Automated Photovoltaic Hotspot Detection Based on I-V Curve Characteristics and Machine Learning Classification

Imam Faried Assalam [1,*,a], M Sahrul Ramadan [2,b], Henra Jasman [3,c], Ihlas [4,d], Fitriani [5,e], Dea Calista [6,f]
[1,2,3,4] Department of Electrical Engineering, State Polytechnic of Ujung Pandang, Makassar, 90234, Indonesia.
[*,a]imamfaried@poliupg.ac.id, [b]m.sahrulramadan@poliupg.ac.id, [c]henra.jasman@poliupg.ac.id, [d]ihlas@poliupg.ac.id, [e]fitriani@poliupg.ac.id, [f]deacalista@poliupg.ac.id

***Abstract*— Hotspot formation in photovoltaic (PV) modules represents a major degradation issue that reduces system efficiency, reliability, and operational lifespan. Conventional diagnostic techniques such as infrared thermography (IRT) and electroluminescence (EL) imaging provide accurate fault detection but are often constrained by high costs and limited scalability for large PV installations. To address these challenges, this study develops a machine learning based diagnostic framework for hotspot fault classification using electrical parameters derived from I-V curve analysis. Field measurements were obtained using I-V curve tracers and environmental sensors, then normalized to Standard Test Conditions (STC). Two classification algorithms Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) were employed with various kernel and distance configurations, including Linear, Quadratic, Cubic, and Gaussian kernels for SVM, as well as Euclidean, Cosine, and Weighted metrics for KNN. Model performance was evaluated through training/testing accuracy, confusion matrices, prediction speed, and classification metrics such as precision, recall, and F1-score. The Fine Gaussian SVM achieved the highest performance with 91.7% training accuracy, 95.0% testing accuracy, 97.2% recall, and an F1-score of 91.10%, demonstrating its strong capability for reliable and cost-effective hotspot diagnosis in PV systems.**

***Keywords*— Classification, Hotspot, I-V curve, KNN, SVM.**

I. Introduction

Photovoltaic (PV) systems have become a cornerstone of the global transition toward clean and sustainable energy generation. However, the long-term performance and reliability of these systems are highly dependent on the health and stability of individual PV modules. One of the most common and detrimental faults affecting PV modules is the hotspot phenomenon, a condition in which localized cells experience significantly elevated temperatures compared to their surrounding areas. This issue often arises from partial shading, microcracks, delamination, or internal cell defects [1][2]. The presence of hotspots not only reduces the output power of PV modules but also accelerates thermal degradation, potentially leading to irreversible damage or even fire hazards.

Early detection of hotspots is therefore crucial for enabling predictive maintenance and improving system reliability. Conventional diagnostic approaches such as infrared thermography (IRT) and electroluminescence (EL) imaging are known for their high diagnostic accuracy; however, their practical implementation in large scale PV arrays remains challenging due to high costs, the need for specialized equipment, and dependence on manual inspection [3]. These limitations underscore the need for an automated, scalable, and cost-efficient alternative for hotspot detection and condition monitoring.

One promising approach for diagnosing PV degradation involves the analysis of current-voltage (I-V) curves. Electrical parameters derived from I-V curves, including open-circuit voltage (Voc), short-circuit current (Isc), maximum power (Pmax), voltage and current at maximum power point (Vmp and Imp), and fill factor (FF) are highly sensitive to the electrical and thermal anomalies induced by hotspots [4][5]. By extracting and analyzing these parameters, it becomes possible to identify distinctive patterns that differentiate healthy modules from those affected by faults.

With the rapid advancement of data-driven techniques, machine learning (ML) has emerged as a powerful tool for automating PV fault diagnosis. Among various ML algorithms, K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) are particularly effective for classification tasks in PV system diagnostics. KNN is a simple yet robust non-parametric algorithm that classifies data based on the predominant class among its nearest neighbors, while SVM constructs an optimal decision boundary that maximizes the separation margin between different classes. Notably, SVM has shown strong

generalization performance, even with limited datasets, especially when kernel functions are utilized [6][8].

