

Review

Artificial-Intelligence-Based Detection of Defects and Faults in Photovoltaic Systems: A Survey

Ali Thakfan ^{1,2} and Yasser Bin Salamah ^{3,*} 

¹ Joint Master's Program in Renewable Energy, Deanship of Graduate Studies, King Saud University, Riyadh 11473, Saudi Arabia; 442105774@student.ksu.edu.sa

² Sustainable Energy Technologies Center, King Saud University, Riyadh 11421, Saudi Arabia

³ Department of Electrical Engineering, King Saud University, Riyadh 11421, Saudi Arabia

* Correspondence: ybinsalamah@ksu.edu.sa

Abstract: The global shift towards sustainable energy has positioned photovoltaic (PV) systems as a critical component in the renewable energy landscape. However, maintaining the efficiency and longevity of these systems requires effective fault detection and diagnosis mechanisms. Traditional methods, relying on manual inspections and standard electrical measurements, have proven inadequate, especially for large-scale solar installations. The emergence of machine learning (ML) and deep learning (DL) has sparked significant interest in developing computational strategies to enhance the identification and classification of PV system faults. Despite these advancements, challenges remain, particularly due to the limited availability of public datasets for PV fault detection and the complexity of existing artificial-intelligence (AI)-based methods. This study distinguishes itself by proposing a novel AI-based approach that optimizes fault detection and classification in PV systems, addressing existing gaps in AI-driven fault detection, especially in terms of thermal imaging and current–voltage (I-V) curve analysis. This comprehensive survey identifies emerging trends in AI-driven PV fault detection, highlights the most advanced methodologies, and proposes a novel AI-based approach to enhance fault detection and classification capabilities. The findings aim to advance the state of technology in this field, offering insights into more efficient and practical solutions for PV system fault management.



Citation: Thakfan, A.; Bin Salamah, Y. Artificial-Intelligence-Based Detection of Defects and Faults in Photovoltaic Systems: A Survey. *Energies* **2024**, *17*, 4807. <https://doi.org/10.3390/en17194807>

Academic Editors: Hubert Seigneur and Kristopher Davis

Received: 31 August 2024

Revised: 15 September 2024

Accepted: 24 September 2024

Published: 25 September 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: solar PV; defect detection; machine learning; thermal images; I-V curves; neural networks; SVM; random forest; decision trees; logistic regression; KNN

1. Introduction

The shift towards sustainable energy globally has elevated the role of photovoltaic (PV) systems as a leading solution in renewable energy. Ensuring the operational efficiency and longevity of these systems necessitates robust mechanisms for fault detection and diagnosis. Traditionally, PV system faults were identified through manual inspections and standard electrical measurements. This study proposes that employing advanced artificial-intelligence (AI) techniques, particularly machine-learning (ML) and deep-learning (DL) models, can significantly improve fault detection by analyzing complex patterns in thermal images and I-V curves, offering greater precision than traditional methods. However, manual approaches have become inefficient for large-scale solar installations due to their labor-intensive nature and potential inaccuracies. Recent advancements in ML and DL have prompted researchers to investigate various computational strategies for the efficient identification and classification of PV system faults. According to the International Renewable Energy Agency, several AI-based systems specifically designed for monitoring PV systems include BeeBryte, which uses AI to optimize energy consumption and equipment health; IBM Watson, which applies AI to manage grid reliability, including PV system performance; and SmartNet, which offers AI tools to enhance coordination and stability in PV operations [1]. Despite the development of numerous AI-based techniques for fault detection in PV systems, determining the most

effective approach remains a challenge. A primary obstacle is the scarcity of publicly available datasets for PV fault detection, which hampers the full exploitation of AI capabilities in this domain. Some existing methods, while accurate, are either excessively time-consuming or too complex for practical application. Additionally, there is a notable research disparity, with certain PV faults receiving extensive examination while others are less explored.

In response to these challenges, our study was initiated. The goal is to offer a comprehensive survey of the latest developments in AI for fault detection and diagnosis in PV systems, with more focus on thermal imaging and I-V curve analytics, bridging the gaps and providing a well-rounded view of the field.

This research study offers several key contributions. First, it identifies the latest PV fault detection trends by leveraging AI techniques. Additionally, it highlights the most advanced AI methods currently employed for identifying PV faults. Finally, the study proposes a novel AI-based approach designed to both detect and classify these faults, advancing the current state of technology in this field.

This review paper is organized as follows:

- **Section 1:** Provides an introduction to the structure and objectives of the study.
- **Section 2:** Offers a general overview and background on solar PV defect classification types and discusses the role of AI in solar PV systems. It also highlights the most commonly used classification algorithms, focusing on data pre-processing techniques and performance evaluation metrics.
- **Section 3:** Outlines the methodology employed in this review.
- **Section 4:** Examines recent studies on AI models and tools for defect detection and fault diagnosis in PV systems.
- **Section 5:** Presents a comprehensive comparison of these models, emphasizing their applicability in utilizing thermal imaging and I-V curves for defect detection.
- **Section 6:** Engages in a detailed discussion.
- **Section 7:** Concludes the paper with key findings and takeaways.

2. Background and Literature Review

2.1. Solar Photovoltaic System Defects and Faults

In recent decades, solar PV systems have gained increasing importance in the sustainable energy sector and electricity production. Ensuring efficient operation and maintenance is critical for achieving optimal performance and longevity. Maintenance of solar PV systems can be classified into three main types [2]: preventive, corrective, and predictive.

Preventive maintenance involves regular inspections, servicing, and cleaning to prevent failures and maintain efficiency. Although this approach can be costlier due to potentially unnecessary site visits, it plays a vital role in averting issues before they occur. Corrective maintenance, on the other hand, addresses problems after they arise, striking a balance between repair costs and minimizing downtime. This type of maintenance includes both urgent and scheduled repairs. Predictive maintenance utilizes real-time data to schedule preventive actions and predict potential failures, thereby reducing the frequency and costs associated with corrective measures.

Several AI-driven corrective maintenance approaches have been applied to restore PV plant systems after malfunctions, such as inverter reconfiguration, which adjusts output to mitigate power losses caused by faults [3]. Dynamic voltage regulators have been employed as compensators to maintain power profiles within permissible limits [4]. Additionally, static synchronous compensators serve as both sources and sinks of reactive power, improving power factor and addressing poor voltage regulation resulting from the power system's inability to meet reactive load demands [5]. Unified power quality conditioners, which function as active filters, offer multifunctional power conditioning by compensating for voltage disturbances and fluctuations and preventing harmonic-loaded currents from entering the power system [6]. Energy storage systems, including batteries, capacitors, and flywheels, are utilized to store excess energy when it is not being consumed [7].

PV modules are subjected to various environmental and physical stresses during their manufacturing, installation, and operational phases. These stresses can negatively impact the efficiency and longevity of the systems. Faults in PV systems can lead to a reduction in annual power generation by approximately 3.6 to 18.9% and a decrease in system efficiency by 0.5 to 1% due to a variety of environmental and operational electrical factors [8,9].

Defects in solar PV systems can be classified based on different perspectives (Figure 1). They can be categorized based on their nature and surroundings. These categories include physical issues, such as cracks, corrosion, and discoloration in the panels; electrical problems, like direct current (DC) and alternating current (AC) faults; and environmental faults, such as hotspots, partial shading, and non-uniform irradiance distribution [10].

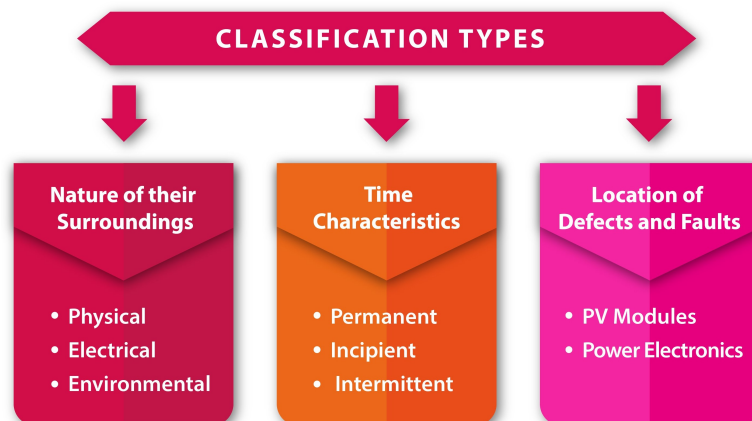


Figure 1. Classification types of solar PV systems.

The second classification for solar PV defects is based on time characteristics: intermittent faults caused by external factors like shading and dust are temporary but reduce efficiency; permanent faults such as short circuits require immediate repair; and incipient faults like cell degradation can worsen over time, making early detection crucial for maintaining long-term system efficiency [11].

Another classification system can be utilized to identify defects or faults in solar PV systems, distinguishing between those that occur in the PV modules, power electronics, or balance-of-system (BOS) components such as inverters and charge controllers [12].

2.2. The Role of AI in Solar PV Systems

Artificial intelligence is a rapidly evolving field in our contemporary society (Figure 2). AI encompasses diverse technologies aimed at solving practical challenges across various industries. As a branch of science and technology, AI develops intelligent machines, algorithms, and programs that perform tasks requiring human-like intelligence, such as learning, pattern recognition, and decision-making [13].

Machine learning, a subset of artificial intelligence, centers on enabling machines to perform tasks that typically require human intelligence. Through a process of learning from their errors, these machines progressively enhance their abilities and improve their performance over time with accumulated experience [14].

