UNIVERSITI TEKNOLOGI MARA

HOTSPOT LOCALIZATION AND SEVERITY CLASSIFICATION IN PHOTOVOLTAIC AERIAL THERMAL IMAGERY USING SEMANTIC SEGMENTATION AND CONVOLUTIONAL NEURAL NETWORK

NURUL HUDA BINTI ISHAK

PhD

July 2026

UNIVERSITI TEKNOLOGI MARA

HOTSPOT LOCALIZATION AND SEVERITY CLASSIFICATION IN PHOTOVOLTAIC AERIAL THERMAL IMAGERY USING SEMANTIC SEGMENTATION AND CONVOLUTIONAL NEURAL NETWORK

NURUL HUDA BINTI ISHAK

Thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering)

Faculty of Electrical Engineering

July 2026

CONFIRMATION BY PANEL OF EXAMINERS

I certify that a Panel of Examiners has met on 15 June 2026 to conduct the final examination of Nurul Huda binti Ishak on her Doctor of Philosophy thesis entitled “Hotspot Localization and Severity Classification in Photovoltaic Aerial Thermal Imagery using Semantic Segmentation and Convolutional Neural Network” in accordance with Universiti Teknologi MARA Act 1976 (Akta173). The Panel of Examiners recommends that the student be awarded the relevant degree. The Panel of Examiners was as follows:

Zainal Hisham Che Soh, PhD
Associate Professor
Faculty of Electrical Engineering
Universiti Teknologi MARA
(Chairman)

Intan Rahayu Ibrahim, PhD
Senior Lecturer
Faculty of Electrical Engineering
Universiti Teknologi MARA
(Internal Examiner)

Mohammad Faridun Naim Tajuddin, PhD
Associate Professor
Faculty of Electrical Engineering and Technology
Universiti Malaysia Perlis
(External Examiner)

PROFESSOR DR HJH ZURAEDA IBRAHIM
Dean
Institute of Postgraduates Studies
Universiti Teknologi MARA
Date: 15 July 2026

AUTHOR'S DECLARATION

I declare that the work in this thesis was carried out in accordance with the regulations of Universiti Teknologi MARA. It is original and is the result of my own work, unless otherwise indicated or acknowledged as referenced work. This thesis has not been submitted to any other academic institution or non-academic institution for any degree or qualification.

I hereby acknowledge that I have been supplied with the Academic Rules and Regulations for Postgraduate Studies at Universiti Teknologi MARA, which regulate the conduct of my study and research.

Name of Student	:	Nurul Huda binti Ishak
Student ID. No.	:	2022602836
Programme	:	Doctor of Philosophy (Electrical Engineering) – CEEE950
Faculty	:	Electrical Engineering
Thesis Title	:	Hotspot Localization and Severity Classification in Photovoltaic Aerial Thermal Imagery using Semantic Segmentation and Convolutional Neural Network

Signature of Student	:	
Date	:	July 2026

ABSTRACT

The rapid expansion of solar photovoltaic (PV) farms has increased the need for reliable automated fault detection systems. To achieve optimal performance and effective energy management in solar PV systems, it is crucial to detect faults in solar PV modules at the early monitoring stage. Hotspot formation is one of the most critical PV faults, often caused by shading, cell damage, or degradation, leading to power loss and potential safety hazards. While thermal imaging effectively detects hotspots, variation in emissivity, reflections, and temperature fluctuations can reduce localization accuracy and affect hotspot severity classification. Moreover, classifying hotspot severity based on temperature distribution remains a significant challenge due to the complexity of thermal colour variations and the lack of standardized reference frameworks or benchmarking criteria. A dual-input feature fusion strategy is introduced to integrate thermal image features and temperature descriptors, improving hotspot severity classification. This study proposes an automated hotspot localization and severity classification for aerial thermal imagery using Convolutional Neural Network (CNN) framework. Initially, a series of image pre-processing techniques such as image annotation and resizing are applied to adapt with the CNN framework and training process. This study proposes a two-tier semantic segmentation approach for the automatic detection of hotspot regions in aerial thermal images of solar PV systems. In the first tier, a semantic segmentation model is developed to isolate solar PV modules from the background. Subsequently, in the second tier, a fine-grained semantic segmentation model is used to segment and detect hotspot regions on the extracted solar PV modules. For this study, three CNN models namely U-Net, ResNet-18 and ResNet-50, were evaluated for hotspot segmentation and severity classification. Next, a novel dual-input feature fusion strategy is introduced by combining CNN-derived features and thermal descriptors for hotspot severity classification. This integration captures complementary thermal information and enables more reliable severity assessment. To further improve classification performance, an enhanced CNN model incorporating the proposed dual-input feature fusion strategy is developed. Specifically, convolutional layers from multiple DL models are concatenated to enhance multi-input hotspot severity classification. Finally, the model's performance is evaluated using classification metrics-based performance and compared with the solar thermographer evaluation to validate the robustness and reliability. Overall, the enhanced CNN-based framework significantly improves hotspot severity classification performance, with ResNet-18 achieving a 23% accuracy improvement compared with the single-input image-based model. The proposed approach extends existing hybrid and traditional methods by integrating pixel-level hotspot localization with structured severity classification in a fully automated framework. The developed framework provides a practical solution for automated hotspot severity classification in PV thermal inspection systems, supporting improved renewable energy monitoring and management.

ACKNOWLEDGEMENT

In the name of Allah S.W.T., the Most Gracious, the Most Merciful. Alhamdulillah, all praise and gratitude be to Allah for His countless blessings, which have granted me the strength, patience, and opportunity to embark on and successfully complete this long and challenging PhD journey.

I would like to express my deepest gratitude to my supervisor, Associate Professor Ir. Dr. Iza Sazanita Isa, and my co-supervisors, Associate Professor Ir. Ts. Dr. Muhammad Khusairi Osman, Ir.Ts. Dr. Kamarulazhar Daud, and Ts. Dr. Mohd Shawal Jadin, for their invaluable guidance, continuous support, patience, and insightful ideas throughout the completion of this research. Their expertise, guidance, and unwavering support have significantly enhanced the quality and depth of this research.

I would also like to extend my sincere appreciation to all faculty members of the Faculty of Electrical Engineering, UiTM Pulau Pinang, who were directly or indirectly involved in this research. Their cooperation, valuable assistance, and professional support were essential in the successful completion of this study.

I am deeply grateful to my beloved husband, Faizal Mohamad Twon Tawi, for his unwavering support, understanding, and sacrifices throughout my daily struggles in completing this journey. My loving appreciation also goes to my sons, Muhammad Nasih ‘Ulwan, Muhammad Thaqif Thamin, and Muhammad Fatihuddin Hayyan, who have been my constant sources of inspiration, strength, and motivation.

To my mentor throughout this journey, Dr. Mohd Khairul Nizam, your mentorship has greatly shaped my growth, sharpened my perspective, and inspired me to move forward with confidence and purpose. I would also like to extend my heartfelt appreciation to my dear friends, Noorfadzilah, Atharah, Atiqah, Alwani, Fadzeelah, Syafika, Asma, Darina, Nik Zuliani, Mardzuliana, Zuhri, Najwa, Anna, Yana and Zafirah. Thank you for your continuous encouragement, meaningful discussions, moral support, and for being part of this journey. Your friendship, understanding, and shared experiences have made this challenging path more manageable and memorable.

Finally, this thesis is dedicated to my beloved family members, especially my mother, for her endless support, encouragement, and prayers throughout this journey. She stood by me during the most challenging moments of this project. Although my late father, Ishak Abdullah, has passed away, his spirit, values, and encouragement have always remained in my heart. I would also like to express my heartfelt appreciation to my parents-in-law for their continuous prayers and support throughout this journey. I am also deeply grateful to my beloved siblings, Khairiah, Shahidah, Faizul, and Fakhrul, for their constant encouragement, understanding, and moral support. This achievement is dedicated to all of you.

Alhamdulillah.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

Symbols

%	Percent
°C	Degree Celsius
A	Ampere
dB	Decibel
GW	Gigawatt
I	Current
K	Kelvin
MWp	Megawatt Peak
P	Electrical Power
R	Electrical Resistance
T	Temperature
V	Volt
W/m2	Watt Per Square Metre
µm	Micrometre
κ	Cohen's Kappa Coefficient
Ω	Ohm

LIST OF ABBREVIATIONS

Abbreviations

Adam	Adaptive Moment Approximation
AI	Artificial Intelligence
CLAHE	Contrast-Limited Adaptive Histogram Equalization
CNN	Convolutional Neural Network
Conv	Convolutional
COSCEM	Centre of Energy Management and Sustainable Campus
DL	Deep Learning
EL	Electroluminescence
EXIT	Exchangeable Image File Format
FCL	Fully Connected Layer
FLIR	Forward-Looking Infrared
FN	False Negative
FP	False Positive
GLCM	Grey-Level Co-occurrence Matrix
GPS	Global Positioning System
GT	Ground Truth
HC	Hand-Crafted
HE	Histogram Equalization
HOG	Histogram of Gradient

IEA	International Energy Agency
IoT	Internet of Things
IoU	Intersection over Union
kNN	k-Nearest Neighbours
LBP	Local Binary Patterns
Max	Maximum
ML	Machine Learning
MSE	Mean Squared Error
NLM	Non-Local Means
PA	Pixel Accuracy
PID	Potential Induced Degradation
PPV	Positive Predictive Value
PSNR	Peak Signal Noise Ratio
PV	Photovoltaic
RE	Renewable Energy
ResNet	Residual Network
RF	Random Forest
ReLU	Rectified Linear Unit
RGB	Red, Green, Blue
RO	Research Objective
ROI	Region of Interest
SDG	Sustainable Development Goals

SegNet	Segmented Neural Network
SN	Sensitivity
SNR	Signal-to-Noise Ratio
SP	Specificity
ST	Solar Thermographer
SVM	Support Vector Machines
TN	True Negative
TNB	Tenaga Nasional Berhad
TP	True Positive
UAV	Unmanned Aerial Vehicle
UiTM	Universiti Teknologi MARA
UniMAP	Universiti Malaysia Perlis

CHAPTER 1
INTRODUCTION

1.1 Research Background

In recent years, the global shift towards renewable energy (RE) has highlighted the critical need for clean and reliable power generation technologies. A report from the International Energy Agency (IEA) shows a rapid increase in the use of photovoltaics (PV) [1], illustrating a growing transition away from fossil fuels toward solar-based solutions. In Malaysia, the national RE target was revised in August 2023 to accelerate the country's transition toward sustainable energy development [2]. Under this policy, the contribution of RE to Malaysia’s electricity generation mix is targeted to increase from 40% in 2040 to 70% by 2050 [3], [4]. This development underscores the need to understand the operational challenges that arise as PV installations continue to grow, particularly in ensuring long-term system reliability and performance.

Figure 1.1 presents the projected global capacity trends for primary energy sources from 2020 to 2030, highlighting the continued expansion of renewable technologies relative to coal-based generation. The rapid expansion of large-scale solar farms requires efficient, scalable monitoring strategies to ensure optimal system operation and minimize degradation. Consequently, the rapid growth of PV installations increases inspection demands and underscores the need for automated, Unmanned Aerial Vehicle (UAV)–based thermal monitoring to efficiently detect hotspots.

In real-world operating conditions, PV systems are exposed to significant environmental and electrical variability, which can lead to localized overheating phenomena known as hotspots. These hotspots are characterized by abnormal temperature elevations within specific regions of PV modules and are commonly associated with cell mismatch, partial shading, material defects, and material degradation [5]. If not detected early, hotspots can reduce PV module efficiency. Temperature differences of approximately 10 K and 18 K may result in efficiency losses of about 4% and 7-10%, respectively, while prolonged thermal stress can compromise long-term operational reliability [6]. In severe cases, persistent hotspot conditions may accelerate module degradation, increase the likelihood of component failure, and pose potential fire safety risks. Therefore, accurate localization and evaluation of hotspot conditions are essential components of effective PV condition assessment.

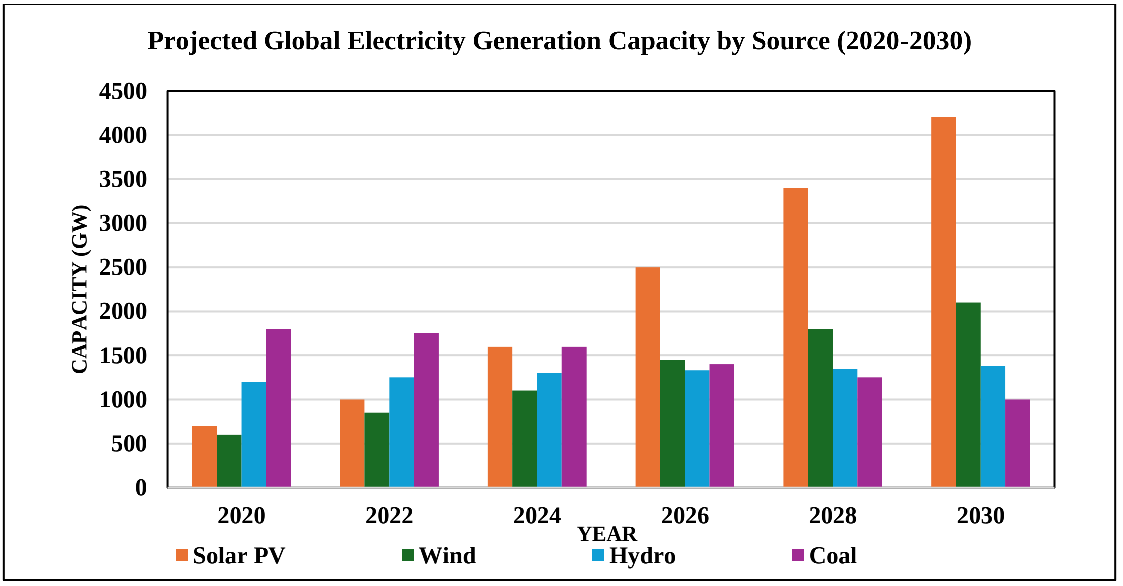


Figure 1.1 Projected Global Electricity Generation Capacity for Solar PV, Wind, Hydro, and Coal From 2020 to 2030 [7]

In this context, infrared (IR) thermal imaging has emerged as a widely adopted non-destructive inspection technique for detecting thermal anomalies in PV systems. Thermal imaging enables visualization of temperature distribution across PV modules, allowing abnormal regions to be identified through thermal contrast. With the advancement of UAVs, aerial-based thermal inspection has significantly improved the efficiency and coverage of large-scale PV monitoring [8]. Nevertheless, the accuracy of UAV-based thermal inspection can be influenced by several factors, including flight altitude, viewing angle, solar irradiance, wind conditions, surface emissivity, reflections and thermal image resolution. In this study, visual imagery was used to provide contextual information during data interpretation and hotspot verification, while thermal imagery served as the primary input for hotspot localization and severity classification. Furthermore, the combination of visual and thermal imagery provides complementary structural and thermal information, thereby supporting more comprehensive anomaly interpretation and verification.

As PV installations grow in complexity, automated hotspot localization and severity assessment have become increasingly significant for automated monitoring systems. Recent advancements in artificial intelligence (AI), particularly convolutional neural networks (CNNs), have demonstrated strong capability in extracting spatial and semantic features from image data [9], [10]. The incorporation of semantic segmentation techniques enables pixel-level classification, allowing precise localization of hotspot regions within aerial imagery [11]. However, conventional hotspot detection methods mostly rely on simple numerical value grouping [8], [12], threshold-based temperature evaluations [13], [14], or manual visual examination [15]. In contrast, semantic segmentation provides detailed spatial mapping of thermal anomalies, facilitating a more comprehensive assessment of PV conditions [16]. Furthermore, integrating severity classification enables hotspot regions to be categorized according to their thermal intensity levels, supporting maintenance prioritization and decision-making. In this study, hotspot severity level was determined based on temperature characteristics, including the temperature differential (ΔT), and subsequently validated through expert solar thermographer assessment.

Therefore, this study investigates hotspot localization and severity classification in PV aerial thermal imagery using semantic segmentation and CNN. By integrating UAV-acquired thermal and visual data, the research establishes a deep learning (DL)–based framework for pixel-level hotspot delineation and multi-class severity assessment. This approach can help reduce energy losses, minimize the risk of severe thermal damage, and improve the reliability of large-scale PV systems. Furthermore, the proposed framework aligns with the growing demand for automated, scalable PV monitoring systems that support large-scale solar deployment.

1.2 Problem Statement

Hotspots in PV modules are localized thermal anomalies that can cause persistent damage to the modules [17]. The presence of hotspots due to partial shading, soiling, or inherent cell defects results in elevated temperatures, reducing the solar panel's efficiency. However, conventional hotspot detection methods remain limited in reliability, as most rely on manual visual inspection [18], simple threshold-based temperature comparisons [19], or basic clustering of thermal values [20]. These conventional approaches often struggle to distinguish true hotspot regions from minor temperature variations caused by irradiance fluctuations, surface contamination, or sensor noise, leading to inconsistent detection results [21]. Therefore, there remains a need for an automated localization method that can distinguish true hotspot regions from background thermal variations in aerial PV imagery.

Reliable thermal inspection of PV modules is essential for maintaining stable energy production and enabling timely preventive maintenance, where hotspots play a critical role in condition assessment. Although DL-based semantic segmentation has been increasingly adopted for hotspot detection, most existing approaches rely on single-stage models that directly segment hotspots, PV module and background classes simultaneously from thermally complex scenes [19], [22]. In large-scale UAV-based inspections, aerial thermal images often contain background interference, structural variability across PV modules, and non-uniform temperature distributions, all of which hinder consistent hotspot delineation. Performing hotspot segmentation without prior isolation of PV modules reduces spatial discrimination and affects localization reliability [23], [24]. However, limited studies have investigated a hierarchical segmentation strategy that first isolates PV modules and subsequently localizes the hotspot regions in UAV-based thermal imagery. These limitations demonstrate that single-model segmentation lacks the structural refinement needed to effectively handle the complexity of aerial thermal imagery. Therefore, a structured two-tier semantic segmentation framework is expected to enhance hotspot localization by progressively isolating relevant PV regions before detailed anomaly detection.

Furthermore, accurately assessing hotspot severity is therefore a significant challenge in thermal-based PV inspection. Severity assessments that rely solely on image appearance may be inaccurate, as the distribution of thermal intensities within segmented hotspot areas can exhibit visual overlap even when the underlying temperature differences are significant [25]. Similarly, temperature-threshold–based approaches are highly sensitive to irradiance variability, ambient temperature fluctuations, and sensor noise, which can obscure the true thermal deviation within hotspot regions and reduce the reliability of severity assessment [19], [26]. These limitations indicate that neither image-based features nor temperature-based features alone provide sufficient discriminative information for robust hotspot severity classification. Therefore, a feature extraction and fusion strategy is required to capture complementary characteristics of the hotspots. By integrating temperature-based features, including maximum temperature (T_{max}), minimum temperature (T_{min}), reference temperature (T_{ref}) and ΔT with CNN-derived thermal features extracted from the segmented region of interest (ROI), a more comprehensive representation of hotspot severity can be achieved.

Following the extraction of complementary features from segmented hotspot regions, an additional challenge arises in developing a robust model for accurate hotspot severity classification [27], [28]. Although CNNs have demonstrated strong capabilities in thermal image analysis, conventional architecture primarily focuses on visual texture and intensity patterns. As a result, these models are not designed to fully exploit temperature-based information and often struggle to discriminate fine-grained hotspot severity levels [29]. Furthermore, class imbalance and overlapping feature distributions reduce the robustness of existing severity classification models. Besides, thermal variations across different PV installations further limit the generalizability of these models [30]. To address these challenges, the proposed framework employs independent testing and expert validation to improve model robustness and generalization performance. These limitations highlight the need for an enhanced CNN model with an additional convolutional layer that incorporates a feature-fusion strategy to integrate temperature-based features with CNN-derived thermal features. By leveraging fused, complementary information, the enhanced model aims to achieve more accurate and reliable classification of hotspot severity in solar PV thermal inspection.

