e-ISSN: 2689-4934

Available online at
http://uppp.uitm.edu.my/online-issues.html

ESTEEM Academic Journal Vol 22, March 2026, 58-72

ESTEEM
Academic Journal
(EAJ)

Performance improvement of CNN-based model for multiclass hotspot severity classification in photovoltaic thermal imagery

Nurul Huda Ishak[1], Iza Sazanita Isa[2]*, Muhammad Khusairi Osman[3], Kamarulazhar Daud[4], Mohd Shawal Jadin[5], Noor Fadzilah Razali[6]

[1, 2, 3,4, 6]*Electrical Engineering Studies, Universiti Teknologi MARA, Cawangan Pulau Pinang, Permatang Pauh Campus, 13500 Pulau Pinang, Malaysia*
[5]*Faculty of Electrical and Electronics Engineering Technology, Universiti Malaysia Pahang Al-Sultan Abdullah, 26600 Pekan, Pahang, Malaysia*

ARTICLE INFO

Article history:
Received 27 January 2026
Revised 13 March 2026
Accepted 28 March 2026
Online first
Published 1April 2026

Keywords:
Photovoltaic
Convolutional Neural Networks
Hotspot
Severity
Classification

DOI:
10.24191/esteem.v22iMarch.10388

ABSTRACT

Intelligent diagnostic models that can identify and categorise thermal anomalies affecting efficiency and safety are essential for the reliable operation of photovoltaic (PV) systems. One of the most serious defects is thermal hotspots, which frequently result in energy loss and potential fire hazards. Convolutional Neural Networks (CNNs) have demonstrated promising performance when applied to aerial thermal images for PV fault detection. However, most existing models are limited to binary fault identification, thereby constraining their applicability for risk-based maintenance. Conventional CNN architectures also lack an optimised flatten-layer representation and fail to capture the full variability in hotspot severity. This study presents an enhanced CNN-based framework for multiclass classification of hotspot severity in PV modules. To improve feature abstraction and class separability across severity levels, the proposed model employs a staged architectural refinement strategy by progressively adding convolutional layers to a baseline CNN. Hotspot regions extracted from unmanned aerial vehicles (UAV)-acquired thermal imagery were categorised into low, medium, and high severity levels for model training and evaluation. The improved model achieved a 9.1% increase in accuracy and a 7.6% increase in F1-score, outperforming the baseline CNN, thus confirming its superior discriminative learning capability and diagnostic robustness. These findings demonstrate that architectural deepening and flatten-layer optimisation can advance automated thermographic inspection toward severity-aware predictive maintenance for PV systems.

[2] Corresponding author. *E-mail address*: izasazanita@uitm.edu.my
https://doi.org/ 10.24191/esteem.v22iMarch.10388

1. INTRODUCTION

The increasing demand for intelligent diagnostic systems to ensure the safe and reliable operation of solar assets aligns with the global growth in photovoltaic (PV) installations [1]. However, in PV operation, localised thermal hotspots remain one of the most critical failure modes, contributing to material degradation and reduced lifespan. These thermal hotspots accelerate PV material ageing, reduce electrical efficiency, and may lead to increased safety risks, including module ignition under severe conditions [2].

Meanwhile, unmanned aerial vehicles (UAVs) have emerged as a practical, non-invasive method for large-scale PV monitoring through aerial thermographic inspection. It provides rapid visualisation of suspected abnormal localised temperature variations across large solar farms [3]. Fig.1 presents the visual and thermal representations of a PV module acquired through UAV-based inspection. While the visual image in Fig. 1(a) conveys structural information, the thermal image in Fig. 1(b) reveals localised temperature anomalies associated with hotspot formation through changes in colour formation.

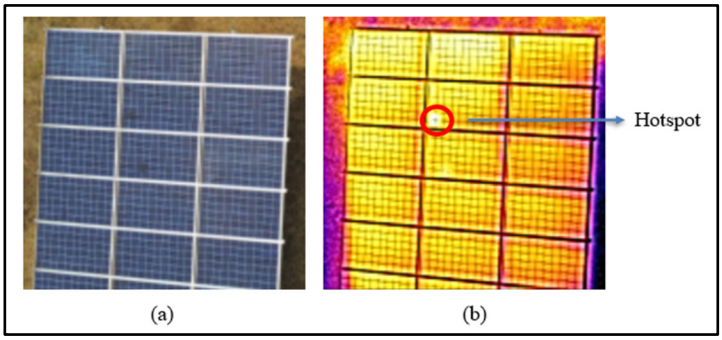

Fig. 1. (a) Visual image of a photovoltaic module captured using UAV; (b) Corresponding thermal image highlighting a localised hotspot region

In recent studies, deep learning (DL) models using Convolutional Neural Networks (CNNs) architecture have proven effective for analysing PV thermal images. The implementation of CNNs using artificial intelligence-based techniques has improved the automation of PV fault diagnosis. However, most existing CNN models are limited to binary classification tasks that only distinguish between defective and non-defective modules [4]. Moreover, conventional CNNs pay limited attention to localised hotspot regions and instead rely on global thermal properties within an image. This constraint consequently limits the model's capacity to detect subtle spatial feature variations and temperature fluctuations within the affected region [5]. This generalisation to assess the PV module disregards the aspects of varying levels of thermal severity, which is crucial for risk-based maintenance, as is prioritised during inspection planning [5].

To address these limitations, this study proposes an enhanced CNN framework that emphasises feature representation at the early convolutional layers. These layers capture local spatial structure and fine-grained thermal variations, which are important for consistent severity differentiation in later processing stages [6]. The proposed architecture strengthens the convolutional hierarchy and refines the flattened feature representations to improve feature abstraction and separability across severity levels. This study also enhances feature representation by incorporating thermal hotspot anomalies as important features for determining PV module defects through multiclass hotspot severity classification.