Deep learning, a branch of machine learning, uses multiple layers of algorithms, including input, hidden, and output layers, to represent different levels of abstraction. It is applied in voice synthesis, image processing, handwriting recognition, object detection, predictive analytics, and decision-making, enabling advanced data analysis and pattern recognition [15].

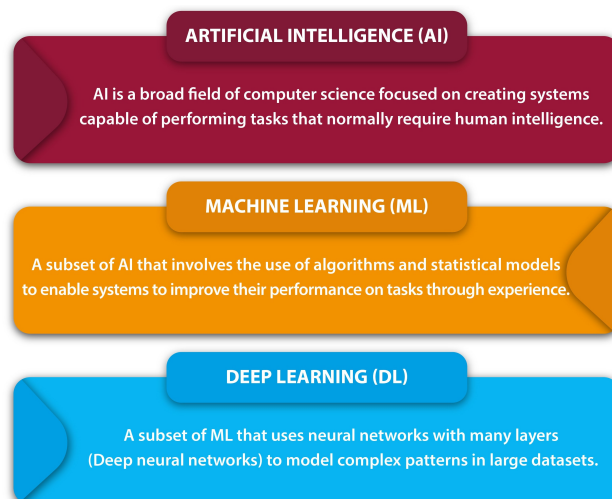


Figure 2. Difference between AI, ML, and DL.

Artificial intelligence can be effectively employed in various aspects of solar PV systems (Figure 3), particularly in predictive maintenance, where it analyzes historical data and identifies patterns to forecast potential failures in critical components such as inverters and modules, thereby enhancing the overall efficiency, reliability, and longevity of solar PV installations [16]. Performance optimization of solar PV systems can be achieved through artificial intelligence, which dynamically adjusts operational parameters in real-time, including the angle of solar panels based on seasonal changes and orientations [17]. Another application of AI in solar PV systems is forecasting energy generation by analyzing weather patterns and historical data, such as cloud coverage, humidity, rainfall, air pressure, temperature, and wind speed, which enables better planning and management of energy resources, ensuring more efficient and reliable solar power generation [18]. Defects and faults in solar PV systems can be detected and diagnosed by analyzing sensor data, such as visual, thermal, and electroluminescent images, and identifying anomalies through various AI techniques, thus enhancing system reliability and maintaining optimal performance [19]. AI can also play a crucial role in energy management systems by employing various techniques to balance demand and supply. Additionally, AI enhances control over storage systems and grid-connected PV systems, ensuring efficient energy distribution and optimal utilization of resources [20]. Modeling and simulation is one of the key features of AI that can help in designing and testing new configurations of solar PV systems by predicting their performance under various weather conditions, such as temperature and sunlight intensity, thus optimizing the system's design and improving overall efficiency [21].

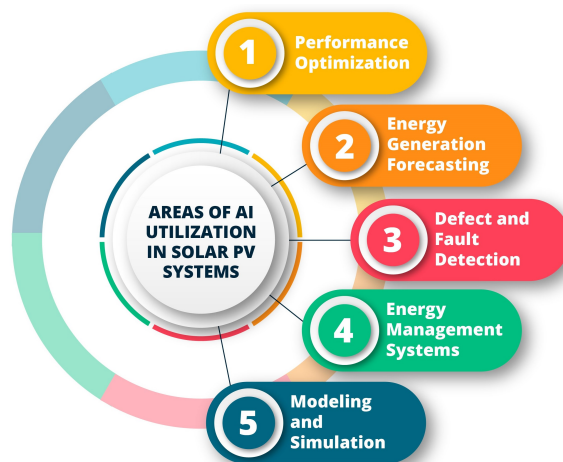


Figure 3. Areas of AI utilization in solar PV systems.

2.3. AI Detection Techniques of Defects and Faults in PV Systems

Numerous AI techniques have been developed and applied for detecting and classifying defects in solar PV systems (Figure 4). Classification, a fundamental concept extensively studied in fields like data mining, machine learning, and information retrieval, involves categorizing ideas, objects, or data into distinct groups based on specific characteristics or features [22]. The following sections will explain the fundamentals of the most common classification techniques.

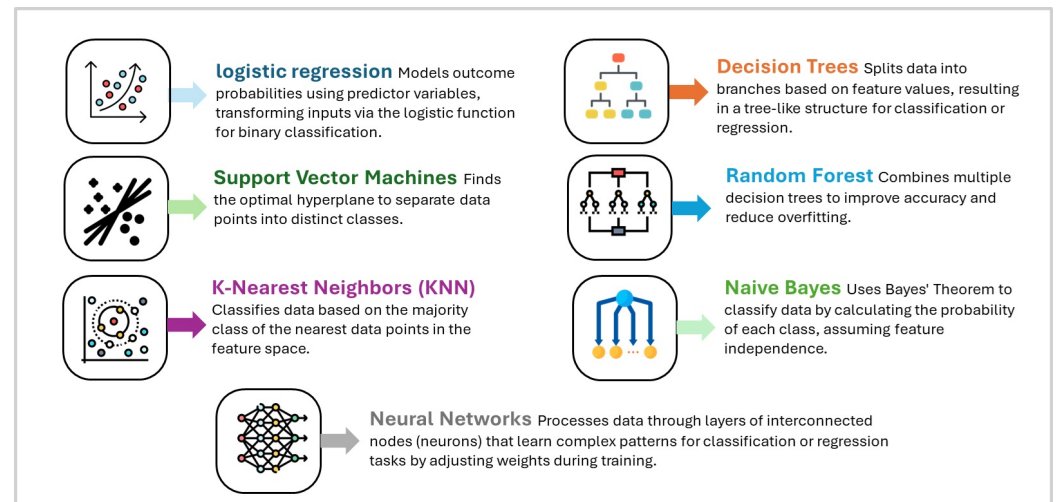


Figure 4. AI detection techniques of defects and faults in photovoltaic systems.

2.3.1. Logistic Regression

Logistic regression is a statistical method for binary classification that models the probability of a categorical dependent variable based on one or more predictor variables, utilizing the logistic (sigmoid) function to transform the linear combination of inputs into a probability between 0 and 1; it is prized for its simplicity and interpretability, as the model's coefficients reflect the change in logarithmic odds of the outcome for a one-unit change in the predictor, offering probabilistic outputs useful for decision-making, and its computational efficiency makes it suitable for large datasets, although it assumes a linear relationship between predictors and log odds [23]. While primarily designed for binary classification, this method can be adapted for multiclass classification challenges using techniques such as multinomial logistic regression or softmax regression [24].

2.3.2. Decision Trees

Decision trees are a widely used classification method that create a flowchart-like structure with decision points (internal nodes) based on attributes, outcomes (branches) of these decisions, and class labels (leaf nodes); they are highly interpretable and visualizable as tree diagrams, manage both numerical and categorical inputs without assuming any specific data distribution, and capture complex, non-linear relationships, making them versatile for various applications, but they can overfit, especially with deep trees, and are sensitive to slight data variations, which can lead to different tree structures and poor performance on unseen data [25,26].

2.3.3. Support Vector Machines

Support vector machines (SVMs) identify the optimal hyperplane to maximize the distance between different classes by establishing a decision boundary that effectively separates them; handle non-linear classification challenges using kernel functions that project input data into higher-dimensional spaces where linear separation becomes feasible; offer flexibility with various kernel functions such as linear, polynomial, and radial basis functions, allowing customization based on different data types; and, despite their

robustness and effectiveness, require significant computational resources, particularly with large datasets, and demand precise parameter tuning, including the regularization parameter and kernel choice, making them more complex to utilize compared to simpler models [27,28].

2.3.4. Random Forest

Random forest, an ensemble learning technique, enhances the accuracy and stability of decision trees by constructing multiple trees during training and aggregating their outputs through majority voting for classification or averaging for regression. Each tree is built from different bootstrap samples and uses random feature subsets for optimal splits, making the model robust and capable of handling large feature sets while reducing overfitting and sensitivity to noise. However, the complexity and computational demands of random forests, particularly with large datasets, can limit interpretability and application [29,30].

2.3.5. K-Nearest Neighbors

The K-nearest neighbor (KNN) algorithm is a straightforward, non-parametric classification method that predicts the class of a data point based on the majority vote of its KNN in the feature space. While KNN is easy to understand and implement—requiring no separate training phase and making predictions directly from stored data—its computational demands can be high due to the need to calculate the distances to all training points for each prediction. This can be particularly problematic with large datasets. Additionally, KNN's performance can be negatively affected by noise and irrelevant features, making feature scaling and the selection of an appropriate k value critical for optimal results [31,32].

2.3.6. Naive Bayes

Naive Bayes, a probabilistic classifier based on Bayes' theorem, assumes independence between predictors and, despite its simplicity, is highly efficient and scalable, making it suitable for large datasets. It performs well in practice, particularly in text classification, by handling both binary and multiclass problems and making probabilistic predictions. However, its strong independence assumption, often violated in real-world scenarios, can limit performance when features are highly correlated [33,34].

2.3.7. Neural Network

Neural networks, inspired by the human brain, consist of interconnected layers of nodes (neurons) that can learn complex data representations, making them highly effective for tasks such as image and speech recognition. Despite their flexibility and applicability to a wide range of problems, including regression and classification, they demand significant computational resources and expertise in tuning parameters like layer count, neuron count, and learning rates. Although neural networks achieve state-of-the-art performance in various challenging tasks in image and text classification [35,36], they are often criticized for their black-box nature, which makes their decision-making process difficult to interpret and understand, potentially limiting their adoption in certain fields [37,38].