Previous research has shown that both Support Vector Machine (SVM) and k-Nearest Neighbors (KNN) are highly effective algorithms for hotspot classification in photovoltaic (PV) modules. Previous studies report high classification accuracy using SVM-based PV fault detection methods [3][4][12]. Likewise, KNN demonstrates strong performance with its effectiveness increasing in larger and more diverse datasets [9][12][13].

Several studies have explored hybrid and ensemble approaches combining SVM, KNN, and other classifiers, demonstrating improved classification accuracy. However, increased model complexity may reduce computational efficiency and scalability for real-time field deployment [2][4][5][10][16][20].

Comparative analyses reveal that SVM generally provides more stable and accurate performance than KNN, particularly in smaller or noise-affected datasets [7][11][15][19]. However, KNN remains advantageous for larger-scale applications and scenarios prioritizing computational simplicity and rapid deployment.

This study proposes a comparative evaluation of multiple SVM kernel functions and KNN variants using real operational data from a utility-scale tropical PV plant, emphasizing accuracy and cost tradeoff analysis.

II. Research Methodology

A. *I-V Curve Analysis in PV Diagnostics*

Faults such as microcracks, shading, bypass diode failures, and hotspots introduce distortions in I–V curve characteristics. These distortions affect key electrical parameters, including Voc, Isc, Pmp, FF, Rs, and Rsh, enabling electrical-based identification of degradation patterns without invasive inspection [26-28].

The occurrence of hotspots has a pronounced impact on PV system performance and energy yield. Localized heating in affected regions elevates carrier conductivity, resulting in disproportionate current flow at specific voltage levels. This phenomenon manifests on the I–V curve as an increase in current accompanied by a decline in voltage, as illustrated in Figure 1. The localized dissipation of electrical energy as heat leads to efficiency losses and accelerates material degradation. In severe cases, persistent hotspots can cause irreversible damage to solar cells, reduce module lifespan, and compromise the overall reliability and safety of the photovoltaic system.

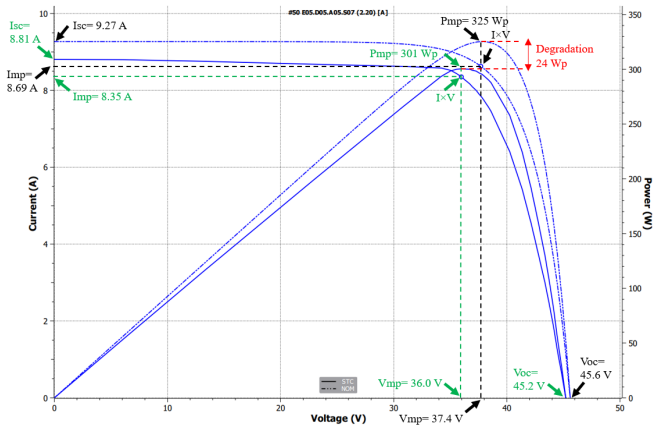


Figure 1. I-V curve parameters under hotspot condition

Table 1 illustrates the variations of electrical parameters derived from the I–V curve under nominal and Standard Test Conditions (STC). The nominal condition represents the ideal operational state in which the photovoltaic module performs according to the manufacturer's rated specifications, whereas the STC condition corresponds to the actual measurement standardized to 1000 W/m² irradiance, 25 °C cell temperature, and an air mass of 1.5. Furthermore, Table 1 provides an example of I–V curve parameter deviations under hotspot conditions. When a hotspot forms on one or more cells, all major electrical parameters including Voc, Isc, Pmp, and FF exhibit degradation, leading to a significant decline in performance and a measurable power loss of approximately 24 Wp.

The examination of I–V curve characteristics has been recognized as a robust and non-invasive electrical diagnostic approach for PV system monitoring. Through systematic comparison between measured I–V curves and baseline reference models, this method effectively detects a wide range of performance anomalies, including hotspots, partial shading, and module aging. When combined with normalization techniques, I–V curve

analysis enables consistent performance evaluation across varying environmental conditions, thereby enhancing scalability and reliability in field applications.