Despite recent advancements in automated hotspot severity classification, a key limitation remains the lack of comprehensive validation against solar thermographer assessments [31], [32], [33]. Existing studies frequently report high accuracy on labelled datasets [34]. However, such performance does not necessarily reflect reliability under real-world operating conditions, where expert interpretation remains the reference standard for severity assessment [8], [40], [41], [42]. Moreover, few studies systematically correlate automated severity predictions with expert-rated severity scores, resulting in limited understanding of how well these models align with domain knowledge during practical deployment [35], [36]. This gap highlights the need for a comprehensive validation framework that evaluates model performance using standard evaluation metrics and agreement analysis, such as Cohen's Kappa, against expert assessments.

In summary, four key research gaps remain in the current literature. First, existing hotspot localization approaches face challenges in accurately distinguishing hotspot regions from complex thermal backgrounds in UAV-based imagery. Second, current severity assessment methods often rely on either image-based or temperature-based information alone, resulting in an incomplete representation of hotspots. Third, conventional CNN architectures have limited capability to effectively integrate complementary thermal information for accurate hotspot severity classification. Finally, the lack of systematic validation against expert solar thermographer assessments limits confidence in the practical applicability of automated hotspot inspection systems. These research gaps motivate the development of the proposed framework presented in this study.

1.3 Research Objectives

The main objective of this research is to develop an automated method for hotspots localization and severity classification in solar PV modules based on aerial thermal imagery. To achieve this objective, the work is divided into several components and carried out systematically with four associated research objectives identified as follows:

i. To develop a new framework based on a two-tier semantic segmentation with a hybrid image enhancement method to localize hotspot regions accurately in solar PV aerial thermal images.
ii. To develop a dual-input feature fusion strategy integrating hand-crafted (HC) thermal features and CNN-based features from the segmented hotspot regions obtained in (1) for hotspot severity classification.
iii. To propose an enhanced CNN model that incorporates the dual-input features in (2) for accurately classifying hotspot severity conditions.
iv. To analyze the performance of the proposed CNN-based classification model using standard evaluation metrics and validate the results through expert solar thermographer assessment and statistical agreement analysis.

1.4 Research Scope

This research focused on developing an automated hotspot localization and severity classification on solar PV aerial thermal imagery. The study's scopes are as follows:

a) Image Acquisition and Data Collection

The dataset used in this study was obtained from a utility-scale solar PV farm located in Kelantan, Malaysia. The site was selected due to the availability of hotspot occurrences, accessibility for aerial data acquisition, and suitability for thermal inspection activities. The site also represents a typical large-scale PV installation operating under the environmental conditions commonly encountered in Malaysia. It comprises two types of imagery: visual and thermal images. Both image types were captured simultaneously from the same aerial perspective, at the same height and angle to ensure spatial alignment. Critical information in the visible-light spectrum is provided by visual images, which enable identification of environmental context and physical elements not visible in thermal imagery. Thermal imaging, on the other hand, reveals temperature anomalies and heat distribution, which are essential for identifying operational issues, such as hotspots. Combining thermal and visual modalities provides a complete dataset that facilitates a more thorough examination of PV infrastructure. This combination enhances inspection accuracy in solar PV system evaluations. It enables a more comprehensive analysis of the site's thermal behaviour and physical layout, thereby supporting informed decision-making.

The images were captured using UAV. Accordingly, this study focuses exclusively on aerial thermal and visual images for training and testing the CNN model. A total of 670 thermal images and 670 corresponding visual images were collected for this study. Although the dataset size is relatively limited, expert-annotated ground truth data and pretrained DL models were utilized to support effective model training and improve generalization performance. The study focused on a single PV installation employing commercial PV modules. Since different PV technologies may exhibit different thermal characteristics and hotspot patterns, the findings of this study are limited to the panel technology available at the selected site. All images were collected at a consistent flight altitude of approximately 15 m under similar image acquisition conditions. Maintaining a consistent altitude helped ensure comparable image scale and sufficient thermal resolution for hotspot detection across the dataset. This imaging distance enables accurate thermal profiling of individual PV modules. As a result, reliable hotspot detection can be achieved. The dataset supports a practical, application-focused assessment of the proposed framework by reflecting actual PV system operating conditions.

b) Methods Selection

The methodological design of this study is guided by the research objectives and focuses on integrating image processing techniques with DL–based analysis. The overall framework is structured into sequential stages, comprising hotspot localization, feature extraction, and severity classification. For hotspot localization, a two-tier semantic segmentation strategy is adopted to first segment PV modules from aerial thermal images, followed by fine-grained segmentation of hotspot regions within the extracted modules. The segmentation performance was evaluated using U-Net, ResNet-18, and ResNet-50 architectures to determine the most suitable model for PV module and hotspot localization. This staged segmentation design enables accurate ROI extraction for subsequent analysis. Following segmentation, a dual-input feature-extraction approach combines HC thermal features, including T_{max}, T_{min}, T_{ref}, and ΔT, with CNN-derived image features. The extracted features are then utilized by an enhanced CNN architecture designed to perform hotspot severity classification. This methodological selection provides a systematic and coherent framework for implementing automated hotspot detection and severity assessment in PV thermal inspection.

c) System Development and Use of Established Methods

The system algorithm was developed offline using MathWorks MATLAB© version R2023a. This algorithm is designed to detect and analyze hotspot regions from thermal images of solar PV modules. Established image pre-processing techniques were applied as preliminary processing steps prior to the proposed semantic segmentation and classification framework. These operations included image annotation for ground truth generation and image resizing to standardize the input data for subsequent hotspot localization and severity classification. The Deep Learning Toolbox was utilized to support the design and implementation of CNN architecture. This study focuses exclusively on hotspot severity conditions rather than other types of solar PV faults. Hotspot regions were identified and analyzed using DL-based feature extraction techniques. Finally, the detected hotspots were classified into low, medium, and high severity levels based on temperature characteristics, including the ΔT, and subsequently validated through expert solar thermographer assessment.

d) Evaluation and Analysis

The performance of each step in the proposed framework is evaluated using a thermal image dataset. Performance assessment for each research objective is conducted using classification metrics, including accuracy, precision, recall, and F1-score. Intersection over Union (IoU) is adopted as the primary metric for hotspot detection and segmentation because it provides a direct measure of the overlap between the predicted and ground truth regions. To further validate the proposed approach, an agreement analysis is conducted by comparing the algorithm's outputs with assessments from experienced solar thermographers. Details regarding the thermographers' qualifications, evaluation procedures, and agreement analysis are presented in Chapter 3.

1.5 Thesis Outline

This thesis report is organized into five chapters, each of which several sections and subsections. It begins with a brief introduction to solar PV systems, providing an overview of the importance and widespread use of these systems in the RE sector. The introduction establishes the research context by highlighting key areas of concern and outlining the challenges associated with maintaining the performance and longevity of solar PV systems. Toward the end of the chapter, problem statements, the objectives, and the scope of the thesis are defined.

Chapter 2 presents an extensive literature review, focusing on the fundamental approaches for identifying and localizing hotspots within solar PV modules. The chapter begins by discussing the fundamental characteristics of hotspot formation in PV modules. Subsequently, the chapter reviews CNN-based approaches to hotspot analysis, with an emphasize on widely adopted architectures such as U-Net and ResNet. The role of CNNs in feature extraction and classification is then examined. In addition, recent advancements in enhanced CNN models that improve detection accuracy and robustness are discussed. The chapter then surveys past research on automated segmentation and hotspot severity classification. This includes a review of the criteria used for severity assessment and the integration of semantic segmentation with classification frameworks. Finally, the limitations and research gaps in existing hotspot detection and severity classification methods are identified. These gaps motivate the development of the proposed methodology presented in subsequent chapters.

Chapter 3 deals with the proposed research methodology framework. The method comprises five main parts, namely image acquisition, hotspot region localization, feature extraction, hotspot severity, and system performance. In the image acquisition section, the methods and equipment used to collect the images are described. The technique begins with a pre-processing phase, which is crucial for investigating localized solar PV hotspot regions in thermal images. This process improves data quality and enhances relevant features, making the data more suitable for accurate and effective hotspot detection algorithms. Next, identical solar PV modules are segmented in an infrared image. This chapter also examines the classification process used to assess the severity of hotspots on solar PV modules. The research methodology for the hotspot severity classification model begins with image acquisition and concludes with the evaluation of model performance.

Chapter 4 presents and analyzes the experimental results of the proposed framework. The training and testing outcomes are explained in detail. This is followed by a performance analysis comparing existing inspection approaches, including manual delineation, with the proposed automated method. The results are organized according to the methodological sections outlined in Chapter 3 to enable a systematic comparison across both datasets. The chapter concludes with an evaluation of system reliability and validity based on assessments provided by experienced solar thermographers.

Finally, Chapter 5 concludes and highlights the research's main contributions and limitations. At the end of the chapter, some possible improvements and directions for future research are suggested.

CHAPTER 2
LITERATURE REVIEW

2.1 Introduction

The presence of faults, such as hotspots, significantly impacts the efficiency and continuous operation of photovoltaic (PV) systems. Hotspots are localized thermal anomalies that arise from internal or external issues within PV modules and serve as early indicators of potential module failure. As PV installations expand in size and complexity, the limitations of conventional defect detection methods have become increasingly apparent. Consequently, researchers are developing intelligent, fully automated systems to address these limitations. Since these systems aim to identify hotspot localizations and classify their severity accurately, this chapter reviews the literature related to the hotspot detection, localization and severity classification in PV systems.

This chapter begins with an overview of hotspot phenomena in solar PV modules and the associated thermal-based inspection techniques. The discussion covers hotspot formation mechanisms, commonly adopted inspection approaches, and the key challenges encountered in hotspot detection. The chapter then critically reviews DL–based methods for hotspot analysis, with a focus on CNN-based approaches. This review examines how CNNs have been applied for automated hotspot detection, localization, and feature learning from thermal images. Subsequently, segmentation-based techniques for hotspot detection in thermal images are reviewed, focusing on their strengths, limitations, and research gaps. Finally, the chapter concludes by summarizing the main findings and identifying research gaps that motivate the proposed methodology in this study.

2.2 Characteristics of Hotspot in Solar Photovoltaic

Hotspots in PV modules are localized regions where the temperature significantly exceeds that of the surrounding cells, primarily due to non-uniform current flow within the array [17], [37]. Such thermal anomalies typically occur when one or more solar cells are partially shaded, cracked, or electrically mismatched, leading to reverse-bias operation in the affected cells. Under these conditions, rather than converting solar irradiance into electrical energy, the affected cells dissipate the absorbed energy as heat, resulting in localized temperature increases. This phenomenon is governed by Ohmic heating and reverse-bias conduction, which increase resistive power losses (I^2R) in the affected regions. In severe cases, excessive heat generation may induce encapsulant browning, solder joint fatigue, or even fire hazards, thereby compromising module integrity and operational safety [38]. Furthermore, elevated operating temperature adversely affects PV performance, with efficiency decreasing by approximately 0.4–0.5% for every 1°C increase in temperature [38]. The following subsections further discuss mechanisms of hotspot formation and thermal-based detection approaches.

2.2.1 Mechanism of Hotspot Formation

The mechanism of hotspot formation can be further explained through the thermal runaway phenomenon, where elevated temperatures increase the cell’s resistance, leading to further heating and a self-reinforcing feedback loop [39]. Over time, these repetitive thermal stresses cause permanent structural deformation and material degradation, including delamination and crack propagation within the encapsulant layers. Such physical and electrical deterioration directly affects the module’s current–voltage (I–V) characteristics, resulting in reduced fill factor and efficiency losses. Under reverse-bias conditions, the affected cell behaves as a resistive element, dissipating electrical energy as heat. The resulting current mismatch within the PV string distorts the I–V characteristics and reduces the output current and maximum power output. Consequently, the fill factor and overall conversion efficiency decrease.

Figure 2.1 illustrates the mechanism of hotspot formation, showing how shaded or faulty cells enter reverse bias and act as resistive elements, dissipating power as heat at the P–N junction. The illustration indicates that internal currents within the hotspot region produce excessive localized heating, raising the temperature well above standard operating values. This not only reduces the PV module's immediate efficiency but also triggers long-term degradation, including glass breakage, delamination, and internal wiring failures.

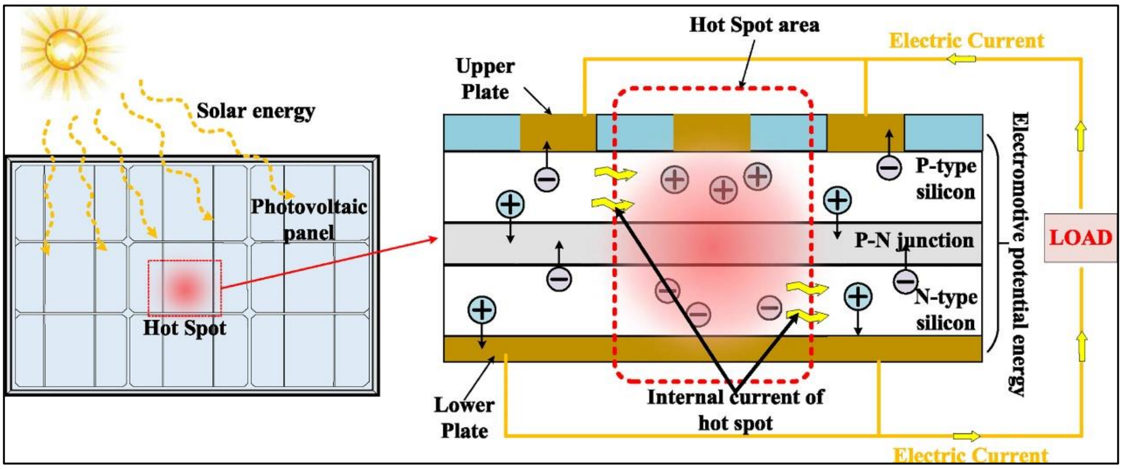


Figure 2.1 A Shaded or Defective Cell Becomes Reverse-biased, Dissipates Electrical Energy as Heat, and Generates Localized Hotspots That Accelerate Material Degradation and Reduce Module Reliability [40]

An electrical mismatch between interconnected solar cells generates localized heating within PV modules creates a phenomenon widely recognized as hotspot formation [41]. The hotspots typically arise when intrinsic material degradation, microcracks, or partial shadowing cause one or more cells to operate at a reduced efficiency. Under these conditions, the affected cells are driven into a reverse-bias state by the current flowing from neighboring cells. Instead of converting solar energy into electrical power, the defective cells act as resistive elements and consequently dissipate electrical energy as heat at the P–N junction. The resulting resistive losses (I^2R) induce thermal anomalies that compromise module performance and long-term reliability [31].

Accurate interpretation of this mechanism is crucial because repeated cycles of intense localized heating may trigger thermal runaway. This feedback process occurs when increased resistance generates additional heat, further accelerating the degradation of affected cells. Although the concept of thermal runaway has been widely cited, most empirical studies [39], [42], [43] focus on short-term effects under controlled shading scenarios. While controlled experiments provide valuable insights into hotspot behaviour, these settings may not fully represent the real-world operating conditions encountered in large-scale PV installations. In UAV-based thermal inspections, hotspot characteristics can be influenced by varying irradiance levels, environmental conditions, dust accumulation, module ageing, and complex fault interactions. These factors introduce thermal variability that is often absent in laboratory studies, making hotspot detection and severity assessment more challenging. Based on research by Al Mahdi et al., the work has further shown that even high-efficiency PV modules are not immune to hotspot formation [44]. In addition, Ghosh et al. [45] have confirmed that mismatched conditions arising from uneven dust deposition or partial shading can cause temperature surges well above standard operational levels. For instance, a thermal imaging study in [45] has documented hotspot areas reaching temperatures exceeding 90°C, even during brief exposure to shade. While these findings are significant, additional investigation into the threshold limits and thermal tolerance of different PV cell materials could refine current understanding and inform material selection for hotspot resistance [46].

The consequences of hotspot formation extend far beyond minor efficiency losses. Repeated cycles of heating and cooling induce fatigue within the glass and encapsulant materials, causing issues such as localized discoloration, interconnect corrosion, and potential-induced degradation (PID) [47]. If left unaddressed, these defects not only reduce energy yield but also shorten the module's service life and increase maintenance costs. In large-scale PV farms, even a small percentage of defective cells can propagate systemic performance degradation and reduce overall energy output. Such degradation may also pose significant financial and safety risks.

Understanding and identifying hotspot behavior is therefore critical for effective PV system monitoring and preventive maintenance. Studies such as [48], [49] have emphasized the integration of time-series thermal analysis with environmental variables, including irradiance, temperature, and wind speed, to predict hotspot initiation and evolution. The spatial intensity, shape, and growth dynamics of hotspots can be characterized using advanced analysis techniques such as thermal differential mapping and spatial intensity modeling. These insights provide a theoretical basis for automated thermal inspection, supporting improved asset management, warranty evaluation, and long-term system reliability.

2.2.2 Thermal-based Inspection for Hotspot Detection and Localization

Infrared (IR) thermography is the main method for detecting hotspots in PV systems. This thermal-based inspection measures the surface temperature of a PV module under normal operating conditions. The accuracy of thermal measurements depends on several operational and environmental factors, including module emissivity, viewing angle, irradiance level, wind speed, and electrical loading conditions. Variations in these parameters can affect the observed thermal distribution and influence the visibility of hotspot regions. Therefore, thermal inspections are typically conducted under stable operating conditions to ensure reliable hotspot detection and localization. The technique is non-contact and non-destructive. Hotspots are localized areas that show higher temperatures than adjacent cells. These areas are clear signs of problems such as electrical mismatch, material degradation, or shading effects [17], [50].

Thermal hotspot identification focuses on detecting abnormal heat generation in PV modules. Detection is commonly achieved by analyzing absolute temperature values, relative thermal contrast, or localized thermal irregularities observed in thermal images. Such thermal signatures are frequently associated with partial shading, cell mismatch, soiling, or micro-cracks, which disrupt current flow and induce localized resistive heating [51]. Due to their ability to directly visualize these anomalies, IR thermography has been widely used as an effective early warning tool to identify potential hotspot formation before irreversible damage occurs [52].

In addition to detection, thermal-based localization aims to precisely delineate the spatial dimensions and limits of hotspot areas. Initial localization methods relied heavily on conventional thermal image-processing techniques, including global or adaptive thresholding, region-growing algorithms, and clustering-based segmentation. These methods carry out pixel-level or region-based hotspot delineation by utilizing thermal gradients and spatial continuity in the temperature field [53], [54]. These techniques exhibit low computational complexity and achieve accurate localization under specified conditions. However, the performance of these methods is highly dependent on image quality and thermal contrast. Thresholding techniques may fail when hotspot boundaries are unclear or when temperature variations are non-uniform. Region-growing approaches are sensitive to seed-point selection and may produce inaccurate boundaries in the presence of thermal noise. Similarly, clustering-based segmentation may struggle to separate hotspot regions from surrounding areas when thermal distributions overlap. These limitations reduce localization reliability in complex field environments and motivate the development of more robust deep learning-based segmentation approaches.

Accurate hotspot localization is important because the quality of the extracted hotspot region directly influences subsequent severity assessment. Inaccurate localization may include irrelevant background information or exclude affected areas, resulting in unreliable temperature measurements and degraded classification performance. Therefore, precise hotspot localization serves as a critical pre-requisite for reliable hotspot severity classification.

Thermal-based inspection relies on the distinctive thermal patterns exhibited by defective PV cells during operation. Hotspots exhibit clear local temperature increases, strong thermal contrast with nearby cells, uneven heat distribution, and sharp temperature changes at their edges [55]. These temperature patterns are the primary indicators of hotspots in IR thermographic analysis. This method identifies unusual areas without physical contact or interruption to system operation [52], [56].