In this study, hotspot regions obtained from collected UAV-acquired thermal imagery were categorised into low-, medium-, and high-severity classes for model training and evaluation. The proposed framework advances automated PV fault diagnostics toward a more severity-aware and predictive maintenance

https://doi.org/ 10.24191/esteem.v22iMarch.10388

strategy. Instead of increasing network depth, the model refines task-relevant convolutional features and improves feature mapping within localised thermal regions.

2. LITERATURE REVIEW

Due to its non-contact capability and efficacy in identifying localised temperature anomalies linked to cell defects, connection failures, and mismatch circumstances, thermal imaging has been widely employed for PV hotspot analysis [7]. In the past, manual thermographic inspection commonly relied on temperature-difference (ΔT) thresholding, where ΔT represents the temperature contrast between a hotspot region and the surrounding healthy module area. This approach is widely used to detect abnormal thermal signals associated with PV module defects.

Further studies have improved this inspection method by incorporating UAV-assisted data acquisition to support rapid, large-scale condition assessments of PV installations [8], [9]. More recent work has introduced learning-based techniques to automate hotspot detection and reduce reliance on manual interpretation. Despite these developments, the interpretation of hotspot behaviour remains inconsistent across the literature, specifically when relating thermal patterns to severity assessment under varying operating conditions [10]. Thermal signatures are strongly influenced by multiple external factors such as irradiance variation, ambient temperature and module ageing. These factors complicate direct severity assessment when image-based information is used [11]. Although several studies recognise these challenges, many existing frameworks continue to rely on simplified binary interpretation strategies. Overall, the literature indicates that while thermal imaging provides a reliable basis for hotspot detection, accurate severity-aware interpretation requires more advanced modelling. Effective severity classification must integrate spatial-thermal characteristics with contextual operating conditions to support informed decision-making in PV inspection and maintenance applications. To address these limitations, deep learning approaches, particularly CNNs, have been widely adopted for their ability to automatically learn discriminative spatial features from thermal imagery without requiring handcrafted feature extraction. CNN-based models have demonstrated strong performance in capturing complex thermal patterns and spatial variations, making them suitable for hotspot detection and severity classification tasks in PV systems [12], [13].

Recent studies have explored multimodal strategies by combining thermal imagery with contextual or sensor-derived parameters to improve diagnostic performance [14], [15]. While these approaches enhance feature representation, they are rarely designed with severity-aware objectives, particularly in distinguishing different levels of hotspot intensity. Consequently, the integration of spatial–thermal descriptors remains essential for enabling more reliable and interpretable hotspot severity classification. Despite these advancements, systematic feature fusion strategies that effectively combine image and temperature-based representations remain underexplored in the current literature, highlighting the need for more structured and application-specific modelling approaches.

Although deep learning approaches combined with thermographic imaging demonstrate strong performance in detecting hotspot occurrences, systematic hotspot severity categorisation remains relatively underexplored in PV research. Many recent studies formulate hotspot detection as a binary classification task, focusing on the presence or absence of anomalies rather than differentiating severity levels [16], [17]. Some works have introduced rule-based criteria or ΔT-based categorisation to assess severity. However, variation in threshold definitions and methodological choices across studies limits reliability and general applicability [11]. More recent research has explored class-based defect classification using CNN-based techniques and machine learning approaches. However, severity labelling is often constrained by limited datasets and simplified feature representations [18]. In addition, image-based severity models frequently struggle to capture minor thermal variations associated with early-stage deterioration, particularly under changing environmental conditions [19]. Overall, the literature indicates that effective severity

https://doi.org/ 10.24191/esteem.v22iMarch.10388

classification requires more structured formulations such as integrating spatial-term features with learning-based models, rather than relying solely on binary detection.

Nevertheless, beyond feature representation, architectural design also plays a critical role in improving classification performance. Several studies have shown that incorporating task-specific architectural refinements, such as adding dedicated convolutional layers, can enhance feature discrimination without requiring a complete network redesign [37], [38]. This approach facilitates deeper interaction among learned representations and enables the extraction of more relevant high-level features associated with fault characteristics [39]. Rather than arbitrarily increasing network depth, strategically integrating additional convolutional layers has been found to be more effective in improving feature abstraction and classification capability. This observation motivates the proposed enhancement strategy, which incorporates additional layers to improve hotspot severity classification performance.

3. METHODOLOGY

To evaluate CNN performance for multi-class hotspot severity classification using PV thermal images, this study implements a structured methodology. The proposed workflow is designed to overcome the limitations of existing image-centric and binary-based approaches by integrating spatial-thermal feature information. Fig. 2 shows the overall workflow of the CNN-based model framework for hotspot severity classification.

The process begins with thermal image acquisition, followed by image pre-processing that consists of labelling and resizing procedures. The CNN-based model is then applied to represent hotspot features. In the final stage, model performance is evaluated using standard quantitative metrics. Overall, the framework is organised into a sequence of dependent stages. Each stage is designed to contribute systematically to the overall improvement in model performance. A performance enhancement module is subsequently introduced to refine feature representation and improve classification robustness. This framework establishes a clear progression from baseline modelling to the proposed improvements.

