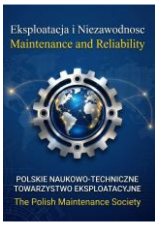

Eksploatacja i Niezawodnosc - Maintenance and Reliability

Volume 29 (2027), Issue 1

https://ein.org.pl

The Polish Maintenance Society

Efficient UAV-based thermal inspection of photovoltaic installations using deep neural networks

Maciej Pawełczyk[1,*], Maciej Plichta[1], Michał Kaczmarczyk[1], Jacek Pawełczyk[1]

[1] SENSE Software Sp. z o.o. Warsaw, Poland
* Corresponding Author: mp@sensesoft.eu

Received: 24 April 2026
Revised: 7 June 2026
Accepted: 9 June 2026
Online: 15 July 2026

Abstract

Reliable operation of large-scale photovoltaic (PV) installations requires inspection methods capable of detecting degradation and failures at an early stage. Conventional ground-based inspections are often impractical for extensive PV farms, leading to the use of unmanned aerial vehicles (UAVs) equipped with infrared thermography for condition monitoring. This paper presents a two-stage inspection methodology supporting reliability-oriented diagnostics of PV modules under real conditions. In the first stage, a convolutional neural network (CNN) with dedicated post-processing extracts individual PV module geometries from radiometric images, including installations with non-uniform orientation and high module density. In the second stage, lightweight machine learning models identify reliability-critical defects at module and string levels. The approach was validated on novel, annotated, and publicly accessible thermographic dataset collected during five years of UAV inspections, demonstrating high detection accuracy with low computational complexity suitable for large-scale deployment.

Keywords: *fault detection, photovoltaic systems, infrared thermography, machine learning, unmanned aerial vehicles*

Article citation:
Pawełczyk M, Plichta M, Kaczmarczyk M, Pawełczyk J, Efficient UAV-based thermal inspection of photovoltaic installations using deep neural networks, Eksploatacja i Niezawodnosc – Maintenance and Reliability 2027: 29(1) http://doi.org/10.17531/ein/224349

Highlights

- Efficient two-stage UAV thermographic inspection framework for photovoltaic installations.
- Deep neural networks for PV module segmentation and defect classification.
- Hybrid CNN–LightGBM architecture for reliability-critical defect detection.
- Low computational complexity suitable for large-scale deployment.
- Validated on a novel, annotated, publicly accessible five-year thermographic datasetighlights.

1. Introduction

The dynamic development of renewable energy systems, driven by industrial growth and environmental requirements, has led to a rapid increase in the installed capacity of photovoltaic (PV) power plants worldwide [1]. As photovoltaic installations become a critical component of modern energy infrastructure, ensuring their reliable operation throughout the entire life cycle has become a key engineering challenge. Maintaining high availability of PV farms requires effective inspection and diagnostic methods capable of identifying failures and degradation processes at an early stage, thereby potentially reducing energy losses, mitigating safety risks, and minimizing unplanned maintenance [2,3].

Conventional ground-based inspection methods are often labor-intensive, time-consuming, and difficult to apply to large-scale installations. Consequently, aerial inspection techniques based on infrared thermography (IRT) have gained increasing importance as practical tools for PV system condition monitoring. UAV-mounted thermal cameras enable non-contact inspection of extensive PV arrays with high spatial coverage and significantly reduced inspection time compared to traditional approaches [4]. The growing availability of radiometric data acquired during such inspections has created a demand for automated analysis methods capable of supporting maintenance decision-making processes [2,5].

UAV-based thermographic inspections enable the detection of a wide range of defects, including hotspots, string-level failures, and bypass diode malfunctions, which are among the most critical factors affecting system reliability and long-term performance. These defects are often difficult to identify using visual inspection alone and may require immediate corrective actions. However, a single UAV mission can generate thousands of thermal images, rendering manual analysis impractical and prone to human error. In this context, artificial intelligence methods, particularly convolutional neural networks (CNNs), have emerged as effective tools for automated anomaly detection, localization, and classification in thermographic imagery [6–8].

In recent years, numerous deep learning–based approaches have been proposed for automated PV defect detection using UAV-acquired data. Object detection frameworks based on YOLO architectures have been widely adopted in real-time inspection scenarios, enabling efficient defect detection in combined thermal and RGB imagery [9], as well as in dedicated thermographic hotspot detection tasks [10]. Similarly, Mask R-CNN–based methods have been applied to PV fault detection and instance segmentation, demonstrating strong localization performance, including handling complex module configurations such as bifacial panels [11], and further improved through attention mechanisms and feature pyramid networks [12].

Semantic segmentation approaches based on U-Net architectures have also demonstrated strong performance in thermal fault detection, particularly when enhanced with attention mechanisms and multi-scale feature extraction [13], [14]. These methods enable detailed spatial characterization of defects; however, they are often associated with higher computational complexity. In parallel, lightweight CNN-based classifiers have been proposed for efficient defect recognition, such as the LIRNet architecture, which combines clustering-based preprocessing with convolutional inference to achieve fast and resource-efficient operation [8].

Despite these advancements, most existing approaches are evaluated on relatively limited datasets, typically comprising several hundred to a few thousand images, often collected under controlled or semi-controlled conditions. Moreover, some methods rely on electroluminescence (EL) imagery rather than UAV-based thermography, which restricts their applicability in real-world inspection scenarios [15]. In addition, many studies focus on isolated tasks—such as detection, segmentation, or classification—without providing a unified framework capable of supporting the full inspection workflow.

A comparative overview of representative state-of-the-art approaches, including object detection, segmentation, and classification methods applied to PV inspection tasks, is presented in Table 1. As shown, existing methods differ significantly in terms of task formulation, dataset scale, and applicability to operational UAV inspection conditions. In particular, the limited size and diversity of available datasets may constrain the generalization capability of these approaches.

An important step toward practical UAV-based inspection systems was presented in the solAIr framework [2], which demonstrated the feasibility of deep learning–based thermographic analysis using UAV-acquired data. Nevertheless, challenges related to scalability, computational efficiency, and robustness under diverse operating conditions remain insufficiently addressed.

In this context, this paper presents an integrated methodology for UAV-based inspection of photovoltaic installations, designed with a focus on operational reliability and practical deployment. The proposed approach combines deep learning–based image segmentation with dedicated post-processing algorithms to accurately extract individual PV module geometries, followed by an ensemble of lightweight machine learning models for defect classification at the module and string levels. The methodology was evaluated using a large dataset comprising 30,983 thermographic images acquired during 29 UAV inspections conducted over a five-year period, covering 416,847 PV modules and 13,139 defect instances under diverse operating conditions. Compared to existing studies summarized in Table 1, the scale and variability of the dataset support improved generalization to real-world inspection scenarios. The presented solution aims to enable reliable condition monitoring of PV farms while maintaining low computational complexity suitable for large-scale and embedded applications.

The main contributions of this work can be summarized as follows. First, a large-scale thermographic dataset for photovoltaic inspection is introduced, comprising 30,983 UAV-acquired images collected over a five-year period and covering more than 416,000 PV modules under diverse real-world conditions. Second, a hybrid defect classification framework is proposed, combining convolutional neural networks with a Light Gradient Boosting Machine in a hierarchical decision scheme, enabling accurate and computationally efficient module-level diagnostics. Third, the study presents a deployment-oriented inspection pipeline that operates directly on individual radiometric images without requiring orthomosaic generation or georeferencing, thereby simplifying practical integration into UAV inspection workflows. Finally, the methodology introduces an explicit module-level representation, ensuring that each detected PV module is uniquely identified and associated with a structured diagnostic output, which aligns with the requirements of industrial inspection and maintenance processes.

Table 1. Comparative overview of representative state-of-the-art deep learning methods for photovoltaic defect detection using thermographic and related imaging modalities, including object detection, semantic segmentation, and classification approaches. The table highlights differences in task formulation, dataset scale, and applicability to UAV-based inspection scenarios, providing context for the proposed method.

Model	Approach Type	Task Type	Dataset Size (Images)	UAV Inspections	Notes
RTPV-YOLO (2026) [9]	YOLO-based detector	Object Detection	Not reported	Not reported	Real-time UAV inspection framework
Awedat et al. (2025) [13]	Enhanced U-Net (Residual + ASPP + Attention)	Semantic Segmentation	1,009	Not reported	Robust segmentation; higher computational cost (2.84 s/image)
Setiawan & Fathurrahman (2025) [11]	Mask R-CNN	Object Detection + Instance Segmentation	660	Not reported	Includes bifacial modules
PV-YOLOv12n (2025) [15]	YOLO-based detector	Object Detection	5,589 / 6,493	Not applicable	Based on electroluminescence (EL), not UAV thermography
Pratticò et al. (2025) [14]	Efficient Attentive U-Net	Semantic Segmentation	100	Not reported	Lightweight (5.24M parameters), near real-time (118 ms)
HOTSPOT-YOLO (2025) [10]	Lightweight YOLO	Object Detection	7,170	Not reported	Optimized for thermal anomaly detection
Lee et al. (2023) – LIRNet [8]	Lightweight CNN	Classification	20,000	Not reported	Two-stage pipeline (K-means + CNN), fast inference (1.3 ms)
Yang et al. (2023) [12]	Improved Mask R-CNN (Attention + FPN)	Object Detection + Instance Segmentation	3,383 + COCO2017	Not reported	Attention-enhanced Mask R-CNN
solAIr (2020) [2]	Mask R-CNN + segmentation models	Segmentation (Instance + Semantic)	1,009	Single UAV campaign (7 days)	Early UAV thermographic DL framework
Proposed method (this work)	CNN + LightGBM (hybrid pipeline)	Segmentation + Classification	30,983 (7,003 annotated)	29 inspections (5-year period)	Large-scale real-world UAV dataset (416k modules); deployment-oriented

2. IRT dataset acquisition process

The effectiveness of thermographic diagnostics of photovoltaic installations strongly depends on the quality and consistency of the acquired measurement data. In UAV-based infrared inspections, data acquisition is influenced by environmental conditions, flight parameters, and sensor characteristics, all of which affect the reliability of defect detection and subsequent diagnostic conclusions. This section outlines the general process of thermographic data acquisition adopted in this study, including inspection planning, measurement conditions, and data collection procedures applied during UAV inspections of operational PV farms. Particular attention is given to practical constraints encountered during field measurements, such as deviations from standardized inspection conditions and variability in acquisition parameters, which are typical for real scenarios. The described acquisition process forms the basis for the construction of the thermographic dataset used for model training and validation in subsequent sections.

It should be explicitly noted that the thermographic dataset used in this study is based on individual radiometric images acquired during UAV inspections. Although selected examples (e.g., Fig. 2A) present visually stitched representations of PV installations, these are intended solely for illustrative purposes and were generated using low-precision mosaicking techniques rather than rigorous photogrammetric reconstruction.

The primary objective of this work is to develop and validate a robust, scalable per-image processing pipeline for PV module detection and defect classification under real operational conditions. As such, the proposed methodology is designed to operate independently of any mapping or orthorectification stage, which enables its direct applicability to raw UAV inspection data and simplifies deployment in practical scenarios.

As such, the methodology does not rely on GPS/RTK-corrected positioning data, which does not significantly affect the accuracy of image-level segmentation or defect classification. None of the 29 inspections used in this study contained RTK-corrected positional metadata, which is consistent with typical operational practice in commercial PV thermographic inspections, where RTK-equipped systems remain uncommon due to hardware cost constraints. The IEC standard does not mandate RTK positioning for thermographic inspections. In operational practice, standard GNSS metadata embedded in image files is sufficient to associate each acquisition with the corresponding plant area for maintenance purposes.

Detailed investigation of photogrammetric processing, orthomosaic generation, georeferencing, and their integration with automated defect detection is considered outside the scope of the present study and will be addressed in a dedicated future publication.

The quality of metadata recorded during UAV inspections was quantitatively assessed across the full dataset to evaluate the feasibility of metadata-based image rectification. GPS coordinates were present in 100% of radiometric images; however, their reliability varied across inspections. Altitude standard deviation ranged from approximately 0.01 m to over 14 m despite planned constant-altitude flights, and exceeded 1 m in approximately 20% of inspections. Inter-image spatial step outliers, identified using an inspection-specific threshold (median + 3·σ of inter-image distances), exceeded 1% in approximately 30% of inspections, with maximum observed jumps exceeding 500 m between consecutive images.

Gimbal orientation metadata were also analyzed. Pitch and roll values were recorded for all images, but the roll component showed minimal variation across the dataset ($\sigma < 0.02°$ in all inspections), which suggests that the recorded values may not fully capture the instantaneous platform roll relative to the world frame. Pitch values exhibited meaningful variability: most inspections operated in near-nadir configuration, yet several inspections contained frames with off-nadir deviation exceeding 5°, and specific inspections were deliberately acquired with fixed oblique pitch as required by the target installation geometry (e.g., rooftop surveys).

These observations indicate that the available UAV metadata are insufficient to support reliable per-image geometric rectification based solely on recorded parameters. A robust rectification workflow would require either RTK/PPK positioning combined with full three-axis IMU data, or photogrammetric reconstruction with bundle adjustment, neither of which is part of this study. Operating directly on raw radiometric images is therefore a deliberate design choice that provides a consistent and deployable solution across heterogeneous acquisition conditions.

2.1. Thermographic measurement standards

Proper thermographic data acquisition is critical to ensure reliable defect detection. Typically, predefined UAV missions were executed under clear-sky conditions with irradiance exceeding 600 W/m², wind speeds below 28 km/h and cloud cover of two octas maximum. Thermal cameras with a minimum resolution of 640 × 512 pixels were used, calibrated according to IEC TS 62446-3 standards [16]. The spatial resolution at ground level, expressed as Ground Sample Distance (GSD), can be computed from the flight parameters and camera specifications using the pinhole camera model:

$$GSD = \frac{h \cdot Sw}{f \cdot W_{img}} \tag{1}$$

where h denotes the relative flight altitude [m], Sw is the sensor width [m], f is the focal length [m], and W_{img} is the image width in pixels. This relationship shows that GSD increases linearly with altitude and inversely with focal length, directly linking acquisition parameters to the achievable spatial resolution. Factors such as altitude, velocity, and camera angle significantly influence the accuracy of the measurements [17].