2.4. Data Acquisition and Pre-Processing

Data acquisition and pre-processing are essential steps in preparing datasets for analysis and modeling, particularly in machine learning and data mining, where pre-processing procedures enhance the dataset's compatibility with machine-learning algorithms [39]. Data can be available in various forms: structured tables, unstructured tables, images, audio files, and videos. The first step for building an effective model is data pre-processing, which takes up 50 to 80% of the entire classification process, proving the importance of pre-processing in building a model [40]. Data pre-processing involves preparing and transforming raw data into a suitable format for data mining, aiming to reduce data size, identify relationships, normalize values, remove outliers, and extract relevant features through techniques such as data cleaning, integration, transformation, and reduction [41].

Data cleaning involves identifying and correcting or removing erroneous, incomplete, or inaccurate data from a database, and data cleaning algorithms are designed to detect these issues and enhance data quality by rectifying detected errors and omissions [42]. Data integration entails combining data from various databases, data cubes, or files. However, this process can be complicated by inconsistencies and redundancies, especially when attributes representing the same concept are labeled differently across databases, requiring more extensive data cleaning at this stage [43]. Data reduction involves creating a smaller, more manageable representation of a dataset that preserves essential features and produces analytical results comparable to the original, with various strategies available to achieve this effectively [44].

Thermal Images and I-V Curves Pre-Processing

For PV system fault detection, data pre-processing is especially critical. Thermal image pre-processing involves essential steps such as image normalization to adjust intensity values across different lighting and environmental conditions [45], noise reduction through Gaussian filters to eliminate sensor noise [46], and segmentation techniques like edge detection to isolate areas affected by faults such as hotspots, cracks, or shading [47].

Similarly, I-V curve pre-processing focuses on the feature extraction of parameters like short-circuit current (I_{sc}), open-circuit voltage (V_{oc}), and maximum power point (MPP), which are critical for identifying issues like shading, while curve rescaling standardizes data for better machine-learning compatibility, and adding Gaussian noise enhances robustness by simulating real-world variations, ensuring the model handles diverse data with greater precision [48].

2.5. Performance Metrics in AI Techniques

Various performance evaluation methods have been employed to assess the accuracy of the developed classification models. These methods often include analysis based on the confusion matrix, probabilistic interpretation of errors, and the model's discriminatory power [49]. Metrics derived from the confusion matrix, such as accuracy, precision recall, and the F-measure, are widely used due to their intuitive nature and are considered collectively in a balanced manner, particularly in imbalanced data classification, to ensure comprehensive model evaluation [50].

2.5.1. Accuracy

Accuracy is one of the most commonly used and easily understood metrics. It quantifies the proportion of correctly classified instances out of the total number of instances. Despite its widespread use, accuracy can sometimes be misleading, particularly in scenarios involving imbalanced datasets where one class significantly outnumbers the others [51].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

where true positives (TP) are instances correctly identified as positive, true negatives (TN) are instances correctly identified as negative, false positives (FP) occur when negative instances are incorrectly predicted as positive, and false negatives (FN) are when positive instances are incorrectly predicted as negative.

2.5.2. Precision

Precision, also referred to as positive predictive value (PPV), measures the proportion of true positive results among all positive predictions made by the model. This metric is particularly useful in contexts where the cost of false positives is high, making it a critical component of model evaluation [52].

$$\text{Precision} = \frac{TP}{TP + FP}$$

2.5.3. Recall (True Positive Rate or Sensitivity)

Recall, also known as sensitivity or true positive rate (TPR), indicates the proportion of actual positive cases that the model correctly identifies. This metric is especially important when the consequences of missing a positive case (false negative) are severe [53].

$$\text{Recall} = \frac{TP}{TP + FN}$$

2.5.4. F-Measure (F1 Score)

The F-measure, or F1 score, serves as the harmonic mean of precision and recall, balancing these two metrics to provide a single performance measure. It is particularly advantageous in situations where the dataset is imbalanced, as it takes into account both false positives and false negatives, offering a more nuanced view of the model's effectiveness [53].

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

For showing the crucial role of performance metrics in AI model evaluation, the next case study focuses on a trained model for PV system fault detection, as detailed in [54]. This model achieves a precision of 92%, effectively minimizing false alarms by accurately distinguishing between anomalies and non-anomalies. It maintains a recall of 93% for anomalies, ensuring most actual faults are promptly identified to prevent system failures. With an F-measure of 92%, the model adeptly balances precision and recall, proficiently handling both false positives and false negatives. An overall accuracy of 92% confirms its reliable and precise fault detection capabilities.

3. Methodology

In this study, we established criteria to select relevant literature for a comprehensive analysis of AI models and tools used in detecting defects and faults in PV systems (Table 1). We focused on research highlighting AI-driven methodologies, including ML, DL, image-based approaches, and other techniques specifically targeting fault detection and diagnosis. Emphasis was placed on studies addressing fault classification advancements, particularly those dealing with micro-cracks, hotspot formation, and module mismatching.

To ensure the review reflected current advancements, we selected papers published in the last five years, prioritizing those using thermal images and I-V curve data. Papers were sourced from reputable academic databases, with a focus on peer-reviewed articles, high citation counts, and journals with strong impact factors, and non-English publications were excluded for consistency.

Table 1. Summary of selection criteria.

Criterion	Details
Focus	AI-based fault detection in PV systems
AI Techniques	ML, DL, image-based, and other AI methods
Key Focus	Fault classification (e.g., micro-cracks, hotspots)
Publication Date	Last 5 years
Data Types	Thermal images, I-V curves
Source and Type	Peer-reviewed journals
Language	English only

4. Survey of AI Models and Tools on Detection of Defects and Faults in Photovoltaic Systems

Significant efforts have been undertaken to harness the power of machine learning to identify various types of defect in solar PV modules. Several studies have utilized different types of algorithms using images, electrical, and environmental data for simulated solar PV systems or existing systems.

The author presented in [55] an innovative algorithm that leverages logistic regression combined with cross-validation techniques. This approach is specifically designed for the early detection and precise identification of faults within the DC components of solar PV modules at string levels. Impressively, the algorithm achieves a 97.11% accuracy rate in identifying the three primary categories of faults: open-circuit faults, short-circuit faults, both permanent and temporary mismatch faults, and undefined faults. This work was done by simulating the solar PV cell circuit using a dataset that has temperature and irradiance values ranging from 1.32 to 35.06 °C and 0.04 to 984.84 W/m², and then looking at the I-V curves of the solar PV cell to find out what kind of faults there are. The multi-class classification was done using either one-vs-one (OVO) or one-vs-rest (OVR) methods.

Another study [56] presents an effective method for detecting DC arc faults in PV systems, utilizing logistic regression as the core analytical tool. The researchers designed their method to operate under conditions that closely resemble those found in real-world PV systems, incorporating copper electrodes with adjustable gap distances. By systematically analyzing the relationship between key electrical variables and the occurrence of arc faults, they were able to achieve an impressive 100% accuracy rate across 80 experimental trials. Moreover, the study provides valuable insights into the behavior of electrical currents during arc faults. The researchers observed distinctive changes in the current waveform, including a noticeable drop in DC current, the emergence of high-frequency impulses, and a “thicker” waveform. These anomalies are associated with specific frequency components, particularly at 100 Hz and 40 kHz, which are critical markers for identifying such faults.

In [45], a fault classification method for PV cells is introduced, using SVMs in combination with thermal imaging. The research focuses on detecting various types of fault in solar PV systems, including cracks, hotspots, soiling, and internal failures. The method is designed to accurately differentiate between defective and non-defective PV cells, achieving an impressive 97% accuracy. The study was carried out on a 5 kWp PV plant comprising twenty monocrystalline PV panels, each with a capacity of 250 Wp, as well as a smaller PV array configured in a 3 × 3 series-parallel connection of solar cells. The procedure involved capturing thermal images of both healthy and faulty PV panels. These images were then processed using MATLAB to extract relevant features, which were analyzed and classified using the SVM model.

The authors in [57] have developed an SVM model utilizing infrared thermography to enhance the detection and classification of hotspots in PV panels. By combining features such as RGB, texture, histogram of oriented gradient (HOG), and local binary pattern (LBP) into a hybrid feature vector, the model effectively categorizes thermal images into healthy, non-faulty hotspots, and faulty modules. This approach achieved notable accuracy, with 96.8% in training and 92% in testing, while also being computationally efficient. The research was based on a dataset from a 42.24-kWp solar PV installation in Lahore, Pakistan, and focused on using thermal imaging to identify degradation and faults in PV systems.

In [58], two methods were developed for detecting solar panels in thermal images captured by unmanned aerial vehicles (UAVs). The first method relies on classical image processing techniques, including image correction, segmentation, and classification using SVMs with optimized texture descriptors, followed by a post-processing step to locate any undetected panels. The second method leverages deep learning, specifically a Mask R-CNN network trained on pre-processed images with data augmentation, and also includes post-processing. The classical method achieved high precision at 99.7%, recall at 97.0%, and an F1 score of 98.3%. The deep-learning approach slightly surpassed this, with a precision of 99.6%,

recall of 98.1%, and an F1 score of 98.9%. Additionally, the deep-learning method reduced false positives by over 60%, demonstrating superior efficiency in panel detection.