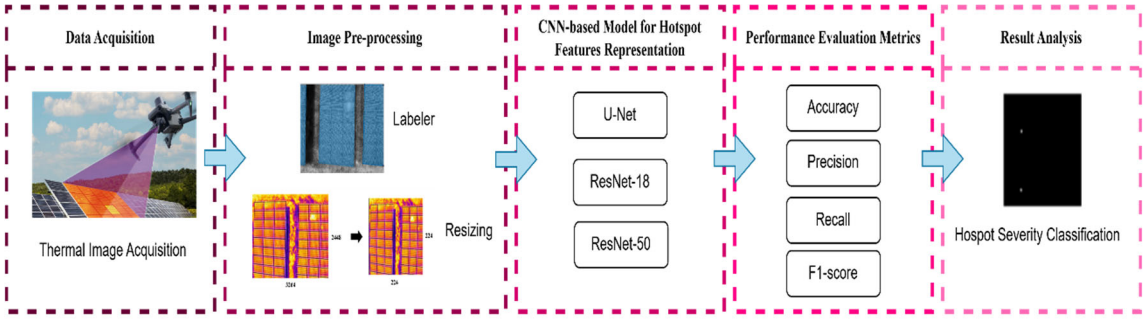

Fig. 2. Framework of the proposed CNN-based model for hotspot severity classification and performance evaluation

3.1 Thermal image acquisition

The dataset used in this study consists of thermal images acquired from a utility-scale solar PV farm located in Kelantan, Malaysia. Thermal image acquisition was conducted using a Parrot ANAFI thermal camera. All the images were recorded at a resolution of 72 dpi. Aerial data collection was performed at a flight height of 15 m. The altitude provided full coverage of the PV array while maintaining sufficient spatial detail for hotspot analysis. During image acquisition, the PV module operated at an average temperature of approximately 34.43 °C. The recorded solar irradiation ranged from 624.72 to 805.70 W/m², which is suitable for detecting anomalies and hotspot-related failures. In total, 670 thermal images were

https://doi.org/ 10.24191/esteem.v22iMarch.10388

collected. Out of these, 536 images (80%) were used for training, while the remaining 134 images (20%) were reserved for testing and validation.

3.2 Image pre-processing

Raw thermal data is often affected by noise and uneven thermal contrast. In addition, background temperature variations may arise from environmental and operational conditions. If left unaddressed, these factors can obscure fine-scale thermal patterns and reduce the reliability of subsequent learning-based analysis. Therefore, image pre-processing was applied to improve the quality and consistency of the acquired thermal images.

Pre-processing steps were performed before model training to prepare the thermal dataset for subsequent segmentation and learning tasks. This step included manual annotation and image resizing. Image annotation was carried out using the MATLAB Image Labeler app to identify solar PV and hotspot regions. This process enabled the generation of reliable ground-truth data.

The annotated regions of interest were reviewed by domain experts in PV thermal inspection. This validation step helped ensure consistent and accurate annotations across the dataset. Regions of interest (ROIs) were manually delineated using polygonal annotations to represent hotspot boundaries and surrounding module areas.

All thermal images were resized to a uniform resolution of 224 × 224 pixels to maintain consistency across samples. This resizing step reduces computational cost and memory usage during model training. In addition, it ensures uniform spatial dimensions for subsequent CNN-based processing.

3.3 CNN-based model for hotspot features representation

A CNN-based model is used to represent hotspot features for severity classification. The proposed architecture is designed to process spatial features extracted from hotspot areas while incorporating additional temperature-based descriptors. In addition, a hybrid, multi-input feature representation architecture is used to facilitate hotspot severity classification. To improve severity classification, the design uses image-based spatial attributes extracted from hotspot areas. The integration of these features with comprehensive temperature-based feature descriptors is illustrated in Fig. 3. This design enables joint learning of spatial and thermal information, thereby improving robustness and interpretability in severity assessment.

Fig. 3 shows the flow of the proposed architecture that consists of two parallel processing branches for temperature-based feature descriptors and image-based spatial features. The image input branch employs a CNN backbone with progressive convolutional refinement that extracts spatial feature representations. In contrast, the temperature feature input branch processes thermal descriptors of each training image through fully connected layers. The outputs from both branches are subsequently fused to form a unified feature representation for hotspot severity classification.

The proposed contributions are highlighted, with the progressive convolutional refinement module serving as an architectural enhancement for improved spatial feature extraction. In addition, the feature fusion process through the concatenation layer integrates CNN-derived spatial features with temperature-based descriptors, enabling more effective severity discrimination. These components collectively form the primary contribution of the proposed framework.