IEC TS 62446-3 standard also defines inspection norms for the camera itself, most important of which are an object temperature within the range of 253 K to 393 K, a spatial resolution of up to 3 cm per pixel at the PV module level and absolute measurement error of ±2 K. Theoretically, all PV module inspections should adhere to this standard. In reality, both environmental conditions and camera specifications often did not meet acceptable thresholds, usually violating resolution per pixel and/or irradiance requirement. Although this significantly complicated the annotation process, such images and inspections were intentionally included in the dataset, as they reflect real-world use cases submitted by field inspectors and thus offer independent practical value.

Parallel visible-light images were also captured to assist in defect validation and contextual interpretation. Environmental parameters including irradiance, ambient temperature, and wind speed were measured using calibrated instruments to ensure consistent data quality [18].

Common thermographic anomalies observed in photovoltaic (PV) installations include open and short-circuited modules, hotspots, sub-string disconnections, and bypass diode malfunctions. The most frequently identified defects are as follows:

- Open-circuit modules exhibit a temperature difference of $\Delta T = 2 - 7$ K at an irradiance of 1000 W/m², with a uniformly heated surface. This large-area anomaly typically results from failures in modules, inverters, cabling, connectors, or fuses.
- Short-circuit modules display an average temperature rise of $\Delta T = 2 - 7$ K across the surface, commonly caused by wiring defects, insufficient insulation resistance, internal cell faults, or poor electrical matching between modules.
- Sub-string disconnections and bypass diode activations are indicated by localized heating of the diode region by approximately $\Delta T = 2 - 7$ K, suggesting damaged electrical isolation within the string.
- Hotspots are characterized by pronounced overheating of individual cells, typically $\Delta T = 10 - 40$ K above the average cell temperature, arising from cell structure damage, contamination, or shading effects.

Table 2 summarizes typical temperature patterns and their probable causes.

The UAV platforms used in this study were commercially available multi-rotor systems equipped with calibrated radiometric thermal cameras compliant with IEC TS 62446-3 requirements.

The dataset comprises imagery collected using several UAV thermal imaging configurations, including integrated thermal platforms such as DJI Matrice 30T and DJI Mavic 3T, as well as external thermal payloads such as FLIR XT2 and DJI Zenmuse H30T. This hardware variability supports improved generalization of the proposed approach under real inspection conditions.

Table 2. Mapping of custom defect classes to the corresponding IEC TS 62446-3 abnormality categories.

Abnormality	IEC abnormalities	Delta T [K]	Thermal Pattern and Characteristics
1) Multi-Hotspot (MH)	Short circuited crystalline module	2 - 7	Pattern like PID, mismatch, cell defects, glass
2) Hotspot (HS)	Single overheated cell	10 - 40 or more	Strong localized hotspot, usually cell breakage
3) Not-working String (NS)	Open circuit module crystalline or thin film	2 – 7	Uniform module heating, junction box with normal temperature
4) Diode (DI)	One or two substring open circuit	2 – 7	Heated substring, bypass diode heat signature, possible arc
5) Hotspot – Vegetation-Induced Shading (TR)	Cells shaded by dirt	Typically > 40	Dirt related shading pattern
6) Short circuit of large surface contamination (ZW)	Increased transfer resistance crystalline	Typically > 40	Surface contamination short-circuit pattern
7) Resistance to Transition (RT)	Short circuited crystalline module	Typically > 10	Increased transfer resistance crystalline

2.2. Challenges of PV defect detection

The detection of defects in photovoltaic (PV) modules presents several unique challenges that distinguish it from conventional object detection tasks. Firstly, all detectable anomalies are visible only in radiometric imagery, which requires a preprocessing pipeline markedly different from that applied to standard RGB or grayscale images. Secondly, the conventional rectangular representation of output objects is insufficient for this task, as PV modules may appear at varying angles or be partially occluded. Finally, the number of PV modules within a single image can vary significantly—from zero (no modules present in the image) to 120 for high-quality images (conforming to IEC norms), and up to 1,173 modules in the most complex cases acquired from lower-quality flyovers. The average high-quality IRT image contained 35.1 PV modules, while all the images combined contained an average of 59.5 modules per image.

The first challenge is inherent to radiometric image processing. Each pixel in a thermal image encodes an absolute temperature value measured by the infrared camera. In theory, temperature values should represent only the PV array surface captured under industry-standard conditions—clear sky, stable irradiance, and moderate ambient temperature—typically within a range of 283 – 323 K (10 – 50 °C). In practice, however, shading effects, solar reflections from metallic or glass surfaces (often exceeding 500 K), and sky reflections cause a much wider distribution of temperature values. Consequently, temperature clipping and normalization thresholds must be applied adaptively. These thresholds depend on environmental conditions, location, season, and time of day, and must ensure consistent contrast between defective and non-defective regions. The selection of appropriate thresholds therefore constitutes an additional hyperparameter in the image preprocessing stage.

To improve robustness to radiometric disturbances such as solar reflections and emissivity variability, a statistical normalization procedure is applied to each radiometric image. To formalize the applied preprocessing, the radiometric image I ∈ R^HxW is normalized using a clipped linear transformation:

$$I' = \frac{min(max(I,T_{min}),T_{max})-T_{min}}{T_{max}-T_{min}} \quad (2)$$

where T_{min} and T_{max} denote lower and upper temperature thresholds, respectively. These thresholds are selected adaptively based on the statistical distribution of pixel intensities within each image. In practice, T_{min} and T_{max} are determined using percentile-based clipping:

$$T_{min} = P_{\alpha}(I) \text{ and } T_{max} = P_{\beta}(I) \quad (3)$$

where $P_{\alpha}(I)$ and $P_{\beta}(I)$ denote the α-th and β-th percentiles of the temperature distribution (e.g. α=1, β= 99).

This approach effectively suppresses extreme temperature outliers caused by solar reflections or emissivity variations while preserving relative thermal contrasts relevant for defect detection. Instead of explicitly modelling emissivity or reflected radiation components, the proposed pipeline relies on data-driven learning to achieve robustness against such radiometric disturbances.

A second, equally critical challenge concerns the geometric representation of PV modules. Fig. 1 illustrates how bounding-box-based detection methods may fail to accurately isolate individual modules. Defects typically occur within small sub-regions of modules that must be analyzed separately, requiring precise extraction of individual module polygons. When PV modules are oriented parallel to the image axes, bounding boxes can provide adequate localization, as shown in the upper pair of images in Fig. 1. However, when modules are inclined relative to the image axes, severe misalignment occurs. In such cases, bounding boxes corresponding to one module may overlap adjacent ones, causing the system to misattribute a defect to neighboring modules and resulting in an unacceptable number of false positives.

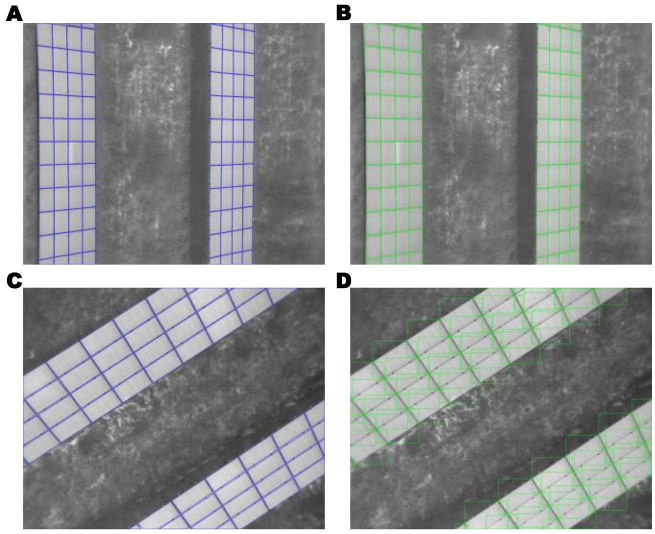

Figure 1. Comparison of bounding-box localization performance under different PV module orientations relative to the image axes. (A) Radiometric image with a diode defect and ground-truth labels, where PV modules are aligned with the image borders. (B) Corresponding bounding-box-based detection result. (C) Radiometric image with ground-truth labels, where PV modules are rotated approximately 45° relative to the image borders. (D) Corresponding bounding-box-based detection result.

One potential mitigation strategy involves rotating or rectifying the image to align the PV module edges with the image axes. Nevertheless, this approach is impractical for many inspection scenarios—particularly for rooftop installations with irregular orientations, tracking PV modules, radiometric images affected by geometric distortion, and farms where panel rows are installed at varying angles. Two representative examples of these limitations are presented in Fig. 2, which shows (A) a stitched farm overview with non-uniform panel orientations and (D) a ground-based inspection captured at a steep angle.

A further difficulty arises from the high density of PV modules visible in a single frame. In the dataset used in this study, up to 1,173 modules were captured in one image, with an average of 59.5 PV modules per IRT image. These extreme cases typically originated from high-altitude or oblique flyovers performed by operators under non-standard conditions, as illustrated in Fig. 3. The resulting images exhibited reduced spatial resolution of individual modules, significantly affecting the precision of defect localization.

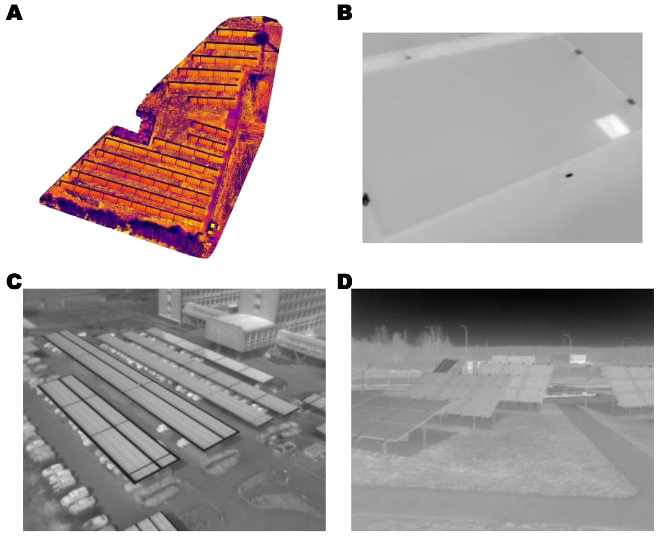

Figure 2. Representative edge cases illustrating scenarios in which global image rectification or rotation is not applicable. (A) Stitched overview of a PV farm composed of multiple inspection images with non-uniform panel orientations. (B) Single PV module captured using a handheld thermal camera. (C) High-altitude image encompassing the entire PV installation. (D) Closely spaced PV panels mounted at different angles, preventing reliable geometric alignment through image rotation. (A) represents a low-precision stitched visualization for illustrative purposes only and does not correspond to a rigorously generated orthomosaic.

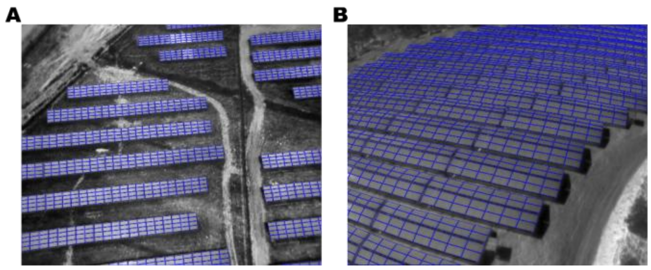

Figure 3. Two examples of low-quality UAV flyovers illustrating challenging acquisition conditions. (A) Image containing multiple PV modules with reduced spatial resolution per module. (B) Image captured at a high camera angle, introducing geometric distortion and reduced localization accuracy.

The described challenges highlight the limitations of conventional detection approaches and justify the need for a segmentation-based methodology capable of reliable PV module identification under non-ideal acquisition conditions.

2.3. IRT annotation process

Accurate annotation of thermographic data is a critical step in the development and validation of automated diagnostic methods for photovoltaic installations. In this study, infrared images were manually annotated at the PV module-level to identify module boundaries and assign defect labels based on observed thermal patterns. The annotation process was performed by trained personnel with experience in PV thermography and was guided by IEC TS 62446-3 recommendations, while accounting for deviations arising from non-standard acquisition conditions. To ensure annotation reliability, selected samples were cross-checked and validated through expert review. The resulting annotations provided the ground truth data required for training and evaluating the proposed detection and classification models.

3. IRT dataset

A key contribution of this work is the establishment of a structured thermographic dataset named Sense-Solar, comprising 30,983 thermal images, 7,003 fully annotated radiometric images, which include 416,847 annotated PV modules and 13,139 module defects, distributed across 12,930 modules. This corresponds to 209 PV modules containing two defects each. No instances with more than two defects on a single module were observed. Annotations were validated by certified PV inspectors and verified by independent auditors to ensure reliability. Each defect instance was categorized as MH (Multi-Hotspot), HS (Hotspot), DI (Diode Defect), TR (Vegetation), ZW (Short Circuit), NS (Not-Working String), REF (Sun Reflex – useful for false positives debugging), which were grouped into four major categories for training purposes: hotspot, bypass diode failure, string defect, or no defect. The REF class explicitly captures solar reflection effects and emissivity-related radiometric artifacts, which are treated as part of real-world measurement variability rather than being removed during preprocessing. This problem formulation reflected typical industrial actions associated with specific failure modes – respectively PV module inspection (that may not end up with module replacement), PV module replacement, string inspection/reinstalment or no action required. The final dataset comprised 7512 hotspot-type defects (2269 regular hotspots and the rest containing MH/TR/ZW defects mapped onto HS), 1133 diode defects, 4494 string defects, and 402,409 non-defective modules, yielding an overall average defect rate of 3.2% across 29 documented inspection missions. Representative examples of the obtained database are presented in Fig. 4.