Table 1. Normal VS Hotspot differences

Parameters	STC	Hotspot	Diff
Open circuit voltage (Voc)	45,6 V	45.2 V	-0.4 V
Voltage at max power (Vmp)	37.5 V	36.0 V	-1.5 V
Short circuit current (Isc)	9.27 A	8.81 A	-0.46 A
Current at max power (Imp)	8.69 A	8.35 A	-0.34 A
Max power (Pmax)	325 Wp	310 Wp	-24 Wp
Fill Factor (FF)	0.77	0.75	-0.02

Moreover, the integration of I–V curve–based diagnostics with machine learning (ML) methods has further improved the automation and accuracy of fault classification. Algorithms such as K-Nearest Neighbor (KNN) and Support Vector Machine (SVM) are particularly well-suited for pattern recognition within electrical datasets, enabling the identification of subtle variations in curve parameters that indicate the presence of faults. This hybrid approach provides a cost-effective, data-driven solution for large-scale PV health monitoring and predictive maintenance.

B. *I-V Curve Measurement*

The evaluation of I-V curve serves as a fundamental procedure for assessing the performance of photovoltaic (PV) modules and detecting early-stage anomalies such as hotspot formation. In this study, I–V curve data were collected using a portable I–V curve tracer directly connected to individual PV modules or strings under investigation. Prior to measurement, environmental parameters specifically solar irradiance and module surface temperature were recorded using a calibrated pyranometer and a surface-mounted thermocouple sensor, respectively. All measurements were performed under clear-sky conditions, with irradiance levels maintained above 700 W/m² and minimal temperature fluctuation, in accordance with IEC 61829 standards for field I–V curve testing [29].

During data acquisition, the I–V tracer executed a rapid voltage sweep from open-circuit to short-circuit conditions within approximately 0.5 seconds. This short acquisition time minimized the impact of transient environmental variations on measurement accuracy. The resulting I–V curves provided critical electrical parameters, including open-circuit voltage (Voc), short-circuit current (Isc), maximum power point (Pmp), fill factor (FF), and internal resistances series resistance (Rs) and shunt resistance (Rsh). All data were digitally stored for subsequent feature extraction and analysis.

To ensure consistency and comparability across different test sessions, all electrical parameters were normalized to Standard Test Conditions (STC: irradiance 1000 W/m², cell temperature 25 °C, and air mass 1.5). The normalization process employed standard correction equations that account for irradiance and temperature dependencies, as outlined in [30]. This approach ensured that the resulting dataset accurately represented module performance under standardized operating conditions, thereby enabling reliable input for subsequent machine learning–based classification and diagnostic analysis.

C. *Proposed Machine Learning-Based Diagnostic Framework*

This study proposes a structured machine learning-based diagnostic framework for automated hotspot detection in photovoltaic modules. The framework consists of four sequential stages: (1) I–V curve acquisition under field conditions, (2) parameter extraction and normalization to Standard Test Conditions (STC), (3) supervised machine learning model development using SVM and KNN algorithms, and (4) performance evaluation using confusion matrix and statistical classification metrics. This systematic architecture ensures reproducibility, scalability, and suitability for real-time PV fault monitoring applications.

D. *Machine Learning for Classifications*

In this study, the classification of photovoltaic (PV) module conditions specifically distinguishing between

normal and hotspot-affected modules was conducted using two supervised machine learning algorithms: Support Vector Machine (SVM) and k-Nearest Neighbors (KNN). These algorithms were selected due to their proven effectiveness in PV fault diagnostics and suitability for small to medium-sized datasets.

SVM and KNN were selected due to their strong generalization capability and suitability for small-to-medium datasets in PV diagnostics [5][7].

The input features used for both models were derived from the measured and STC-normalized I–V curve parameters, including open-circuit voltage (Voc), short-circuit current (Isc), maximum power point (Pmp), fill factor (FF), and the series and shunt resistances (Rs and Rsh). For the SVM model, a Radial Basis Function (RBF) kernel was employed to enable non-linear decision boundary formation. Hyperparameter optimization was performed through a grid search approach, tuning the penalty parameter (C) and kernel coefficient (γ) to achieve optimal classification accuracy.