https://doi.org/ 10.24191/esteem.v22iMarch.10388

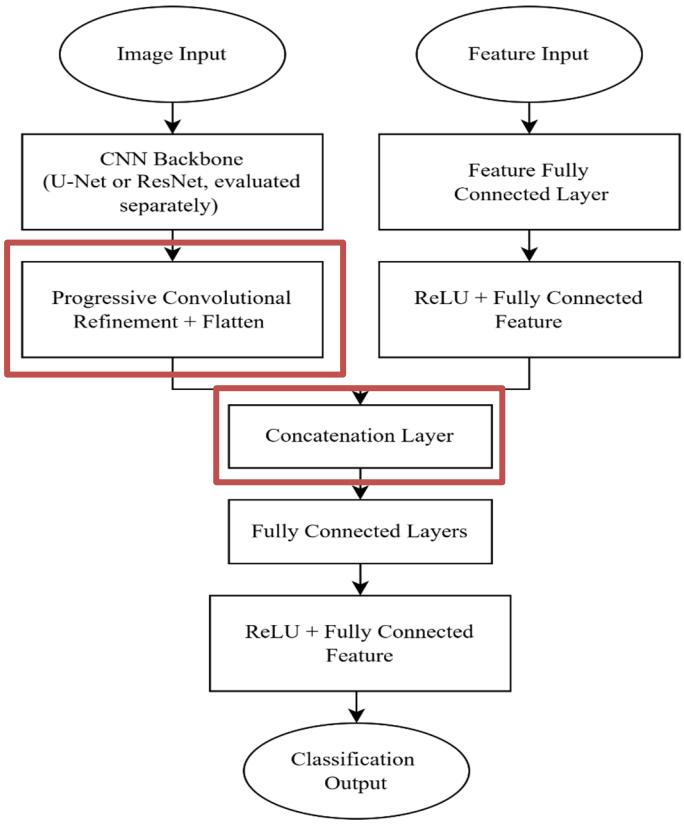

Fig. 3. Hybrid multi-input CNN architecture integrating CNN-derived spatial features and temperature-based descriptors through progressive convolutional refinement and feature fusion for hotspot severity classification

Baseline CNN architecture and dual-input feature representation

The baseline CNN architecture in this study is derived from standard backbone networks, namely U-Net and ResNet, which consist of convolutional feature extraction layers followed by feature flattening and fully connected layers for classification. In this configuration, no additional convolutional refinement layers were incorporated, and the extracted features were directly used for classification after the backbone processing stage.

To enhance feature representation capability, additional convolutional layers were introduced in the proposed framework. These layers were appended after the backbone feature extraction stage to further refine spatial features prior to classification. The effect of architectural refinement was systematically investigated by progressively increasing the number of additional convolutional layers, enabling a comparative analysis between the baseline and enhanced configurations.

In the image-based domain, pre-processed hotspot ROIs were used as input to a convolutional backbone utilising U-Net and ResNet architectures. U-Net was employed due to its encoder-decoder structure with skip connections, which enables effective localisation of spatial features in segmentation tasks [20], [21]. ResNet facilitates robust feature extraction through residual learning, allowing deeper networks to improved representation capability [22], [23].

https://doi.org/ 10.24191/esteem.v22iMarch.10388

The backbone learns hierarchical spatial representations associated with hotspot morphology and thermal distribution patterns. These features also capture complex thermal distribution patterns, as demonstrated in convolutional neural network-based image analysis studies [12], [13]. A concise spatial feature vector that encapsulates essential image-based characteristics relevant to severity classification is generated by combining and flattening the resulting feature maps.

Additional thermal descriptors were processed concurrently by the temperature feature branch via a fully connected encoding layer and a nonlinear activation function [24]. The thermal input was formulated as a temperature feature vector, denoted as x_T. This temperature feature summarises quantitative temperature information extracted from hotspot regions. In this study, x_T is defined as Eq. (1):

$$x_T = \left[T_{Max}, T_{Min}, T_{Avg}, \Delta T\right] \tag{1}$$

where T_{Max} represents the maximum temperature, T_{Min} represents the minimum temperature, T_{Avg} represents the average temperature and ΔT denotes the relative temperature difference between the hotspot and a reference region. These descriptors provide explicit thermal information that may not be fully captured by image-based spatial features alone.

The thermal feature vector was subsequently encoded using a fully connected layer followed by a nonlinear activation function, expressed as Eq. (2):

$$z_T = \emptyset(W_T x_T + b_T) \tag{2}$$

where W_T and b_T denote the weight matrix and bias term respectively, while $\emptyset$ represents a nonlinear activation function. The encoding step converts the raw temperature descriptors into a compact and discriminative form. These thermal features are subsequently combined with image-based spatial features to support joint spatial–thermal learning in hotspot severity classification. The integration of CNN-derived spatial features with temperature-based descriptors represents a key contribution to the proposed framework. Unlike conventional CNN-based approaches that rely solely on image features, this dual-input representation enables the model to incorporate both spatial and thermal characteristics for improved hotspot severity classification. This integration enhances the model's ability to distinguish visually similar hotspots with different thermal intensities, thereby improving classification robustness and interpretability.

Architectural refinement via progressive convolutional layer addition

In this study, a progressive architectural refinement process was used to improve feature abstraction and class separability across different hotspot severity levels. The CNN architecture was initially configured with the baseline backbone described in Section 3.3.1. This baseline CNN architecture in Section 3.3.1 was first used to establish a reference level of performance. Additional convolutional layers were progressively integrated into the image-based branch to refine spatial feature representations in a controlled and systematic manner.

Under this progressive refinement strategy, multiple refined configurations were evaluated by progressively increasing the number of additional convolutional layers from one to five. The progressive convolutional refinement configurations are summarised in Table 1 to clearly illustrate the incremental addition of convolutional layers in the image-based branch. This staged design encourages stronger interaction among learned feature maps and supports the extraction of higher-level spatial features related to hotspot severity. Stepwise refinement also enables a systematic examination of the effect of convolutional depth on classification performance.

https://doi.org/ 10.24191/esteem.v22iMarch.10388

Table 1. Progressive convolutional refinement configurations

Model Variant	Backbone	Additional Convolutional Layers	Description
Baseline	U-Net/ResNet	0	Standard backbone without refinement
Variant 1	U-Net/ResNet	1	One additional convolutional layer added after feature extraction
Variant 2	U-Net/ResNet	2	Two additional convolutional layers for deeper feature refinement
Variant 3	U-Net/ResNet	3	Three additional convolutional layers for enhanced feature abstraction
Variant 4	U-Net/ResNet	4	Four additional convolutional layers for increased feature representation capacity
Variant 5	U-Net/ResNet	5	Five additional convolutional layers for further feature abstraction and refinement

All configurations were trained and evaluated under identical experimental conditions. This ensures a fair comparison between the baseline and refined models. Any observed performance differences are attributed to the inclusion of additional convolutional layers.