Despite these advantages, thermal-based localization remains highly influenced by environmental and operational variations. Factors such as non-uniform cooling conditions, ambient temperature fluctuations, and surface emissivity variations can significantly affect thermal contrast, leading to inconsistent hotspot boundaries and potential false detections [57]. As a result, the reliability of hotspot localization may vary across different PV installations and operating scenarios. In addition, many classical thermal-based approaches rely on fixed threshold values or manual parameter tuning, which limits their robustness and repeatability under changing environmental conditions. These limitations restrict their applicability for large-scale PV monitoring, particularly in utility-scale systems operating under diverse and dynamic conditions. Table 2.1 summarizes key thermal characteristics commonly used in hotspot analysis, together with their associated limitations that affect accurate hotspot localization.

Table 2.1
Summary of Thermal Characteristics and Limitations in Thermal-based Hotspot Inspection

Aspect	**Description**	**Limitation**	**Implications for the proposed method**
Localized temperature elevation [31], [58]	Hotspots appear as regions with significantly higher temperature than surrounding cells	Magnitude affected by irradiance and cooling conditions	Two-tier segmentation framework improves hotspot localization accuracy
Thermal contrast [8]	The temperature difference between the hotspot and neighbouring cells enables detection	Reduced contrast under non-uniform illumination	CLAHE and image enhancement improve hotspot visibility

Aspect	Description	Limitation	Implications for the proposed method
Thermal gradients [6]	Steep gradients at hotspot boundaries support spatial localization	Gradient distortion due to wind and emissivity variation	Semantic segmentation enables more precise hotspot boundary delineation
Non-uniform heat distribution [41]	Irregular heat patterns indicate electrical or material defects	May be confused with benign thermal variations	Dual-input image and temperature feature fusion supports severity classification

To address these challenges, thermal-based hotspot localization techniques have gradually evolved from rule-based image processing methods toward data-driven learning frameworks. Early approaches primarily relied on classical segmentation techniques, such as thresholding, edge detection, and morphological operations. These methods offered simplicity and computational efficiency. However, limited robustness was observed under real-world operating conditions, as illustrated in Figure 2.2.

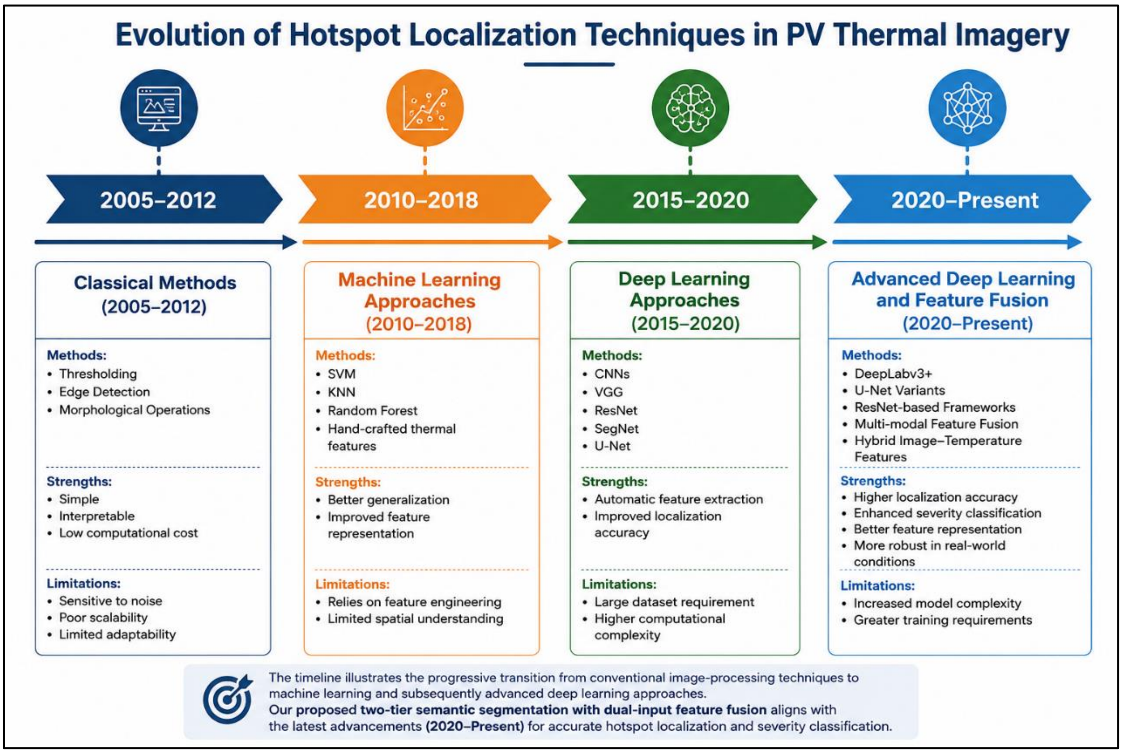


Figure 2.2 Evolution of Hotspot Localization Techniques in PV Thermal Imagery [28], [35], [59]

More recent studies have increasingly adopted Machine Learning (ML) and DL approaches to improve hotspot localization accuracy and automation [14], [60], [61].

ML-based methods introduced HC feature representations to enhance generalization, whereas DL-based frameworks enabled automatic feature learning directly from thermal images. These advances significantly improved localization performance and reduced dependency on manual parameter selection.

Despite the progress made by ML and DL-based approaches, several fundamental challenges remain unresolved. Hotspot localization accuracy can still degrade under low thermal contrast, non-uniform irradiance, and complex background conditions commonly encountered in field inspections. Moreover, many existing DL-based methods focus primarily on spatial localization performance, with limited consideration of severity-aware discrimination and boundary refinement under minor thermal variations.

These unresolved issues highlight the need for more robust hotspot localization strategies that can effectively suppress background interference, refine hotspot boundaries, and provide reliable inputs for subsequent severity classification. The following section further discusses these challenges and motivates the development of the proposed segmentation and classification framework.

2.2.3 Issue and Challenges in Hotspot Detection and Localization based on Thermal Imaging

Despite significant progress in thermal-based inspection and automated hotspot localization, several unresolved issues continue to limit the reliability and scalability of existing approaches. Current methods often struggle in real-world operational conditions, where environmental variability, data acquisition constraints, and data limitations introduce uncertainty into hotspot detection and interpretation. This section critically reviews the key challenges associated with hotspot detection and localization in PV systems, highlighting the limitations of manual inspection, UAV-based thermal imaging, temperature variability, and data-driven learning frameworks. These challenges include:

a) Limitations of Manual and Conventional Inspection Methods

Manual hotspot inspections rely on visual assessment and portable imaging equipment [62]. These procedures are labour-intensive, time-consuming, and heavily reliant on the operator's expertise [63], [64]. Such approaches are impractical for large-scale PV installations. Additionally, this method is also prone to subjective interpretation, leading to inconsistent detection and localization, especially under varying environmental conditions [64]. Differences in operator experience, camera viewing angle, and subjective severity judgements may lead to varying interpretations of the same hotspot condition, thereby reducing inspection consistency and repeatability. Therefore, automated hotspot localization and severity classification methods are required to reduce subjectivity and improve consistency.

b) Challenges Associated with UAV-based Thermal Inspection

UAV thermal imaging can enhance inspection capabilities and reduce labour. However, data quality can also be affected by several operational factors. Inconsistencies in thermal patterns may be caused by variations in flight altitude, flight speed and flight stability, as well as limitations in image resolution, camera focus, positioning accuracy and viewing angle [65]. These factors may influence temperature measurements, hotspot visibility and spatial localization accuracy, making reliable hotspot detection more challenging in large-scale PV installations. Consequently, maintaining consistent acquisition parameters is essential to ensure reliable hotspot localization and subsequent severity classification.

c) Environmental and Temperature-Induced Variability

Thermal-based hotspot detection is highly sensitive to external environmental factors, including irradiance fluctuations, wind-induced cooling, and ambient temperature changes [66]. These factors can distort thermal contrast and obscure inherent hotspot signatures, leading to false positives or missed detections. Such variability remains a major obstacle to achieving robust and repeatable hotspot localization across different operating scenarios. Furthermore, environmental fluctuations may influence temperature-based descriptors and thermal distributions extracted from hotspot regions. This thermal-feature instability may reduce the reliability of severity assessment and increase the risk of misclassifying hotspot severity.

d) Limitations in Hotspot Localization and Severity Interpretation

Most existing studies focus on hotspot detection without explicitly quantifying severity levels [67]. The absence of standardized severity interpretation limits the practical value of thermal inspection, as visually similar hotspots may correspond to significantly different degradation risks. This challenge is made worse when localization methods rely solely on pixel intensity without integrating contextual or temperature-based descriptors. Furthermore, hotspot localization alone does not provide sufficient information for maintenance prioritization, as the severity and potential impact of the detected hotspot remain unknown. Therefore, there is a need for integrated approaches that combine accurate hotspot localization with severity classification using both image and temperature-based features. Such integration can provide more informative assessments to support condition monitoring and maintenance decision-making in PV systems. This limitation highlights the need for a comprehensive framework that integrates hotspot localization and severity classification using complementary image and temperature information, which forms the primary motivation of the present study.

e) Generalization and Dataset-Related Challenges

Data-driven hotspot detection models often exhibit limited generalization across different PV sites due to site-specific training data and controlled experimental conditions [31]. Furthermore, tropical climatic conditions, such as those encountered in Malaysia, remain underrepresented in existing datasets [68]. This lack of diversity restricts model adaptability and reduces confidence in real-world deployment across varying climatic regions.

Collectively, these challenges highlight the limitations of existing hotspot detection and localization frameworks in addressing real-world operational complexity. The lack of robust severity interpretation, sensitivity to environmental variability, and limited cross-dataset generalization underscores the need for more advanced automated inspection frameworks. These research gaps motivate the development of the proposed framework, which integrates a two-tier semantic segmentation approach for hotspot localization, a dual-input feature fusion strategy combining thermal image and temperature information for severity classification, and expert validation through solar thermographer assessment. The proposed framework aims to provide a more reliable and practical solution for PV hotspot monitoring under real-world operating conditions.

2.3 Convolutional Neural Network (CNN) for Hotspot Classification in Thermal Images

CNNs have been widely adopted for hotspot classification in thermal images due to their ability to automatically learn discriminative features from complex spatial and intensity patterns [21], [31]. In solar PV applications, CNN-based models have demonstrated strong potential for analyzing thermal anomalies by capturing localized temperature variations associated with hotspot regions. Compared with conventional image-processing techniques, CNNs enable end-to-end learning without relying on HC rules, making them well-suited for automated hotspot analysis. Despite their effectiveness, applying CNNs to thermal hotspot classification presents several challenges, including limited texture variation, visual similarity across severity levels, and sensitivity to imaging conditions [69]. These challenges highlight the need for appropriate model architectures, feature-extraction strategies, and enhancement techniques, which are reviewed in the following subsections.

2.3.1 CNN Model

CNNs are a type of DL model made to automatically learn features from images. In PV thermal imagery, CNNs have demonstrated strong capability in identifying thermal patterns associated with hotspot defects and variations in severity. Through convolutional operations, CNN models can extract discriminative spatial features that support automated hotspot detection, localization, and classification tasks [70]. Figure 2.3 illustrates the basic architecture of CNN-based classification.

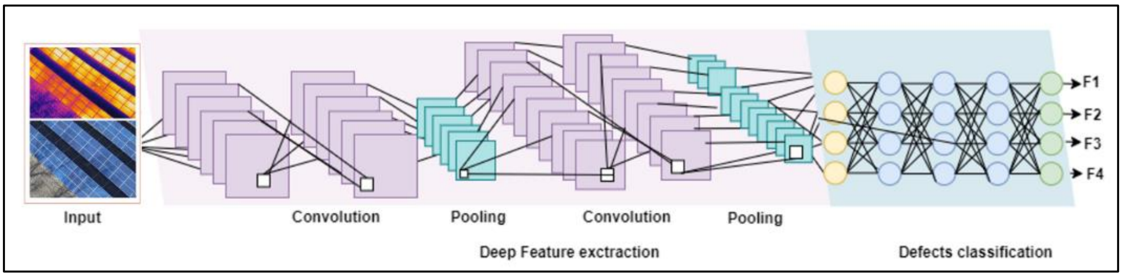


Figure 2.3 The Basic Architecture of CNN-Based Classification [61]

A typical CNN architecture consists of convolutional, activation, pooling, and fully connected layers. These layers operate sequentially to extract features of increasing complexity, ranging from low-level edges and textures to higher-level thermal patterns associated with PV faults. The extracted features are subsequently used for classification through the final output layer. Table 2.2 summarizes the functions of the main CNN layers and their relevance to PV thermal image analysis.

Table 2.2
Roles and Functions of CNN Layers in Thermal Imagery Analysis for PV Systems [71]

CNN Layer	**Description**	**Relevance to Thermal Imagery**
Convolution Layer	Extracts feature from the input image using kernels. It captures spatial hierarchies in the data.	Detects thermal edges and shapes important for identifying PV faults.
Pooling Layer	Reduces spatial dimensions (width and height) of feature maps while retaining essential features.	Helps focus on hot or cold zones by summarizing spatial regions.
Activation Function Layer	Introduces nonlinearity into the model using activation functions such as ReLU or Sigmoid to enable learning of complex patterns.	Crucial for differentiating minor temperature gradients present in hotspots.
Batch Normalization Layer	Standardizes the output of a preceding activation layer, speeding training and offering a degree of regularization.	Stabilizes training across varying brightness levels in thermal images.
Fully Connected Layer (FCL)	Flattens feature maps into a single vector by connecting all neurons in one layer to all neurons in another.	Transforms learned features for final decision-making (e.g., fault or no fault).
Classification Layer	Final layer that maps feature representation to output class probabilities using SoftMax or sigmoid.	Classifies input as hotspot, crack, shading, or normal zone in PV modules.

Despite the promising performance of CNNs in PV thermal inspection, several limitations remain. CNN models are highly dependent on the quality and quantity of training data, which may affect model generalization under varying environmental and operating conditions. In addition, conventional CNN architectures primarily rely on image-based information and may not fully exploit temperature-related descriptors that are directly associated with hotspot severity. Consequently, visually similar thermal patterns may correspond to different degradation levels, leading to uncertainty in interpreting severity. Furthermore, variations in irradiance, ambient temperature, and thermal contrast may affect the reliability of feature extraction and classification. These limitations highlight the need for more advanced approaches that integrate complementary thermal information to improve the assessment of hotspot severity. Therefore, although CNNs provide effective feature extraction, additional approaches are required to enhance hotspot localization and severity assessment in PV systems.

2.3.2 Model Architecture

CNNs are widely used to detect and classify hotspots in PV thermal imagery [72], [73], [74]. These models learn hierarchical spatial features directly from image data, enabling effective representation of complex thermal patterns. Unlike conventional rule-based or HC feature approaches, CNN-based models can capture complex thermal patterns and contextual information that are critical for accurate hotspot localization and severity interpretation. This section reviews the CNN architecture used across various research areas, focusing on its structural design and feature representation capabilities.

2.3.2.1 U-Net Architecture

In many vision applications, including image processing, deep neural networks are crucial, especially for segmenting high-resolution images. For example, research by [75], [76] proposed a U-Net–based approach to segment hotspot regions in solar PV modules. U-Net is a widely adopted DL–based architecture for image segmentation, particularly in applications that require accurate pixel-level localization [77]. Its symmetric encoder–decoder structure facilitates hierarchical feature learning during segmentation. Skip connections further preserve both low-level spatial details and high-level semantic information.

The encoder path performs successive down-sampling operations to extract hierarchical feature representations. The decoder path restores spatial resolution through up-sampling to generate dense pixel-wise predictions. The characteristic U-shaped architecture facilitates effective feature fusion across multiple scales, which has been shown to improve segmentation performance in complex imaging scenarios [78] .

Figure 2.4 presents a representative U-Net architecture, illustrating the encoder–decoder structure and skip connections commonly employed for pixel-level segmentation tasks. This architectural design has been widely adopted in the literature for its effectiveness in preserving spatial information during segmentation.

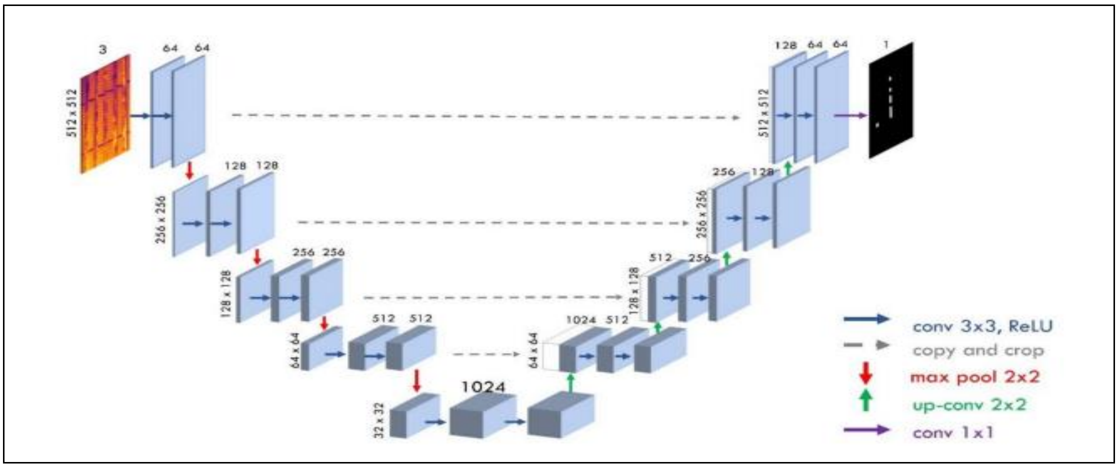


Figure 2.4 U-Net Architecture for Hotspot Region Segmentation in Solar PV thermal Images [79]

To further enhance segmentation accuracy, several studies have proposed modified U-Net architectures that incorporate attention mechanisms within the skip connections. These attention-based variants aim to emphasize relevant feature activations while suppressing less informative responses. This mechanism contributes to improved model robustness and generalization. Previous studies indicate that attention-based architectural enhancements improve the robustness of segmentation in thermal imagery. This is especially important in the presence of low-intensity temperature variations and background noise [81], [82].

Although attention-based U-Net variants have demonstrated improved segmentation performance in several studies, the additional attention modules increase model complexity and computational requirements. Furthermore, the objective of this study is not to develop a new segmentation architecture, but to evaluate and compare representative segmentation models within a unified framework for PV hotspot localization. Therefore, the standard U-Net architecture was selected as the baseline segmentation model due to its established performance, simplicity, and widespread adoption in thermal image segmentation studies.

2.3.2.2 ResNet Architecture

Residual Networks (ResNet) are deep neural network architectures that introduce shortcut connections, including identity mappings, to facilitate the training of very deep models [80]. These shortcut connections enable more effective gradient propagation during backpropagation. As a result, the vanishing gradient problem commonly encountered in deep CNN is reduced.

The central concept underlying ResNet is residual learning, in which the network learns a residual mapping rather than a direct transformation. By allowing information to bypass one or more layers through identity connections, ResNet architectures enable deeper networks to be trained without significant performance degradation. As a result, ResNet models have demonstrated strong performance in image recognition and classification tasks, maintaining high accuracy even as network depth increases.

Previous studies have extensively adopted ResNet architectures for image classification due to their ability to learn deep and discriminative feature representations efficiently [81]. More recent research has explored the application of ResNet-based models in thermal image analysis and PV fault classification. In these applications, robust feature learning is required under varying imaging conditions and limited texture information.

Figure 2.5 illustrates a representative ResNet residual block, which serves as the fundamental building unit of the ResNet architecture. Within each residual block, the input is transformed through a series of weighted layers and then combined with the original input via an identity shortcut connection.