For the KNN model, the optimal number of neighbors (k) was empirically determined through 5-fold cross-validation, balancing model bias and variance. Both algorithms were trained and tested using the same dataset to ensure a fair comparison under controlled conditions.

To evaluate classification performance, confusion matrices were generated for each algorithm. The confusion matrix provides a detailed visualization of correct and incorrect predictions for each class, enabling a granular analysis of model behavior. Previous PV diagnostic studies have demonstrated the importance of this metric across multiple algorithms, including artificial neural networks, SVM, random forest, and deep learning. For instance, confusion matrices were used to evaluate classification performance, as commonly adopted in PV fault diagnosis studies [31–36].

Performance assessment in this study further incorporated precision, recall, and F1-score metrics to provide a comprehensive evaluation of classification reliability. The SVM model exhibited superior accuracy, particularly in datasets characterized by non-linear class boundaries, whereas KNN demonstrated competitive performance with lower computational complexity, especially in regions with well-separated feature distributions. These findings are consistent with prior research, which highlighted the strong generalization capability of SVM for PV fault detection [2], and the suitability of KNN for lightweight, real-time monitoring systems [10]. Consequently, both algorithms were deemed appropriate and complementary tools for hotspot detection based on I–V curve feature analysis in photovoltaic modules.

III. Result and Discussion

The experimental results presented in this section evaluate the effectiveness of the proposed machine learning-based diagnostic framework in classifying hotspot-affected PV modules. The comparison between SVM and KNN models demonstrates the robustness of the framework in handling nonlinear I–V curve characteristics and distinguishing faulty modules from normal operating conditions.

The comparative results of training and testing accuracies, as illustrated in Figure 2 and Figure 3, Polynomial SVM models showed increasing training accuracy with higher order kernels; however, overfitting was observed in the cubic configuration. Among Gaussian kernels, the Fine Gaussian SVM achieved the best generalization performance (95.0% testing accuracy), outperforming medium and coarse variants.

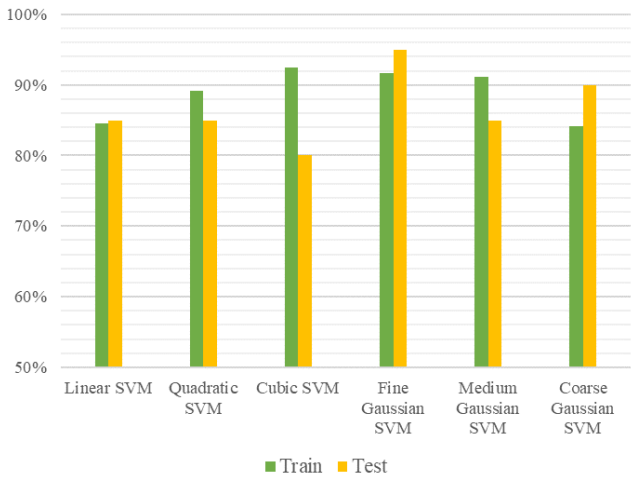


Figure 2. Summary SVM Classifications

These findings confirm that while simpler models like Linear and Quadratic SVMs provide stable baseline performance, the Fine Gaussian SVM offers the optimal trade-off between accuracy and generalization. Conversely, the KNN family particularly its coarse and cosine variants perform adequately but remains more prone to data dependency and overfitting under fine-grained configurations.

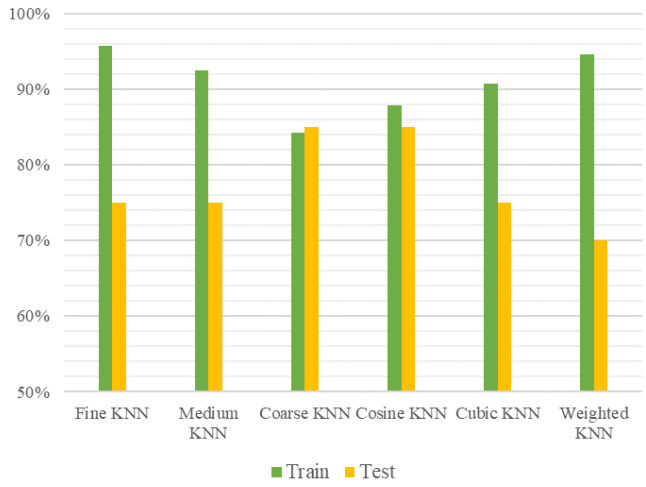


Figure 3. Summary KNN Classifications

Training time analysis (Table 2) indicates that Gaussian SVM variants offer the best trade-off between computational efficiency and accuracy. KNN models exhibit slower prediction rates due to their instance-based learning mechanism.