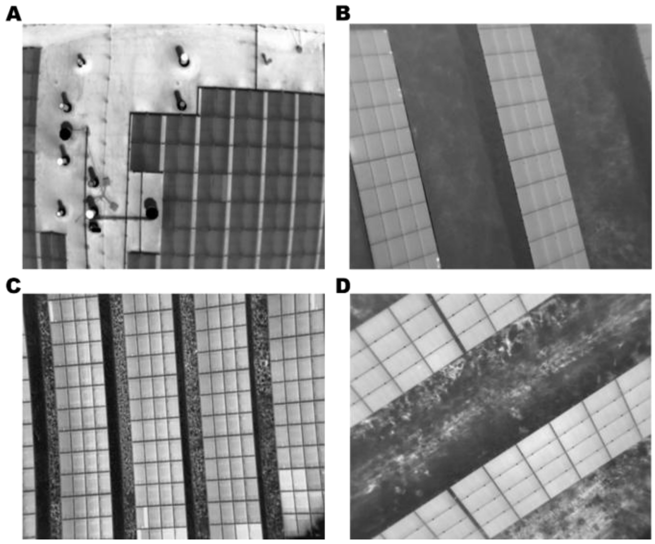

Figure 4. Representative radiometric images from the Sense-Solar dataset. (A) Rooftop PV installation. (B) Ground-mounted installation with visible hotspot defects. (C) Ground-mounted installation with a diode defect. (D) Rotated radiometric image of a ground-mounted installation.

To improve model generalization, the dataset includes imperfect images affected by UAV shadows, sun reflections, occlusions and distortions – all of which have occurred and are expected to happen during normal operation of the defect detection system. This ensures that models trained on this dataset can handle non-ideal conditions frequently encountered in field inspections.

The IEC TS 62446-3 inspection standard was used as a qualitative basis, although approximately 20% of images that did not adhere to that standard, as illustrated in Fig. 5.

To quantify the extent of deviation from the IEC TS 62446-3 requirements, we decided to investigate details of spatial resolution. The GSD was computed for each image across all 29 inspections using the formula defined in Section 2.1, based on EXIF metadata extracted from the acquired radiometric images. The analysis revealed that GSD values ranged from 0.3 cm/pixel to over 32 cm/pixel across the entire dataset. Only two inspections consistently maintained GSD below the 3 cm/pixel threshold. The majority of inspections operated in the 3.2–8.2 cm/pixel range, while several missions conducted at higher altitudes reached median GSD values of 8–12 cm/pixel. Furthermore, six inspections exhibited intra-mission GSD standard deviations exceeding 1 cm/pixel, indicating substantial altitude variation during a single survey. These observations indicate that strict adherence to the 3 cm/pixel standard is rarely achieved in operational PV inspections.

However, since those images were a part of UAV flyovers and were submitted by operators as acceptable these images were retained in the dataset in order to form the basis for model.

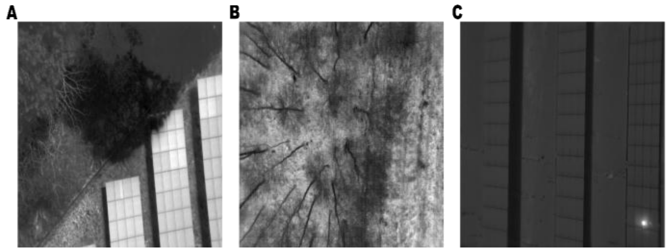

Figure 5. Selected challenging radiometric images illustrating non-ideal inspection conditions. (A) Shadow partially obscuring PV modules affected by NS defects. (B) Image of a non-PV area captured in the immediate vicinity of a PV installation. (C) Temperature artifact caused by solar reflection on one of the PV modules.

To ensure high-quality of predictions and to maximize generalization properties of the resultant model the training dataset is consistently being expanded with new annotated images with a strong emphasis in the images where prediction quality using existing models was imperfect.

4. Pipeline design

Although pre-trained neural network architectures constitute a viable baseline for thermographic image analysis, a dedicated processing pipeline was designed and developed within the scope of this study to better address the specific requirements of photovoltaic inspections under real operational conditions. The overall problem was decomposed into two complementary tasks: PV module detection and PV module defect classification. The adopted two-stage processing concept is illustrated in Fig. 6 and further detailed in Fig. 7.

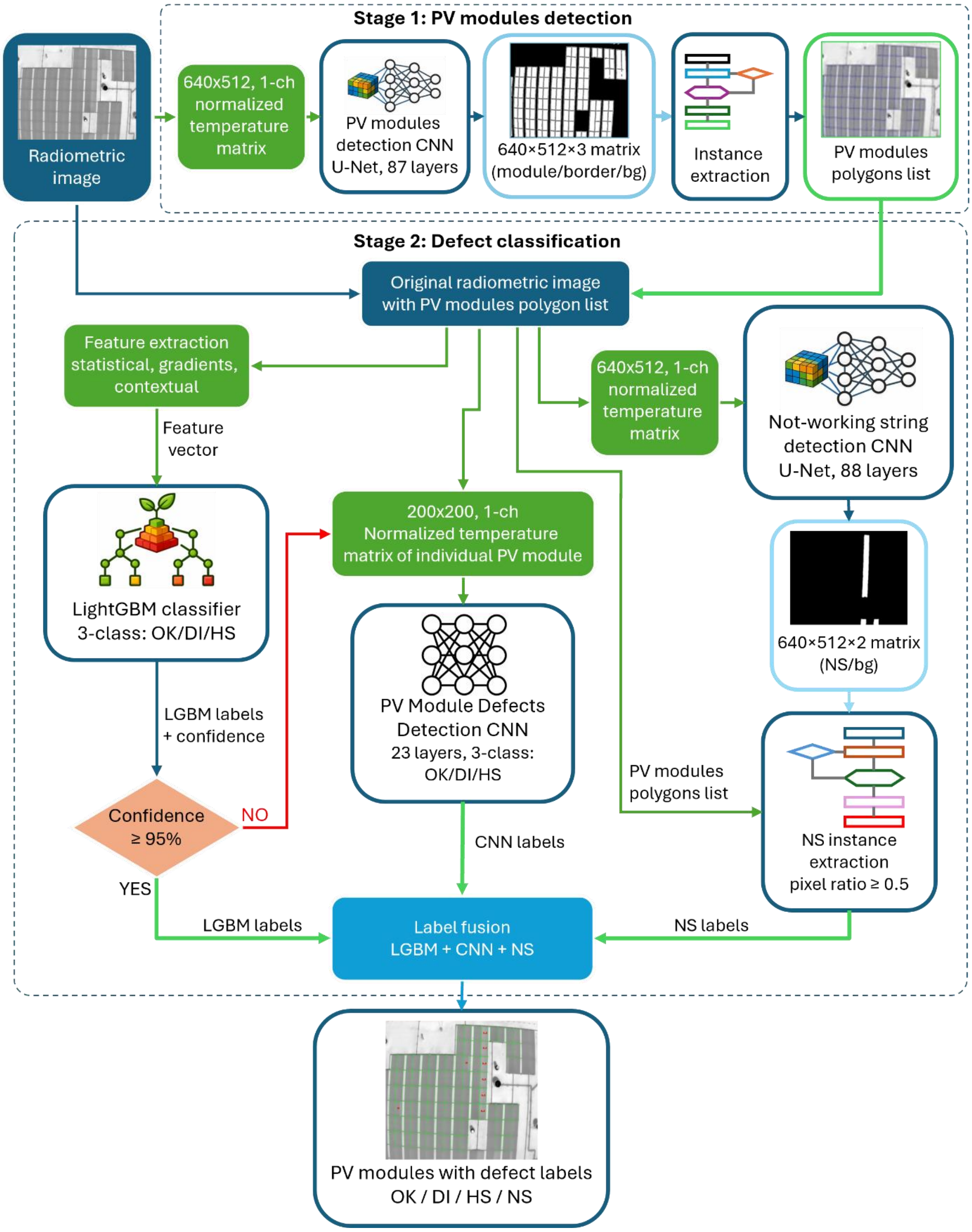

Figure 6. Block diagram of the proposed two-stage inspection pipeline. In Stage 1 (upper section), the input radiometric image is normalized to 640 × 512, single-channel temperature matrix. Such input is processed by a U-Net-based semantic segmentation CNN to produce a three-class segmentation map (module/border/background), which is subsequently refined through classical image processing (contour detection, morphological operations, convex hull extraction, and geometric filtering) to yield a list of individual PV module polygons. In Stage 2 (lower section), the original radiometric image and the detected module polygons serve as inputs to two classification components: (1) a feature extraction module computes statistical, gradient, and contextual descriptors for each detected PV module, which are classified by a LightGBM gradient boosting model; if prediction confidence falls below 95%, the module is forwarded to a dedicated a multi-class CNN for further analysis; (2) the same input image is simultaneously processed by a Not-Working String (NS) detection CNN, producing a binary segmentation map from which per-module NS labels are derived via pixel-ratio thresholding (threshold >= 0.5). A label fusion step combines outputs from the LightGBM, multi-class CNN, and NS classifier to assign a final defect label (OK, DI, HS, or NS) to each detected PV module.

To ensure experimental robustness and to avoid data leakage between processing stages, identical training, validation, and test image partitions were used for all CNNs and feature-based components employed in the pipeline. This approach guarantees consistent evaluation conditions and prevents cross-contamination between models performing different functions.

As shown in Fig. 7 each input radiometric image is first normalized and processed by a CNN to generate a semantic segmentation map of PV modules. The resulting segmentation is subsequently refined using classical image processing techniques to obtain instance-level module representations. In parallel, the same input image is analyzed by a dedicated CNN to produce an image-wide binary segmentation map identifying defects spanning multiple modules, such as not-working strings.

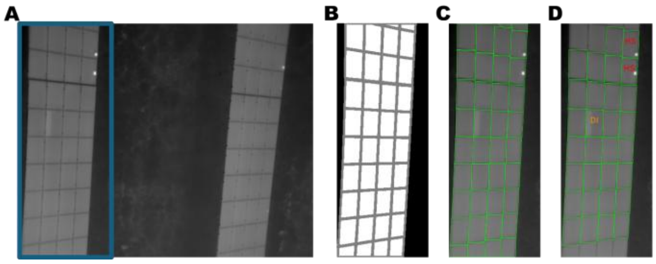

Figure 7. Overview of the proposed multi-network inspection framework. The input radiometric image (A) is processed to generate a semantic segmentation map (B) and corresponding PV module instance segmentation (C), followed by module-level defect classification (D). Defect classes are color-coded as follows: Diode (DI) – orange; Hotspot (HS) – red.

The extracted PV module instances are then processed independently by a trained LightGBM model using a set of temperature-derived statistical features computed for each module. If the confidence of the boosting model is not high enough for a given PV module, a dedicated CNN-based classifier processes that module to identify potential defects. Finally, the outputs of the feature-based LightGBM, module-level classifiers and the image-wide segmentation model are combined to assign a final defect label to each detected PV module.

To ensure consistency across the formulation of individual components, the following notation is adopted throughout the paper. Each detected photovoltaic module is denoted as M_i, where $i \in \{1, \dots, N\}$. The corresponding radiometric input is represented as an image $I \in \mathbb{R}^{HxW}$, while module-level features are expressed as vectors $x_i \in \mathbb{R}^d$. Model predictions are denoted as $\hat{y}_i$ and ground-truth labels as y_i.

4.1. Two-stage model design

Preliminary experiments with single-stage neural network architectures, including both publicly available models and custom designs developed within the project, revealed limitations in detection accuracy, scalability, and computational efficiency under real inspection conditions. To address these constraints, a two-stage model was adopted. In the proposed design, the first stage is dedicated to robust detection and geometric localization of individual PV modules, while the second stage performs defect classification at the module and string levels. This separation allows each stage to be optimized for its specific task, improving reliability of the diagnostic process and enabling efficient deployment in large-scale photovoltaic inspections.

4.2. Evaluation metrics

The way in which inspection results should be presented significantly affects the usefulness of the metrics typically used to evaluate algorithms that process data. All time benchmarks were calculated using i7-13700K CPU, NVIDIA GeForce RTX 4070 GPU and 64GB RAM at 3600MT/s. The metrics most frequently used to evaluate and rate object detection algorithms include:

- Intersection over Union (IoU): measuring the overlap between the predicted bounding polygon and the ground truth bounding polygon for a given object. It is calculated as the ratio of the area of intersection to the area of the union between the two bounded objects.
- Precision and Recall: Precision is the ratio of true positive predictions to the sum of true positive and false positive predictions, while recall is the ratio of true positive predictions to the sum of true positive and false negative predictions. These metrics provide insights into the algorithm's ability to correctly identify objects and avoid false positives or false negatives.
- F1 Score – the harmonic mean of precision and recall. It provides a single metric that balances both precision and recall and is particularly useful when the dataset is imbalanced.
- Jaccard Index – the ratio of the intersection to the union of the predicted and ground-truth label sets. It measures

the degree of overlap between predictions and annotations and is particularly useful as a strict classification metric, as it penalizes both false positives and false negatives simultaneously.