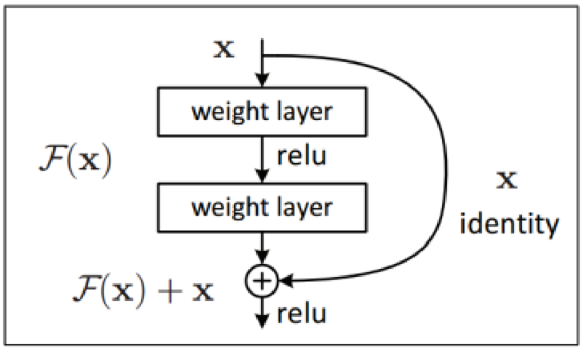


Figure 2.5 ResNet Residual Block Illustrating the Identity Shortcut Connection [82]

The residual mapping can be mathematically expressed as follows:

$$y = F(x) + x \tag{2.1}$$

Where F(x) represents the transformation learned by the weight layers, and x is the original input passed through the shortcut connection. This structure helps mitigate the vanishing-gradient problem, enabling deeper networks to learn effectively without performance degradation.

By focusing on learning residual representations, ResNet architectures simplify the optimization process and improve training stability in deep networks. These characteristics make ResNet-based models particularly well-suited to hotspot severity classification tasks, where effective discrimination is required despite visual similarity among thermal patterns.

a) ResNet-18

ResNet-18 is a widely used variant of the Residual Network architecture, characterized by its relatively shallow depth of 18 layers. It employs basic residual blocks consisting of two convolutional layers with batch normalization and ReLU activation, followed by identity shortcut connections [83]. This design enables effective gradient propagation while maintaining a lower computational complexity compared to deeper ResNet variants [81]. Due to its balanced trade-off between representational capacity and computational efficiency, ResNet-18 has been frequently adopted in image classification tasks. This is particularly the case when training data or computational resources are limited. Several studies have reported that ResNet-18 achieves competitive performance while offering faster training and inference times [84], [85].

These characteristics make it well-suited for time-critical and resource-constrained applications. In thermal image analysis, ResNet-18 has been explored for anomaly and fault classification tasks. These applications require robust feature extraction despite limited texture variation and low contrast in thermal images. Its relatively shallow architecture helps reduce the risk of overfitting while still enabling effective discrimination of thermal patterns associated with different fault conditions. As a result, ResNet-18 is often regarded as an attractive baseline model for hotspot detection and severity classification in solar PV systems. This is particularly relevant in scenarios that demand efficient and reliable performance.

b) ResNet-50

ResNet-50 is a deeper and more complex variant with 50 layers. It employs bottleneck residual blocks, which consist of three layers: a 1x1 convolution layer for dimensionality reduction, a 3×3 convolution for feature extraction, and another 1x1 convolution layer for dimensionality restoration. This bottleneck design reduces the number of parameters while enabling deeper network construction and improved representational capacity. Due to its greater depth, ResNet-50 can capture more complex and fine-grained feature representations than shallower variants. Studies by [86] have shown that the enhanced representational capacity of ResNet-50 leads to superior performance in applications requiring high precision and detailed feature learning. However, the increased architectural complexity also results in higher computational demands and longer training and inference times.

In image classification and visual recognition tasks, ResNet-50 has been widely adopted where sufficient training data and computational resources are available. In segmentation applications, ResNet-50 is commonly employed as an encoder backbone to extract discriminative features prior to pixel-wise prediction. However, this increased depth also meant that ResNet-50 required more computational resources and longer training and inference times than ResNet-18. These characteristics make ResNet-50 suitable for hotspot severity classification tasks that demand higher precision under challenging thermal imaging conditions.

The ResNet-50 model, like ResNet-18, benefited from transfer learning by utilizing weights pre-trained on the ImageNet dataset. Although ImageNet consists primarily of natural RGB images rather than thermal imagery, transfer learning remains beneficial because the lower layers of CNN models learn generic visual features such as edges, textures, and spatial patterns. These features can still provide useful initialization for thermal image analysis, particularly when the available thermal dataset is limited. However, domain differences between RGB and thermal imagery may reduce feature transferability, necessitating fine-tuning using domain-specific thermal images.

It was trained on the same annotated thermal images and optimized using similar techniques, including cross-entropy loss and data augmentation. Despite its higher resource demands, ResNet-50 achieves superior performance in hotspot identification and classification, making it suitable for applications requiring high precision.

2.3.3 Feature Extraction

Feature extraction is a crucial phase in CNN-based hotspot classification. This process transforms an unprocessed image into representations appropriate for learning and decision-making. In contrast to traditional HC descriptors, DL utilizes layered convolutional networks for automatic feature extraction [87]. Consequently, it allows the network to promptly acquire key features and spatial patterns from the data. For PV thermal images, where hotspot characteristics may vary in size, shape, and condition depending on operational conditions, this capacity is particularly important.

Despite these advantages, CNN-derived features mainly rely on visual and spatial information from the image [87]. This means that areas that look the same might have different severity levels if the temperature distribution is different. This issue is important in thermal-based analysis, where severity is usually determined by actual temperature values rather than appearance alone. Therefore, relying solely on image-based features can make it difficult for CNN models to consistently distinguish severity levels across different situations. Visually similar hotspot regions may correspond to substantially different temperature distributions, as reflected by variations in T_{max}, T_{min}, T_{ref}, and ΔT. Since hotspot severity is closely associated with the magnitude of temperature elevation rather than visual appearance alone, image-based CNN features may not adequately capture the thermal characteristics required for reliable severity assessment. Consequently, incorporating temperature-based descriptors alongside image features can provide complementary information and improve discrimination of hotspot severity.

DL-based feature extraction demonstrates superior performance compared with conventional HC features in hotspot detection and classification tasks [88]. By automatically learning hierarchical feature representations from thermal imagery, CNN models can capture complex spatial patterns, temperature gradients, and contextual information that are difficult to represent using manually engineered descriptors alone [89]. These capabilities contribute to improved hotspot localization, reduced false detections, and enhanced severity classification performance across different PV operating conditions [5]. Consequently, CNN-based approaches have become the dominant paradigm for automated hotspot analysis in PV thermal imagery.

Despite these advantages, CNN-derived features remain primarily driven by visual and spatial information contained in the thermal image. As a result, visually similar hotspot areas may be associated with different severity levels when the underlying temperature distribution is different. This limitation is particularly relevant in thermal-based analysis, where severity assessment often relies on quantitative temperature information rather than appearance alone. Therefore, relying on image-based features may limit the ability of CNN models alone to achieve consistent severity discrimination across a variety of operational scenarios.

This finding underscores the need for a feature representation that extends beyond visual descriptors. Adding data such as temperature-related or contextual features can help identify features and improve classification accuracy. This motivation forms the basis for developing an enhanced or hybrid CNN model, discussed in the next subsection.

2.3.4 CNN Feature Fusion Strategy

Accurate classification of hotspots in PV thermal imagery requires feature representations that capture both spatial characteristics and underlying thermal behaviour. DL models demonstrate strong capabilities for learning specific visual patterns. However, hotspot severity is closely associated with temperature-related characteristics that may not be adequately represented by image appearance alone. As a result, image-based representations may provide insufficient information for reliable severity assessment across diverse operational and environmental conditions [90]. In practical PV monitoring scenarios, visually similar hotspot patterns may correspond to different levels of degradation risk, which underscores the need for richer feature representations. As a result, enhanced feature-extraction strategies have attracted increasing attention to improve robustness and operational relevance in hotspot analysis. Together, these considerations motivate the exploration of feature representations that extend beyond single-modality learning.

Feature fusion has emerged as an effective strategy for improving feature representation by integrating complementary information from multiple data sources. In PV thermal imagery, integrating CNN-derived image features with temperature-based descriptors can provide both spatial and thermal information relevant to assessing hotspot severity. Such a dual-input representation has the potential to improve severity discrimination by capturing information that may not be adequately represented by either modality alone. This motivation forms the basis of the dual-input feature fusion strategy investigated in the present study.

CNN-only approaches rely on spatial and intensity-based image features. These models localize hotspots effectively but may struggle to distinguish severity levels when visual patterns appear similar, even if the temperature distribution differs [91]. This limitation reduces the reliability of severity classification. Particularly when thermal contrast is influenced by environmental conditions or sensor variability. As a result, visually comparable hotspot regions may be assigned similar severity labels even though they represent different degradation risks. Such constraints highlight the need for complementary feature representations to support more accurate and robust assessment of hotspot severity.

In contrast, HC feature-based methods use predefined temperature features and statistical measures. At the same time, it offers interpretability but limited adaptability. These approaches are often sensitive to threshold selection and domain-specific assumptions, which restrict generalization across different PV systems and climatic conditions [19]. Furthermore, the complex spatial information in thermal images may not be fully represented by HC features alone. As a result, neither CNN-only nor HC-only approaches consistently meet the requirements for accurate, scalable hotspot severity classification.

In classical ML approaches, HC features are commonly employed to represent texture, intensity, and statistical characteristics extracted from thermal images. However, these approaches rely heavily on manually designed feature representations and may not fully capture the complex spatial and thermal patterns associated with PV hotspot analysis. Consequently, the effectiveness of HC features often depends on domain expertise and feature selection strategies, which may limit generalization across different operating conditions. In contrast, this study employs deep extraction via CNN-based architectures and integrates temperature-related features to assess their impact on severity classification performance.

Table 2.3 summarizes the characteristics of CNN-only, HC-only, and hybrid feature-fusion approaches for hotspot severity classification. The comparison indicates that hybrid feature fusion combines the strengths of both approaches by integrating spatial information extracted by CNN models with explicit temperature-based descriptors, thereby providing a more comprehensive representation of hotspot severity.

Table 2.3
Comparison of CNN-only, HC-only, and Hybrid Feature Fusion Approaches for Hotspot Severity Classification

Aspect	**CNN-only Features**	**HC-only Features**	**Hybrid Feature Fusion**
Feature source	Image-based deep features	Temperature descriptors and manually designed features	CNN-derived image features and temperature descriptors
Spatial information	Excellent	Limited	Excellent
Temperature information	Indirectly represented	Explicitly represented	Explicitly represented
Feature extraction	Automatic	Manual	Automatic and manual
Interpretability	Moderate	High	Moderate to high
Generalization capability	Good	Limited	Improved
Severity discrimination	May be limited when hotspots appear visually similar	May miss complex spatial patterns	Improved through complementary feature representation

Several studies have demonstrated that combining HC and DL features can boost model performance by leveraging the strengths of both approaches [92], [93]. Both HC and DL feature-extraction methods have advantages. HC features are computationally efficient and generalize well across domains but require domain expertise and may miss complex patterns. DL features, while computationally expensive, excel at capturing intricate patterns in data, especially in large datasets. Combining the two approaches often leads to superior performance by leveraging complementary feature representations.

The complementary strengths of DL-derived spatial features and temperature-aware HC descriptors motivate the development of hybrid CNN-based frameworks. Hybrid models integrate CNN-extracted feature maps with explicit thermal descriptors. This approach jointly represents visual structure and quantitative temperature information. Recent studies show that hybrid strategies improve discrimination between visually similar hotspots with different severity implications [94], [95]. This integration also enhances robustness under environmental variability and improves cross-dataset generalization performance. Moreover, hybrid feature representations enhance interpretability and yield more reliable assessments of hotspot severity. These advantages provide a strong foundation for employing enhanced CNN models in advanced PV hotspot classification systems.

2.4 Review of Past Research in Automated Segmentation and Severity Classification of Hotspot Thermal-based Images

Research on automated analysis of PV hotspot thermal imagery has advanced significantly over the past decade, driven by advances in image processing, ML, and DL. Early studies focused mainly on conventional image processing methods for hotspot detection [96], [97] . Building on these methods, more recent work has explored semantic segmentation frameworks to achieve accurate localization of the hotspot region [16], [53]. At the same time, increasing attention has been directed toward hotspot severity classification to support condition assessment and maintenance decision-making. This section reviews prior research on automated hotspot analysis, with emphasis on conventional techniques, semantic segmentation approaches, and severity classification strategies applied to thermal images. The review critically examines methodological trends, performance achievements, and existing limitations, thereby identifying research gaps that motivate the proposed framework.

2.4.1 Conventional and Learning-Based Techniques for Hotspot Segmentation

In solar PV hotspot detection, localization is the process of identifying the exact areas in thermal or infrared images that show unusual temperature signatures [68]. Accurate localization is essential not only for isolating defects but also for subsequent tasks such as performance optimization, preventive maintenance, and severity classification [68], [98].

Traditional image processing methods are widely adopted due to their interpretability, low computational complexity, and suitability for real-time deployment in resource-constrained environments [99]. Common techniques include global and adaptive thresholding, Otsu-based segmentation, region growing, edge detection, and morphological operations applied to temperature-differential maps [100], [101], [102] . However, these methods often exhibit limited robustness to temperature variations, sensor noise, and non-uniform thermal gradients, which restrict their performance under real-world operating conditions [103].

To address these limitations, learning-based approaches such as Support Vector Machines (SVM), k-Nearest Neighbors (k-NN), Random Forest, and Artificial Neural Networks (ANN) have been introduced to enhance feature-based classification capability. More recently, DL architecture including CNNs and U-Net-based semantic segmentation models have been applied for automated hotspot localization. Nevertheless, overall performance remains highly dependent on the availability of well-annotated datasets and the quality of HC feature design [104], [105].

With the introduction of DL approaches, the debate surrounding hotspot localization has intensified. Several studies report that CNN-based models achieve superior precision and generalization compared to classical and traditional ML techniques [99], [105]. However, these models usually require large-scale annotated datasets and substantial computational resources and may also exhibit lower interpretability [106]. To address these issues, hybrid systems that combine traditional pre-processing with ML or DL-based feature extraction have emerged as promising compromise solutions, effectively balancing efficiency and performance [88], [107].

The advantages and limitations of each localization method reveal complementary characteristics. Classical approaches offer superior processing speed but exhibit reduced robustness in complex backgrounds. Conventional ML techniques demonstrate improved generalizability. However, scalability remains limited without robust engineering features. Although DL-based models require substantial computational resources, these architectures automatically learn hierarchical feature representations that address such limitations.

In conclusion, classical, ML, and DL-based localization strategies each offer distinct advantages. However, persistent challenges remain in integrating severity, real-time deployment, and dataset diversity. These limitations highlight the need for a more comprehensive framework that combines accurate hotspot localization with reliable severity assessment. Therefore, this study proposes a two-tier semantic segmentation framework for precise hotspot localization, followed by a severity-aware CNN classification framework that integrates image and temperature information for hotspot severity assessment. This integrated approach forms the foundation of the methodology presented in this thesis.

2.4.2 Semantic Segmentation for Hotspot Localization based on Thermal Image

Semantic segmentation has become a widely adopted and effective technique in the analysis of PV systems, especially for identifying and localizing hotspot defects in thermal imagery. Unlike traditional object detection methods that rely on bounding boxes, semantic segmentation assigns class labels to each pixel, thereby enabling precise delineation of boundaries between healthy and defective regions. This pixel-level accuracy is particularly valuable for detecting small, localized faults such as delamination, microcracks, or emerging hotspots, which may not be captured adequately using conventional inspection techniques.

Recent DL architectures, such as U-Net, DeepLabv3+, and SegNet, have been successfully applied to segment hotspot regions from thermal infrared and electroluminescence (EL) images. While thermal imaging is effective for identifying temperature anomalies associated with hotspot formation, defects such as microcracks and early-stage delamination are often more effectively detected using EL imaging. Thermal imaging becomes useful when these defects generate measurable thermal signatures that manifest as localized temperature variations. U-Net, with its symmetric encoder–decoder structure and skip connections, effectively preserves spatial detail and demonstrates high accuracy in detecting small-scale anomalies in PV modules [108]. This architecture is particularly effective when applied to high-resolution aerial thermal images collected via drone platforms, enabling real-time, non-invasive PV inspection.

An enhanced version of the original DeepLab architecture, DeepLabv3+, combines an encoder–decoder framework with atrous spatial pyramid pooling to collect multi-scale context and perform fine boundary detection. This design allows DeepLabv3+ to perform robustly even with noisy or low-contrast thermal images, making it particularly valuable in dense module arrays where hotspot boundaries are difficult to distinguish [109]. ResNet-based architectures are also widely adopted as encoder backbones in semantic segmentation frameworks due to their residual learning capability. The residual connections facilitate the training of deeper networks by mitigating gradient degradation and enabling more effective feature extraction. In PV thermal imagery, ResNet backbones have demonstrated strong performance in capturing thermal patterns and object boundaries, making them suitable for hotspot localization tasks. Consequently, ResNet-18 and ResNet-50 have been extensively integrated into segmentation frameworks such as DeepLabv3+ to improve localization accuracy and feature representation.

In contrast, SegNet employs a symmetric encoder-decoder architecture that prioritizes computational efficiency and memory optimization. Unlike U-Net, SegNet utilizes pooling indices from the encoder during the up-sampling process, reducing the number of trainable parameters while preserving spatial information [110]. These characteristics make SegNet suitable for real-time and resource-constrained applications. However, the absence of skip connections may limit its ability to recover fine boundary details, particularly when segmenting small hotspot regions in thermal imagery.

Additionally, the accuracy of boundary delineation between hotspot and non-hotspot regions has been improved by incorporating edge-aware modules, attention mechanisms, and multi-resolution inputs into PV systems' semantic segmentation. Such improvements have broadened the application of semantic segmentation beyond hotspot detection, enabling tasks such as module-level labelling, fault classification, and shading analysis [109], [110]. In these processes, image pre-processing techniques such as contrast enhancement (e.g., CLAHE) and denoising filters (e.g., Non-Local Means) are commonly applied to improve image quality and increase model robustness. An essential step in thermal imaging is detecting temperature gradients, which are often minor and therefore sensitive to external interference. A detailed comparison of the main semantic segmentation models applied in PV hotspot detection is presented in Table 2.4

Table 2.4
Comparison of Semantic Segmentation Models for PV Hotspot Detection

Model	**Dataset Type**	**Evaluation Metrics**	**Performance**	**Advantages**	**Limitations**	**Relevance to PV Hotspot Localization**	**Study**
Customized U-Net	UAV thermal infrared orthomosaics from 31 PV sites	Mean IoU	Mean IoU = 0.905	Accurate pixel-level segmentation of overheated PV modules in aerial thermal imagery	Performance affected by blurry thermal boundaries and thermal image resolution limitations	Directly relevant to UAV-based hotspot localization in PV thermal imagery	[112]
U-Net, PSPNet, DeepLabv3+	593 EL images of PV cells	mIoU, Recall	DeepLabv3+ achieved highest average defect mIoU (0.28) and crack recall (0.86)	Effective pixel-level defect localization and boundary delineation	Crack segmentation remains challenging due to sparse and irregular defect morphology	Demonstrates the effectiveness of semantic segmentation for PV defect localization	[113]
Attention-Based SegNet (VGG16 + CBAM)	29-class EL image dataset	Dice, IoU, mIoU, Precision, Recall, F1-score	Dice = 94.08%, mIoU = 91.01%, F1-score = 95.38%	Attention mechanism improves feature representation and fine defect segmentation	Performance remains sensitive to annotation quality and very small defects	Shows the potential of attention-enhanced segmentation for accurate defect localization	[114]
U-Net with VGG16 Encoder	EL images (In-house + ELPV dataset)	Precision, Recall, F1-score, Expert Validation	F1-score = 0.9064 (Dark Zone), 0.9083 (Busbar), 0.6557 (Crack)	Includes expert validation and multiple datasets	Crack detection remains difficult and evaluation is limited to laboratory conditions	Highlights the importance of accurate segmentation and expert validation	[115]

In summary, semantic segmentation provides a high-resolution and automated approach for hotspot localization in PV systems. By combining advanced DL architectures with image pre-processing techniques, these methods improve the accuracy and reliability of hotspot detection. However, most existing studies primarily focus on hotspot localization and defect segmentation, with limited attention given to hotspot severity classification and expert-based validation. Furthermore, the integration of temperature-related information into severity assessment remains underexplored. These limitations highlight the need for a comprehensive framework that combines accurate hotspot localization, severity classification, and expert validation, which motivates the proposed methodology in this study.