Another study [59] employing an SVM model focuses on detecting micro-cracks and hotspots in solar PV modules. This research introduces a technique to estimate the percentage of power loss (PPL) resulting from hotspots in PV modules, using six key input parameters—PPL, Voc, Isc, and irradiance, among others—to accurately identify faults. The data utilized in this study were gathered from polycrystalline silicon (Poly-Si) PV panels, each with a capacity of 10 W. The SVM model is implemented using a multi-class binary tree architecture, organized into four stages for fault classification. The results demonstrate that this SVM-based approach is highly effective, achieving an impressive average accuracy of 99% in detecting and classifying micro-cracks and hotspots in PV modules.

Thermal images were utilized and analyzed in [60] through an SVM to classify three types of defect in PV systems: single hotspot, multiple hotspots, and string defects. The process involved image segmentation, where binary masks were employed using image analysis to identify and categorize the hotspots into these classes. The dataset used in this research was collected from a thermographic inspection of a 66 MW ground-based PV plant in Tombourke, South Africa, and consisted of 3336 images, including 1007 images of damaged PV cells. The dataset was divided into 70% training and 30% testing sets to evaluate the method's performance. The proposed SVM-based method achieved an accuracy of 72.54% in classifying these defects.

In [61], the authors developed machine-learning models for diagnosing faults in PV systems using I-V curve measurements, with decision tree being one of the key models explored. The dataset, generated at RElab, Jijel University, in Algeria, involved a small 960 W PV array and was designed to classify five specific types of defect. These defects included partial shading combined with degraded PV modules, dust accumulation with short-circuited diodes, partial shading of one PV module coupled with dust accumulation, an open circuit with partial shading, and line-to-line faults between two PV modules with degradation. The dataset consisted of 575 samples, and key features such as t_{Isc} , Voc, maximum power current (I_{mp}), maximum power voltage (V_{mp}), maximum power point (MPP), and fill factor (FF) were analyzed to classify these defects. The decision-tree model achieved an accuracy of 73.91%, whereas the artificial neural network (ANN)-based model outperformed other machine-learning classifiers tested in the study.

An advanced technique for monitoring the condition of PV panels and diagnosing faults was developed in [62] using a combination of a U-Net neural network and a decision-tree classifier. This approach intelligently analyzes the infrared thermal images of PV panels captured by drones or other remote operating systems. The study focused on detecting three specific types of defect: safety-glass cracks, defects in PV power units, and contamination of the safety glass. To enhance the reliability of the condition monitoring and fault diagnosis process, the method incorporated the use of true-color image masks. The decision-tree classifier, known for its rule-based classification, demonstrated higher accuracy compared to KNN and SVM. The study was conducted using a dataset of 295 samples collected over a short period of time. The combined use of the trained U-Net neural network and the decision-tree classifier proved highly effective, achieving an impressive fault detection accuracy of 99.8%.

In another study [63], researchers developed a decision-tree-based fault diagnostic method for solar PV systems, utilizing I-V curve measurements to detect faults such as short circuits, abnormal aging, partial shading, and hybrid faults. The methodology involves three key steps: extracting optimal fault features from the I-V curves, standardizing these features to standard test conditions (STC) using a combination of trust-region-reflective (TRR) deterministic and particle-swarm-optimization (PSO) algorithms, and establishing a fault diagnostic model using a multi-class adaptive boosting (AdaBoost) algorithm with the SAMME-CART algorithm. This process also includes a parameter normalization technique involving a nonlinear PSO-TRR least-squares method. The proposed method demonstrated exceptional accuracy, achieving a fault classification accuracy of over 99.70%.

High impedance faults (HIF) were detected through an in-depth analysis in [64], where various faulty and non-faulty events were introduced to evaluate the performance of different classifiers. The research involved simulating an IEEE 13-bus system integrated with a solar PV network using MATLAB/Simulink, under various fault conditions. The methodology centered around feature extraction using discrete wavelet transform (DWT) analysis, followed by the training and testing of several intelligent classifiers, including long short-term memory (LSTM) networks, KNN, SVM, J48-based decision trees, and naive Bayes. The classifiers were rigorously evaluated using metrics such as kappa statistic, precision, recall, and F-measure. Among these, the LSTM classifier demonstrated superior performance, outperforming others, including the decision tree, particularly in terms of classification accuracy, specific detection of HIF events, and faster detection times.

A study conducted in [65] applied the random forest algorithm to detect faults and partial shading in PV systems. In this research, the array's current, voltage, and irradiance were measured to detect and classify various faults, including those that often go undetected under low irradiance, such as line-to-line faults. The study involved creating a comprehensive dataset by extracting features under both normal and partial shading conditions. These features were then analyzed using the random forest algorithm, known for its ability to handle complex and varied datasets effectively. The accuracy of the classification was assessed using a confusion matrix, a standard method for evaluating the performance of classification models. The experimental results, derived from a 100 W PV module with a 4×4 array configuration, showed an impressive fault classification accuracy of 99.98%.

The random forest classifier was combined with a modified grey-wolf optimization (MGWO) algorithm in [66] to enhance fault detection in solar PV systems. The methodology involved two key steps: first, extracting the five critical parameters of the one-diode model by transforming three arbitrary I-V curves into a reference curve using a current-voltage translation method and MGWO algorithm; second, simulating the PV array to determine MPP coordinates and build operational databases through co-simulations in PSIM/MATLAB. The RFC was then used to detect anomalies, achieving an impressive accuracy rate of 99.4%.

In [67], a compound fault diagnosis method for PV systems is proposed, focusing on detecting faults such as short-circuit, open-circuit, partial shading, and degradation. The approach involves pre-processing I-V curves by smoothing and normalizing voltage and current, followed by feature extraction. Two novel multi-label classification models are used: k-nearest neighbor combined with random forest (ML-RFKNN) and a simple residual network (ML-SResNet). The results demonstrate that ML-RFKNN and ML-SResNet achieve high diagnosis accuracy, with exact match ratios of 99.17% and 97.38%, respectively, outperforming other methods.

The random forest algorithm was utilized in [68] to develop a decision-tree model for analyzing PV panel operating data. The model processes data such as PV array current, output power, and temperature readings to detect five types of fault: open circuit, short circuit, hotspot phenomena, abnormal aging, and partial shading. The dataset comprised 5395 normal and 1527 abnormal operating records. Using two-thirds of the data for training and one-third for testing, the model demonstrated high accuracy in predicting the causes of PV panel failures, effectively identifying faults with significant precision.

A novel fault detection approach combining random forest and modified-independent component analysis was proposed in [69] to identify various PV system faults, including inverter faults, voltage sags, partial shading, and open circuits. The research utilized a publicly available dataset containing seven PV fault types alongside normal conditions, with 14 features such as time, PV array current, and DC voltage. The study employed a feature importance tree to prioritize these features, selecting the top seven. To address data imbalance, two strategies were implemented: SMOTE for oversampling and random undersampling, achieving accuracy rates of 99.88% and 99.43% for the two scenarios, respectively.

A machine-learning-based technique in [70] was proposed for diagnosing hotspots in solar PV modules using thermal images and a naive Bayes classifier. The study focuses

on a 42.24-kWp PV system, capturing thermal images with a handheld camera. Texture features are extracted from these images using the gray-level co-occurrence matrix (GLCM) and HOG. These features are combined to train a naive Bayes-based classifier, which categorizes PV modules into three groups: defective, non-defective with hotspots (NDH), and non-defective without hotspots (NDNH). The method achieves a mean recognition rate of approximately 94.1%

The authors in [71] developed a DL model to detect anomalous cells in images captured by UAVs equipped with thermal infrared sensors. They utilized a mask region-based convolutional neural network (Mask R-CNN) for object detection and instance segmentation, comparing three deep neural networks: UNet, FPNNet, and LinkNet. The dataset, collected from a 66 MW solar PV plant over seven days, was categorized into single anomalous cells, multiple anomalous cells, and contiguous series of anomalous cells, with temperatures ranging from 2.249 to 103.335 °C. UAVs scanned the PV system, and the captured frames were annotated and stored. The region-based convolutional neural network (RCNN) was trained on part of this dataset, with the remaining data used for testing. Performance was evaluated using metrics like the Dice coefficient and intersection over union (IoU). The study found the model effective and suitable for diagnosing thermal images of PV plants, indicating its potential to enhance solar power system maintenance and monitoring.

In another study [72], the authors present a methodology for detecting hotspots on solar PV panels using RCNN. The proposed model comprises three stages. Initially, an image database is created with labeled regions of interest (ROIs) for panels and hotspots to train the R-CNN detectors. Following this, thermographic images and telemetry data are captured, prepared, and processed to identify panels and hotspots. Finally, GPS data, telemetry data, and hotspot detections are integrated into a comprehensive report. The results indicate that the proposed model is highly effective, achieving a detection accuracy of 99.02% and a precision of 91.67% in identifying and localizing hotspots on solar PV panels.

In [73], physics-based simulations of string-level I-V curves were used to detect three statuses of solar PV modules: no faults, partial soiling, and cell crack system modes. A one-dimensional (1D) convolutional neural network (CNN) was employed to explore the relationship between various features, including I-V, power, finite difference, and current difference curve geometries. The findings demonstrated a perfect classification accuracy of 100% when simulated curves were used to classify measured curves from the test split.