It is important to acknowledge that the initial metric mentioned, IoU (Intersection over Union), serves as the fundamental basis for the subsequent metrics. This is primarily because the analysis of object occurrence (detection) within an image typically involves a degree of uncertainty or arbitrariness. Factors such as out-of-focus images or object occlusion by other elements can contribute to situations where correct object localization may not precisely align with the arbitrarily defined ground truth. The application of IoU value thresholding allows for a flexible evaluation of object detection, considering the extent of overlap between predicted and actual object locations. However, it should be noted that in the case of the photovoltaic installation analyzed here, the modules do not overlap in the images, and the localization can be considered successful if the detector unambiguously indicates given module.

4.3. PV module detection network

The PV module detection problem is particularly challenging due to the inclination of the module in relation to image edges as well as the large possible number of PV modules in the image, specifically partially occluded ones (image edges). To counteract this challenge, the PV module detection problem was represented in the form of image semantic segmentation in conjunction with image post-processing to obtain instance segmentation based results. Segmentation was achieved by subdividing the image into three semantic classes: PV module, non-PV module, and a border between those two. The third class was introduced not only to emphasize the difference between the first two classes but also to distinguish individual modules from each other thus allowing effective instance segmentation, but supporting full scalability from 0 to hundreds of PV modules on a single image. Starting with the original ground-truth polygons, their edges were broadened to create a third, "border" class. This represents a ground-truth model data upon which the segmentation CNN is trained.

The entire PV module detection process has been visualized in Fig. 8 which represents one of the test set images.

To obtain train/validation/test split each PV inspection was separately divided according to 80%/10%/10% ratio. Ultimately train set consisted of 5,572 images, validation set contained 692 images and the test set was composed of 739 images. During assigning to sets images were chosen in a deterministically seeded random selection.

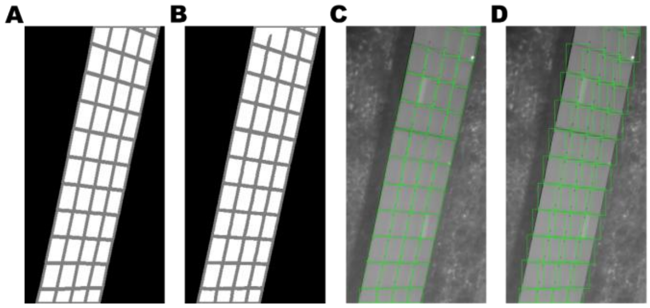

Figure 8. Example of the PV module detection process (cropped for clarity). (A) Ground-truth semantic segmentation map. (B) Predicted semantic segmentation output. (C) Instance segmentation obtained after post-processing of the segmentation map. (D) Bounding-box representation derived from the instance segmentation results.

To increase the number of examples used for training of semantic segmentation CNN image augmentation strategies were used. Those included flipping, clipping and their permutations, strategically chosen to ensure high model generalization capabilities. Moreover images containing defects were additionally augmented with use of flipping, noise addition and their permutations proportionally to the number of defects present. This ensured that defective PV modules were adequately represented amongst the dataset.

Additionally, absolute temperature shifts were added to PV modules or surrounding areas to increase or decrease the difference between PV module surface and surrounding areas. These mimicked conditions of high and low intensity sunlight which are often present in inspections that do not meet IEC standards – as shown in Fig. 5.

After combining all the images from the augmentation methods mentioned above the train set size has grown 22.2 times.

For PV module detection problem a segmentation CNN U-Net like encoder-decoder architecture with 4 skip connections has been chosen, after extensive testing of other solutions. This model consisted of 87 layers which were subdivided into 23 submodules. Input to the neural network consists of 640 × 512 image, while the output (consisting of 3 classes - PV module, border and background) results in the matrix of a 640 × 512 × 3 shape. This resulted in a model with 252,291 trainable parameters. After training, the neural network model was optimized for performance by quantization. This resulted in CNN size reduction from 1292 kB to 248 kB and inference time of 12.1 ms per image. On the target deployment hardware, the end to end inference time (image normalization, preprocessing, and model inference) of the PV module detection segmentation CNN stage is approximately 219 ms per image.

The architecture of the semantic segmentation CNN is shown in Fig. 9.

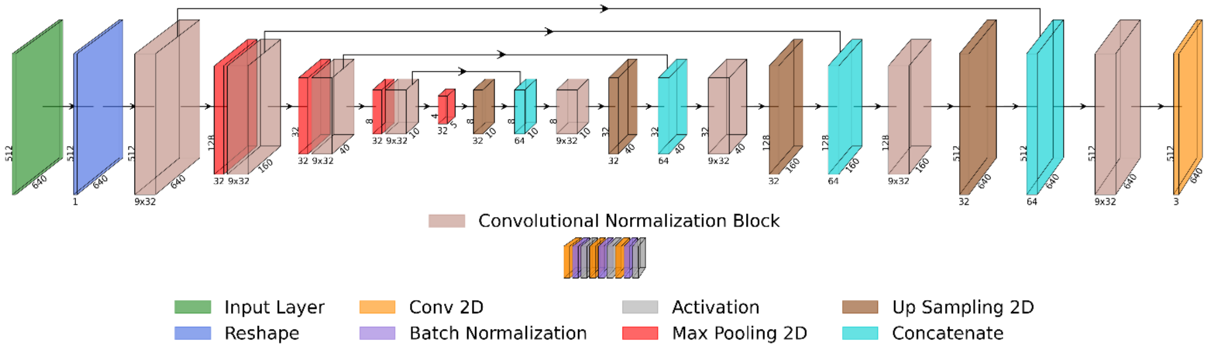

Figure 9. Architecture of the U-Net–like convolutional neural network optimized for PV module semantic segmentation.

During training, at the end of each training step, the network is trained using a pixel-wise multi-class cross-entropy loss function defined as:

$$\mathcal{L}_{\text{seg}} = -\sum_{u=1}^{H}\sum_{v=1}^{W}\sum_{c=1}^{C} y_{u,v,c} \log \hat{y}_{u,v,c} \qquad (4)$$

where H and W denote the spatial dimensions of the input image, C is the number of semantic classes, $y_{u,v,c}$ represents the ground-truth label for pixel $(u,\ v)$, and $\hat{y}_{u,v,c}$ is the predicted probability for class c. This formulation enables dense supervision of the segmentation task and supports accurate delineation of PV module boundaries.

During training phase model performance is evaluated on data from the validation set. To obtain a quantitative measure of training progress a set of classification metrics (accuracy, precision, specificity, F1 score and others) are computed - both on training and validation sets, with subsampling to ensure a performant training process. F1 score is used as the primary metric to designate best model checkpoint to store as the potential model candidate. Finally, the top 5 checkpoints are evaluated on the – previously not used - test set. An example of such a model confusion matrix for per pixel semantic segmentation, at a scale of 1,000,000 to 1 for readability, is shown in Fig. 10.

To obtain instance segmentation outputs from 3-class segmentation custom image post-processing system has been designed. First steps consist of rough estimation of median area of PV modules present on the segmentation mask with use of contour detection on binarized segmentation mask. Next step enhances borders between neighboring PV modules by use of dilation and erosion algorithms on "border" class on original segmentation map. This is skipped if the estimated median area of PV module is relatively small and use of dilation and erosion algorithms would lead to the loss of those instances. After borders correction, contours are detected and convex hull extraction is performed.

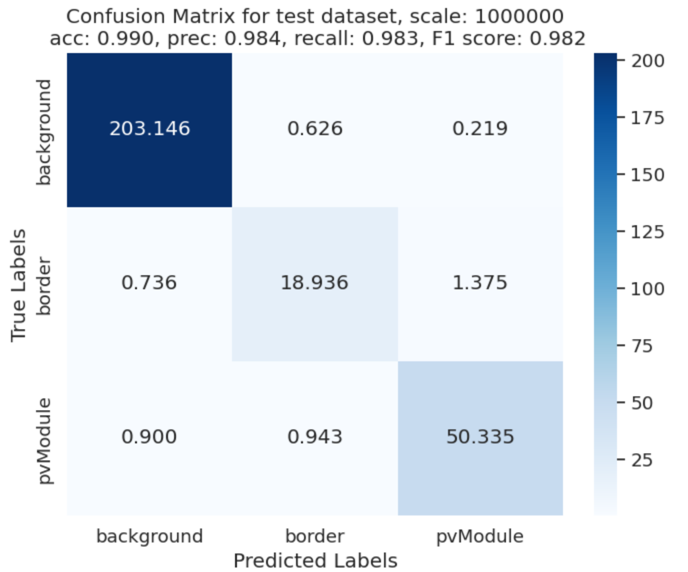

Figure 10. Confusion matrix of the semantic segmentation CNN evaluated on the test set. Matrix values are scaled at a ratio of 1,000,000:1 for readability. For example, a value of 203.146 corresponds to 203,146,000 pixels correctly classified as background across the entire test dataset.

Next step in the algorithm is thresholding which removes contours with less than 3 vertices and surface area below specified level. Filtered contours are then refined by building quadrangle for a specified convex contour and if needed simplifying the resulting shape by performing polygonal curve approximation using Ramer–Douglas–Peucker algorithm. After that, resulting contours are again filtered to remove area outliers using a modified Z-score computed from median of only polygons with 4 vertices with interior angles between 60 and 120 degrees and not directly adjacent to image borders.

After that remaining contours are enlarged to compensate for size reduction during extraction and cropped to image boundaries. Example image processing steps are visualized in Fig. 11.

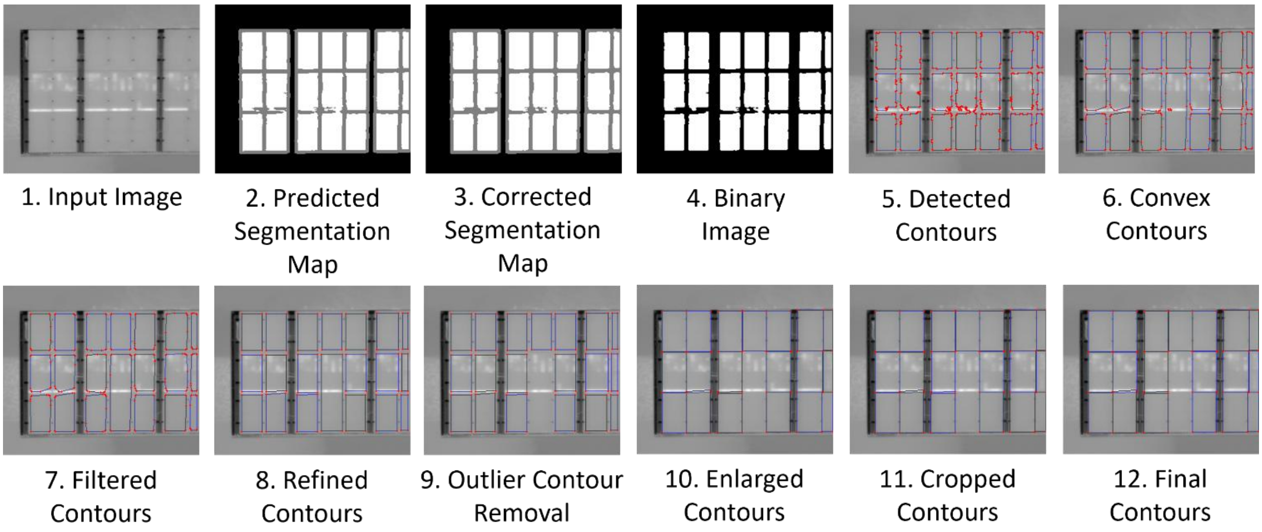

Figure 11. Main stages of the instance segmentation post-processing algorithm used to convert the semantic segmentation map into individual PV module instances.

Instance segmentation evaluation requires predicted modules to be unambiguously matched with their ground truth definitions. As mentioned earlier, IoU is used for this matching process. This creates matching of each predicted module to its "best" ground truth counterpart resulting in 1:1 pairs which are treated as true positives. All predicted and ground truth modules that had their "best" match with IoU below the assumed threshold (0.5) were marked as false positives and false negatives respectively. Additionally images that did not contain any ground truth PV modules and for which detection pipeline did not detect any PV modules were marked as true negatives.

Instance segmentation combined with image post-processing and ground truth matching approach was evaluated on the aforementioned test set and shows F1 score of 90.0% and accuracy of 81.9%. Confusion matrix is presented in Fig. 12.

Combination of semantic segmentation CNN and instance segmentation algorithm results in complete PV module detection pipeline, which is capable of overcoming critical issues present in single-stage approach. Orientation of detected PV modules is deterministic as it is derived from per pixel module location. Furthermore this approach is also applicable to PV modules of various sizes, which is advantageous for real-world applications, which are often not fully aligned with IEC recommendations.

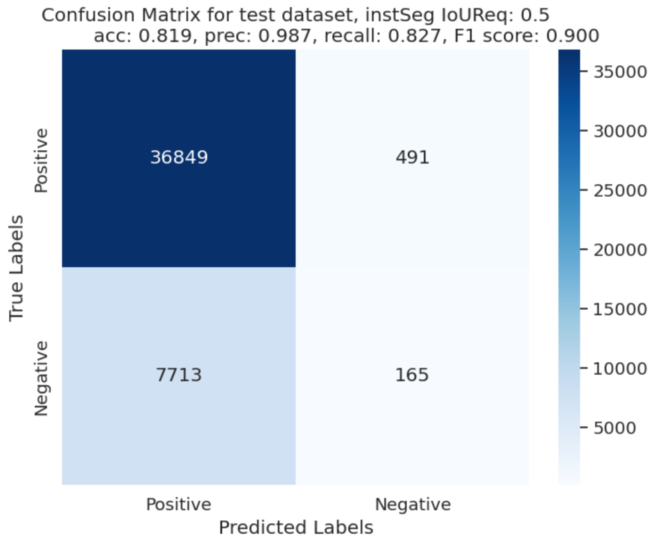

Figure 12. Confusion matrix of the best-performing PV module detection pipeline, combining semantic segmentation and instance-level post-processing. Each quadrant represents the number of PV modules predicted by the CNN and subsequently extracted by the instance segmentation algorithm.