2.4.3 Hotspot Severity Classification based on Enhanced CNN Model

The severity classification of hotspots in solar PV systems is critical in ensuring the operational reliability and long-term performance of solar PV modules. Furthermore, it contributes to optimizing overall system efficiency and extending its service lifespan. Hotspots are localized areas on solar PV modules that exhibit higher temperatures compared to the surrounding areas, often indicative of underlying defects or inefficiencies. Solar PV hotspot classification involves the identification, analysis, and categorization of hotspots in solar PV systems to prevent performance degradation and ensure the longevity of the system [111]. Accurately classifying the severity of these hotspots is essential for effective maintenance and preventing potential damage or performance loss in the solar PV module.

The classification of hotspot severity has undergone multiple separate stages of development. Basic thermography and manual inspections were the mainstays of early research, though these methods were constrained by subjectivity and inefficiency, despite providing valuable information on thermal conditions [112]. Between 2010 and 2015, infrared thermography became widely adopted, enabling non-invasive thermal imaging of PV modules with greater precision [113]. From 2015 to 2020, the introduction of UAVs revolutionized PV monitoring by allowing large-scale inspection at reduced cost and higher efficiency [114], [115]. Since 2020, studies have been focusing more on AI and DL methods, such as CNNs and semantic segmentation, which aid in automatic detection and severity classification [116], [117]. The progression from manual processes to AI-driven approaches illustrates the field's movement towards more robust, scalable, and adaptive frameworks, as summarized in Table 2.5.

Table 2.5
Timeline of Hotspot Severity Classification in PV Systems

Period	**Key Development**	**Description**	**Representative Studies**
Before 2010	Manual inspection and basic thermography	Hotspots detected visually or with handheld infrared cameras. Methods were labour intensive, subjective, and limited in scalability.	[112]
2010–2015	Infrared thermography integration	Systematic use of infrared cameras for anomaly detection in PV modules. Enabled non-invasive and more reliable detection compared to manual inspection.	[6], [86]
2015–2020	UAV-based imaging	UAVs deployed for aerial thermal imaging, allowing large-scale PV plant inspections with reduced cost and higher efficiency.	[114], [115]
2020–2025	AI-driven classification (CNN, DL, semantic segmentation)	Application of machine learning and deep learning techniques for automated hotspot detection and severity classification. Enabled multi-class categorisation and predictive maintenance frameworks.	[116], [117]

Despite these advances, several challenges remain unresolved. Most recent studies focus primarily on hotspot detection or localization, while limited attention has been given to comprehensive hotspot severity assessment. Existing AI-based approaches often rely solely on image features and may not adequately incorporate temperature-related information that is directly associated with hotspot severity. Furthermore, variations in severity definitions, environmental conditions, and dataset characteristics continue to affect model generalization and result consistency across different PV installations. The lack of expert-based validation further limits confidence in the practical deployment of automated severity classification systems. These limitations indicate that accurate hotspot localization alone is insufficient and highlight the need for integrated frameworks that combine localization, severity classification, and expert validation.

Two dominant perspectives are reflected in the literature. One group of studies argues that binary classification provides sufficient information for maintenance decision-making [61]. According to a different group, multi-class severity categorization, which differentiates between mild, severe, and critical irregularities, has greater practical utility [118]. Similar debates exist regarding rule-based threshold methods, which are interpretable and straightforward, and AI-driven approaches, which are more precise but computationally intensive [118].

Despite being practical and straightforward to use, binary classification techniques lack the severity granularity required to prioritize maintenance interventions. Binary classification only differentiates between normal and faulty conditions, providing limited information regarding the severity of hotspot degradation. In practical PV maintenance, hotspots with different severity levels may require different intervention priorities and response times. Therefore, multi-class severity classification offers greater practical value by supporting condition-based maintenance and more effective resource allocation. On the other hand, multi-class severity systems rely on high-quality annotated datasets and computational resources to deliver targeted maintenance recommendations [27], [119]. Moreover, AI-driven models often struggle to generalize, as training data from specific environments does not transfer well to other climates [120]. A lack of alignment in severity classification criteria remains a persistent challenge, as varying definitions and threshold values across studies limit the consistency and comparability of reported findings [121].

Clear patterns emerge when results from several sources in the literature review are integrated. When paired with AI frameworks, UAV-based imagery improves scalability and cost-effectiveness, enabling real-time monitoring of huge PV plants [114], [115]. Hybrid methods that combine thermal and electrical imaging further improve reliability [121]. Existing research contributes to SDG 7 and SDG 9 through improved PV system monitoring and operational efficiency.

Finally, research on hotspot severity classification has progressed considerably, moving from manual inspections to AI-powered frameworks. Binary classification methods are efficient but provide limited information regarding hotspot severity, whereas multi-class approaches support more informed maintenance decisions. Nevertheless, reliable assessment of hotspot severity requires three key elements: accurate hotspot localisation, reliable temperature reference data, and robust severity classification. In addition, expert-based validation remains essential to ensure that automated predictions are consistent with practical thermographic assessment. These requirements highlight the need for an integrated framework that combines semantic segmentation, temperature-aware severity classification, and expert validation, which directly motivates the methodology proposed in this study. In order to fulfil the requirement, there are two main important aspects to be explored as follows:

2.4.3.1 Criteria for Hotspot Severity Assessment

Hotspot severity classification in solar PV systems is essential for effective maintenance and optimization of solar panel performance. Several key criteria are used to assess the severity of hotspots. One of the primary indicators of hotspot severity is the temperature differential, which measures the difference in temperature between the hotspot and the normal operating temperature of the panel [56]. A larger temperature differential typically signifies a more severe hotspot, as it indicates a significant deviation from the panel's optimal operating conditions. High-temperature differentials can lead to increased thermal stress on the panel's materials, potentially causing long-term damage and reducing the panel's efficiency [17].

Typically, the severity of defects is determined by calculating the temperature difference between overheated and healthy areas within a module [122]. The hotspot temperature is the temperature of a malfunctioning component that is higher than that of a reference location. The reference location is a component of the same type, load, or repetition with the lowest temperature. In the present study, the reference temperature is determined from thermally healthy regions identified within the PV module image. This image-based reference temperature is subsequently used to calculate the temperature differential (ΔT), a key indicator of assessing hotspot severity.

To prevent future occurrences of the hotspot, it's critical to examine its underlying causes in addition to addressing the immediate problem. This could entail locating and repairing any shading or damage to the solar PV module, enhancing the system design to reduce the risk of hotspots, or implementing routine maintenance and inspection procedures to detect and address hotspots before they worsen. The most relevant study by [56] suggest three levels of hotspot severity which are low, moderate, and high.

Table 2.6

Comparative Review of Hotspot Severity Assessment Techniques in PV Systems

Severity Basis	Dataset Type	Severity Class	Expert Validation	Limitation	Study
Temperature Differential (ΔT)	Thermal IR images	3 Classes: Overheated, Potential Defect, Normal	No	Sensitive to irradiance, ambient temperature, and emissivity variation	[57]
Temperature Gradient	UAV Thermal Images	4 Classes: Extremely Severe, Severe, Heated Cell, Normal	No	Influenced by altitude, viewing angle, illumination, and wind-induced blur	[128]
Captured IR Images	Thermal Images	4 Classes: Healthy, Low, Medium, High	No	Sensitive to dataset size and hyperparameter tuning	[129]
Time-Series Waveform + Fuzzy Inference	Electrical and Environmental Monitoring Data	3 Classes: Severe, Slight, Medium	No	Requires stable weather conditions and fine-grained monitoring data	[12]

Table 2.6 presents a comparative analysis of PV hotspot severity assessment approaches, organised by severity criteria, dataset type, classification scheme, and reported limitations. Existing studies employ various severity indicators, including ΔT, thermal gradients, image-based analysis, and fuzzy inference methods. The comparison shows that severity definitions and classification schemes vary considerably across studies, ranging from three-class to four-class categorization systems. In addition, different data sources, including thermal images, UAV thermal imagery, and time-series monitoring data, influence the applicability and generalizability of the reported methods. Although several approaches demonstrate promising performance in severity assessment, expert-based validation is rarely reported, limiting confidence in the practical reliability of automated severity classification systems. Among the reviewed approaches, ΔT-based methods remain widely adopted because relative temperature differences reduce dependence on absolute temperature measurements and partially mitigate the effects of environmental variations and camera calibration. However, ΔT remains affected by factors such as irradiance, ambient temperature, and measurement

conditions. These observations highlight the need for a standardized severity assessment framework that combines reliable temperature references, robust classification methods, and expert validation.

For each hotspot severity level, different maintenance actions may be required [123]. Low-severity hotspots generally require routine monitoring to ensure that the condition does not deteriorate over time. Moderate-severity hotspots typically require corrective maintenance, such as addressing shading effects, repairing damaged components, or improving operating conditions. In contrast, high-severity hotspots demand immediate intervention because they may lead to accelerated degradation, significant power loss, or safety risks. In such cases, repair, replacement, or temporary system shutdown may be necessary. Therefore, hotspot severity classification plays an important role in maintenance prioritization by enabling resources and corrective actions to be allocated according to the level of risk and urgency.

In addition to addressing the current issue, it is essential to examine the root causes of the hotspot to prevent future recurrence. This may involve identifying and repairing shading or damage to the solar PV module. It can also include improving the system design to reduce the likelihood of hotspot formation. In addition, regular maintenance and inspection procedures can be established to detect and address hotspots before they develop. Beyond identifying the causes and preventive measures, the physical characteristics of a hotspot are also important for assessment.

The thermal size or area of the hotspot is another crucial criterion for evaluation. Larger hotspots can affect a greater number of solar cells within a panel, leading to more extensive areas of reduced efficiency and power output [17]. The size of the hotspot is typically measured in terms of the number of pixels in thermal images, which corresponds to the affected area on the panel [124]. Larger hotspots may also indicate underlying issues that could spread or worsen if not addressed promptly.

The location of the hotspot on the solar PV module is a significant factor in determining its severity. Hotspots located near critical components, such as electrical junctions, busbars, or connectors, pose a higher risk of causing electrical failures or fires. Additionally, hotspots in central areas of the panel can disrupt the energy flow more significantly than those on the periphery. The spatial proximity to critical system components and the potential for secondary damage make location an important factor in severity assessment.

The persistence of a hotspot over time is another important criterion. Hotspots that appear intermittently may indicate transient issues, such as temporary shading or minor dirt accumulation, which might not be as severe. However, hotspots that remain constant over extended periods are generally more concerning, as they suggest ongoing issues like cell degradation, soldering defects, or permanent shading. Persistent hotspots can lead to continuous performance degradation and accelerated wear of the affected cells.

Maintenance teams can prioritize interventions by evaluating this temperature differential. They can also allocate resources more effectively by considering hotspot size, location, and duration criteria. Addressing the most severe hotspots first helps mitigate potential damage and maintain the overall efficiency and longevity of the solar PV system. Accurate severity assessment enables proactive maintenance, reducing downtime and optimizing the energy output of solar installations [56] [123].

2.4.3.2 Integration of Semantic Segmentation and Convolutional Neural Network for Severity Classification

Recent years have seen a remarkable evolution in the use of DL in hotspot severity identification. CNNs were primarily used in early DL-based research to classify hotspot occurrences in a binary manner [21]. Compared with manual thermographic analysis, the detection accuracy of these early attempts improved significantly. To accomplish pixel-level localization of hotspot regions, researchers subsequently incorporated semantic segmentation architectures such as SegNet, U-Net, and DeepLab [76], [110], [125]. More recent research has expanded these frameworks by incorporating attention mechanisms, residual connections, and lightweight CNN models, which improve both computational efficiency and accuracy [126]. Emerging approaches are increasingly focusing on real-time applications using UAV-mounted thermal cameras, hybrid feature fusion, and transfer learning for improved generalizability [105].

In contrast to conventional image processing, DL techniques show excellent accuracy in identifying and categorizing hotspot severity. There are still a few restrictions, though. Robust model training requires large, annotated datasets, but these are rare and often limited to specific conditions [105]. Additionally, DL models are prone to overfitting, especially when trained on imbalanced severity classes or small datasets. Another difficulty is computational complexity, as models with complex structures require substantial resources and are challenging for real-time inspections [71], [105]. A further weakness lies in the lack of universally accepted severity metrics, which hinders comparability across studies and slows down industry adoption [127].

A summary of previous research reveals several recurring themes. Hybrid frameworks that integrate segmentation and CNN-based classification provide more accurate severity evaluation than stand-alone methods [128]. Furthermore, combining data augmentation with transfer learning has proven effective in addressing limited dataset challenges and enhancing model generalization [128]. UAV-enabled real-time inspections coupled with DL algorithms are increasingly validated for industrial-scale PV monitoring [129]. Collectively, these advancements demonstrate that progress in DL improves performance in hotspot detection and severity classification while contributing to broader sustainability objectives.

In conclusion, a synergistic framework for PV hotspots severity assessment can be achieved by integrating semantic segmentation and CNN-based classification. Semantic segmentation enables accurate hotspot localization and region-of-interest extraction, while CNN models support automated severity classification. However, existing studies rarely integrate segmentation-driven hotspot extraction with temperature-aware feature representations for severity assessment. In addition, expert-based validation remains limited, restricting confidence in the practical reliability of automated classification systems. These gaps motivate the development of the proposed framework, which combines two-tier semantic segmentation, dual-input feature fusion, and thermographer-based validation to improve the classification of hotspot severity in PV thermal imagery.

2.4.4 Research Gaps and Limitations in Hotspot Detection and Severity Classification in Thermal Imaging

There remains a significant research gap in assessing hotspot severity, despite advances in automated hotspot detection using PV thermal imaging. Most current methods rely primarily on binary or coarse classification to differentiate between healthy and defective PV modules, without explicitly identifying their severity levels. Visually identical hotspots may correlate with different degradation risks and maintenance priorities, limiting the practical applicability of such systems. Several recent reviews highlight that severity-aware analysis remains insufficiently addressed in current thermal-based hotspot assessment frameworks, particularly in the context of predictive maintenance and safety assessment [130], [131].

The hotspot characteristic representation is another important area of exploration that remains underexplored. DL-based techniques, which mainly rely on spatial and intensity-based features extracted from thermal images, frequently overlook quantitative temperature descriptors. Image-based features improve localization accuracy. The intensity of hotspots is fundamentally linked to absolute temperature measurements, temperature gradients, and thermal differentials, which are not entirely represented by visual appearance alone [55]. In contrast, methods relying solely on HC temperature features lack spatial context and robustness. The lack of a unified framework integrating both image-derived and temperature-based information remains a key limitation in existing hotspot analysis studies [132].

Another unsolved issue is robustness and generality across various operating situations. Many reported models are trained and evaluated using site-specific or controlled datasets collected from a limited number of PV installations under fixed conditions, which limits their ability to generalize across different PV installations and climatic regions [133], [134]. Environmental elements that have significant effects on thermal signatures and reduce model reliability include variations in ambient temperature, wind-induced cooling, and irradiance fluctuations [38]. Additionally, tropical climates, such as those in Southeast Asia, remain underrepresented in publicly accessible datasets, limiting the practicality of current approaches [135], [136].

Limited validation in real-world operational settings is another gap in the study, in addition to methodological limitations. Numerous studies report promising performance in offline evaluations but lack validation through expert annotation, long-term field testing, or large-scale deployment. Issues related to interpretability, scalability and real-time implementation are often not sufficiently examined, reducing confidence in operational acceptance [137]. These gaps underscore the need for automated PV monitoring frameworks capable of accurate hotspot localization, robust generalization, severity-aware evaluation, and reliable validation in operational environments. Based on the critical gaps identified in the preceding review, Table 2.7 presents the alignment between the identified research gaps, the corresponding research objectives, and the associated research contributions.

Table 2.7
Alignment of Identified Research Gaps with Research Objectives and Contributions

Identified Research Gap	**Related Research Objective**	**Corresponding Research Contribution**
Limited capability of existing approaches to perform precise pixel-level hotspot localization in aerial thermal images.	To develop a two-tier semantic segmentation framework for precise hotspot localization in solar PV aerial thermal images.	A two-tier semantic segmentation framework that accurately segments PV modules and localizes hotspot regions at pixel level.
Inadequate integration of spatial image features and temperature-based characteristics for hotspot severity analysis.	To model a dual-input feature extraction strategy that integrates CNN-based and temperature-based features from segmented hotspot regions.	A hybrid dual-input feature fusion strategy combining deep thermographic features with temperature descriptors.
Limited focus on hotspot severity classification beyond binary detection.	To develop an enhanced CNN-based classification model that integrates the dual-input feature extraction strategy for accurate hotspot severity classification.	A severity-aware CNN-based hotspot classification model capable of distinguishing multiple hotspot severity levels.
Limited validation of automated hotspot severity classification against expert assessment.	To validate the performance of the proposed model using standard evaluation metrics and expert-based statistical agreement analysis.	Expert-validated hotspot localization and severity classification supported by quantitative evaluation and statistical correlation analysis.

These identified gaps collectively justify the need for a robust and severity-aware hotspot analysis framework, which is detailed in the subsequent methodology chapter.

2.5 Summary

Infrared thermography's ability to detect irregularities, including hotspots, has established it as a crucial tool for monitoring PV systems. The reviewed literature demonstrates a clear transition from labour-intensive manual inspection and electrical testing to automated, learning-based hotspot assessment frameworks. Advances in image processing, ML, and DL have significantly improved the accuracy and scalability of hotspot detection and localization in large-scale PV installations.

Despite these advances, several limitations in the research still exist. The robustness of many techniques is reduced in real-world operating conditions due to the high sensitivity to environmental variability, including fluctuations in ambient temperature, cooling effects, and solar irradiance. Furthermore, binary hotspot identification is the primary focus of most described frameworks, with little attention given to severity characterization. The lack of integrated, temperature-aware, and severity-sensitive analysis limits the effectiveness of current methods in supporting risk mitigation and predictive maintenance.

Overall, the literature highlights a significant research need for developing robust, scalable, and severity-based hotspot assessment frameworks that can precisely locate hotspots and evaluate their severity across a variety of operating conditions. Addressing these limitations provides the foundation for the methodology presented in the subsequent chapter. To address the identified research gaps, Chapter 3 proposes a framework comprising two-tier semantic segmentation for hotspot extraction, dual-input image and temperature feature fusion, enhanced CNN-based severity classification, and expert thermographer validation to assess practical reliability.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 Introduction

This study employs deep learning (DL) methods to develop a hotspot localization and classification model for solar photovoltaic (PV) modules using thermal imagery. The proposed model is designed based on thermal images captured from solar PV modules and verified by a solar thermographer. The model aims to classify the detected hotspots according to the corresponding severity levels. This chapter outlines the comprehensive research framework for developing an automated hotspot localization and severity classification model. The proposed framework is designed to enhance the accuracy of hotspot detection and severity classification. The proposed methodology is systematically organized into four main stages that involve aerial thermal image pre-processing, semantic segmentation for hotspot identification, severity classification using convolutional neural networks (CNNs), and performance evaluation. This study enhanced the CNN classification model through dual-input feature fusion and additional convolutional refinement layers. The overall research design presented in this study aims to improve the efficiency of PV module monitoring and maintenance systems. This is achieved by minimizing energy losses associated with undetected anomalies, with a specific focus on hotspot-related anomalies.

3.2 Research Design

This research outlines the overall technique adopted in this study, presented in four stages with corresponding research objectives (ROs). Figure 3.1 illustrates the overall framework of the proposed methodology for hotspot localization and its severity classification model. Each stage corresponds to the respective ROs, and the detailed explanation is presented at the end of each corresponding section, where the red blocks indicate the main research contributions.