In [74], the researchers developed an isolated convolutional neural model (ICNM) to classify thermal images of solar PV modules. The images were categorized into three main groups: healthy, hotspot, and faulty. Additionally, hotspot and faulty images were further divided into five defect sub-classes: bird drops, single, block, patchwork, and HA string. A three-class ICNM was constructed, trained, and validated using transfer learning with pre-trained networks such as ShuffleNet, GoogLeNet, and SqueezeNet. This method capitalized on the ICNM's rapid response time, straightforward architecture, and high accuracy to effectively classify the five defect types in PV panels. Initially, the ICNM achieved an accuracy of 96% in classifying the solar PV panels. Applying the transfer learning approach further improved the accuracy to 97.62%.

In their study, the authors in [75] developed two CNN models to classify defects in solar PV panels and identify the regions of interest in faulty panels. Thermal images were first processed and then analyzed using a deep transfer learning CNN model to identify the types of defect. Subsequently, the locations of hot spots on the panels were determined using the Faster R-CNN technique. The defects were categorized into five classes: single cell hotspot, multi cell hotspot, diode fault, PID defect, and dust and shadow hotspot. The ResNet-50 model achieved the highest accuracy rate for defect classification at 85%, while the KNN model had the lowest accuracy at 48%. For object detection, the Faster R-CNN technique proved to be the most effective, with a mean average precision (mAP) of 0.67.

I-V curves were utilized in [76] by developing a methodology that enhances fault detection in PV systems. This approach incorporated a thorough pre-processing phase,

involving feature extraction following the correction and resampling of I-V curves using advanced techniques such as Gramian angular difference fields and recurrence plots. The dataset comprised 12,000 simulated I-V curves, designed to replicate various fault conditions, with additional validation performed using real-world field data. The study evaluated six machine-learning models—ANN, SVM, decision tree, random forest, KNN, and naive Bayes classifier—to classify eight distinct operational conditions. These conditions included normal operation, two types of partial shading, two types of short circuits, open circuit faults, and degradation in series and shunt resistances. Among these models, the ANN emerged as the most effective, achieving a perfect classification accuracy of 100% for both the simulated dataset and the field data.

A CNN framework was proposed in [54] to identify and classify 12 operational conditions of PV systems based on thermal images. The dataset utilized in the research consists of 10,000 infrared (IR) images, each representing different fault categories. Notably, some classes within the dataset had fewer samples than others, prompting the use of data augmentation techniques, such as vertical and horizontal flips, to balance the dataset while maintaining the defect patterns. The CNN framework extracts critical features like edges and shapes for each fault category through its convolutional layers. The proposed methodology comprises four main steps: defect detection, defect localization and classification, measurement of defect extent, and the prediction of the solar PV module's lifetime. The study's results indicate that the CNN framework achieved a testing accuracy of 92.5% for anomaly detection and 78.85% for defect classification.

The authors of [77] present an in-depth analysis of the classification of field-collected string-level I-V curves, specifically focusing on baseline conditions, partial soiling, and cracked failure modes. The study evaluates various neural network-based architectures by employing domain-specific parameters across different sections of the I-V curve and varying irradiance thresholds, despite the limited dataset of approximately 400 samples. The findings indicate that both multi-headed long short-term memory (LSTM) networks and one-dimensional CNNs achieve high classification accuracies, consistently exceeding 99%. Notably, 1D CNNs slightly outperform the multi-headed LSTMs, demonstrating their effectiveness in analyzing and classifying I-V curves in PV systems under different failure modes.

4.1. Comparative Analysis of AI Models for Defect Detection in Photovoltaic Systems

This section provides an in-depth overview of the seven primary AI models commonly used for detecting defects and faults in PV systems. It evaluates their strengths, limitations, and suitability for different PV systems.

4.1.1. Neural Networks

Neural networks are highly effective for detecting faults in PV systems, handling complex patterns in large datasets like thermal images and I-V curves. While offering high accuracy, they require significant computational resources and are less interpretable, making them ideal for large-scale, complex applications.

Advantages:

- High accuracy in complex, large datasets.
- Excellent for image-based fault detection.
- Effective for detailed analysis of PV systems.

Limitations:

- High computational requirements.
- Less interpretable than other models.

Applicability: Best suited for large-scale PV systems where detailed and complex fault detection is needed.

4.1.2. Support Vector Machines

SVM models have demonstrated strong performance in detecting faults using thermal images and electrical data in PV systems. They are particularly effective in classifying multiple types of defect, such as cracks and hotspots. SVM models are efficient with smaller datasets but require careful parameter tuning.

Advantages:

- High accuracy in various PV fault types.
- Works well in smaller datasets.

Limitations:

- Performance can be affected by noisy data.
- Requires careful tuning of parameters.

Applicability: Suitable for PV systems where binary or multi-class classification is required, particularly for detecting specific defects like hotspots.

4.1.3. Random Forest

Random forest is known for its robustness and ability to handle complex datasets with varied features. It performs well in analyzing data from current, voltage, and irradiance measurements and is highly effective for detecting multiple types of fault, such as shading, short circuits, and system degradation.

Advantages:

- Handles complex datasets effectively.
- High accuracy in diverse fault types.

Limitations:

- Computationally intensive for very large datasets.
- Less interpretable due to its ensemble nature.

Applicability: Best suited for large PV systems that generate a variety of data types, such as electrical and environmental data.

4.1.4. Decision Trees

Decision tree models are highly interpretable and are effective for classifying faults based on specific features, such as I-V curve data. While they may not reach the accuracy of ensemble models like random forest, they offer reliable fault detection when used in simpler scenarios.

Advantages:

- High interpretability.
- Effective in analyzing I-V curves.

Limitations:

- Can be prone to overfitting with complex data.
- Lower accuracy compared to ensemble models like random forest.

Applicability: Useful for PV systems where interpretability is crucial, particularly in fault diagnostics involving I-V curves.

4.1.5. Logistic Regression

Logistic regression has been successfully used for fault detection, especially in smaller datasets. It is effective in early fault detection, particularly in DC component analysis in PV systems. While being simple and efficient, logistic regression is limited to linear relationships and struggles with more complex fault patterns.

Advantages:

- Simple and efficient for small datasets.
- High accuracy in early fault detection.

Limitations:

- Limited to linear relationships.
- Less effective for complex fault patterns.

Applicability: Ideal for small-scale PV systems or specific components where early fault detection is required.

4.1.6. K-Nearest Neighbors

KNN is a simple model that performs well when combined with other advanced techniques. It is often used in smaller datasets for detecting faults like shading or short circuits but generally lags behind in terms of accuracy and scalability compared to more advanced models.

Advantages:

- Easy to implement.
- Effective in combination with other models.

Limitations:

- Lower accuracy in large or complex datasets.
- Computationally expensive for large datasets.

Applicability: Best suited for smaller PV systems or when combined with more advanced models for better accuracy.

4.1.7. Naive Bayes

Naive Bayes is fast and computationally efficient, making it suitable for quick classification tasks. It has been applied effectively in detecting faults using thermal images, but its assumption of feature independence limits its accuracy in more complex datasets.

Advantages:

- Fast and computationally efficient.
- Effective for quick fault classification.

Limitations:

- Assumes feature independence, which can reduce accuracy.
- Less effective for complex datasets.

Applicability: Suitable for fast, low-complexity fault detection in smaller PV systems.

5. Comparison of AI Models on Detection of Defects and Faults in Photovoltaic Systems

Different AI models were reviewed thoroughly in the previous section, focusing on their application in fault detection and diagnosis in PV systems. The following Table 2 summarizes the key findings, ranking and comparing the effectiveness of various AI methods when applied to thermal images and I-V curves.

Table 2. Ranked AI methods for solar PV defect detection with defect types.

Rank	AI Method	Thermal Images	I-V Curves	Defects Types
1	Neural Networks	Highly effective in detecting complex patterns and defects in thermal images.	Capable of learning complex relationships in I-V curves.	Hotspots, cracks, shading, soiling, panel degradation.
2	Support Vector Machines	Effective in high-dimensional feature spaces; good for distinguishing clear boundaries.	Works well for distinguishing between normal and defective curves with clear margins.	Cracks, hotspots, micro-cracks, shading, power loss estimation.
3	Random Forest	Robust against overfitting, effective with extracted features from thermal images.	Suitable for feature-based classification tasks; handles anomalies in I-V curves well.	Hotspots, shading, partial shading, short circuits, degradation.
4	Decision Trees	Easy to interpret; useful when features are clearly defined, but less effective alone.	Good for simple decision-making tasks; requires pruning to avoid overfitting.	Shading, dust accumulation, diode faults, line-to-line faults.

Table 2. Cont.

Rank	AI Method	Thermal Images	I-V Curves	Defects Types
5	Logistic Regression	Less commonly used with thermal images; primarily focused on analyzing electrical characteristics rather than thermal data.	Works mainly for binary and multi classification in simple cases; less effective with complex patterns.	Open-circuit faults, short-circuit faults, mismatch faults, undefined faults.
6	K-Nearest Neighbors	Applicable in simpler thermal image classification tasks but less favored for high-dimensional, complex data.	Less commonly used for I-V curve data; applicable for simple fault classification tasks but less effective for complex scenarios.	Simple classification of hotspots, shading, and other visible defects.
7	Naive Bayes	Less commonly used due to the assumption of feature independence, which often does not hold in complex thermal images.	Least commonly used for I-V curves; suffers from assumptions that do not typically apply to I-V curve data.	Hotspots detection, primarily less effective with complex faults.