Because the segmentation CNN is trained and evaluated on the same non-rectified image distribution, with ground-truth annotations prepared on raw radiometric images, the reported segmentation metrics (IoU, Dice/F1 score, precision, recall) are internally consistent with the input distribution. Perspective distortions present in the training data are learned as part of the natural distribution rather than acting as a source of systematic bias with respect to ground truth.

4.4. Defect classification system

Once individual PV modules have been localized within the radiometric image, a dedicated defect classification stage is applied. To ensure reliable and scalable automated diagnostics in thermographic photovoltaic inspections, a multi-component machine learning pipeline was designed and experimentally validated. The proposed approach combines structured dataset preparation, dedicated preprocessing procedures, and a unified evaluation framework with three complementary machine learning models. Two CNNs are specialized for distinct defect categories observed under real field conditions, while an additional gradient-boosting decision tree classifier is incorporated to enhance robustness and computational efficiency. This hybrid design enables accurate module-level and image-level defect identification while maintaining suitability for large-scale and resource-constrained deployment environments.

4.4.1. Multi-class defect detection

The first component of the classification system is a multi-class pipeline operating on cropped and normalized radiometric images of individual PV modules. Each module instance is assigned to one of the predefined defect categories, including a non-defect class representing healthy modules and typical anomaly types such as hotspot patterns or shunted bypass diode malfunctions. The objective of this stage is to provide reliable module-level discrimination between normal thermal behaviour and the most frequently observed defect signatures.

The complete thermographic dataset was divided into training, validation, and test subsets using an 80%/10%/10% split defined at the image-level using an identical approach as for PV module detection pipeline. Consequently, all module crops originating from a single infrared image were always assigned to the same subset, preventing information leakage between partitions. Each subset consisted of normalized module-level radiometric crops with associated ground-truth defect labels. Data augmentation and class balancing procedures were applied exclusively to the training subset, while the validation and test sets preserved the natural class distribution to ensure unbiased performance assessment.

To mitigate class imbalance, defect-dependent augmentation factors were introduced, resulting in more aggressive upsampling of rare defect categories relative to healthy modules. In practical terms, additional healthy module samples were drawn from the same image whenever a defect instance was present, whereas images without defects contributed a limited number of randomly selected healthy modules. Augmentation operations included geometric transformations such as horizontal and vertical flips, 90-degree rotations, optional additive Gaussian noise, and controlled absolute temperature shifts analogous to those used in the module detection pipeline. Following augmentation, the image-level dataset comprised 4,451 training images, 554 validation images, and 585 test images, with the training subset expanded by a factor of approximately 1.94 which resulted in 8,616 augmented training images. Images that did not contain any PV modules were excluded from the defect detection pipeline. This is why the total number of images in those two pipelines is slightly different.

To improve generalization to real deployment conditions, the training, validation, and test sets were supplemented with correctly detected module instances produced by the upstream semantic segmentation network. In addition, all module contours were filtered in order to remove incomplete or geometrically invalid detections, including contours located too close to image borders, non-quadrilateral shapes, excessively acute internal angles, or areas below a predefined threshold. This ensured that only well-defined module instances were used for classifier training and evaluation.

This preprocessing yielded the following PV module–level dataset sizes. After merging manually annotated ground-truth instances with module candidates generated by the upstream detection stage, the final dataset comprised 211,405 modules in the training subset, 55,997 modules in the validation subset, and 62,563 modules in the test subset. Within the training partition, this total consisted of 24,149 original ground-truth modules, 164,533 augmented training modules, and 22,723 modules predicted by the module detection network. For the validation and test partitions, the reported totals included 30,118 and 32,940 original ground-truth modules, respectively, supplemented by 25,879 and 29,623 module instances produced by the upstream detection stage.

The resulting normalized and augmented module crops were used to train the convolutional classifier. During inference, the same preprocessing workflow was applied without augmentation to obtain per-module defect predictions.

Following an extensive architecture search, the final model was selected as a lightweight CNN composed of small-kernel convolutional layers interleaved with batch normalization, max-pooling, and dropout regularization. The architecture remains intentionally shallow to balance representational capacity with computational efficiency, comprising 23 effective layers organized into four convolution-pooling stages. Each module crop is resized to a 200 × 200 single-channel input and mapped to three output classes representing one healthy state (OK) and two defect categories (DI and HS).

This per-module resampling approach to a fixed 200 × 200 pixel representation, combined with the upstream geometric validity filter (requiring four detected corners, a minimum distance of 2 pixels from the image boundary, and the absence of anomalous internal angles), decouples the classification stage from absolute module size in pixels and from image-level perspective distortions. Any moderate geometric distortion is applied uniformly within a module region and does not affect the relative temperature distribution on which the classifier operates. The final network architecture is illustrated in Fig. 13.

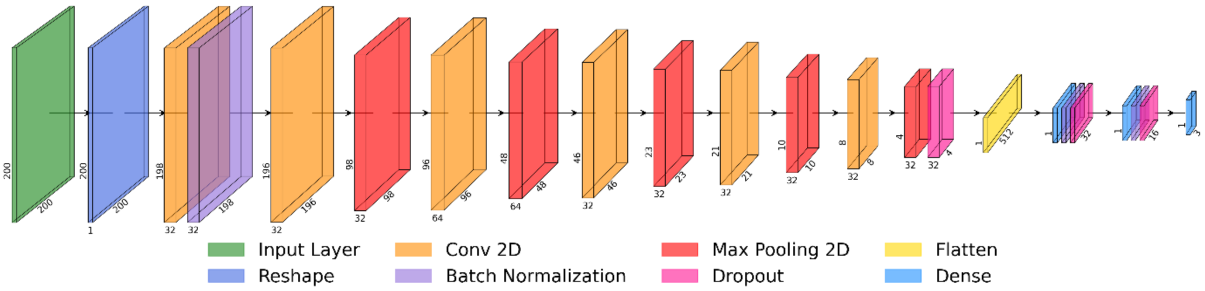

Figure 13. Architecture of the multi-class CNN model for PV module defect classification.

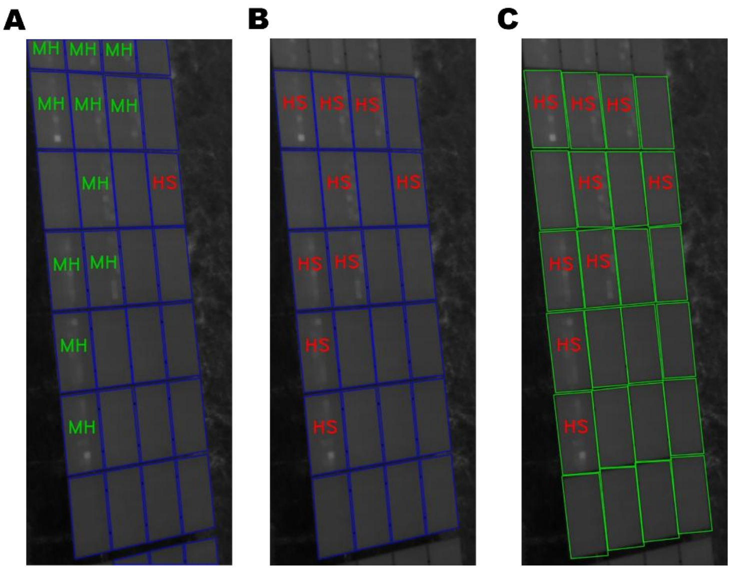

Figure 14. Representative results of the multi-class PV module defect classification model. (A) Manually annotated ground-truth labels. (B) Defect predictions generated for ground-truth PV module regions. (C) Defect predictions obtained for PV modules automatically detected by the module detection pipeline. Defect classes are color-coded as follows: Multi-Hotspot (MH) – light green; Hotspot (HS) – red.

The CNN-based defect classifier is trained using a categorical cross-entropy loss defined as:

$$\mathcal{L}_{\text{cls}} = -\sum_{i=1}^{N}\sum_{k=1}^{K} y_{i,k}\log\hat{y}_{i,k} \qquad (5)$$

where N is the number of training samples, K denotes the number of defect classes, $y_{i,k}$ is the ground-truth indicator for class k, and $\hat{y}_{i,k}$ is the predicted probability. This loss formulation supports multi-class discrimination between healthy and defective PV modules.

The final classification network comprises approximately 67,158 trainable parameters, which enables an average inference time of about 0.2 ms per PV module, corresponding to approximately 11.9 ms for a typical image containing 59.5 modules. The compact architecture is also reflected in a modest storage footprint of 384 kB, which is further reduced to 96 kB after model quantization, making the solution suitable for high-throughput processing and deployment on resource-constrained embedded platforms.

When executed on the target deployment hardware, the complete processing time of the neural defect detection stage—including image normalization, preprocessing, and model inference—amounts to approximately 237 ms per image, or about 4 ms per PV module assuming an average of 59.5 modules per frame. Representative examples of multi-class defect classification results are presented in Fig. 14.

The multi-class defect classifier achieved a precision of 96.4%, recall of 95.6%, Jaccard index of 92.5%, fallout of 4.4%, F1 score of 96.0%, and an overall accuracy of 99.7%. The corresponding multi-class confusion matrix illustrating the distribution of predictions across defect categories is presented in Fig. 15.

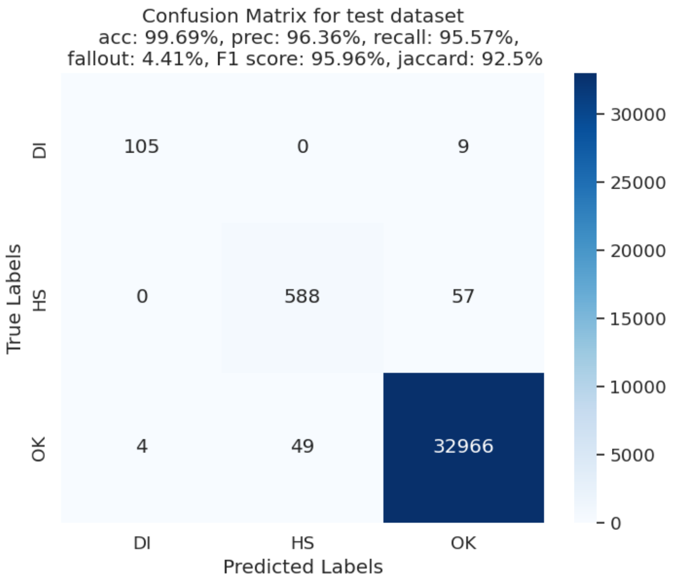

Figure 15. Confusion matrix of the multi-class PV module defect classification model evaluated on the test set.

4.4.2. Light gradient boosting machine

The introduction of the Light Gradient Boosting Machine (LightGBM) was motivated by the need to increase both performance and robustness of the learning-based defect classification stage. This feature-driven classifier complements neural models by explicitly capturing fine-grained radiometric characteristics of individual PV modules through the use of handcrafted descriptors derived directly from raw temperature data. Such an approach enables precise modelling of temperature distributions, gradients, and local variability within module contours. LightGBM represents an efficient implementation of the gradient boosting decision tree algorithm employing gradient-based one-side sampling [19].

For each detected photovoltaic module M_i, a feature vector is constructed based on the radiometric temperature distribution and its local context:

$$\mathbf{x}_i = [\mu_i, \tilde{\mu}_i, \sigma_i, p_{10,i}, p_{50,i}, p_{90,i}, r_i, g_i, h_i] \tag{6}$$

where μ_i and $\tilde{\mu}_i$ denote the mean and median temperatures, σ_i is the temperature standard deviation, $p_{10,i}, p_{50,i}, p_{90,i}$ represent selected percentiles, r_i captures range- or threshold-based descriptors, g_i represents gradient-based descriptors capturing spatial temperature changes and h_i denotes higher-order or local variability features derived from second-order statistics.

The defect classification is performed using an additive gradient boosting model:

$$\hat{y}_i = \sum_{t=1}^{T} \alpha_t \, h_t(\mathbf{x}_i) \tag{7}$$

where $h_t(\cdot)$ denotes the t-th decision tree, α_t is its corresponding weight, and T is the total number of trees.

The model is trained by minimizing a regularized objective function:

$$\mathcal{L}_{\text{GBM}} = \sum_{i=1}^{N} l(y_i, \hat{y}_i) + \sum_{t=1}^{T} \Omega(h_t) \tag{8}$$

where $l(y_i, \hat{y}_i)$ denotes the classification loss function (in our case Jaccard index loss) and $\Omega(h_t)$ is a regularization term controlling model complexity.

The LightGBM model was trained using an 80%/10%/10% train/validation/test split defined at the image-level to prevent data leakage between subsets. In contrast to the CNNs, no image augmentation was applied, as geometric or radiometric transformations do not provide meaningful benefits for statistics-based decision tree models. At the image-level, this partition resulted in 4,187 training images, 522 validation images, and 550 test images.

Following dataset preparation, PV modules were extracted from normalized images and feature vectors were computed either per module or at the image-level for background and neighborhood comparisons. Prior to feature extraction, module contours were filtered using the same geometric validity criteria as in the multi-class CNN training pipeline. The resulting module-level datasets comprised 248,491 samples for training, 31,173 samples for validation, and 33,460 samples for testing. Model hyperparameters—including the number of trees, maximum tree depth, leaf count, subsampling ratios, and regularization coefficients—were optimized using an Optuna-based search strategy, with model selection guided by task-specific validation metrics. Representative prediction results obtained with the optimized LightGBM model are presented in Fig. 16.