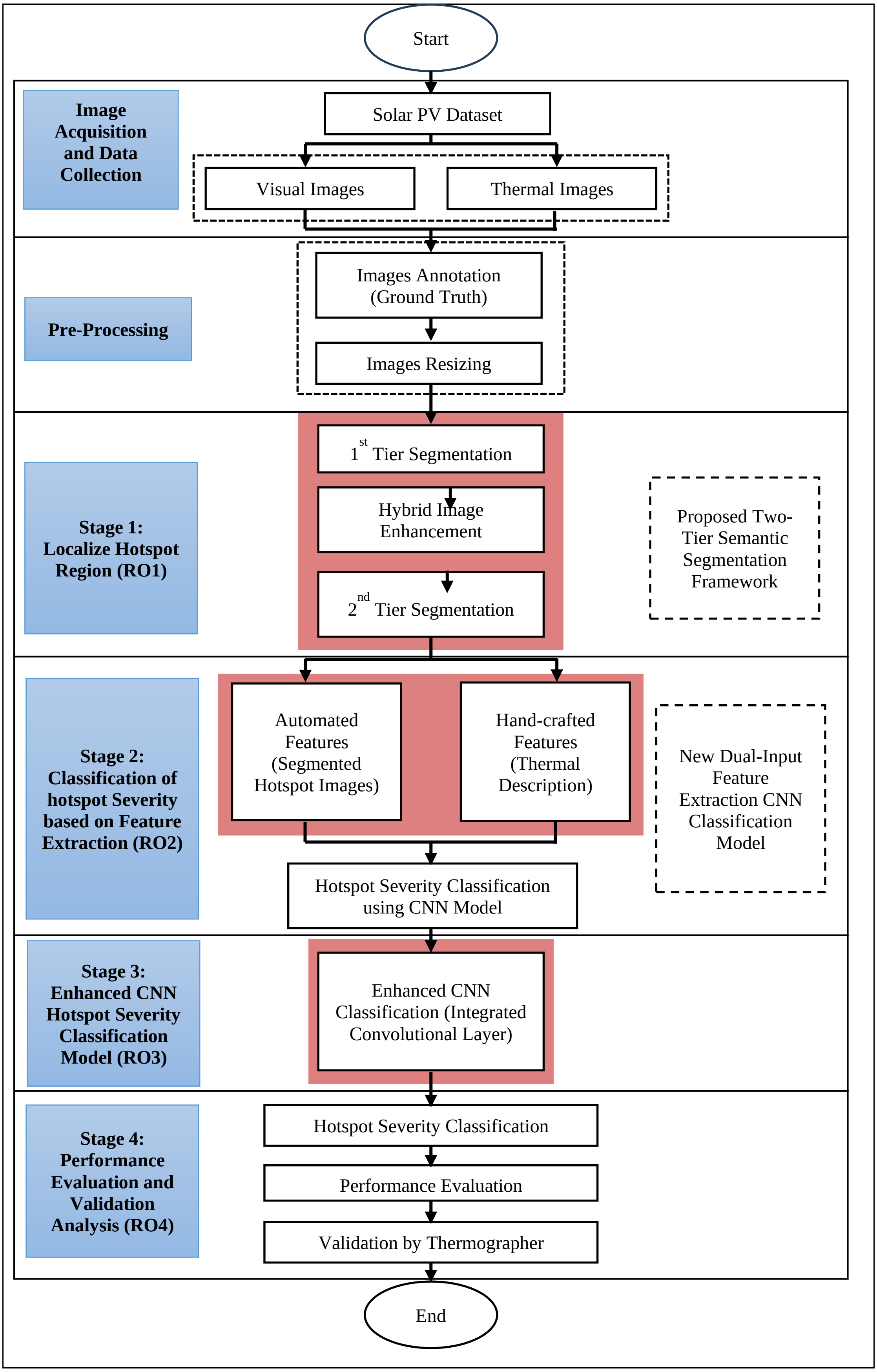


Figure 3.1 Overall Proposed Methodology for Hotspot Localization and Severity Classification Model with Highlighted Contributions for Each Research Objective

This study began with image acquisition and data collection. All the collected datasets contain two types of single-capture images, namely, visual and thermal images. All images are pre-processed using an image annotation task and resized to a standard size, which is then used as input to stage 1. Details of these processes are further explained in Sections 3.3 and 3.4.

The first stage aims to localize all the hotspot regions in the solar PV thermal images dataset as proposed in RO1. The pre-processed images are automatically segmented using a new proposed segmentation technique, namely two-tier semantic segmentation, to localize the hotspot regions in thermal images. The proposed techniques aim to segment hotspot regions effectively, enabling precise delineation. In this method, the first-tier segmentation is used to segment the solar PV and further enhances the image to localize the hotspot region during the second-tier segmentation. The segmented thermal images are further used to verify potential hotspot regions and to extract temperature distribution features associated with each hotspot for the next stage. The overall process of Stage 1 is further explained in Section 3.5

Furthermore, the second stage focuses on extracting features from the segmented hotspot regions to classify severity using a CNN-based classification model. This stage proposes a dual-input feature extraction and fusion strategy for hotspot severity classification. The framework combines CNN-derived image features extracted from segmented hotspot regions with hand-crafted (HC) thermal descriptors. Feature fusion is performed at the feature representation level after independent extraction, in which the two feature vectors are concatenated to form a unified representation. The fused feature vector is then fed into the fully connected classification layers to classify hotspot severity. This integration exploits complementary spatial and thermal information to improve classification performance. This technique aims to improve the accuracy of automated hotspot analysis by leveraging CNNs' ability to extract complex and hierarchical visual representations. This phase directly supports RO2 by strengthening the characterization of features to achieve accurate classification performance. The detailed formulation of this process is further explained in Section 3.6.

The next phase is stage 3, which involves the enhancement of the CNN model architecture to further improve the classification of severity levels based on related temperature features. At this stage, hotspot severity is classified using an enhanced CNN architecture that integrates the model's architecture and adds convolutional layers.

As in the stage 2 classification process, the segmented thermal images are split into two datasets for training and testing. The classification model categorizes hotspots into three severity classes, namely low, medium, and high. The model is trained on the training dataset and evaluated on the test set to assess its classification accuracy and robustness. The details of CNN optimization and parameter tuning are elaborated in Section 3.7.

Lastly, the final stage focuses on the proposed model performance evaluation and validation, which corresponds to RO4. The model is evaluated using standard metrics and analysis, with validation by certified solar thermographers. Overall, the results were compared statistically with model predictions to validate the proposed system. In addition, an agreement analysis was conducted to assess the agreement between the thermographer's evaluation and the automated classification outcomes.

3.3 Image Acquisition and Data Collection

This section describes the image acquisition and data collection process conducted before the development of the automated hotspot detection framework, as outlined in Figure 3.1. The data were collected at a solar farm located in Pasir Mas, Kelantan, Malaysia. The solar farm occupies approximately 6 hectares and has an installed capacity of 4 MWp. Aerial imaging was employed to capture thermal images of solar PV modules across the site.

The dataset was collected over multiple UAV flight sessions conducted between September 2022 and August 2023. The flights were conducted on different days under clear-sky conditions between 11:00 a.m. and 1:00 p.m. to ensure consistent irradiance levels and adequate thermal contrast for hotspot detection. GPS coordinates embedded in the Exchangeable Image File Format (EXIF) metadata indicate an approximate location of 5.9924°N, 102.1099°E. During data acquisition, the average surface temperature was approximately 307.58 K (34.43°C), with recorded irradiance values ranging from 624.72 to 805.70 W/m^2. The thermal aerial images were captured using a Parrot Anafi thermal camera operating in the long-wave infrared spectral band (8–14 μm). The camera employs an uncooled Forward-Looking Infrared (FLIR) Lepton 3.5 microbolometer detector with a thermal sensitivity of less than 0.05°C. In addition, the platform provides a high-resolution visual imaging sensor capable of capturing RGB images at 3264 × 2448 pixels. Both thermal and visual images are acquired simultaneously and spatially aligned through the dual-imaging system. These specifications ensure accurate temperature measurement and reliable thermal representation of PV module surfaces.

For image acquisition, the UAV platform was operated at a low altitude of approximately 15 m to ensure sufficient spatial resolution while providing full coverage of the PV modules, as illustrated in Figure 3.2. The PV modules were installed at an inclination of approximately 30°, which is typical for outdoor installations. This mounting configuration influences heat dissipation behaviour and the observed thermal distribution patterns. Therefore, the panel orientation and installation geometry were considered during image acquisition and subsequent analysis.

To ensure measurement accuracy and repeatability, key environmental parameters, including ambient temperature, target emissivity, and irradiance, were recorded during data collection. The captured thermal images were stored in sRGB format, which is suitable for subsequent digital image processing and ML applications. Although the thermal images were stored in sRGB format for image processing and CNN analysis, the associated temperature information embedded within the thermal image files was extracted in MATLAB to obtain T_{max}, T_{min}, T_{ref}, and ΔT. Therefore, the HC thermal descriptors were derived from temperature information available in the thermal image files rather than solely from colour-mapped image intensities.

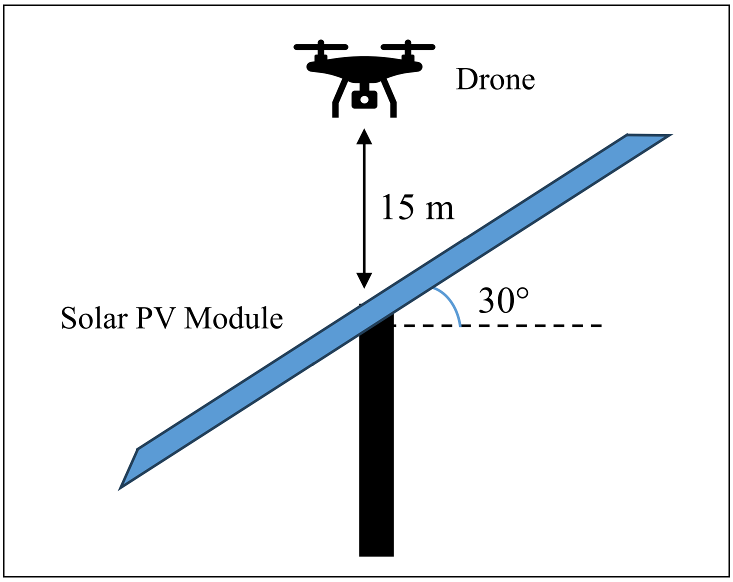


Figure 3.2 Height for the Image Capture Procedure

Figure 3.3 illustrates the overall workflow of the dataset acquisition process employed in this study. The process begins with image capture at the solar PV module site using a UAV-based dual-imaging system that simultaneously acquires thermal and visual imagery. This dual-modality capture ensures that thermal anomalies can be accurately correlated with the corresponding visual structures of the PV modules.

Following image acquisition, the collected thermal and visual images are organized and prepared for subsequent processing. Ground-truth annotations were conducted by certified solar thermographers through expert thermal interpretation of hotspot regions and corresponding severity levels. This expert-driven annotation process establishes reliable reference labels for subsequent model training and evaluation. The finalized, annotated dataset forms the solar PV dataset used throughout this study for hotspot localization, severity classification, and performance validation.

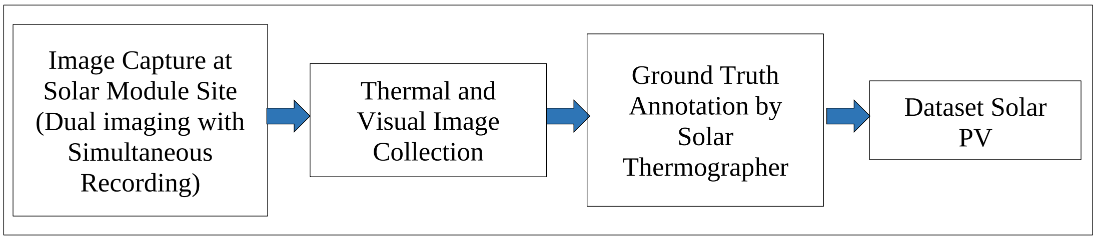


Figure 3.3 The General Flow of Dataset Acquisition

Table 3.1 presents the key specifications of the Parrot Anafi thermal camera, which is equipped with a FLIR Lepton 3.5 radiometric sensor that captures high-precision thermal data. This specification ensures adequate spatial and thermal resolution for detecting minor temperature variations across PV module surfaces.

Table 3.1
Parrot Anafi Thermal Camera Specification

Parameter	**Specification**
Test Detector Type	FLIR Lepton 3.5 microbolometer (radiometric)
Photo resolution	3264 x 2448
Thermal sensitivity	Less than 0.1 °C at 30°C
Temperature measurement range	-20 °C to +350 °C
Field of view	57°
Infrared spectral band	8 µm to 14 µm

Table 3.2 presents the dataset composition and summarizes the distribution of visual and thermal images used for training, testing, and validation. A total of 670 thermal images were acquired for this study. Although the dataset comprises 670 thermal images, the use of transfer learning with pre-trained CNN architectures mitigates the limitations of a relatively small dataset. The pre-trained networks provide robust feature representations learned from large-scale image datasets, enabling effective model training and improved generalization performance. The dataset was partitioned into 80% (536 images) for training, 10% (67 images) for testing, and 10% (67 images) for validation. Only thermal images were used as model inputs during training. The visual images were used for verification purposes, particularly to validate annotation quality and classification consistency.

Table 3.2
Dataset Distribution of Images for Model Execution

Process	**Thermal Images**
Training	536
Testing	67
Validation	67
Total	670

This structured data allocation ensures that the model receives sufficient training samples for learning and generalization while maintaining dedicated sets for evaluation and fine-tuning. The testing dataset was used to assess model performance on unseen data, and the validation dataset was used during training to prevent overfitting. This distribution strategy promotes and supports evaluation under the field conditions observed.

3.4 Image Pre-processing

Pre-processing techniques are essential for improving the quality of thermal imagery before segmentation processes. These techniques improve the visibility of hotspots, reduce noise, and enhance overall image contrast. In this study, there are two procedures conducted that are image annotation and image resizing. Image annotation is performed to create a ground-truth dataset for supervised learning. While image resizing is used to ensure that the thermal images are properly standardized before performing the segmentation process. Figure 3.4 presents the overall image pre-processing process, which includes image annotation and resizing.

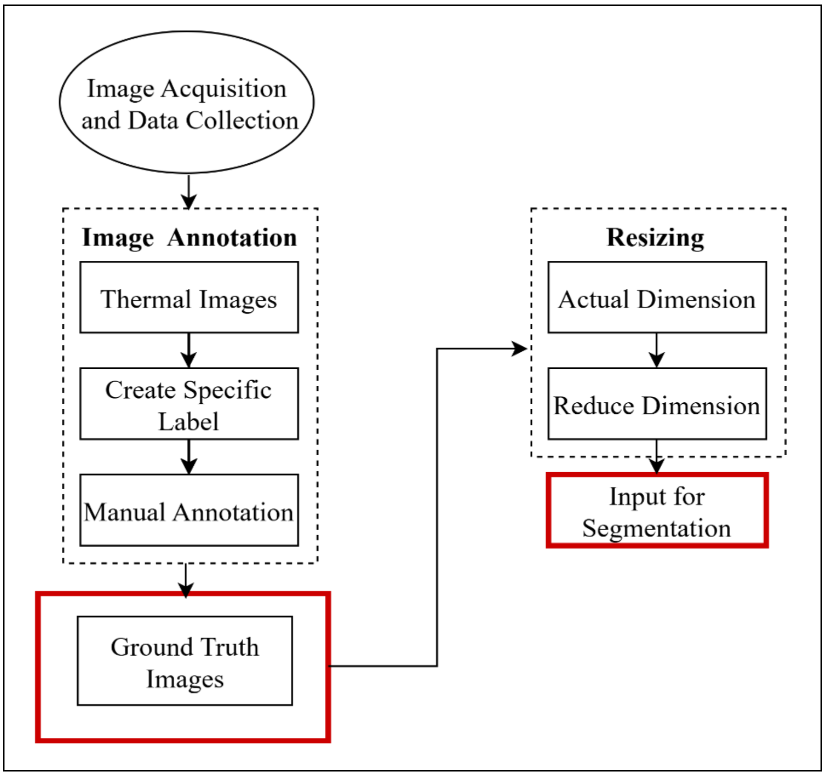


Figure 3.4 Image Pre-processing Techniques Performed using Image Annotation and Resizing Techniques

The pre-processing workflow begins with manual annotation of thermal images to generate pixel-level ground-truth masks for supervised segmentation. The annotated images are then resized to standardized dimensions to ensure consistent input size during model training. This sequential process ensures both accurate labelling and computational uniformity prior to segmentation. The next subsections detail the processes for image annotation and resizing, respectively.

3.4.1 Image Annotation

In this process, an experienced solar thermographer evaluates all collected images and verifies the presence of hotspots in accordance with established PV thermographic inspection practices. The thermographer examines the thermal characteristics and anomaly patterns within each image to distinguish actual hotspot regions from non-defective areas. This verification procedure serves as the dataset's annotation process and provides reliable ground-truth labels for subsequent model training and evaluation. The use of expert assessment enhances annotation consistency and improves the reliability of the developed dataset.

In this step, the MATLAB application Image Labeler tool is used by the thermographer for labelling and segmentation of the thermal images’ dataset. This labelled image facilitates the creation of ground truth datasets and ROI detection. In the annotation process, the thermal images are uploaded into the Image Labeler, and specific labels are assigned to the PV modules and hotspot regions. Manual annotation is then performed to define ROI objects, such as rectangular or polygonal ROIs, to mark the hotspot areas. At the same time, visual images are used to guide the marking of hotspot locations since sun reflections can appear like hotspots. To distinguish reflections from true hotspots, the thermographer cross-referenced the thermal images with the corresponding visual images during the annotation process.

Thermal anomalies were labelled as hotspots only when they corresponded to actual PV module locations and exhibited thermal patterns consistent with PV defects. Regions caused by sunlight reflection or reflective surfaces that did not correspond to physical hotspot characteristics on the PV modules were excluded from the hotspot annotations. This step produces both the image and the associated coordinate information. The manually annotated polygonal ROIs were subsequently converted into pixel-level binary masks using the MATLAB Image Labeler export function. In the generated masks, hotspot pixels were assigned to the target class, while all remaining pixels were assigned to the background class. These binary masks served as ground truth labels for training and evaluating the semantic segmentation models. By clearly defining hotspot zones, annotation enables precise identification of defective areas in the solar PV modules.

Once completed, the application exports all tagged data to create a dataset that includes the images and the corresponding coordinates. The annotated images serve as training data for supervised machine learning (ML) models to recognize patterns and classify the PV modules and hotspots correctly. The system examines these particulars within the framework of performance evaluations and operational assessments.

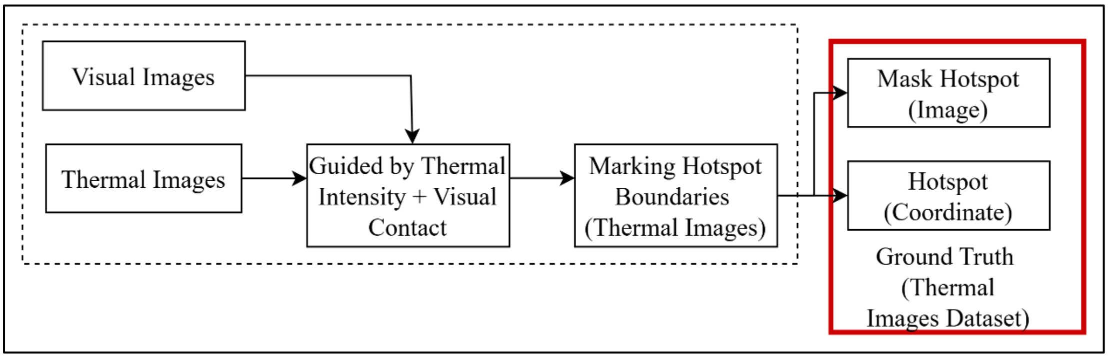


Figure 3.5 Image Annotation Process using Image Labeller, Manually Performed by a Thermographer

Figure 3.5 illustrates the label procedure performed by a thermographer to develop ground truth for segmenting the hotspot in a solar PV module. Once the labelling process is complete, the annotated data are exported to form a dataset comprising images, labels, and coordinates. Next, the dataset is further performing an image resizing step as discussed in Section 3.4.2.