6. Discussion

The integration of AI classification techniques for defect detection in solar PV modules has seen significant advancements, primarily through the application of methods such as neural networks, SVM, random forest, decision trees, logistic regression, KNN, and naive Bayes.

6.1. Neural Networks

Neural networks, particularly deep-learning models, have emerged as the most popular and effective AI method for detecting defects in solar PV modules, both through thermal imaging and I-V curve analysis. Studies have shown that neural networks achieve near-perfect classification accuracy, especially in identifying complex faults like hotspots, cracks, shading, and soiling. For instance, the use of a Mask R-CNN network trained on pre-processed thermal images yielded an F1 score of 98.9%, demonstrating the model's robustness in handling diverse defect types. Similarly, studies employing CNNs for analyzing I-V curves reported 100% classification accuracy, further underscoring the versatility and effectiveness of neural networks in PV defect detection.

The key advantage of neural networks lies in their ability to learn complex patterns from large datasets without extensive feature engineering. This makes them particularly suitable for real-time monitoring and automated fault detection in large-scale solar installations. However, their performance is heavily dependent on the availability of large labeled datasets and substantial computational resources for training, which can be a limiting factor in some cases.

6.2. Support Vector Machines

SVMs are widely utilized for defect detection in solar PV modules, particularly for identifying cracks, hotspots, micro-cracks, and other internal failures. SVM models have been effectively applied to both thermal images and I-V curve data, achieving high accuracy rates—often above 97%—in various studies. For example, a study leveraging SVM with thermal imaging to classify defects in PV cells achieved a remarkable 99% accuracy, while another using SVM for I-V curve analysis reported similar success in fault classification.

SVM's effectiveness stems from its ability to handle high-dimensional spaces and create clear decision boundaries between different classes. This makes it an excellent choice for detecting subtle defects that might not be easily distinguishable using simpler models. Nevertheless, SVM requires careful tuning of parameters and extensive feature engineering, which can be a challenge in dynamic and unstructured environments.

6.3. Random Forest and Decision Trees

Random forest and decision trees are both effective for defect detection in solar PV modules, with random forest generally outperforming decision trees due to its ensemble nature, which reduces overfitting and enhances generalization. Random forest has been applied successfully to both thermal images and I-V curve data, achieving high accuracy in defect classification, including faults such as shading, partial shading, short circuits, and degradation. Studies have reported accuracy rates as high as 99.98% with random forest, making it a reliable method for various fault detection tasks.

Decision trees, on the other hand, are valued for their interpretability and ease of implementation. They have been particularly useful when combined with other models, such as neural networks, to enhance detection accuracy. However, they tend to overfit when used alone, especially with noisy or unbalanced data, which limits their standalone applicability.

6.4. Logistic Regression

Logistic regression has shown promising results, particularly in the early detection of faults in DC components of solar PV modules. The use of logistic regression combined with cross-validation techniques has achieved high accuracy rates (e.g., 97.11%) in identifying different types of fault, such as open-circuit faults, short-circuit faults, mismatch faults, and undefined faults. This method is especially effective when the relationship between the input features and the output labels is linear or near-linear.

However, logistic regression is less commonly used with thermal images, as it primarily focuses on analyzing electrical characteristics rather than visual data. Its application is more suited to scenarios where a binary or multi-class classification of faults is required, making it less versatile than other AI methods in handling complex image data.

6.5. K-Nearest Neighbors and Naive Bayes

KNN and naive Bayes are less popular for defect detection in solar PV modules due to their limitations in handling complex, high-dimensional data. KNN is mainly applicable in simpler classification tasks involving thermal images, where computational efficiency and simplicity are more important than model sophistication. However, it is less favored for analyzing I-V curve data or more complex scenarios due to its sensitivity to the choice of K and the curse of dimensionality.

Naive Bayes, on the other hand, is rarely used in solar PV defect detection, as its assumption of feature independence often does not hold in complex datasets like thermal images or I-V curves. While it can be useful in specific cases with well-defined, simple features, its application is largely limited compared to more robust methods like neural networks or SVM.

Overall, neural networks and SVMs are the most popular and effective AI methods for detecting defects in solar PV modules, reflecting a significant focus by researchers. These methods offer high accuracy and adaptability to various fault types and conditions, with a particular advantage for neural networks due to their effectiveness in detecting complex patterns and their ability to utilize pre-trained models through transfer learning rather than developing models from scratch. Random forest and decision trees also make valuable contributions, especially when integrated with other methods. While logistic regression is effective in specific contexts, it is generally less versatile. Techniques like KNN and naive Bayes are less suitable for complex defect detection tasks.

Data plays a crucial role in AI-based solar PV defect detection, where increasing dataset size generally enhances model performance, though the benefits tend to diminish with larger datasets. Even smaller datasets (e.g., 800 instances) can provide comparable defect coverage to larger ones, making them useful for testing, with defect coverage serving as a reliable evaluation metric [78]. To ensure validation precision, 75–100 test samples are typically required, and comparing classifiers may need hundreds of independent samples or may be theoretically impossible [79].

To efficiently run AI-based PV monitoring systems, the required hardware configuration can vary depending on the complexity of the AI models being used. In general, systems should have sufficient random access memory (RAM), processing power from a central processing unit (CPU), and sometimes a dedicated graphics processing unit (GPU) for tasks like image processing. Neural networks, especially those involved in analyzing thermal images and I-V curves, usually demand more advanced hardware. Basic setups may begin with 16 GB of RAM, a multi-core processor, and solid-state drive (SSD) storage, but more powerful systems are often necessary for larger datasets and complex models, as indicated by previous studies on AI hardware requirements.

The high accuracy of over 90% achieved by these models in detecting PV defects can be attributed to several factors. These include the use of high-quality data, effective feature engineering, pre-processing techniques like data augmentation, and the complexity of models such as deep-learning architectures, which capture non-linear relationships. Additionally, combining models through hybrid approaches, rigorous cross-validation, and hyperparameter tuning also significantly enhance performance.

Nowadays, these essential AI techniques greatly enhance PV systems' performance and lifespan. By detecting and classifying subtle defects missed by traditional methods, AI enables prioritized maintenance, preventing minor issues from becoming major failures. Early detection minimizes downtime and costs, while predictive analytics enable proactive maintenance, ensuring peak efficiency, reliability, and maximizing return on investment.

Future research should explore hybrid AI approaches that combine machine learning, deep learning, and reinforcement learning to enhance detection accuracy and efficiency in solar PV systems. Integrating data from thermal imaging, drones, and IoT sensors can provide comprehensive system assessments. Developing explainable AI models increases transparency and trust, while transfer learning can improve performance with limited labeled data.

7. Conclusions

In this study, we have explored the current landscape of AI-driven fault detection and diagnosis techniques in PV systems, identifying the latest trends and the most advanced methodologies for detecting faults based on thermal images and I-V curve analysis, and ranking these detection techniques based on their applicability, strengths, and limitations, highlighting the significant potential of AI in enhancing the efficiency and accuracy of PV fault management. However, challenges such as the scarcity of publicly available datasets and the complexity of existing methods indicate there is still considerable room for improvement.

Our findings suggest that different AI models, including neural networks, SVMs, random forest, decision trees, logistic regression, KNN, and naive Bayes, each offer unique strengths for fault detection, but no single model fully addresses all the challenges faced by PV systems. For instance, neural networks and random forest models excel in large, complex datasets, while simpler models like logistic regression and naive Bayes are more suitable for small-scale or less complex systems. The limitations of these individual approaches point to the need for more integrated solutions.

As a step forward, we propose the implementation of a hybrid approach that combines the strengths of multiple AI techniques and defect diagnosis methods to overcome the limitations of individual methods. This hybrid model aims to enhance fault detection accuracy, reduce computational complexity, and improve the practicality of these techniques for large-scale applications. Future work will focus on experimentally validating this hybrid approach under real-world conditions, with an emphasis on optimizing the balance between accuracy, speed, and resource efficiency. By conducting extensive experimental trials, we aim to refine the approach and further contribute to the advancement of PV system fault detection and diagnosis.

Author Contributions: Conceptualization, Y.B.S. and A.T.; methodology, Y.B.S. and A.T.; software, Y.B.S. and A.T.; validation, A.T. and Y.B.S.; formal analysis, A.T.; investigation, A.T.; resources, Y.B.S.; data curation, Y.B.S.; writing—original draft preparation, Y.B.S. and A.T.; writing—review and

editing, A.T.; visualization, A.T.; supervision, Y.B.S. and A.T.; project administration, Y.B.S.; funding acquisition, Y.B.S. All authors have read and agreed to the published version of the manuscript.

Funding: The authors would like to express their profound gratitude to King Abdullah City for Atomic and Renewable Energy (K.A.CARE) for their financial support in accomplishing this work.