Feature fusion and severity classification

After the convolutional feature extraction and refinement, features from the image-based and temperature-based branches were integrated using a fusion layer, as shown in Fig. 3. This fusion step produces a unified feature vector that represents spatial features and hotspot-related thermal information. The unified feature vector was then processed through a series of fully connected layers to support learning prior to classification. These layers enhance feature interaction and representation capacity. A SoftMax layer was subsequently applied to generate the hotspot severity classification output.

The framework integrates dual inputs and refined feature representations. This study focuses on evaluating classification performance rather than introducing architectural novelty. Model effectiveness was assessed based on the robustness and reliability of hotspot severity classification enabled by the fused spatial–thermal features.

Implementation details and training consistency

Consistent training and optimisation settings were applied across all experiments to ensure a fair among architectural configurations. The same input data, optimisation parameters, and evaluation procedures were used for both baseline and refined CNN models. The initial learning rate was set to 1×10^{-3} and adjusted during training using a learning rate drop factor of 0.3, as summarised in Table 2.

Table 2. Training configuration and optimisation parameters for CNN model variant

Parameter	Settings
Optimizer	Adam
Max Epoch	80
Mini Batch Size	10
Learning Rate	1×10^{-3}
Learning Rate Drop Factor	0.3
Threshold	1×10^{-4}

The additional convolutional layers introduced during refinement followed the design principles of the baseline backbone. Activation functions and kernel sizes were kept consistent to avoid introducing confounding effects. Under these controlled conditions, performance differences were mainly attributed to convolutional depth.

https://doi.org/ 10.24191/esteem.v22iMarch.10388

Model variants for progressive convolutional refinement

Multiple model variants were evaluated using a progressive convolutional refinement strategy. The baseline CNN configuration was first established, followed by models incorporating an increasing number of additional convolutional layers, ranging from one to five. These variants enable a systematic examination of the relationship between convolutional depth and hotspot severity classification performance. The performance of the different configurations was analysed through comparative evaluation. This analysis also reveals the limitations of controlled convolutional layer augmentation in capturing spatial–thermal feature representations. The findings from this evaluation provide the basis for subsequent performance analysis and discussion.

3.4 Performance evaluation metrics

The performance of the proposed hotspot severity classification framework was evaluated using standard quantitative metrics. These metrics were selected to assess classification performance at the pixel level, with particular emphasis on hotspot detection accuracy and reliability. Accuracy is used to evaluate the overall correctness of pixel-level classification, as defined in Eq. (3):

$$Accuracy = \frac{TP + TN}{(TP + FN + TN + FP)} \tag{3}$$

Precision is used to assess the reliability of hotspot predictions by measuring the proportion of correctly identified hotspot pixels, as expressed in Eq. (4):

$$Precision = \frac{TP}{(TP + FP)} \tag{4}$$

The recall process evaluates the model's ability to detect actual hotspot pixels without omission, as given in Eq. (5):

$$Recall = \frac{TP}{(TP + FN)} \tag{5}$$

The F1-score provides a balanced assessment by jointly considering precision and recall, as formulated in Eq. (6):

$$F1 - score = 2x \frac{PrecisionxRecall}{Precision + Recall} \tag{6}$$

4. RESULTS AND DISCUSSION

4.1 Overview of experimental evaluation

The experimental results of hotspot severity classification are presented and discussed in this section. The baseline performance of U-Net, ResNet-18, and ResNet-50 is evaluated before architectural refinement. The impact of progressively adding additional convolutional layers is then analysed to examine the changes in classification performance.

https://doi.org/ 10.24191/esteem.v22iMarch.10388

4.2 Baseline performance before architectural refinement

The baseline classification performance of the evaluated models before the introduction of additional convolutional is summarised in Table 3. Among the tested architectures, ResNet-18 achieved the highest baseline performance, with an F1-score of 98.67% and an accuracy of 96.00%. ResNet-50 also showed strong performance, recording 96.00% accuracy. In contrast, U-Net showed significantly lower baseline classification results, with 33.33% accuracy and an F1 score of 16.67%.

Table 3. Classification performance before architectural refinement

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
U-Net	33.33	11.11	33.33	16.67
ResNet-18	98.67	98.67	98.67	98.67
ResNet-50	96.00	96.08	96.00	96.01

Overall, these baseline results highlight the importance of tailoring network architecture to task-specific objectives. Residual networks are well suited for severity quantification due to their focus on classification accuracy. Partition-oriented architecture, such as U-Net, require further architectural refinement to achieve comparable classification accuracy. This observation drives the refinement strategies explored in the next subsection.

4.3 Effect of additional layers on u-net performance

This section examines the effect of increasing the number of additional convolutional layers on the U-Net architecture. Based on the baseline results presented in Section 4.2, the analysis evaluates whether a deeper convolutional configuration improves U-Net's ability to classify hotspot severity. Configurations with one, two, and three additional convolutional layers are compared in terms of classification performance. The impact of architectural refinement on the segmentation-oriented U-Net backbone is evaluated in this analysis. The classification performance of U-Net under different convolutional layer configurations is summarised in Table 4.

Table 4. Classification performance of U-Net

Layers	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
+1 Layer	33.33	11.11	33.33	16.67
+2 Layers	33.33	11.11	33.33	16.67
+3 Layers	33.33	11.11	33.33	16.67

The addition of one, two, and three convolutional layers does not result in measurable performance improvement. Accuracy, precision, recall, and F1-score remained unchanged at 33.33%, 11.11%, 33.33%, and 16.67%, respectively. These results indicate that increasing the convolution depth alone is not sufficient to improve classification performance in U-Net.