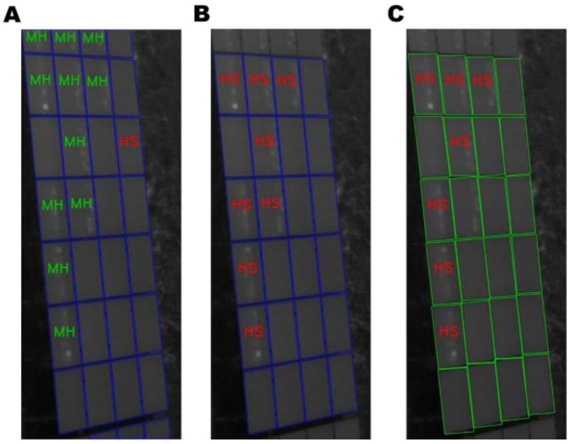

Figure 16. Representative defect classification results obtained using the LightGBM model. (A) Manually annotated ground-truth labels. (B) Defect predictions for ground-truth PV module regions. (C) Defect predictions for PV modules automatically detected by the module detection pipeline. Defect classes are color-coded as follows: Multi-Hotspot (MH) – light green; Hotspot (HS) – red.

The best-performing model identified during the optimisation trials was subsequently evaluated on the independent test set to determine its final performance. The outcomes of this evaluation are illustrated in Fig. 17. The optimized model achieved an accuracy of 99.5%, an F1 score of 93.1%, and a precision of 95.6%.

On the target deployment hardware, image normalization, feature extraction and inference performed by the LightGBM classifier requires approximately 90 ms per image, corresponding to about 1.5 ms per PV module assuming an average of 59.5 modules per frame. This execution time is more than 2.5 times shorter than that of the neural defect classification network. Owing to this computational efficiency, the LightGBM model can be employed as an initial screening stage that rapidly evaluates all detected modules and forwards only low-confidence or ambiguous cases—defined by a prediction confidence below a 95% threshold—to the CNN for further analysis.

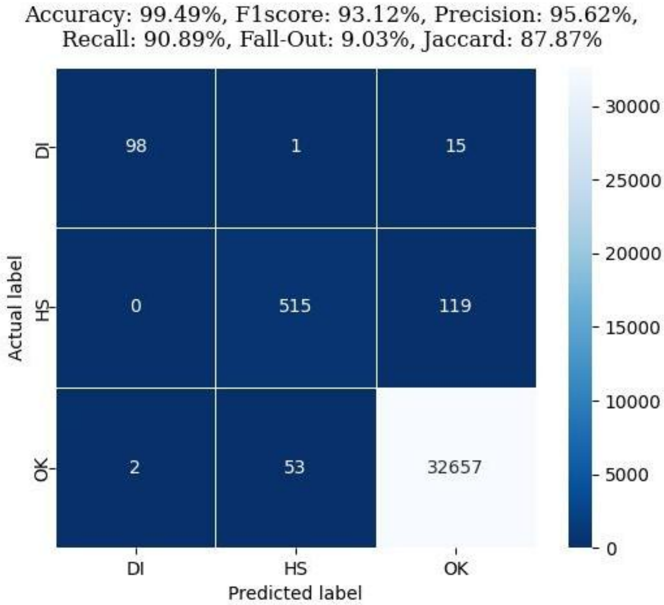

Figure 17. Confusion matrix of the LightGBM-based PV module defect classification model evaluated on the test set.

The 95% confidence threshold was determined through a systematic sensitivity analysis performed on the test subset, with the Jaccard index used as the primary decision metric. The end-to-end performance of the combined LightGBM–CNN classification pipeline was evaluated across a range of confidence thresholds, from 0% (LightGBM only) to 100% (CNN only). For each threshold, the resulting Jaccard index, F1 score, accuracy, precision, recall, and average inference time per image were recorded. The results, averaged over five independent timing runs, are summarized in Table 3.

The Jaccard index increases monotonically with the threshold, from 81.88% (LightGBM only) to 90.73% (CNN only). At the 95% threshold the Jaccard index reaches 90.46%, capturing 96.9% of the maximum achievable improvement relative to the LightGBM-only baseline, while the average defect classification inference time is reduced by approximately 67% compared to the CNN-only configuration (90.3 ms vs. 274.1 ms per image). Above 95%, the gain in classification quality becomes negligible — between 95% and the CNN-only configuration the Jaccard improves by only 0.27 pp — while below 95% it drops more substantially. Consequently, the 95% threshold was selected as the operating point that preserves the diagnostic quality of the hybrid pipeline while retaining the computational advantage of the LightGBM screening stage. The reported inference times are representative measurements obtained on the training workstation and may vary depending on the hardware configuration. This threshold is automatically re-determined after each model retraining cycle, ensuring adaptability to changes in the underlying data distribution.

Table 3. Sensitivity analysis of the LightGBM confidence threshold governing the redirection of samples to the CNN classifier. Classification metrics are computed on the held-out test subset (585 images). Inference times are averaged over five independent runs and reported as mean ± standard deviation. The selected operating point (95%) is highlighted in bold.

Confidence threshold	F1 score (%)	Jaccard (%)	Accuracy (%)	Precision (%)	Recall (%)	Avg. inference time per image (ms)
0% (LightGBM only)	89.05	81.88	99.13	88.29	90.25	70.4 ± 0.6
50%	92.05	86.19	99.39	92.49	91.74	77.6 ± 2.8
70%	93.05	87.73	99.46	93.26	92.91	78.8 ± 0.3
80%	93.81	88.93	99.53	94.55	93.16	81.1 ± 0.6
85%	94.07	89.35	99.55	94.85	93.39	82.5 ± 0.2
90%	94.51	90.06	99.58	95.10	93.97	85.3 ± 0.7
95% (selected)	**94.75**	**90.46**	**99.60**	**95.51**	**94.05**	**90.3 ± 2.0**
99%	94.76	90.47	99.60	95.15	94.40	105.6 ± 0.5
100% (CNN only)	94.92	90.73	99.61	94.79	95.04	274.1 ± 1.5

In practical deployment scenarios, this hierarchical processing strategy reduces the overall inspection time by approximately one half while maintaining diagnostic reliability.

4.4.3. NS semantic segmentation

The second deep learning component of the defect classification system is a specialized network dedicated to detecting anomalies that span multiple adjacent PV modules, referred to as the NS classifier. This model is implemented as a semantic segmentation CNN that performs binary segmentation to identify image regions associated with large-scale defects, such as not-working strings.

The NS classifier operates exclusively on samples labelled with the NS defect category and formulates the task as a pixel-wise segmentation problem. For each input image, the ground-truth annotation is provided in the form of a binary mask indicating areas affected by the defect. An example of such a mask is presented in Fig. 18.

The exclusive focus on a single defect category required a modified dataset preparation strategy. As in the preceding models, the dataset was partitioned into 80%/10%/10% training, validation, and test subsets; however, only images containing NS defects were considered, which substantially reduced the total number of samples. To mitigate class imbalance at the image-level, the training subset maintained an approximate OK-to-defect ratio of 2.8, ensuring a controlled predominance of healthy images while preserving sufficient representation of NS cases. This NS- centred selection resulted in 840 training, 103 validation, and 117 test images. To compensate for the limited dataset size, augmentation procedures were applied exclusively to the training subset and included geometric transformations such as flips, rotations, and cropping operations. In total, 13,312 augmented samples were generated from the original images. Additionally absolute temperature–shift augmentation, previously employed in the PV module detection pipeline, was also applied to the NS classifier and extended training dataset to 16,040.

The NS detector is implemented as a lightweight U-Net–like encoder–decoder semantic segmentation network with four skip connections. The architecture comprises 88 layers organized into 24 submodules. The input consists of a normalized single-channel radiometric image resized to 640 × 512, and the network produces a pixel-wise probability map through a final sigmoid-activated convolutional layer, yielding a binary NS likelihood mask at the same spatial resolution as the input.

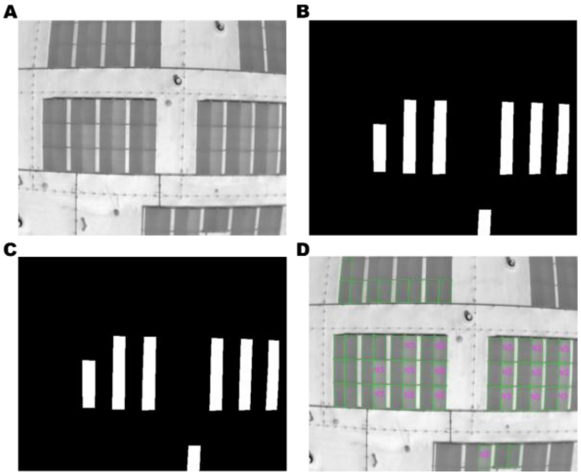

Figure 18. Example of image processing performed by the NS semantic segmentation classifier. (A) Normalized radiometric image. (B) Ground-truth binary segmentation mask. (C) Predicted NS segmentation mask. (D) Module-level NS defects mapped onto the corresponding detected PV modules. Defect class is color-coded as follows: Not-Working String (NS) – purple.

The final NS semantic segmentation model contains 251,713 trainable parameters, requires approximately 915 kB of storage and 244 kB after quantization, and achieves an average inference time of 12.3 ms per image. On the target deployment hardware, the end-to-end processing time—including image normalization and preprocessing—amounts to approximately 234 ms per image. The architecture of the final NS classifier is presented in Fig. 19.

To derive module-level NS decisions from the predicted segmentation map, the network's probability output is first thresholded using an empirically determined confidence value. The resulting binary mask is subsequently mapped to individual PV modules by evaluating each detected module contour independently. For every contour, the proportion of pixels classified as NS within its boundaries is computed, and the module is labelled as Not-Working String if this ratio exceeds 0.5; otherwise, it is classified as healthy. This procedure effectively translates pixel-wise semantic segmentation into the module-level decisions required by the inspection pipeline, allowing the model to capture spatially localized NS patterns while still producing actionable diagnostic labels for individual modules.

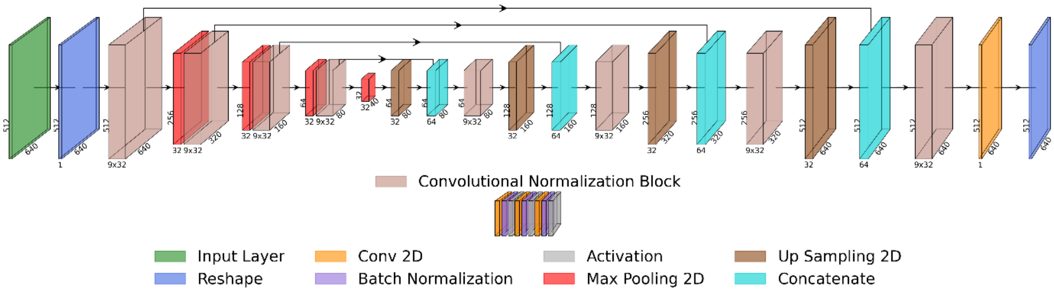

Figure 19. Architecture of the NS semantic segmentation CNN model.

The NS semantic-segmentation-based classifier achieved an accuracy of 98.2%, an F1 score of 86.7%, and a precision of 91.0% in per-module classification. The corresponding confusion matrix is presented in Fig. 20.

Taken together, the two learning pipelines and the gradient-boosting decision tree classifier constitute a complementary diagnostic framework capable of identifying both common defect categories and more specialized failure modes. The obtained results indicate that the adopted preprocessing procedures, dataset organization, and model architecture selections enable a PV module defect classification system that achieves accurate and computationally efficient high-throughput inference on radiometric imagery.

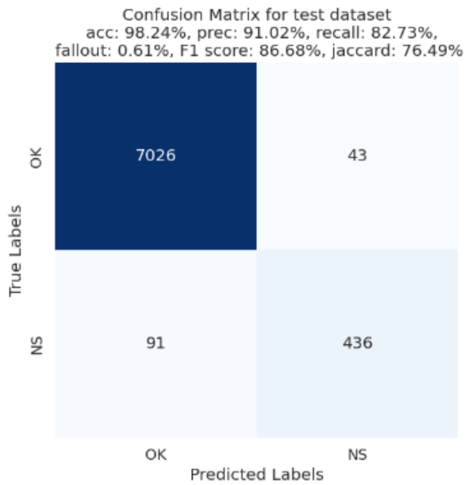

Figure 20. Confusion matrix of the NS semantic segmentation model evaluated at the PV module level on the test set.

5. Results and discussion

Since the proposed inspection framework operates as a two-stage process, comprising PV module detection followed by module and image-level defect classification—the overall performance was evaluated in an end-to-end manner to reflect realistic operational use. The assessment was conducted on a held-out test dataset using identical inspection-level data partitions as those employed during model development, thereby ensuring consistency and preventing information leakage between the detection and classification stages.

From the perspective of photovoltaic system operation and maintenance, the evaluation emphasized metrics relevant to reliability-oriented diagnostics rather than confidence-based object-detection indicators. The adopted measures included PV module detection recall, defect detection recall, missed-defect ratio (false-negative rate), false-alarm rate (false-positive rate), and overall inspection coverage at both image and inspection levels. In addition, computational indicators such as average processing time per image and achievable inspection throughput rates were analyzed to determine suitability for UAV-based deployments.

Representative inspection outcomes illustrating combined module detection and defect classification results are presented in Fig. 21, as well as Table 4 and Table 5. The reported results were obtained across 29 independent PV inspections, reflecting diverse environmental and operational conditions.