3.4.2 Image Resizing

After annotation, the thermal images are resized to meet the input specifications of segmentation models. Standardizing image dimensions ensures uniformity, reduces computational complexity, and supports reliable model training. This process minimizes variability and provides structured input for segmentation. At the same time, improving the accuracy and robustness of hotspot localization and classification.

In the collected dataset, the thermal images used for image processing were stored at a resolution of 3264 × 2448 pixels. This resolution corresponds to the exported thermal image format generated by the Parrot ANAFI Thermal system for visualization and analysis. The thermal images are resized to ensure consistent dimensions after labelling. The main objective of this phase is to reduce image size, thereby minimizing computational cost and memory usage during training. In addition, this step ensures that the proposed CNN model receives suitable patch divisions. For this study, all 670 thermal images were resized from an original resolution of 3264×2448 to 224×224 pixels to facilitate efficient segmentation.

Consistent with previous studies on CNN-based thermal image analysis, the input images were resized to 224×224 pixels [67], [138]. This resolution was chosen because it balances the preservation of PV modules and hotspot features with computational efficiency. Besides, it is widely adopted in DL models for PV fault and thermal anomaly detection. Figure 3.6 illustrates the resizing process, showing how high-resolution images are transformed into standardized dimensions suitable for segmentation tasks. Figure 3.6(a) depicts the original high-resolution image, while Figure 3.6(b) shows the standardized resized image used as input to the CNN model.

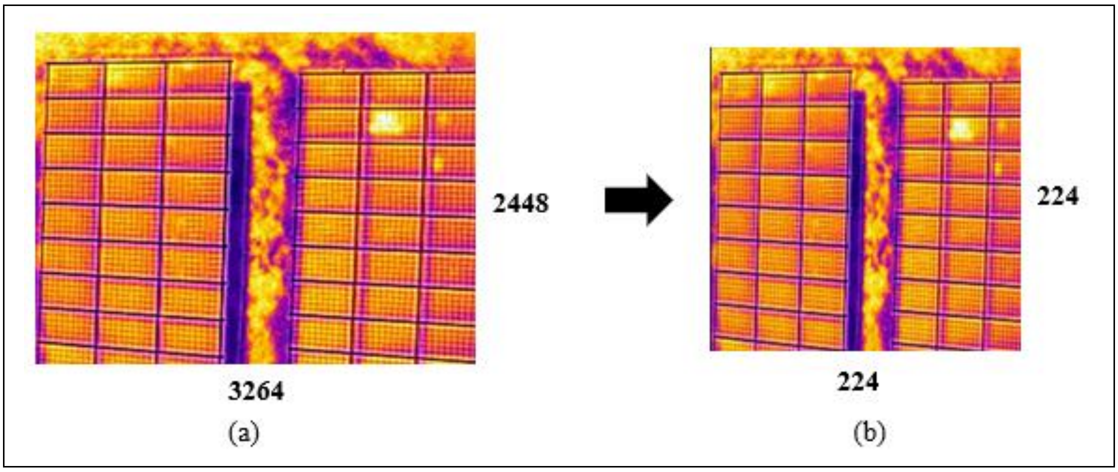


Figure 3.6 Example of Image Resizing of Thermal Image (a) Original Resolution Before Resizing (3264x2448) (b) Standardize Dimension After Resizing (224x224)

Resizing from 3264 × 2448 to 224 × 224 introduces some aspect-ratio distortion. However, the distortion was applied consistently to all images, and the key thermal characteristics required for hotspot localization and severity classification were preserved. Visual inspection confirmed that the PV module structures and hotspot regions remained clearly distinguishable after resizing. Furthermore, the resized dimensions were selected to satisfy the input requirements of the CNN architectures while reducing computational complexity during model training.

In summary, resizing ensures uniform image dimensions across the dataset, simplifying training and improving segmentation efficiency. By balancing computational requirements with image quality, this pre-processing step provides a reliable foundation for the framework’s accurate hotspot detection.

3.5 Stage 1: Localization of Hotspot Region using A New Two-tier Semantic Segmentation Framework

This section outlines the proposed methodology at Stage 1 using a two-tier semantic segmentation framework that is aligned with RO1. The primary aim of this stage is to enhance the localization and detection of hotspot region through a two-tier segmentation approach. As shown in Figure 3.7, the proposed two-tier semantic segmentation framework is designed into three subsequent processes. The first-tier segmentation is aimed at isolating the PV module area from the background. Next, after the PV modules are identified, an image enhancement process is applied to improve visual contrast and structural detail before the next step. The details of the proposed two-tier semantic segmentation framework are further discussed in the next subsections.

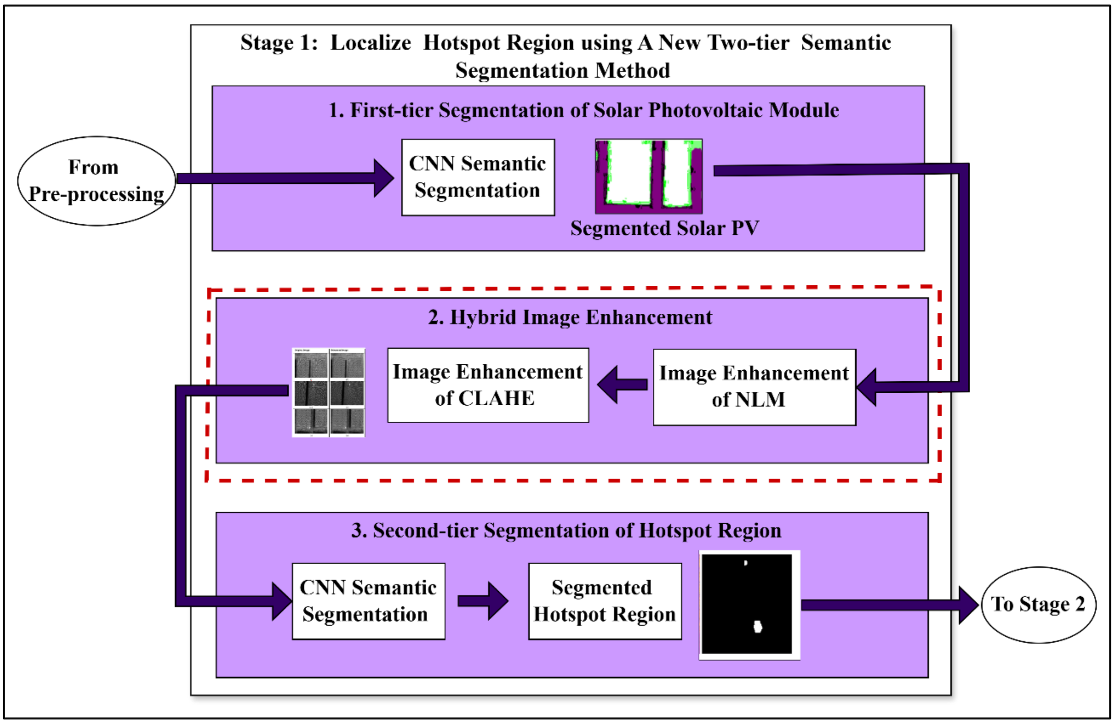


Figure 3.7 Overall Method of Proposed Two-tier Semantic Segmentation at Stage 1(RO1)

The first stage focuses on the semantic segmentation for segregating solar PV module from the surrounding background. This step isolates the PV module area by removing irrelevant areas such as vegetation, ground surface and structural elements. This creates a focus area of interest for further processing and reduces background interference.

After PV modules are segmented, a hybrid image enhancement process is applied to the segmented images, which represents a key contribution of the proposed framework. This stage integrates non-local means (NLM) filtering and contrast-limited adaptive histogram equalization (CLAHE) to enhance thermal contrast while suppressing noise. The enhancement process enhances the visibility of minor hotspot signatures and structural boundaries, improving feature quality for next second-tier process.

The final process uses enhanced PV module images as input to perform semantic segmentation of hotspot areas. The improved image quality enables more accurate recognition of local thermal anomalies under various imaging conditions. The following subsections discussed further details of all three main processes at Stage 1.

3.5.1 First-tier Segmentation

The initial phase of the proposed two-tier semantic segmentation procedure involves first-tier semantic segmentation. This phase aims to identify and separate PV modules from the surrounding background in thermal images. At this step, the pre-processed thermal image dataset is used to train and evaluate semantic segmentation in a supervised learning framework. Three semantic segmentation architecture were evaluated, namely U-Net, DeepLabV3+ with a ResNet-18 and ResNet-50 backbone. These models are widely used in image analysis due to their strong feature learning capabilities and high pixel-level classification accuracy [77], [83], [139]. This process supports consistent and accurate PV module segmentation across a wide range of background scenes and environmental conditions.

The training hyperparameters adopted for the CNN-based two-tier semantic segmentation framework are summarized in Table 3.3. A maximum of 80 epochs was selected to allow sufficient convergence while minimizing overfitting. A mini-batch size of 10 was employed based on computational resource constraints and training stability considerations. An initial learning rate of 0.001 was used in conjunction with the Adaptive Moment Estimation (Adam) optimizer to ensure stable and efficient convergence during training. The validation frequency was set to 50 iterations to enable consistent monitoring of model performance. These hyperparameter settings were standardized across both the first-tier and second-tier segmentation stages to ensure consistency and fair performance comparison among the evaluated architectures. All experiments were conducted in MATLAB on a high-performance personal computer running Windows 11, equipped with an Intel Core i5 2.5 GHz processor, 16 GB of RAM, and an NVIDIA GeForce RTX 3060 GPU with 8 GB of memory.

Table 3.3
Training Hyperparameters for CNN-Based Two-tier Semantic Segmentation

Hyperparameter setting	**Set Value**
Maximum epoch	80
Mini-batch size	10
Learning rate	0.001
Validation frequency	50
Solver	Adaptive Moment Estimation (Adam)

Figure 3.8 presents the first-tier semantic segmentation block diagram for analyzing and detecting solar PV modules.

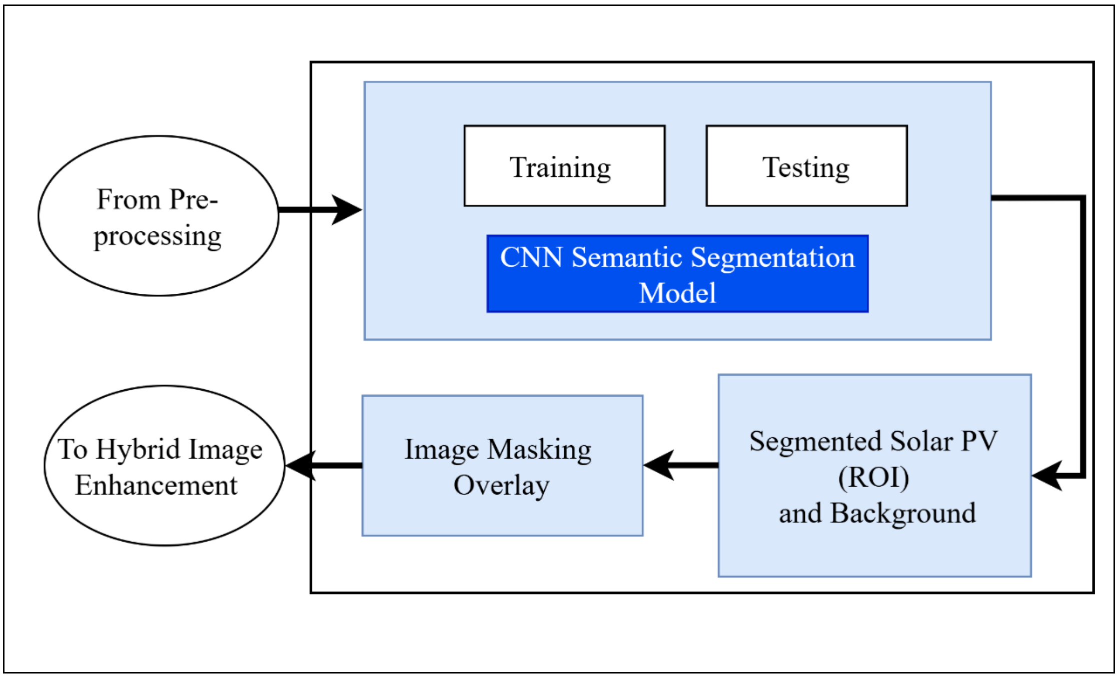


Figure 3.8 First-tier Semantic Segmentation Process to Segregate Solar PV and Background in Thermal Image

The coarse-to-fine segmentation process is facilitated by this hierarchical structure, which ensures that only the most essential module areas are sent to the next phase. The PV module areas are effectively segmented, thereby removing irrelevant background components such as mounting structures, shadows and background noise. This separation improves computational efficiency and reduces false detections during hotspot identification. Additionally, the segmentation results obtained at this hotspot stage serve as the foundation for the fine localization of hotspots in second-tier segmentation. As a result, the proposed two-tier framework remains precise and consistent.

Once the solar PV modules are segmented, the next step is to overlay the segmented solar PV image with its corresponding thermal image. This step provides a comprehensive view of the PV modules' condition for subsequent image enhancement, as explained in Section 3.5.2.

3.5.2 Hybrid Image Enhancement using NLM and CLAHE

After initial segmentation, the PV module regions undergo image enhancement. This phase is essential for the fine-tuning of image quality by enhancing contrast, reducing noise, and emphasizing module boundaries that may not be apparent after segmentation. By enhancing segmented images before hotspot identification, the method reduces the possibility of misclassifying hotspot locations and enhances feature visibility, such as thermal intensities. The step aims to improve image clarity, particularly for blurred or low-light images affected by both low contrast and noise. The concept is to balance noise suppression and contrast enhancement by integrating these two techniques. A hybrid enhancement strategy that combines Non-Local Means (NLM) denoising with Contrast Limited Adaptive Histogram Equalization (CLAHE) is proposed to leverage the strengths of both techniques to optimize overall image quality.

This step aims to improve image clarity, particularly for blurred or low-light images affected by both contrast and noise. To achieve the aim, NLM denoising is employed since it is highly effective in preserving edges and suppressing noise. However, excessive filtering can reduce contrast or blur image details. Therefore, to avoid such conditions, CLAHE is used since it effectively increases local contrast. Figure 3.9 illustrates the proposed hybrid image enhancement method, which integrates NLM with CLAHE or (NLM-CLAHE).

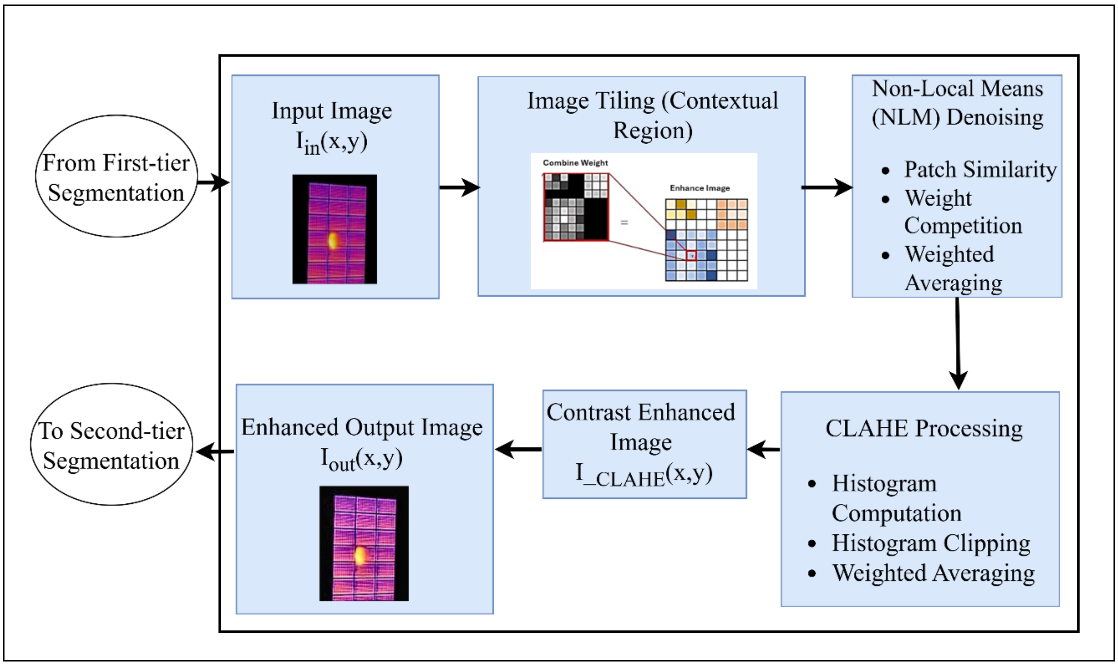


Figure 3.9 Overall Process of the Proposed Hybrid NLM-CLAHE Enhancement

The segmented image from first-tier segmentation serves as input to the proposed enhancement pipeline. The proposed hybrid enhancement method begins with the original thermal image $I_{in}(x, y)$, where *x* denotes the horizontal (column-wise) pixel coordinate and *y* denotes the vertical (row-wise) pixel coordinate. This two-dimension representation allows for pixel-level processing through two sequential stages. Denoising procedure using NLM and contrast enhancement using CLAHE. This combination is to produce an enhanced output $I_{out}(x, y)$, optimized for the subsequent segmentation task. The following are the procedures conducted to perform hybrid NLM-CLAHE enhancement:

1. The grayscale conversion process is expressed in Equation (3.1), which transforms the first-tier output image $I_{in}(x, y)$ into an intensity-based representation.

$$I_{gray}(x, y) = Grayscale(I_{in}(x, y)) \tag{3.1}$$

Subsequently, the resizing and double-precision normalization steps are defined in Equation (3.2) and (3.3) respectively.

$$I_{resized}(x,y) = Resize(I_{gray}(x,y), 50\%) \tag{3.2}$$

$$I_{in}(x,y) = ConvertToDouble(I_{resized}) \tag{3.3}$$

2. The pre-processed image I_{in} is filtered using the NLM denoising algorithm, which preserves fine structural details while reducing pixel-level noise. The NLM denoising operation is formulated in Equation (3.4).

$$I_{NLM}(x,y) = \sum_{(i,j)\in\Omega} w(x,y,i,j).I_{in}(i,j) \tag{3.4}$$

The similarity-based weighting function governing the NLM filtering process is defined in Equation (3.5), where patch-wise intensity differences are used to quantify structural similarity between pixels.

$$w(x,y,i,j) = \frac{1}{Z(x,y)}.\exp\left(-\frac{\|P(x,y) - P(i,j)\|_2^2}{h^2}\right) \tag{3.5}$$

where:
$P(x,y), P(i,j)$: patch neighborhoods around pixels.
h: filtering parameter.
$Z(x,y)$: normalization factor.