Data Availability Statement: All data are available within the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

PV	Photovoltaic
AI	Artificial Intelligence
ML	Machine learning
DL	Deep Learning
DC	Direct Current
AC	Alternating Current
BOS	Balance-of-System
SVMs	Support Vector Machines
KNN	K-Nearest Neighbors
I_{sc}	Short-circuit current
V_{oc}	Open-circuit voltage
P_{mp}	Maximum power
TP	True Positives
TN	True Negatives
FP	False Positives
FN	False Negatives
TRP	True Positive Rate
OVO	One-vs-One
OVR	One-vs-Rest
HOG	Histogram of Oriented Gradient
LBP	Local Binary Pattern
UAV	Unmanned Aerial Vehicles
PPL	Percentage of Power Loss
IMP	Maximum power current
VMP	Maximum power voltage
IMP	Maximum power point
FF	Fill Factor
ANN	Artificial Neural Network
STC	Standard test conditions
TRR	Trust-Region-Reflective
PSO	Particle-Swarm-Optimization
AdaBoost	Adaptive Boosting
HIF	High Impedance Faults
DWT	Discrete Wavelet Transform
LSTM	Long Short-Term Memory
MGWO	Modified Grey-Wolf Optimization
ML-RFKNN	K-Nearest Neighbor combined with Random Forest
ML-SResNet	Simple residual network
GLCM	Gray-Level Co-Occurrence Matrix
NDH	Non-Defective with Hotspots
NDNH	Non-Defective without Hotspots
Mask R-CNN	Mask Region-based Convolutional Neural Network
RCNN	Region-Based Convolutional Neural Network
IoU	Intersection over Union
ROIs	Regions of Interest

CNN	Convolutional Neural Network
ICNM	Isolated Convolutional Neural Model
IR	Infrared
RAM	Random Access Memory
CPU	Central Processing Unit
GPU	Graphics Processing Unit
SSD	Solid-State Drive

References

- International Renewable Energy Agency (IRENA). Artificial Intelligence and Big Data. 2019. Available online: <https://www.irena.org/publications/2019/Sep/Artificial-Intelligence-and-Big-Data> (accessed on 14 September 2024).
- Abubakar, A.; Almeida, C.F.M.; Gemignani, M. A review of solar photovoltaic system maintenance strategies. In Proceedings of the 2021 14th IEEE International Conference on Industry Applications (INDUSCON), Sao Paulo, Brazil, 15–18 August 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 1400–1407.
- Rendroyoko, I.; Rusli, M. Development of power quality control procedures and standards to control the connection of non-linear loads in electric power systems. In Proceedings of the 22nd International Conference and Exhibition on Electricity Distribution (CIRED 2013), Stockholm, Sweden, 10–13 June 2013.
- Choi, S.S.; Li, B.; Vilathgamuwa, D. Dynamic voltage restoration with minimum energy injection. *IEEE Trans. Power Syst.* **2000**, *15*, 51–57. [\[CrossRef\]](#)
- Sumathi, S.; Bansilal. Artificial neural network application for control of STATCOM in power systems for both voltage control mode and reactive power mode. In Proceedings of the 2009 International Conference on Power Systems, Kharagpur, India, 27–29 December 2009; IEEE: Piscataway, NJ, USA, 2009; pp. 1–4.
- Kamran, F.; Habetler, T.G. Combined deadbeat control of a series-parallel converter combination used as a universal power filter. *IEEE Trans. Power Electron.* **1998**, *13*, 160–168. [\[CrossRef\]](#)
- Mahmoud, M.; Ramadan, M.; Olabi, A.G.; Pullen, K.; Naher, S. A review of mechanical energy storage systems combined with wind and solar applications. *Energy Convers. Manag.* **2020**, *210*, 112670. [\[CrossRef\]](#)
- Firth, S.K.; Lomas, K.J.; Rees, S.J. A simple model of PV system performance and its use in fault detection. *Sol. Energy* **2010**, *84*, 624–635. [\[CrossRef\]](#)
- Selvaraj, T.; Rengaraj, R.; Venkatakrisnan, G.; Soundararajan, S.; Natarajan, K.; Balachandran, P.; David, P.; Selvarajan, S. Environmental fault diagnosis of solar panels using solar thermal images in multiple convolutional neural networks. *Int. Trans. Electr. Energy Syst.* **2022**, *2022*, 2872925. [\[CrossRef\]](#)
- Venkatakrisnan, G.; Rengaraj, R.; Tamilselvi, S.; Harshini, J.; Sahoo, A.; Saleel, C.A.; Abbas, M.; Cuce, E.; Jazlyn, C.; Shaik, S.; et al. Detection, location, and diagnosis of different faults in large solar PV system—A review. *Int. J. Low-Carbon Technol.* **2023**, *18*, 659–674. [\[CrossRef\]](#)
- Abdulmawjood, K.; Refaat, S.S.; Morsi, W.G. Detection and prediction of faults in photovoltaic arrays: A review. In Proceedings of the 2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018), Doha, Qatar, 10–12 April 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1–8.
- Hijawi, U.; Lakshminarayana, S.; Xu, T.; Fierro, G.P.M.; Rahman, M. A review of automated solar photovoltaic defect detection systems: Approaches, challenges, and future orientations. *Sol. Energy* **2023**, *266*, 112186. [\[CrossRef\]](#)
- Buchanan, B.G. A (very) brief history of artificial intelligence. *Ai Mag.* **2005**, *26*, 53.
- Samuel, A.L. Some studies in machine learning using the game of checkers. *IBM J. Res. Dev.* **2000**, *44*, 206–226. [\[CrossRef\]](#)
- Kuang, P.; Cao, W.N.; Wu, Q. Preview on structures and algorithms of deep learning. In Proceedings of the 2014 11th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP), Chengdu, China, 19–21 December 2014; IEEE: Piscataway, NJ, USA, 2014; pp. 176–179.
- Qureshi, M.S.; Umar, S.; Nawaz, M.U. Machine Learning for Predictive Maintenance in Solar Farms. *Int. J. Adv. Eng. Technol. Innov.* **2024**, *1*, 27–49.
- Liu, W.; Shen, Y.; Aungkulanon, P.; Ghalandari, M.; Le, B.N.; Alvarez-Meza, A.; Cárdenas-Escrocia, Y. Machine learning applications for photovoltaic system optimization in zero green energy buildings. *Energy Rep.* **2023**, *9*, 2787–2796. [\[CrossRef\]](#)
- Scott, C.; Ahsan, M.; Albarbar, A. Machine learning for forecasting a photovoltaic (PV) generation system. *Energy* **2023**, *278*, 127807. [\[CrossRef\]](#)
- Meribout, M.; Tiwari, V.K.; Herrera, J.P.P.; Baobaid, A.N.M.A. Solar panel inspection techniques and prospects. *Measurement* **2023**, *209*, 112466. [\[CrossRef\]](#)
- Al Smadi, T.; Handam, A.; Gaeid, K.S.; Al-Smadi, A.; Al-Husban, Y.; Khalid, A. Artificial intelligent control of energy management PV system. *Results Control Optim.* **2024**, *14*, 100343. [\[CrossRef\]](#)
- Assareh, E.; Hoseinzadeh, S.; Agarwal, N.; Delpisheh, M.; Dezhdar, A.; Feyzi, M.; Wang, Q.; Garcia, D.A.; Gholamian, E.; Hosseinzadeh, M.; et al. A transient simulation for a novel solar-geothermal cogeneration system with a selection of heat transfer fluids using thermodynamics analysis and ANN intelligent (AI) modeling. *Appl. Therm. Eng.* **2023**, *231*, 120698. [\[CrossRef\]](#)
- Moore, A.W.; Zuev, D. Internet traffic classification using bayesian analysis techniques. In Proceedings of the 2005 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems, Banff, AB, Canada, 6–10 June 2005; pp. 50–60.