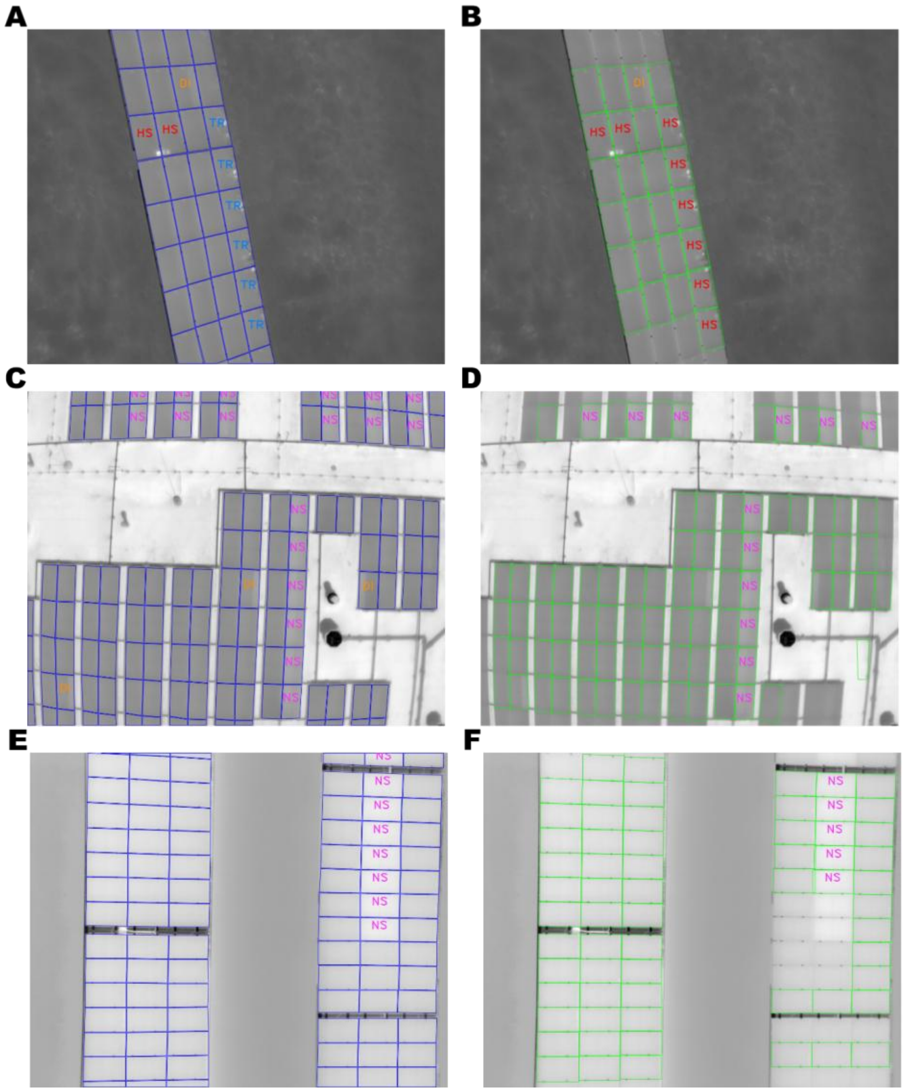

Figure 21. Representative results of the complete two-stage inspection pipeline combining PV module detection and defect classification. Each pair presents manually annotated ground-truth labels (A, C, E) and the corresponding final model outputs (B, D, F). Defect classes are color-coded as follows: Not-Working String (NS) – purple; Diode (DI) – orange; Hotspot (HS) – red; Vegetation/Shading (TR) – light blue.

Owing to the sequential structure of the proposed pipeline, the end-to-end inspection accuracy can be expressed as:

$$Acc_{e2e} = R_{PVmD} \cdot Acc_{PVmDC} \quad (9)$$

where R_{PVmD} denotes the recall of the PV module detection stage and Acc_{PVmDC} represents the conditional accuracy of the defect classification stage. A sample is considered correctly processed only if it is first detected as a valid module and subsequently classified without error. This joint formulation provides an unbiased estimate of practical system performance, as it simultaneously accounts for both missed detections and misclassifications, avoiding overly optimistic evaluations based on individual stage results alone.

The end-to-end per-image processing time was determined as the aggregate execution time of both stages averaged over the 734 radiometric images contained in the test subset. The obtained results indicate that the proposed dual-stage architecture achieves a favorable balance between diagnostic effectiveness and computational efficiency, supporting its applicability to large-scale UAV-based photovoltaic inspections.

Table 4. Performance summary of the proposed two-stage PV module detection and defect classification pipeline.

Overall performance metric	Result
PV module detection recall	88.0%
Defect detection recall	95.0%
Defect detection false-negative rate	8.6%
Defect detection fallout	7.4%
End-to-end inspection accuracy (PV Module Detection Recall × Defect Classification Accuracy)	83.9%
Average end-to-end inference time per image	478.8 ms

The false-negative rate is computed at the module level and includes undetected modules from the detection stage.

Limitations of proposed approach in practical deployment

Energy consumption constitutes an additional practical consideration for UAV-based inspection systems, particularly in the context of onboard or edge deployment. Based on the measured average end-to-end processing time of 478.8 ms per image and the full-load power consumption of the target deployment hardware (CPU + GPU platform), the approximate processing energy can be estimated as 64.6 J per image, corresponding to approximately 0.018 Wh/image. This value represents an engineering estimate derived from nominal hardware power characteristics and should not be interpreted as a direct measure of onboard UAV energy usage.

It is important to emphasize that the proposed system is primarily intended for offline or ground-based processing rather than execution directly onboard the UAV. In practical inspection workflows, radiometric data are typically processed after acquisition, often in conjunction with additional computationally intensive tasks such as large-scale data aggregation, visualization, and reporting. Consequently, the presented energy estimate serves mainly as a reference for computational efficiency rather than a constraint on UAV flight endurance.

Additionally, it is important to note that the scope of this work is limited to PV module detection and thermographic defect classification, and does not encompass quantitative panel efficiency analysis. The proposed pipeline is designed to operate independently of module-specific electrical parameters such as nominal power rating, temperature coefficients, or cell technology. The defect classification stage, covering hotspot, diode, and string-level fault categories, relies entirely on characteristic thermal patterns observable in radiometric imagery rather than on electrical specifications of individual panel models. The dataset encompasses 29 inspections across multiple PV installations with diverse module types and manufacturers, including cases where several module models coexisted within a single site. The consistent performance of both the detection and classification stages across this heterogeneous dataset confirms the module-agnostic nature of the approach.

Although the proposed pipeline operates on individual images, detected defects can be associated with approximate GPS coordinates available in image metadata, enabling practical localization at the string or row level. Fine-grained mapping is left for future work. The integration of lightweight inference models with autonomous flight planning mechanisms represents a promising direction for future photovoltaic inspection systems.

An important aspect of the obtained results concerns the impact of missed PV module detections on the overall inspection performance. The PV module detection recall of 88.0% implies that approximately 12% of modules are not explicitly identified during the first stage of the pipeline, which directly propagates to the end-to-end accuracy. However, from the perspective of photovoltaic system diagnostics, the practical impact of these missed detections is mitigated by several factors.

First, undetected modules are predominantly associated with low-quality acquisition conditions, such as extreme viewing angles, low spatial resolution, or partial occlusions, where reliable defect identification would be inherently uncertain even under manual inspection. Second, the statistical distribution of defects within the dataset indicates that failure occurrences are sparse and typically localized, which reduces the probability that missed modules systematically bias energy yield assessments at the installation level. Third, the proposed framework operates on large inspection datasets, where redundancy between overlapping UAV images further limits the likelihood of consistently missing the same defective module across multiple frames.

As a result, the impact of the 12% missed detections on overall energy efficiency analysis remains limited in practice. Nevertheless, this effect represents a known limitation of the current approach and constitutes an important direction for future work, particularly in the context of improving robustness under non-ideal acquisition conditions and enhancing module detection recall in edge-case scenarios.

Impact of Ground Sample Distance on Detection Performance

To evaluate the sensitivity of the proposed pipeline to varying spatial resolution, a per-inspection analysis was performed by correlating the computed GSD statistics with PV module detection recall and defect classification accuracy measured on the test set. For this analysis, only non-frame-adjacent modules were considered to eliminate edge effects, and defect classification was evaluated on modules detected by the upstream neural network to reflect realistic pipeline operation. Inspections were grouped into five categories based on their median GSD and intra-mission variability: below 3 cm/pixel, 3–5 cm/pixel, 5–8 cm/pixel, 8–12+ cm/pixel, and a separate group for inspections with GSD standard deviation exceeding 1 cm/pixel regardless of absolute value. The resulting per-group performance is summarized in Fig. 22.

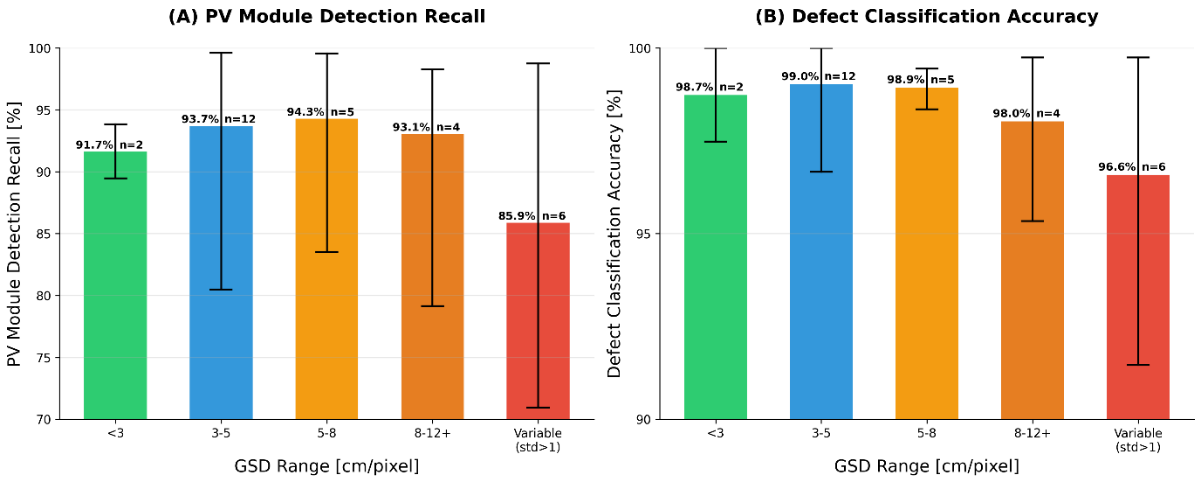

Figure 22. Impact of Ground Sample Distance on pipeline performance across 29 inspections. (A) PV module detection recall. (B) Defect classification accuracy. Inspections are grouped by median GSD range; the "Variable" group contains inspections with intra-mission GSD standard deviation exceeding 1 cm/pixel. Error bars indicate the range (min–max) within each group. A Kruskal-Wallis test across all five groups yielded $p = 0.58$ for detection recall and $p = 0.39$ for classification accuracy, indicating no statistically significant effect of GSD on pipeline performance within the examined resolution range.

To formally assess whether the observed differences between GSD groups are statistically meaningful, a Kruskal-Wallis test was applied across the five groups. The test yielded $H = 2.85$, $p = 0.58$ for PV module detection recall and $H = 4.14$, $p = 0.39$ for defect classification accuracy. Both p-values are well above the $\alpha = 0.05$ significance threshold, indicating that GSD variation within the examined range does not have a statistically significant effect on pipeline performance. This result suggests that the proposed approach is largely insensitive to GSD across the full range of spatial resolutions encountered in the dataset.

Although the group characterized by high intra-mission GSD variability (standard deviation > 1 cm/pixel) exhibits the lowest average module detection recall in Fig. 22, with individual inspection recall as low as 70.9%, this difference did not reach statistical significance at the group level due to high within-group variability. The observed tendency nevertheless indicates that altitude changes during a single mission—resulting in mixed high- and low-resolution frames within the same dataset—affect module detection more strongly than the absolute GSD value. The classification stage proved less sensitive to this effect, maintaining high accuracy even for the most challenging inspections.

These findings indicate that the absolute GSD value is not the primary determinant of pipeline performance within the examined range. The model's robustness to elevated GSD can be attributed to the deliberate inclusion of non-standard-compliant imagery in the training dataset, which spans the full range of spatial resolutions encountered in real-world UAV inspections. This design choice ensures that the proposed approach remains applicable even when the IEC TS 62446-3 resolution requirement of 3 cm/pixel is not met, as is frequently the case in operational scenarios.

Comparison With State-Of-The-Art

To further contextualize the obtained results and address the requirement for a broader comparative analysis, a benchmark comparison with selected state-of-the-art approaches is presented in Table 4. The included methods represent object detection, semantic segmentation, and classification frameworks applied to photovoltaic defect detection using thermographic or related imaging modalities, as introduced in Section 1.

Table 5. Comparative overview of selected state-of-the-art methods for photovoltaic defect detection, including object detection, segmentation, and classification approaches, with corresponding dataset sizes and reported performance metrics.