Cost analysis indicates that the Fine Gaussian SVM demonstrates the most efficient balance between cost and accuracy, followed by Quadratic SVM and Weighted KNN. Other models exhibit higher computational cost with moderate performance gains. This cost-based analysis underscores the importance of balancing learning performance and generalization for practical deployment.

In Figures 4 and 5, precision and recall analysis confirms, precision and recall analysis confirms that Fine Gaussian SVM and Weighted KNN provide the most balanced performance. Fine Gaussian SVM achieved the highest recall (97.2%), while Weighted KNN exhibited superior precision (98.9%). Other models remained within acceptable classification ranges.

Table 2. Summary time & cost classifications

Method	Train time	Prediction	Cost (Train)	Cost (Test)
Linear SVM	11.642 sec	1400 obs/sec	37	3
Quadratic SVM	1.813 sec	3600 obs/sec	26	3
Cubic SVM	1.589 sec	3300 obs/sec	18	4
Fine Gaussian SVM	1.747 sec	3400 obs/sec	20	1
Medium Gaussian SVM	1.531 sec	3100 obs/sec	21	3
Coarse Gaussian SVM	1.538 sec	2500 obs/sec	38	2
Fine KNN	5.881 sec	1200 obs/sec	10	5
Medium KNN	1.857 sec	2000 obs/sec	18	5
Coarse KNN	1.597 sec	2100 obs/sec	38	3
Cosine KNN	1.793 sec	1800 obs/sec	29	3
Cubic KNN	1.868 sec	1900 obs/sec	22	5
Weighted KNN	1.424 sec	2100 obs/sec	13	6

Overall, the results confirm that the proposed diagnostic framework successfully integrates I–V curve analysis and machine learning classification into a practical automated solution for hotspot detection. The framework not only achieves high classification accuracy but also maintains computational efficiency suitable for real-time deployment in large-scale PV systems.

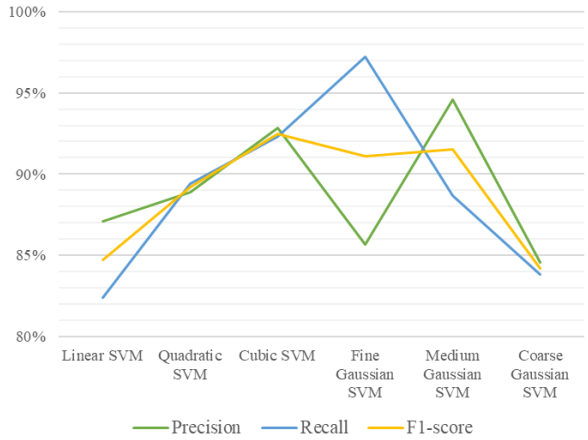


Figure 4. Summary SVM Evaluations

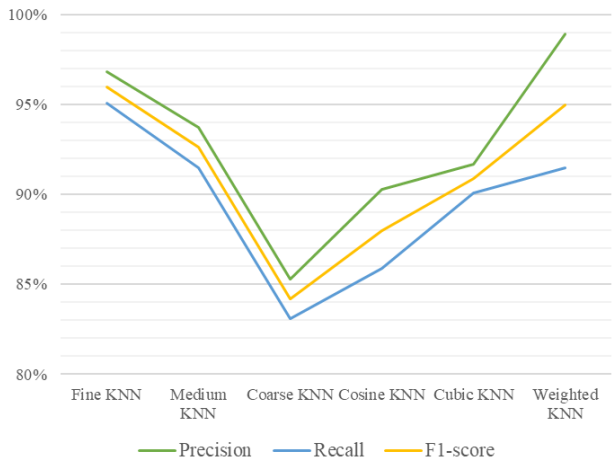


Figure 5. Summary KNN Evaluations

IV. Conclusion

This study proposed a machine learning-based diagnostic framework for PV hotspot detection using I–V curve parameters. Experimental results demonstrate that Fine Gaussian SVM achieved the highest testing accuracy (95.0%), while Weighted KNN provided competitive precision. The framework offers a reliable and computationally efficient solution suitable for practical PV monitoring applications.