23. Hosmer, D.W., Jr.; Lemeshow, S.; Sturdivant, R.X. *Applied Logistic Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2013.
24. Chiang, J. Multinomial logistic regression algorithms via quadratic gradient. *arXiv* **2022**, arXiv:2208.06828.
25. Breiman, L. *Classification and Regression Trees*; Routledge: London, UK, 2017.
26. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [[CrossRef](#)]
27. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
28. Schölkopf, B.; Smola, A.J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*; MIT Press: Cambridge, MA, USA, 2002.
29. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
30. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
31. Cover, T.; Hart, P. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. [[CrossRef](#)]
32. Fix, E. *Discriminatory Analysis: Nonparametric Discrimination, Consistency Properties*; USAF School of Aviation Medicine: Dayton, OH, USA, 1985; Volume 1.
33. McCallum, A.; Nigam, K. A comparison of event models for naive bayes text classification. In Proceedings of the AAAI-98 Workshop on Learning for Text Categorization, Madison, WI, USA, 26–27 July 1998; Volume 752, pp. 41–48.
34. Domingos, P.; Pazzani, M. On the optimality of the simple Bayesian classifier under zero-one loss. *Mach. Learn.* **1997**, *29*, 103–130. [[CrossRef](#)]
35. Sharma, S.; Guleria, K. Deep learning models for image classification: Comparison and applications. In Proceedings of the 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), Greater Noida, India, 28–29 April 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1733–1738.
36. Abreu, J.; Fred, L.; Macêdo, D.; Zanchettin, C. Hierarchical attentional hybrid neural networks for document classification. In Proceedings of the International Conference on Artificial Neural Networks, Munich, Germany, 17–19 September 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 396–402.
37. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
38. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
39. Chahid, I.; Elmiad, A.K.; Badaoui, M. Data preprocessing for machine learning applications in healthcare: A review. In Proceedings of the 2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA), Casablanca, Morocco, 22–23 November 2023. [[CrossRef](#)]
40. Kadhim, A.I. An evaluation of preprocessing techniques for text classification. *Int. J. Comput. Sci. Inf. Secur.* **2018**, *16*, 22–32.
41. Tamilselvi, R.; Sivasakthi, B.; Kavitha, R. An efficient preprocessing and postprocessing techniques in data mining. *Int. J. Res. Comput. Appl. Robot* **2015**, *3*, 80–85.
42. Srivastava, P.; Kaur, N. An Overview on Data Cleaning on Real World Data. *Authorea Prepr.* **2023**.
43. Baskar, S.; Arockiam, L.; Charles, S. A systematic approach on data pre-processing in data mining. *Computsoft* **2013**, *2*, 335.
44. Maharana, K.; Mondal, S.; Nemade, B. A review: Data pre-processing and data augmentation techniques. *Glob. Transit. Proc.* **2022**, *3*, 91–99. [[CrossRef](#)]
45. Natarajan, K.; Bala, P.K.; Sampath, V. Fault detection of solar PV system using SVM and thermal image processing. *Int. J. Renew. Energy Res.* **2020**, *10*, 967–977.
46. Wang, N.; Sun, Z.L.; Zeng, Z.; Lam, K.M. Effective segmentation approach for solar photovoltaic panels in uneven illuminated color infrared images. *IEEE J. Photovolt.* **2020**, *11*, 478–484. [[CrossRef](#)]
47. Deshmukh, S.; Moh, T.S. Fine object detection in automated solar panel layout generation. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 1402–1407.
48. Lin, W.T.; Chang, C.M.; Huang, Y.C.; Wu, C.C.; Kuo, C.C. Fault Diagnosis in Solar Array IV Curves Using Characteristic Simulation and Multi-Input Models. *Appl. Sci.* **2024**, *14*, 5417. [[CrossRef](#)]
49. Ferri, C.; Hernández-Orallo, J.; Modroiu, R. An experimental comparison of performance measures for classification. *Pattern Recognit. Lett.* **2009**, *30*, 27–38. [[CrossRef](#)]
50. Gong, M. A novel performance measure for machine learning classification. *Int. J. Manag. Inf. Technol.* **2021**, *13*. [[CrossRef](#)]
51. Sokolova, M.; Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [[CrossRef](#)]
52. Davis, J.; Goadrich, M. The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA, USA, 25–29 June 2006; ACM: New York, NY, USA, 2006; pp. 233–240.
53. Powers, D.M. Evaluation: From Precision, Recall and F-measure to ROC, informedness, markedness and correlation. *J. Mach. Learn. Technol.* **2011**, *2*, 37–63.
54. Alves, R.H.F.; de Deus Junior, G.A.; Marra, E.G.; Lemos, R.P. Automatic fault classification in photovoltaic modules using Convolutional Neural Networks. *Renew. Energy* **2021**, *179*, 502–516. [[CrossRef](#)]
55. Voutsinas, S.; Karolidis, D.; Voyiatzis, I.; Samarakou, M. Development of a machine-learning-based method for early fault detection in photovoltaic systems. *J. Eng. Appl. Sci.* **2023**, *70*, 27. [[CrossRef](#)]
56. Jia, F.; Luo, L.; Gao, S.; Ye, J. Logistic regression based arc fault detection in photovoltaic systems under different conditions. *J. Shanghai Jiaotong Univ. (Sci.)* **2019**, *24*, 459–470. [[CrossRef](#)]

57. Ali, M.U.; Khan, H.F.; Masud, M.; Kallu, K.D.; Zafar, A. A machine learning framework to identify the hotspot in photovoltaic module using infrared thermography. *Sol. Energy* **2020**, *208*, 643–651. [[CrossRef](#)]
58. Vega Díaz, J.J.; Vlamincx, M.; Lefkaditis, D.; Orjuela Vargas, S.A.; Luong, H. Solar panel detection within complex backgrounds using thermal images acquired by UAVs. *Sensors* **2020**, *20*, 6219. [[CrossRef](#)]
59. Winston, D.P.; Murugan, M.S.; Elavarasan, R.M.; Pugazhendhi, R.; Singh, O.J.; Murugesan, P.; Gurudhachanamoorthy, M.; Hossain, E. Solar PV's micro crack and hotspots detection technique using NN and SVM. *IEEE Access* **2021**, *9*, 127259–127269. [[CrossRef](#)]
60. Raorane, A.; Magare, D.; Mistry, Y. Performance of fault classification on Photovoltaic modules using Thermographic images. In Proceedings of the ITM Web of Conferences, Craiova, Romania, 29 June–2 July 2022; EDP Sciences: Les Ulis, France, 2022; Volume 44, p. 03065.
61. Mellit, A.; Kalogirou, S. Assessment of machine learning and ensemble methods for fault diagnosis of photovoltaic systems. *Renew. Energy* **2022**, *184*, 1074–1090. [[CrossRef](#)]
62. Wang, X.; Yang, W.; Qin, B.; Wei, K.; Ma, Y.; Zhang, D. Intelligent monitoring of photovoltaic panels based on infrared detection. *Energy Rep.* **2022**, *8*, 5005–5015. [[CrossRef](#)]
63. Huang, J.M.; Wai, R.J.; Gao, W. Newly-designed fault diagnostic method for solar photovoltaic generation system based on IV-curve measurement. *IEEE Access* **2019**, *7*, 70919–70932. [[CrossRef](#)]
64. Veerasamy, V.; Wahab, N.I.A.; Othman, M.L.; Padmanaban, S.; Sekar, K.; Ramachandran, R.; Hizam, H.; Vinayagam, A.; Islam, M.Z. LSTM recurrent neural network classifier for high impedance fault detection in solar PV integrated power system. *IEEE Access* **2021**, *9*, 32672–32687. [[CrossRef](#)]
65. Sowthily, C.; Senthil Kumar, S.; Brindha, M. Detection and classification of faults in photovoltaic system using random forest algorithm. In *Evolution in Computational Intelligence: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020)*; Springer: Berlin/Heidelberg, Germany, 2020; Volume 1, pp. 765–773.
66. Amiri, A.F.; Oudira, H.; Chouder, A.; Kichou, S. Faults detection and diagnosis of PV systems based on machine learning approach using random forest classifier. *Energy Convers. Manag.* **2024**, *301*, 118076. [[CrossRef](#)]
67. He, Z.; Chu, P.; Li, C.; Zhang, K.; Wei, H.; Hu, Y. Compound fault diagnosis for photovoltaic arrays based on multi-label learning considering multiple faults coupling. *Energy Convers. Manag.* **2023**, *279*, 116742. [[CrossRef](#)]
68. Yun, L.; Bofeng, Y.; Dan, Q.; Fengshuo, L. Research on fault diagnosis of photovoltaic array based on random forest algorithm. In Proceedings of the 2021 IEEE International Conference on Power Electronics, Computer Applications (ICPECA), Shenyang, China, 22–24 January 2021; IEEE: Piscataway, NJ, USA, 2021; pp. 194–198.
69. Yang, N.C.; Ismail, H. Robust intelligent learning algorithm using random forest and modified-independent component analysis for PV fault detection: In case of imbalanced data. *IEEE Access* **2022**, *10*, 41119–41130. [[CrossRef](#)]
70. Niazi, K.A.K.; Akhtar, W.; Khan, H.A.; Yang, Y.; Athar, S. Hotspot diagnosis for solar photovoltaic modules using a Naive Bayes classifier. *Sol. Energy* **2019**, *190*, 34–43. [[CrossRef](#)]
71. Pierdicca, R.; Paolanti, M.; Felicetti, A.; Piccinini, F.; Zingaretti, P. Automatic faults detection of photovoltaic farms: SolAIR, a deep learning-based system for thermal images. *Energies* **2020**, *13*, 6496. [[CrossRef](#)]
72. Herraiz, Á.H.; Marugán, A.P.; Márquez, F.P.G. Photovoltaic plant condition monitoring using thermal images analysis by convolutional neural network-based structure. *Renew. Energy* **2020**, *153*, 334–348. [[CrossRef](#)]
73. Hopwood, M.W.; Stein, J.S.; Braid, J.L.; Seigneur, H.P. Physics-based method for generating fully synthetic iv curve training datasets for machine learning classification of pv failures. *Energies* **2022**, *15*, 5085. [[CrossRef](#)]
74. Ahmed, W.; Hanif, A.; Kallu, K.D.; Kouzani, A.Z.; Ali, M.U.; Zafar, A. Photovoltaic panels classification using isolated and transfer learned deep neural models using infrared thermographic images. *Sensors* **2021**, *21*, 5668. [[CrossRef](#)]
75. Pathak, S.P.; Patil, S.; Patel, S. Solar panel hotspot localization and fault classification using deep learning approach. *Procedia Comput. Sci.* **2022**, *204*, 698–705. [[CrossRef](#)]
76. Li, B.; Delpha, C.; Migan-Dubois, A.; Diallo, D. Fault diagnosis of photovoltaic panels using full I–V characteristics and machine learning techniques. *Energy Convers. Manag.* **2021**, *248*, 114785. [[CrossRef](#)]
77. Hopwood, M.W.; Gunda, T.; Seigneur, H.; Walters, J. Neural network-based classification of string-level IV curves from physically-induced failures of photovoltaic modules. *IEEE Access* **2020**, *8*, 161480–161487. [[CrossRef](#)]
78. Chandrasekaran, J.; Feng, H.; Lei, Y.; Kacker, R.; Kuhn, D.R. Effectiveness of dataset reduction in testing machine learning algorithms. In Proceedings of the 2020 IEEE International Conference on Artificial Intelligence Testing (AITest), Oxford, UK, 3–6 August 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 133–140.
79. Beleites, C.; Neugebauer, U.; Bocklitz, T.; Krafft, C.; Popp, J. Sample size planning for classification models. *Anal. Chim. Acta* **2013**, *760*, 25–33. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.