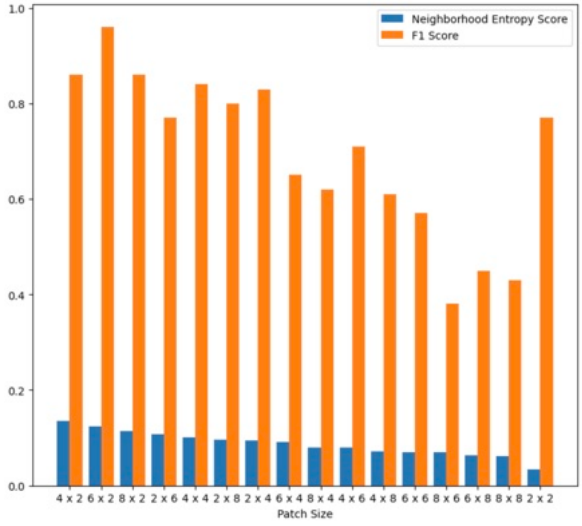


Fig. 8. Bar plot showing best to worst patch size based on Neighborhood Entropy Score

Figure 8 depicts the F1 scores achieved by different patch sizes, sorted according to their corresponding neighborhood entropy scores. To identify promising candidates for the optimal patch-size, we employed an 80th percentile filter on the neighborhood entropy scores. This process resulted in identifying patch sizes of 4 x 2, 6 x 2, and 8 x 2 as candidates for rigorous HPO. By prioritizing only these three patch sizes instead of 16 patch sizes for rigorous HPO, the total runtime could be reduced to under 30 minutes. Even while accounting for the time taken for manual labeling, the total time would be approximately 1 hour, representing a substantial reduction compared to the 12-hour timeframe required for the rigorous HPO with 16 patch sizes.

V. Conclusion

Early detection of bright spots in PV panels during manufacturing is essential for their operational safety. Undetected bright spots can pose a significant safety risk. Our prior work [2], extensively benchmarked with state-of-the-art approaches, demonstrated impressive performance on bright spot detection in EL images of solar PV panels. This paper improves that work by proposing an approach that yields a higher accuracy at a reasonable computational cost. Unlike our previous work, in this paper, we explicitly optimize the patch-size used to split the EL PV cell images leading to more discriminative features, thus improving model accuracy.

Using real EL images captured directly from a functioning PV solar panel manufacturing unit, a rigorous HPO was first conducted, yielding an optimal patch size that led to an impressive F1-score of 0.96, which is higher than the accuracy quoted in our previous work. However, since the extensive HPO was very time-consuming, we proposed a faster alternative based on a greedy heuristic, that yielded the same optimal patch size as the HPO but in significantly less time (approximately 12x faster). The enhanced accuracy in detecting bright spots in EL PV panel images due to the patch-size optimization, thus translates to improved safety and reliability of PV panels.

Future work will focus on investigating the impact of using synthetic data (e.g., generated by GANs) on model accuracy while simultaneously optimizing the strategy to split EL PV cell images for discriminative feature engineering.

References

[1] PVEducation (n.d.). Photovoltaics Education. Retrieved May 2, 2024, from https://www.pveducation.org/

[2] A. Thirwani, R. Nair and K. Kulkarni, "*A Generative Adversarial Network Based Approach for Accurate Detection of Bright Spots in Photovoltaic Panels,*" 2024 11th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2024, pp. 349-354, doi: 10.1109/SPIN60856.2024.10511515

[3] Sergiu Deitsch, Vincent Christlein, Stephan Berger, Claudia Buerhop-Lutz, Andreas Maier, Florian Gallwitz, Christian Riess, "*Automatic classification of defective photovoltaic module cells in electroluminescence images*," Solar Energy,Volume 185,2019,Pages 455-468,ISSN 0038-092X, doi: 10.1016/j.solener.2019.02.067

[4] B. Su, H. Chen and Z. Zhou, "*BAF-Detector: An Efficient CNN-Based Detector for Photovoltaic Cell Defect Detection*," in IEEE Transactions on Industrial Electronics, vol. 69, no. 3, pp. 3161-3171, March 2022, doi: 10.1109/TIE.2021.3070507

[5] B. Du, Y. He, Y. He, J. Duan and Y. Zhang, "*Intelligent Classification of Silicon Photovoltaic Cell Defects Based on Eddy Current Thermography and Convolution Neural Network*," in IEEE Transactions on Industrial Informatics, vol. 16, no. 10, pp. 6242-6251, Oct. 2020, doi: 10.1109/TII.2019.2952261

[6] L. Liu, Y. Zhu, M. R. Ur Rahman, P. Zhao and H. Chen, "*Surface Defect Detection of Solar Cells Based on Feature Pyramid Network and GA-Faster-RCNN*," 2019 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI), Xi'an, China, 2019, pp. 292-297, doi: 10.1109/CCHI.2019.8901952

[7] Liu J and Ji N (2023), "*A bright spot detection and analysis method for infrared photovoltaic panels based on image processing*," Front. Energy Res. 10:978247, doi: 10.3389/fenrg.2022.978247

[8] Ziyao Meng, Shengzhi Xu, Lichao Wang, Youkang Gong, Xiaodan Zhang, Ying Zhao 2022. "*Defect object detection algorithm for electroluminescence image defects of photovoltaic modules based on deep learning,*" Jan, 2022, Energy Science and Engineering, doi: 10.1002/ese3.1056

[9] T. Chen, C Guestrin, "*XGBoost: A Scalable Tree Boosting System,*" KDD '16 Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 785–794, doi: 10.1145/2939672.2939785

[10] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta and A. A. Bharath, "*Generative Adversarial Networks: An Overview,*" in IEEE Signal Processing Magazine, vol. 35, no. 1, pp. 53-65, Jan. 2018, doi: 10.1109/MSP.2017.2765202

[11] C E Shannon, "*A Mathematical Theory of Communication,*" Bell System Technical Journal, vol. 27, no. 3, pp. 379-423, July 1948, doi: 10.1002/j.1538-7305.1948.tb01338.x

[12] M.L. Menéndez, J.A. Pardo, L. Pardo, M.C. Pardo, "The Jensen-Shannon divergence," Journal of the Franklin Institute, vol. 334, no. 2, pp. 307-318, March 1997, doi: 10.1016/S0016-0032(96)00063-4

[13] J. Bergstra, D. Yamins, and D. D. Cox, "*Hyperopt: a Python library for model selection and hyperparameter optimization*," Journal of Machine Learning Research, vol. 13, pp. 281-286, 2015, doi: 10.1088/1749-4699/8/1/014008