Acknowledgement

The authors acknowledge the support of the 21 MWp Likupang Solar Power Plant, North Sulawesi, for providing facilities and measurement access for this study.

References

[1] M. Dhimish, "Defining the best-fit machine learning classifier to early diagnose photovoltaic solar cells hot-spots," Case Studies in Thermal Engineering, p. 100980, 2021.
[2] Assalam, I. Gunadin, and S., "Application of Support Vector Machine (SVM) Method for Photovoltaic Condition Classification Based on Characteristic Curve Indicators," Proc. 2023 Int. Conf. Artificial Intelligence Robotics, Signal and Image Processing (AIRoSIP), pp. 188–193, 2023.
[3] D. Winston et al., "Solar PV's Micro Crack and Hotspots Detection Technique Using NN and SVM," IEEE Access, vol. 9, pp. 127259–127269, 2021.
[4] M. Ali, H. Khan, M. Masud, K. Kallu, and A. Zafar, "A machine learning framework to identify the hotspot in photovoltaic module using infrared thermography," Solar Energy, 2020.
[5] N. S., D. Sripada, S., and M. Aghaei, "Detection of visual faults in photovoltaic modules using a stacking ensemble approach," Heliyon, vol. 10, 2024.
[6] A. Amiri, H. Oudira, A. Chouder, and S. Kichou, "Faults detection and diagnosis of PV systems based on machine learning approach using random forest classifier," Energy Conversion and Management, 2024.
[7] F. Wang, Z. Zhen, B. Wang, and Z. Mi, "Comparative study on KNN and SVM based weather classification models for day-ahead short-term solar PV power forecasting," Applied Sciences, vol. 8, p. 28, 2017.
[8] G. Divya, "Analyzing the performance of SVM and KNN in supervised learning for classification," Int. J. for Research in Applied Science and Engineering Technology, 2025.
[9] H. Limam, A. Zouhair, and W. Oueslati, "A new hybrid multiclass approach based on KNN and SVM," J. Inf. Knowl. Manag., vol. 21, no. 2250061, pp. 1–16, 2022.
[10] E. Boateng, J. Otoo, and D. Abaye, "Basic tenets of classification algorithms K-Nearest-Neighbor, Support Vector Machine, Random Forest and Neural Network: A review," vol. 8, pp. 341–357, 2020.
[11] L. Demidova, "Two-stage hybrid data classifiers based on SVM and kNN algorithms," Symmetry, vol. 13, p. 615, 2021.
[12] P. Noi and M. Kappas, "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine classifiers for land cover classification using Sentinel-2 imagery," Sensors, vol. 18, 2017.
[13] M. Sha'abani, N. Fuad, N. Jamal, and M. Ismail, "kNN and SVM classification for EEG: A review," pp. 555–565, 2020.
[14] G. Dlamini et al., "Meta-analytical comparison of SVM and KNN for text classification," Proc. 2021 Int. Conf. Nonlinearity, Information and Robotics (NIR), pp. 1–6, 2021.
[15] Y. Jusman et al., "Comparison of dental caries level images classification performance using KNN and SVM methods," Proc. 2021 IEEE Int. Conf. Signal and Image Processing Applications (ICSIPA), pp. 167–172, 2021.
[16] A. Roy et al., "Comparative analysis of KNN and SVM in multicriteria inventory classification using TOPSIS," Int. J. Inf. Technol., vol. 15, pp. 3613–3622, 2023.
[17] P. Princess, S. Silas, and E. Rajsingh, "Classification of road accidents using SVM and KNN," pp. 27–41, 2020, doi: 10.1007/978-981-15-3514-7_3.
[18] M. Kushalatha, H. Prasantha, and B. Shetty, "Comparison of image classification techniques and its applications," Asian J. of Convergence in Technology, 2021.