This limitation is particularly evident in architectures that are not optimised for image-level feature abstraction. The U-Net encoder-decoder structure is specifically designed to preserve spatial resolution for segmentation. As a result, the effectiveness of additional convolutional layers for severity classification is limited.

4.4 Effect of additional layers on ResNet-18 performance

ResNet-18 demonstrates a consistent improvement in performance with the addition of convolutional layers, as shown in Table 5. The accuracy increased from 97.87% with one additional layer to 98.13% with

https://doi.org/ 10.24191/esteem.v22iMarch.10388

two additional layers. The highest accuracy of 99.47% was achieved with three additional convolutional layers. Increasing the number of layers to four did not result in additional improvement, as the accuracy showed a slight decrease to 99.20%. Similarly, the configuration with five additional layers achieved an accuracy of 99.47%, which is comparable to the three-layer configuration. This indicates that increasing the convolutional depth beyond this point does not yield significant performance gains.

Table 5. Classification performance of ResNet-18

Layers	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
+1 Layer	97.87	97.88	97.87	97.87
+2 Layers	98.13	98.17	98.13	98.14
+3 Layers	99.47	99.48	99.47	99.47
+4 Layers	99.20	99.20	99.20	99.20
+5 Layers	99.47	99.47	99.47	99.47

The observed performance trend suggests that increasing the convolutional depth improves the feature interaction and representation capacity up to an optimal point. The skip connections in the ResNet architecture preserve feature integrity while enabling deeper network configurations. This design allows for effective feature learning without training instability. However, excessive depth may introduce diminishing returns due to increased model complexity and redundancy in learned representations [25]. Therefore, the three-layer configuration is considered optimal, as it achieves the best balance between classification performance and computational efficiency. This configuration enhances the discriminative capability of ResNet-18, particularly in distinguishing hotspot severity associated with fine thermal differences.

4.5 Effect of additional layers on ResNet-50 performance

A similar performance trend for ResNet-50 with the addition of convolutional layers is shown in Table 6. The classification accuracy increased from 96.53% with one additional layer to 98.93% with two layers. The best performance was achieved with three additional convolutional layers, obtaining an accuracy of 99.20%. Increasing the number of layers to four, results in a slight improvement, with the highest accuracy of 99.47%. However, when five additional layers were introduced, the performance decreased to 99.20%, indicating that further depth does not consistently improve classification performance.

Table 6. Classification performance of ResNet-50

Layers	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
+1 Layer	96.53	96.56	96.53	96.52
+2 Layers	98.93	98.93	98.93	98.93
+3 Layers	99.20	99.21	99.20	99.20
+4 Layers	99.47	99.48	99.47	99.47
+5 Layers	99.20	99.20	99.20	99.20

These results indicate that the deeper residual architecture of ResNet-50 benefits from additional convolutional layers that further refine the representation of higher-level features. However, the performance gain becomes marginal as the number of layers increases, suggesting diminishing returns with excessive convolutional depth. Compared to ResNet-18, the improvement trend is less pronounced, as ResNet-50 already has greater representational capacity prior to refinement. Therefore, although additional layers can enhance feature representation, increasing depth beyond a certain point does not consistently yield performance improvements. This behaviour indicates that ResNet-50 approaches representation

https://doi.org/ 10.24191/esteem.v22iMarch.10388

saturation. As a result, further architectural refinement introduces additional complexity without proportional gains in classification performance.

4.6 Discussion with existing studies

Previous studies on PV hotspot analysis have largely focused on binary fault classification or hotspot identification using CNN-based approaches [26], [27]. Although these methods show promising performance, severity differentiation is often treated as a secondary task. This limitation reduces the effectiveness of maintenance-oriented applications. In addition, many existing frameworks rely on image-centric features [28], [29]. These approaches do not explicitly consider refinement strategies that support robust discrimination of severity under real-world thermal variability.

In contrast, the results of this study show that multiclass hotspot severity classification benefits from improved feature representation supported by controlled architectural refinement. The proposed approach highlights the role of task-aligned feature refinement. Additional convolutional layers are selectively introduced to strengthen feature interactions without adding unnecessary complexity [30]. This strategy allows the network to better capture severity-related features while maintaining architectural efficiency.

5. CONCLUSION

The effectiveness of feature refinement via additional convolutional layers for hotspot severity classification using CNNs has been examined in this current study. Experimental results show that performance is strongly influenced by both the underlying network architecture and the level of feature abstraction. Residual-based models consistently outperform segmentation-oriented networks, highlighting their suitability for image-level severity classification tasks.

Further analysis shows that modest architectural refinement leads to progressive performance improvements. This analysis further confirms that progressive architectural refinement improves classification performance. In particular, optimal results are achieved using a moderate number of additional convolutional layers. In contrast, U-Net does not show any performance improvement after refinement. This result is consistent with its design focus on pixel-level segmentation rather than global severity discrimination. From a practical perspective, the results show that accurate hotspot severity classification relies on task-aligned architectural refinement. The proposed approach supports reliable severity-aware classification for solar PV thermal inspection and maintenance.

This study was conducted under controlled experimental conditions using a selected CNN backbone and a predefined architectural refinement strategy. Future work may explore alternative architectures and validate the approach using larger and more diverse PV thermal datasets.