Model	Approach Type	Task	Task Type	Dataset Size (Images)	Metrics
RTPV-YOLO (2026) [9]	YOLO-based detector	UAV thermal + RGB defect detection	Object Detection	Not reported	mAP: 91%
Awedat et al. (2025) [13]	Enhanced U-Net (Residual + ASPP + Attention)	Thermal PV fault detection	Semantic Segmentation	1,009	F1: 0.868; IoU: 0.790; Precision: 0.911; Recall: 0.840
Setiawan & Fathurrahman (2025) [11]	Mask R-CNN	PV fault detection (thermography)	Object Detection + Instance Segmentation	660	mAP: 84.27%
PV-YOLOv12n (2025) [15]	YOLO-based detector	PV defect detection (EL dataset)	Object Detection	5,589 / 6,493	mAP@50: 0.91; mAP@50–95: 0.58–0.75
Pratticò et al. (2025) [14]	Efficient Attentive U-Net	Thermal PV fault diagnosis	Semantic Segmentation	100	mIoU: 81.4%; F1: 87.5%; Pixel Accuracy: 96.9%
HOTSPOT-YOLO (2025) [10]	Lightweight YOLO	Hotspot detection (thermal)	Object Detection	7,170	mAP@0.5: 90.8%
Lee et al. (2023) – LIRNet [8]	Lightweight CNN	PV defect classification (thermal)	Classification	20,000	Accuracy: 89.1%
Yang et al. (2023) [12]	Improved Mask R-CNN (Attention + FPN)	PV defect detection (thermal)	Object Detection + Instance Segmentation	3,383 + COCO2017	Accuracy: 89%
solAIr (2020) [2]	Mask R-CNN + segmentation models	Thermal anomaly detection & segmentation	Segmentation (Instance + Semantic)	1,009	Jaccard: 0.741 (U-Net); 0.748 (LinkNet); 0.499 (Mask R-CNN)
Proposed method (this work)	CNN + LightGBM (hybrid pipeline)	PV module detection + defect classification	Segmentation + Classification	30,983 (7,003 annotated)	Module detection recall: 88.0%; Defect recall: 95.0%; F1: 90.0%; End-to-end accuracy: 83.9%

It should be emphasized that a direct comparison between the reported results is inherently limited due to differences in datasets, sensor modalities (e.g. UAV thermography vs. electroluminescence imaging), annotation strategies, and evaluation protocols. In particular, the scale and diversity of datasets vary significantly—from fewer than 100 images to several thousand samples—while acquisition conditions, defect distributions, and ground-truth definitions are not standardized. Consequently, the reported metrics should be interpreted as indicative of general performance trends rather than strictly comparable quantitative benchmarks.

Despite these limitations, several important observations can be made. Object detection approaches based on YOLO architectures [9,10,15] achieve high detection accuracy and are well suited for real-time applications. However, their reliance on bounding-box representations introduces fundamental limitations in photovoltaic inspection scenarios. In practical deployments, PV modules are often densely packed, partially occluded, or arbitrarily oriented, which leads to bounding boxes overlapping multiple modules or failing to precisely isolate defect locations. From an operational perspective, this limitation directly affects the usability of inspection results, as industry workflows require each PV module to be uniquely identifiable, spatially consistent, and individually traceable within inspection reports.

Similarly, Mask R-CNN–based approaches [2,11,12] improve localization through instance segmentation; however, they typically involve higher computational complexity and reduced scalability when applied to large-scale UAV inspections. Moreover, their deployment in real-world scenarios often remains constrained by processing time, resource requirements, and limited validation on large and diverse datasets.

Semantic segmentation approaches based on U-Net variants [13,14] enable detailed spatial characterization of defects and demonstrate strong performance metrics. Nevertheless, these methods are frequently evaluated on relatively small datasets and focus primarily on pixel-level accuracy rather than end-to-end inspection usability. Lightweight classification models such as LIRNet [8] provide efficient inference; however, they operate on pre-cropped inputs and do not address the upstream challenge of robust PV module detection and localization.

From the perspective of industrial deployment, an additional critical requirement concerns the structure and interpretability of inspection outputs. In practical photovoltaic maintenance workflows, inspection results must support interactive analysis, where each module is explicitly delineated, uniquely identifiable, and linked to its physical location within the installation. This enables inspectors and asset owners to review, validate, and, if necessary, correct automated predictions, as well as to integrate results into maintenance planning and reporting systems. Methods based solely on bounding boxes or pixel-wise segmentation without structured module representation may not fully satisfy these requirements.

An additional practical consideration concerns deployment constraints. Recent YOLO-based implementations (e.g., YOLOv8 and later) are often distributed under the AGPL-3.0 license, which may impose additional licensing constraints in certain deployment scenarios, particularly when the software is used within network-accessible services. In such cases, compliance may involve source code availability requirements or the acquisition of a commercial license. Although this aspect is rarely discussed in the technical literature, it represents an important factor in real-world industrial applications.

In contrast to the above approaches, the proposed method introduces a unified two-stage inspection framework explicitly designed for operational deployment in large-scale UAV-based photovoltaic inspections. The first stage employs semantic segmentation combined with dedicated post-processing to extract geometrically consistent, instance-level representations of PV modules, ensuring that each module is uniquely identified regardless of orientation, scale, or spatial arrangement. This directly addresses a key limitation of bounding-box–based methods and enables the generation of structured, module-level inspection outputs required in industrial practice.

The second stage integrates a hybrid classification strategy combining LightGBM and CNN models, allowing efficient and accurate identification of reliability-critical defects at both module and string levels. This design provides a balance between computational efficiency and diagnostic performance, while also enabling hierarchical decision-making and reduced inference time through selective model invocation.

A further distinguishing feature of the proposed approach is its validation on a large-scale, real-world dataset. As summarized in Table 1, this dataset exhibits significantly greater scale, variability, and operational realism than those employed in most existing studies.

As summarized in Table 4, the proposed method achieves high defect detection recall (95.0%) and competitive end-to-end accuracy (83.9%) while maintaining low computational complexity suitable for near real-time processing. Importantly, unlike many existing approaches that address only individual components of the inspection problem, the presented framework covers the complete processing pipeline—from module detection and geometric localization to defect classification and reporting—thereby providing a comprehensive and deployment-oriented solution aligned with practical requirements of photovoltaic system inspection and maintenance.

From a practical deployment perspective, the scalability of the proposed methodology has been validated through its application in real industrial inspection workflows. The system has been applied to multiple large-scale PV installations, where it supported automated processing of full UAV inspection campaigns comprising thousands of radiometric images. In these scenarios, the two-stage architecture enabled stable high-throughput analysis while maintaining consistent diagnostic performance across diverse environmental conditions and acquisition parameters. Importantly, the structure of the generated outputs—based on explicitly delineated and uniquely identifiable PV modules—allowed direct integration with existing reporting and asset management systems used by inspectors and asset owners.

Feedback obtained from industrial users indicates that the ability to associate each detected defect with a specific module, combined with the possibility of manual verification and correction, significantly improves the usability and trustworthiness of automated inspection results. In particular, operators emphasized the importance of structured, module-level outputs for maintenance planning, warranty claims, and energy yield analysis. These observations confirm that the proposed approach not only achieves high detection performance but also satisfies key practical requirements of scalable, deployment-oriented PV inspection systems, supporting early fault detection and reducing the risk of undetected degradation processes in operational environments.

6. Conclusions

This paper presented an integrated UAV-based thermographic inspection methodology for photovoltaic installations, designed to support reliability-oriented condition monitoring under real operational conditions. The proposed framework combines structured data acquisition, large-scale dataset preparation, and a two-stage analytical pipeline comprising PV module detection and defect classification. By explicitly separating geometric localization from diagnostic inference, the approach addresses key limitations of existing methods, particularly in scenarios involving non-uniform module orientation, high module density, and deviations from standardized acquisition conditions.

The obtained results demonstrate that the proposed two-stage architecture enables effective identification of reliability-critical defects while maintaining low computational complexity. In contrast to single-stage detection approaches, the use of semantic segmentation combined with dedicated post-processing allows for accurate, instance-level extraction of PV modules, ensuring their geometric consistency and individual traceability. This capability is essential for practical deployment, as real-world inspection workflows require each module to be uniquely identifiable and directly linked to maintenance and reporting processes.

From the perspective of industrial application, the proposed framework provides structured and interpretable inspection outputs that support interactive analysis and decision-making. The ability to assign defects to individual modules enables inspectors and asset owners to verify automated predictions, introduce manual corrections where necessary, and integrate results into maintenance planning systems. In this context, the presented solution goes beyond conventional detection or classification models by addressing the complete inspection pipeline, including module-level localization, defect identification, and result usability.

An additional contribution of this work lies in its validation on a large-scale, real-world dataset comprising 30,983 thermographic images collected during 29 UAV inspections over a five-year period. Compared to existing studies, which are typically based on smaller and less diverse datasets, the scale and variability of the data used in this study contribute to improved robustness and generalization under practical operating conditions. The achieved defect detection recall of 95.0% and end-to-end inspection accuracy of 83.9%, combined with near real-time processing capability, confirm the suitability of the proposed approach for deployment in large-scale UAV-based inspection systems.

Overall, the presented methodology provides a scalable and deployment-oriented solution for early fault detection in photovoltaic installations, supporting improved operational reliability, reduced maintenance costs, and enhanced decision-making in asset management.

Future work will focus on extending the proposed framework toward fully integrated inspection systems, including the incorporation of georeferencing and mapping capabilities, advanced radiometric correction methods, and adaptive UAV mission planning. In addition, further validation across different climatic conditions and installation types will be conducted to enhance the generalization and long-term applicability of the approach.

Acknowledgment

This research was partially supported by the National Centre for Research and Development (NCBR) under grant no. PPOIR.01.01.01-00-0491/20-00. The authors express gratitude to partner PV farm operators for their collaboration and data contributions.

References

1. Ang TZ, Salem M, Kamarol M, Das HS, Nazari MA, Prabaharan N. A comprehensive study of renewable energy sources: classifications, challenges and suggestions. *Energy Strategy Reviews* 2022; 43: 100939. https://doi.org/10.1016/j.esr.2022.100939.
2. 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(24): 6496. https://doi.org/10.3390/en13246496.
3. Hussain M, Sultan M, Uzma F, Longsheng C, Malik MY, Butt AR, Sajjad A, Younis I, Imran M. A comparative analysis of renewable and non-renewable energy generation to relegate CO_2 emissions and general costs in household systems. *Environmental Science and Pollution Research* 2022; 29(52): 78795-78808. https://doi.org/10.1007/s11356-022-21121-0.
4. Thapar S. Centralized vs decentralized solar: a comparison study (India). *Renewable Energy* 2022; 194: 687-704. https://doi.org/10.1016/j.renene.2022.05.117.
5. de Oliveira AKV, Aghaei M, Rüther R. Automatic inspection of photovoltaic power plants using aerial infrared thermography: a review. *Energies* 2022; 15(6): 2055. https://doi.org/10.3390/en15062055.
6. Ling H, Liu M, Fang Y. Deep edge-based fault detection for solar panels. *Sensors* 2024; 24(16): 5348. https://doi.org/10.3390/s24165348.
7. Li X, Yang Q, Lou Z, Yan W. Deep learning based module defect analysis for large-scale photovoltaic farms. *IEEE Transactions on Energy Conversion* 2019; 34(1): 520-529. https://doi.org/10.1109/TEC.2018.2873358.
8. Lee SH, Yan LC, Yang CS. LIRNet: a lightweight inception residual convolutional network for solar panel defect classification. *Energies* 2023; 16(5): 2112. https://doi.org/10.3390/en16052112.
9. Su Z, Zhao K, Fan Z, Guo X. RTPV-YOLO: real-time photovoltaic detection with UAV-based thermal and RGB imaging. *IEEE Transactions on Aerospace and Electronic Systems* 2025; 61(5): 13799-13810. https://doi.org/10.1109/TAES.2025.3580391.
10. Dhimish M. HOTSPOT-YOLO: a lightweight deep learning attention-driven model for detecting thermal anomalies in drone-based solar photovoltaic inspections. *arXiv* 2025. https://doi.org/10.48550/arXiv.2508.18912.
11. Setiawan EA, Fathurrahman M. Exploring insights on deep learning-based photovoltaic fault detection for monofacial and bifacial modules using thermography. *International Journal of Cognitive Computing in Engineering* 2025; 6: 495-507. https://doi.org/10.1016/j.ijcce.2025.04.001.
12. Yang W, Deng Z, Hu E, Zhang Y. Improved Mask R-CNN network method for PV panel defect detection. In: *Proceedings of the 2023 Asia Conference on Computer Vision, Image Processing and Pattern Recognition* (CVIPPR '23); Phuket, Thailand. New York, NY: Association for Computing Machinery; 2023. Article 1, p. 1-7. https://doi.org/10.1145/3596286.3596287.
13. Awedat K, Comert G, Ayad M, Mrebit A. Advanced fault detection in photovoltaic panels using enhanced U-Net architectures. *Machine Learning with Applications* 2025; 20: 100636. https://doi.org/10.1016/j.mlwa.2025.100636.
14. Pratticò D, Laganà F, Versaci M, Franković D, Jakoplić A, La Foresta F. Efficient attentive U-Net for fault diagnosis and predictive maintenance of photovoltaic panels through infrared thermography. *Energies* 2025; 18(24): 6472. https://doi.org/10.3390/en18246472.
15. Mohamed A, Yassa N, Bouzida A, Ali T, Belabbaci EO, Rabehi A, Alsabah YA, Benghanem M. Optimized YOLO based model for photovoltaic defect detection in electroluminescence images. *Scientific Reports* 2025; 15(1): 32955. https://doi.org/10.1038/s41598-025-13956-7.
16. International Electrotechnical Commission. IEC TS 62446-3:2017 *Photovoltaic (PV) systems – Requirements for testing, documentation and maintenance – Part 3: Photovoltaic modules and plants – Outdoor infrared thermography*. Geneva: International Electrotechnical Commission; 2017.
17. Vourkos EG, Christoforou EG, Panayides AS, Kalogirou SA, Agathokleous RA. Optimal aerial imaging parameters for UAV-based inspection and maintenance of photovoltaic installations. *Energies* 2025; 18(21): 5818. https://doi.org/10.3390/en18215818.

18. Reges JP, Moreira FDL, Bezerra LDS, de Alexandria AR, Rebouças Filho PP. Thermographic image processing application in solar followers. *IEEE Latin America Transactions* 2015; 13(10): 3350-3358. https://doi.org/10.1109/TLA.2015.7387242.
19. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu TY. LightGBM: a highly efficient gradient boosting decision tree. In: *Advances in Neural Information Processing Systems 30* (NIPS 2017); Long Beach, CA. Red Hook, NY: Curran Associates; 2017. p. 3146-3154.