[19] R. Pharswan and J. Singh, "Performance analysis of SVM and KNN in breast cancer classification: A survey," pp. 133–140, 2020.
[20] Y. Lin and J. Wang, "Research on text classification based on SVM-KNN," Proc. 2014 IEEE 5th Int. Conf. Software Engineering and Service Science, pp. 842–844, 2014.
[21] N. Azkona, A. Llaria, O. Curea, and F. Recart, "Detection, characterization and modeling of localized defects and thermal breakdown in photovoltaic panels from thermal images and I-V curves," Electronic Materials, vol. 3, no. 2, pp. 154–172, 2022.
[22] G. Li, F. Wang, F. Feng, and B. Wei, "Hot spot detection of photovoltaic module based on distributed fiber Bragg grating sensor," Sensors, vol. 22, no. 13, p. 4951, 2022.
[23] M. Santhakumari and N. Sagar, "A review of the environmental factors degrading the performance of silicon wafer-based photovoltaic modules: Failure detection methods and essential mitigation techniques," Renew. Sustain. Energy Rev., vol. 110, pp. 83–100, 2019.
[24] IEC, IEC TS 63126:2020 ED1 Guidelines for qualifying PV modules, components and materials for operation at high temperatures, International Electrotechnical Commission, 2020.
[25] M. Bressan, Y. El Basri, A. G. Galeano, and C. Alonso, "A shadow fault detection method based on the standard error analysis of I-V curves," Renew. Energy, vol. 99, pp. 1181–1190, 2016.
[26] W. Hermann, W. Wiesner, and W. Vaassen, "Hot spot investigations on PV modules—New concepts for a test standard and consequences for module design with respect to bypass diodes," Proc. 26th IEEE Photovoltaic Specialists Conf., Anaheim, CA, USA, pp. 1123–1132, 1997.
[27] M. Köntges et al., "Performance and reliability of photovoltaic systems subtask 3.2: Review of failures of photovoltaic modules," IEA PVPS Task 13 External Final Report, 2014.
[28] J.-H. Lin and Y.-K. Wu, "Simulation and fault diagnosis using current-voltage characteristics of photovoltaic systems—A case study," Engineering Proceedings, vol. 92, no. 1, p. 79, 2025.
[29] S. Gulkowski and E. Krawczak, "Thin-film photovoltaic modules characterisation based on I-V measurements under outdoor conditions," Energies, vol. 17, no. 23, p. 5853, 2024.
[30] IEC, IEC TR 61853-4:2021 Performance testing and energy rating of photovoltaic modules – Part 4: Standard reference climatic profiles, 2021.
[31] A. Ul-Haq, H. Sindi, S. Gul, and M. Jalal, "Modeling and fault categorization in thin-film and crystalline PV arrays through multilayer neural network algorithm," IEEE Access, vol. 8, pp. 102235–102255, 2020.

[32] K. Zhan, Z. Wang, and W. Shen, "Deep learning-based automatic detection of multitype defects in photovoltaic modules and application in real production line," Prog. Photovolt. Res. Appl., vol. 29, pp. 471–484, 2021.
[33] N. Mekhaznia and R. Khenfer, "Diagnosis of PV module based on neural network using performance indices," Int. J. Power Electron. Drive Syst. (IJPEDS), 2023.
[34] R. Mandal, N. Anand, N. Sahu, and P. Kale, "PV system fault classification using SVM accelerated by dimension reduction using PCA," Proc. 2020 IEEE 9th Power India Int. Conf. (PIICON), pp. 1–6, 2020.
[35] C. Sowthily, S. Kumar, and M. Brindha, "Detection and classification of faults in photovoltaic system using random forest algorithm," pp. 765–773, 2020.
[36] R. Eltuhamy et al., "Fault detection and classification of CIGS thin-film PV modules using an adaptive neuro-fuzzy inference scheme," Sensors, vol. 23, 2023.