6. ACKNOWLEDGEMENTS/FUNDING

The authors wish to thank the Faculty of Electrical Engineering, Universiti Teknologi MARA, Cawangan Pulau Pinang for administrative and facilities support. The authors also extend their appreciation to Universiti Malaysia Pahang Al-Sultan Abdullah, Pekan, Malaysia. This research is supported by the Ministry of Higher Education through Fundamental Research Grant Scheme (FRGS) with code FRGS/1/2023/ICT02/UITM/02/12.

https://doi.org/ 10.24191/esteem.v22iMarch.10388

7. CONFLICT OF INTEREST STATEMENT

The authors agree that this research was conducted in the absence of any self-benefits, commercial or financial conflicts and declare the absence of conflicting interests with the funders.

8. AUTHORS' CONTRIBUTIONS

Nurul Huda Ishak: Writing - Original Draft, Data Curation, Formal analysis; **Iza Sazanita Isa:** Supervision, Project administration; **Muhammad Khusairi Osman:** Conceptualization, Methodology; **Mohd Shawal Jadin:** Visualization, Investigation and Resources; **Kamarulazhar Daud:** Software, Validation; **Noor Fadzilah Razali**: Review and Editing.

9. REFERENCES

[1] M. I. Kolawole and B. L. Ayodele, "Smart electronics in solar-powered grid systems for enhanced renewable energy efficiency and reliability," *International Journal of Science and Research Archive*, vol. 13, no. 2, pp. 2910–2930, 2024. Available: https://doi.org/ 10.30574/ijsra.2024.13.2.2512

[2] E. H. Sepulveda-Oviedo, "Impact of environmental factors on photovoltaic system performance degradation," *Energy Strategy Reviews*, vol. 59, no. October 2024, p. 101682, 2025. Available: https://doi.org/10.1016/j.esr.2025.101682

[3] N. Prajapati, R. Aiyar, A. Raj, and M. Paraye, "Detection and Identification of faults in a PV Module using CNN based Algorithm," *2022 3rd International Conference for Emerging Technology, INCET 2022*, pp. 1–5, 2022. Available: https://doi.org/ 10.1109/INCET54531.2022.9825452

[4] N. Drir, A. Mellit, and M. Bettayeb, "A Novel Ensemble CNN Framework With Weighted Feature Fusion for Fault Diagnosis of Photovoltaic Modules Using Thermography Images," *IEEE J. Photovolt.*, vol. 15, no. 1, pp. 146–154, 2024. Available: https://doi.org/ 10.1109/JPHOTOV.2024.3492283

[5] W. M. Elmessery, A. Habib, M. Y. Shams, T. Abd El-Hafeez, and T. M. El-Messery, "Deep regression analysis for enhanced thermal control in photovoltaic energy systems," *Sci. Rep.*, vol. 14, no. 1, pp. 1–22, 2024. Available: https://doi.org/ 10.1038/s41598-024-81101-x

[6] Q. Guo, B. Han, P. Chu, Y. Wan, and J. Zhang, "MF-FusionNet: A Lightweight Multimodal Network for Monitoring Drought Stress in Winter Wheat Based on Remote Sensing Imagery," *Agriculture (Switzerland)*, vol. 15, no. 15, pp. 1–28, 2025. Available: https://doi.org/ 10.3390/agriculture15151639

[7] M. Dhimish, M. Theristis, and V. d'Alessandro, "Photovoltaic hotspots: A mitigation technique and its thermal cycle," *Optik (Stuttg).*, vol. 300, no. November 2023, p. 171627, 2024. Available: https://doi.org/ 10.1016/j.ijleo.2024.171627

[8] G. E. Mustafa Abro, A. Ali, S. Ali Memon, T. Din Memon, and F. Khan, "Strategies and Challenges for Unmanned Aerial Vehicle-Based Continuous Inspection and Predictive Maintenance of Solar Modules," *IEEE Access*, vol. 12, no. December, pp. 176615–176629, 2024. Available: https://doi.org/10.1109/ACCESS.2024.3505754

[9] M. Aghaei, M. Kolahi, A. Nedaei, N. S. Venkatesh, and S. M. Esmailifar, "Autonomous Intelligent Monitoring of Photovoltaic Systems: An In-Depth Multidisciplinary Review," *Progress in Photovoltaics: Research and Applications*, vol. 33, no. 3, pp. 381–409, 2025. Available: https://doi.org/10.1002/pip.3859

https://doi.org/ 10.24191/esteem.v22iMarch.10388

[10] V. K. Y. Swatilekha Ghosh, Santosh Kumar Singh, "Experimental investigation of hotspot phenomenon in PV arrays under mismatch conditions," *Solar Energy*, vol. 253, 2023. Available: https://doi.org/10.1016/j.solener.2023.02.033

[11] O. E. Olabode, I. K. Okakwu, D. O. Akinyele, T. O. Ajewole, S. Oyelami, and O. V. Olisa, "Effect of Ambient Temperature and Solar Irradiance on Photovoltaic Modules' Performance," *Iranica Journal of Energy and Environment*, vol. 15, no. 4, pp. 402–420, 2024. Available: https://doi.org/ 10.5829/ijee.2024.15.04.08

[12] A. Raorane, D. Magare, and Y. Mistry, "Thermal image-driven CNN for predicting solar photovoltaic module lifespan from hotspots," *Solar Energy*, vol. 302, no. July, p. 113965, 2025. Available: https://doi.org/ 10.1016/j.solener.2025.113965

[13] A. K. Vahid Sinap, "CNN-based automatic detection of photovoltaic solar module anomalies in infrared images: a comparative study," *Neural Comput. Appl.*, 2024. Available: https://doi.org/10.1007/s00521-024-10322-y

[14] L. Zhang, J. Kuang, Y. Teng, S. Xiang, L. Li, and Y. Zhou, "A Lightweight Infrared and Visible Light Multimodal Fusion Method for Object Detection in Power Inspection," *Processes*, vol. 13, no. 9, pp. 1–18, 2025. Available: https://doi.org/10.3390/pr13092720

[15] Z. Barraz, I. Sebari, H. Oufettoul, K. Ait el kadi, N. Lamrini, and I. Ait Abdelmoula, "A holistic multimodal approach for real-time anomaly detection and classification in large-scale photovoltaic plants," *Energy and AI*, vol. 21, no. May, p. 100525, 2025. Available: https://doi.org/ 10.1016/j.egyai.2025.100525

[16] M. Z. A. Hamid, K. Daud, Z. H. C. Soh, M. K. Osman, I. S. Isa, and N. H. Ishak, "Review of Deep Learning-Based Hotspot Detection in Solar Photovoltaic Arrays," *2024 IEEE 4th International Conference in Power Engineering Applications: Powering the Future: Innovations for Sustainable Development, ICPEA 2024*, no. March, pp. 332–337, 2024. Available: https://doi.org/ 10.1109/ICPEA60617.2024.10498989

[17] M. U. Ali, A. Zafar, W. Ahmed, M. Aslam, and S. H. Kim, "Enhancing Photovoltaic Reliability: A Global and Local Feature Selection Approach with Improved Harris Hawks Optimization for Efficient Hotspot Detection Using Infrared Imaging," *Int. J. Energy Res.*, vol. 2024, no. 1, 2024. Available: https://doi.org/10.1155/2024/5586605

[18] E. Tarigan, "Identification of early operational defects in photovoltaic modules: A case study of a 24.9 MWp solar PV system in Sumatra, Indonesia," *Unconventional Resources*, vol. 6, no. February, p. 100156, 2025. Available: https://doi.org/10.1016/j.uncres.2025.100156

[19] K. Masita, A. Hasan, T. Shongwe, and H. A. Hilal, "Deep learning in defects detection of PV modules: A review," *Solar Energy Advances*, vol. 5, no. September 2024, p. 100090, 2025. Available: https://doi.org/10.1016/j.seja.2025.100090

[20] D. Gertsvolf, M. Horvat, D. Aslam, A. Khademi, and U. Berardi, "A U-net convolutional neural network deep learning model application for identification of energy loss in infrared thermographic images," *Appl. Energy*, vol. 360, no. January, p. 122696, 2024. Available: https://doi.org/ 10.1016/j.apenergy.2024.122696

[21] Y. Zefri, I. Sebari, H. Hajji, and G. Aniba, "An Encoder-Decoder U-Net Based Model for Overheated Photovoltaic Modules Extraction from Orthorectified Remotely Sensed Thermal Infrared UAV Imagery," *2022 IEEE Mediterranean and Middle-East Geoscience and Remote Sensing Symposium, M2GARSS 2022 - Proceedings*, pp. 25–28, 2022. Available: https://doi.org/ 10.1109/M2GARSS52314.2022.9840254

[22] M. Abdelsattar, A. AbdelMoety, and A. Emad-Eldeen, "ResNet-based image processing approach for precise detection of cracks in photovoltaic panels," *Sci. Rep.*, vol. 15, no. 1, pp. 1–22, 2025. Available: https://doi.org/10.1038/s41598-025-09101-z

[23] Z. Li, W. Liu, G. Wu, and S. Yang, "Semantic segmentation of unmanned aerial vehicle image based on Resnet-Unet," *2023 8th International Conference on Intelligent Informatics and*

https://doi.org/ 10.24191/esteem.v22iMarch.10388

Biomedical Sciences (ICIIBMS), vol. 8, pp. 295–299, 2023. Available: https://doi.org/10.1109/iciibms60103.2023.10347663

[24] G. Mannino *et al.*, "Photovoltaic Module Degradation Forecast Models for Onshore and Offshore Floating Systems," *Energies (Basel).*, vol. 16, no. 5, 2023. Available: https://doi.org/10.3390/en16052117

[25] M. Shafiq and Z. Gu, "Deep Residual Learning for Image Recognition: A Survey," *Applied Sciences (Switzerland)*, vol. 12, no. 18, pp. 1–43, 2022. Available: https://doi.org/10.3390/app12188972

[26] M. Z. A. Hamid, K. Daud, Z. H. C. Soh, M. K. Osman, and I. S. Isa, "Deep learning-driven thermal imaging hotspot detection in solar photovoltaic arrays using YOLOv10," 2024. Available: https://doi.org/10.24191/esteem.v20iSeptember.18

[27] B. Eri, "Early Hotspot Detection in Photovoltaic Modules using Deep Learning Methods," vol. 09, no. 09, pp. 5065–5072, 2024. Available: https://doi.org/10.47191/etj/v9i09.10

[28] S. Goyal and J. C. Rajapakse, "Self-supervised learning for hotspot detection and isolation from thermal images," *Expert Syst. Appl.*, vol. 237, no. PB, p. 121566, 2024. Available: https://doi.org/10.1016/j.eswa.2023.121566

[29] U. Pruthviraj, Y. Kashyap, E. Baxevanaki, and P. Kosmopoulos, "Solar Photovoltaic Hotspot Inspection Using Unmanned Aerial Vehicle Thermal Images at a Solar Field in South India," *Remote Sens. (Basel).*, vol. 15, no. 7, 2023. Available: https://doi.org/10.3390/rs15071914

[30] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, and Y. Duan, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J. Big Data*, vol. 8, no. 1, 2021. Available: https://doi.org/10.1186/s40537-021-00444-8

https://doi.org/ 10.24191/esteem.v22iMarch.10388