3. The denoised image is then enhanced using CLAHE to Improve local contrast and highlight minor thermal features. The CLAHE operation applied to the denoised image is mathematically expressed in Equation (3.6).

$$I_{CLAHE}(x,y) = CLAHE(I_{NLM}, clip\ limit, tile\ grid\ size) \tag{3.6}$$

In this study, the NLM filtering parameter h was set to 1.5, while the CLAHE enhancement stage employed a clip limit of 0.01 and a tile grid size of [3 × 3]. These parameter settings were adopted based on the default MATLAB implementation and were retained throughout the study to ensure consistent image enhancement across the dataset. The selected settings provided a balanced enhancement of local thermal contrast while preserving structural details and limiting excessive noise amplification. Finally, the output for second-tier input is expressed in Equation (3.7).

$$I_{out}(x, y) = I_{CLAHE}(x, y) \tag{3.7}$$

The hybrid NLM–CLAHE approach aimed to improve image quality by first suppressing noise with NLM denoising, then enhancing local contrast with CLAHE. The NLM stage reduces pixel-level noise by averaging structurally similar patterns across the image while preserving fine textures and edge details. Subsequently, CLAHE enhances local contrast by partitioning the denoised image into contextual regions and adjusting intensity distributions within each tile, thereby preventing excessive amplification of residual noise. This sequential integration produces a refined output with improved structural clarity, reduced artefacts, and enhanced suitability for subsequent thermal hotspot analysis.

The quality of the enhanced images is measured to determine the robustness of NLM-CLAHE enhancement method. The image quality metrics based on MSE, PSNR and SNR were calculated before enhancement (I_{in}) and after enhancement (I_{out}) and are expressed in Equations (3.8), (3.9) and (3.10), respectively.

Mean Squared Error (MSE):

$$MSE = \frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}[I_{in}(i, j) - I_{out}(i, j)]^2 \tag{3.8}$$

Peak Signal-to Noise Ratio (PSNR)

$$PSNR = 10.\,log_{10}(\frac{MAX^2}{MSE}) \tag{3.9}$$

Signal-to-Noise Ratio (SNR)

$$SNR = 10.\,log_{10}\left(\frac{mean(I_{in}^2)}{mean((I_{in} - I_{out})^2)}\right) \tag{3.10}$$

where:

- $I_{in}(i, j)$ denotes the pixel intensity at position (i, j) in the original (reference) image.

- $I_{out}(i,j)$ represents the pixel intensity at position (i,j) in the processed image.
- m and n correspond to the image height and width, respectively.
- MAX denotes the maximum possible pixel intensity value of the image.

The quantitative metrics such as MSE, PSNR, and SNR are calculated for each image to evaluate enhancement performance. The average improvement across the dataset represents the effectiveness of the proposed hybrid approach in enhancing image contrast and reducing noise prior to subsequent segmentation.

3.5.3 Second-tier Segmentation

The second-tier segmentation stage focuses on detecting hotspot regions within the solar PV modules obtained from the first-tier segmentation process. This stage is applied after image enhancement using the hybrid NLM–CLAHE method to improve image contrast and suppress noise. Figure 3.10 illustrates a second-tier segmentation process for analyzing solar module images to detect hotspots ROI in the enhanced images output from the previous process.

In this stage, the enhanced images are used to generate ROI–based pixel labels that serve as ground-truth masks for training the semantic segmentation model. The ROI-labelled images are then provided as inputs to a CNN-based semantic segmentation model to distinguish hotspot regions from non-hotspot areas within the PV modules. Model performance is subsequently evaluated based on segmentation accuracy and related metrics to assess the effectiveness of the second-tier segmentation process. This stage concludes the hotspot localization workflow before proceeding to subsequent analysis stages.

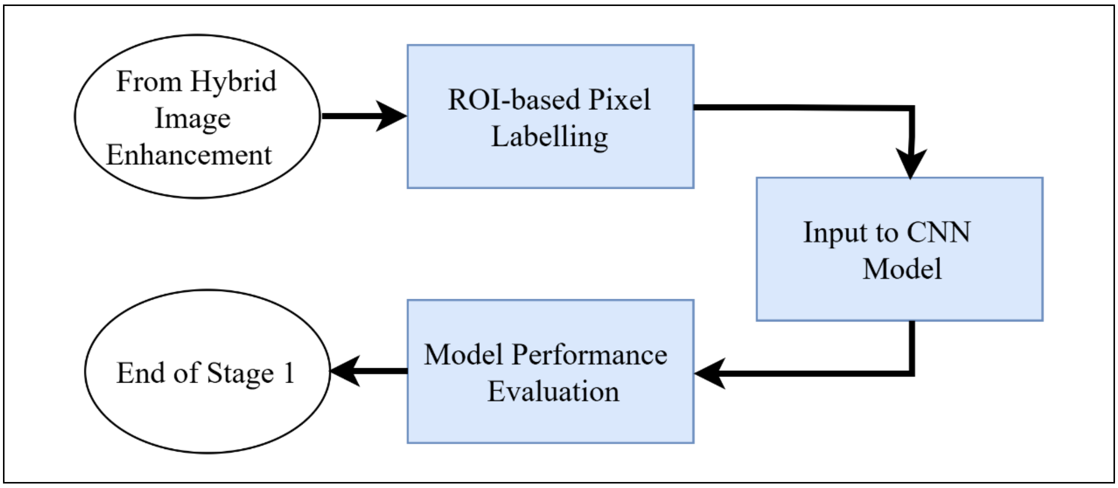


Figure 3.10 Stage 1: Localize Hotspot Region using Deep Learning at the Second-tier Semantic Segmentation

This study evaluates three semantic segmentation architectures, namely U-Net, DeepLabV3+ with a ResNet-18 and ResNet-50 backbones. These architectures were selected for their strong ability to learn hierarchical image features and accurately delineate hotspot boundaries. All models were evaluated using pixel accuracy, Intersection over Union (IoU), precision, and recall.

The training process involves feeding the model enhanced images to learn to distinguish between normal and hotspot regions in solar PV modules. The CNN model will segment the hotspots regions within the modules, providing detailed information on their location and extent. Figure 3.11 shows an overall process of hotspot detection performed at Stage 1 using the proposed two-tier semantic segmentation method.

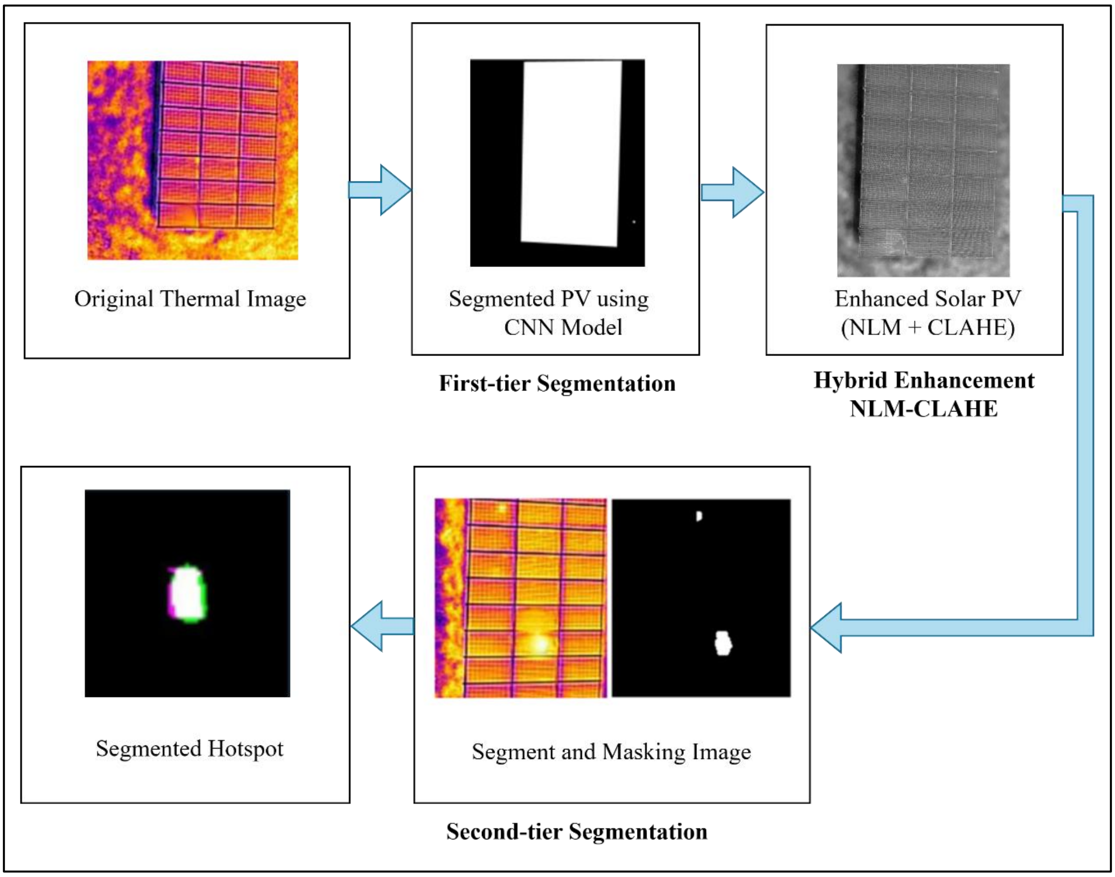


Figure 3.11 Overall Output of Two-tier Semantic Segmentation Framework for Hotspot Localization

In conclusion, Stage 1 presents a structured methodology for detecting hotspots using a new semantic segmentation method with the aim of accurately localizing hotspots within solar PV modules. The integration of image enhancement and pixel-level segmentation ensures reliable boundary preservation and stable network training. This stage establishes a robust foundation for subsequent feature extraction and severity classification processes.

3.6 Stage 2: Classification of Hotspot Severity with Dual-input Feature Fusion Technique

This section elaborates the overall methodology at Stage 2, which corresponds to RO2. The output from Stage 1, which is the segmented hotspot ROI, is used as input for this stage. At this stage, a new dual-feature extraction method is proposed for a CNN classification model to determine hotspot severity levels. The proposed method adopts a dual-path strategy in which thermal images and HC thermal descriptors are processed in parallel by feature-extraction branches. The automatically learned CNN features and encoded thermal features are subsequently fused and fed into the classification model. This design leads to a more comprehensive and resilient feature representation within the training pipeline. The overall works of Stage 2, including the detailed method flow for the proposed dual-input feature extraction method is shown in Figure 3.12.

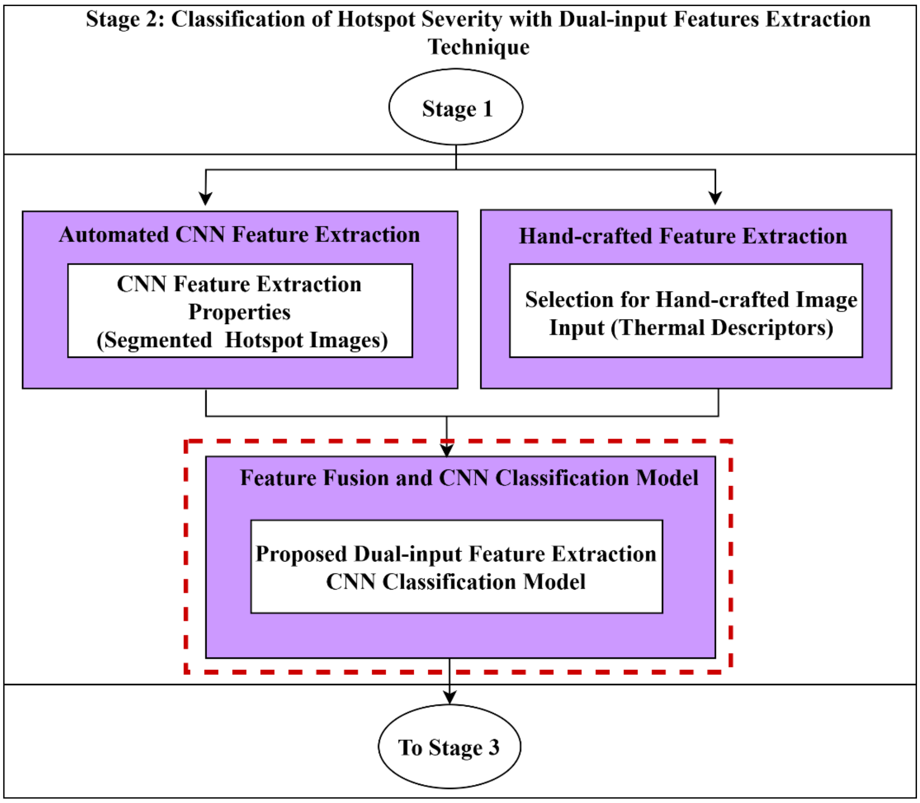


Figure 3.12 Overall Method Flow of Stage 2 for Classifying Hotspot Severity using Feature Extraction of CNN Model

The detailed extraction techniques for both CNN and HC streams are further described in Subsections 3.6.1 and 3.6.2, respectively. The extracted CNN features and thermal descriptors are subsequently combined through a feature fusion stage prior to classification. The detailed fusion mechanism and classification architecture are presented in Subsection 3.6.3. This integration lays a strong foundation for the classification process.

3.6.1 Automated CNN Feature Extraction (Image-based Feature)

This subsection discussed the process of extracting optimal CNN features from thermal images. The process starts with selecting the best CNN architecture using the segmented hotspot region from Stage 1. The CNN is then trained on thermal image datasets using the corresponding CNN parameters.

In this stage, three types of CNN architecture were selected for automated feature extraction due to their robustness in learning deep hierarchical representations, especially when working with complex thermal imagery of PV systems. The CNN models are U-Net, ResNet-18 and ResNet-50. Figure 3.13 illustrates the architectural structures of the three CNN models utilized in this study. U-Net, a widely used encoder-decoder architecture, is implemented for the task of semantic segmentation to detect and localize hotspot regions in PV thermal imagery at the pixel level. This architecture enables precise delineation of hotspot boundaries, essential for accurate downstream classification.

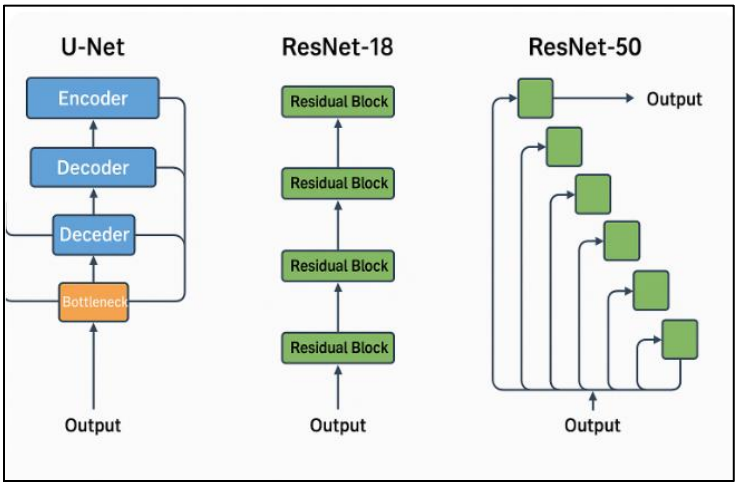


Figure 3.13 Architectural U-Net, ResNet-18 and ResNet-50 for Comparison of Deep Learning Models Used

In contrast, ResNet-18 and ResNet-50 serve as backbone models for feature extraction from segmented thermal regions. ResNet-18, a shallower residual network, is lightweight and computationally efficient, making it suitable for rapid inference in real-time applications. ResNet-50, a deeper and more expressive model, offers greater feature discrimination and is used for detailed severity classification of identified hotspots. The inclusion of residual connections in both ResNet architectures enhances gradient flow and model stability, especially when dealing with complex thermal image inputs. Furthermore, both ResNet-18 and ResNet-50 benefit from pretrained weights learned on datasets, which enhance feature representation capability when adapted to thermal imagery [84]. This transfer learning strategy provides better initialization and accelerates convergence, offering improved feature discrimination compared to U-Net models trained from scratch.

Although the pretrained weights were originally learned from RGB natural images in the ImageNet dataset, the lower convolutional layers capture generic visual features such as edges, textures, and spatial patterns that remain useful for thermal image analysis. During training, the network weights were fine-tuned using the PV thermal image dataset, allowing the pretrained feature representations to adapt to the thermal characteristics and hotspot patterns specific to PV inspection.

The selection of U-Net, ResNet-18, and ResNet-50 was driven by their strengths when evaluated in a typical end-to-end CNN-based framework. U-Net is used to achieve accurate pixel-level localization of the hotspot region. This allows for accurate spatial delineation prior to feature extraction. ResNet-18 was chosen for its computationally efficient architecture, suitable for fast inference. Conversely, ResNet-50 is employed to extract more comprehensive and discriminative features necessary for an accurate classification of hotspot severity.

Examining multiple architectures enables a thorough comparison of localization accuracy, feature effectiveness, and computational efficiency. The resulting analysis provides insights into the suitability of each model within the proposed framework. All the architectures adopted in this study support both spatial localization and hotspot severity interpretation within an end-to-end CNN-based framework. A systematic comparison of their performance characteristics is subsequently performed.

For consistency in training and performance comparison, all CNN architectures are trained using the same hyperparameter settings. The hyperparameter configurations used to extract CNN-based features for each of the evaluated architectures are summarized in Table 3.4.

Table 3.4
Hyperparameter Settings Set for CNN Feature Extraction on Thermal Datasets

Hyperparameter setting	**Set Value**
Maximum epoch	50
Mini-batch size	20
Learning rate	0.001
Validation frequency	30
Solver	Adaptive Moment Estimation (Adam)

A fixed number of training epochs and a consistent mini-batch size are used to ensure stable convergence and fair performance evaluation. The learning rate is selected to balance training stability and convergence speed. At the same time, the validation frequency allows periodic performance monitoring during training. The dataset was partitioned into 80% training, 10% validation, and 10% testing subsets. This data allocation provides sufficient samples for model learning while preserving independent datasets for hyperparameter monitoring and final performance evaluation. Adaptive Moment Approximation (Adam) optimization is effective in tackling complex optimization problems and accelerating convergence. This unified hyperparameter setting ensures consistency and comparability across all experimental evaluations.

3.6.2 Hand-crafted Feature Extraction (Thermal-based Feature)

This subsection delineates the procedure for extracting optimal HC features. HC feature extraction focuses on deriving explicit and interpretable descriptors from the localized hotspot region to facilitate hotspot severity analysis. The HC features are analytically computed from the segmented hotspot region based on temperature based. Temperature features are derived from pixel-level temperature extrema within the hotspot region. The process begins by isolating the hotspot region obtained from Stage 1, followed by the computation of temperature, statistical, and texture descriptors that characterize severity behaviour within the affected PV cells. The subsequent subsections present the mathematical formulations and selection criteria for each category of HC features.

Thermal specifications and deviation metrics are essential for attaining severity classification. HC feature extraction is crucial for characterizing hotspot shape and temperature variation under different PV operating conditions. The following paragraph discusses the development of thermal-related HC characteristics that are the primary indicators of hotspot severity. Table 3.5 summarizes the thermal features used in this study and provides their physical interpretation in the context of hotspot severity analysis.

Table 3.5
Thermal Features and Their Interpretation

Thermal Features	**Interpretation**
T_{max}	Maximum pixel temperature within the hotspot region, representing the hotspot's peak intensity.
T_{min}	Minimum temperature within the hotspot region, used as an internal contrast reference.
T_{ref}	Reference temperature derived from the median or average of surrounding healthy PV regions.
ΔT	The difference between T_{max} and $T_{ref.}$

Temperature-based HC features are computed from pixel-level temperature value within the segmented hotspot region obtained from Stage 1. The minimum and maximum temperature values are extracted to characterize the thermal extrema of the hotspot region, as defined in Equations (3.11) and (3.12), respectively.

$$T_{min} = min(T_i), \qquad i \in \Omega_h \tag{3.11}$$

$$T_{max} = max(T_i), \qquad i \in \Omega_h \tag{3.12}$$

where:
Ω_h: the set of pixels corresponding to the hotspot region.
T_i : the temperature value of each pixel within this region.

In this study, the reference region Ω_n was defined as all non-hotspot pixels within the same segmented PV module after excluding the hotspot region Ω_h. The reference temperature was computed from the remaining healthy region of the corresponding PV module to provide a local thermal baseline for severity evaluation. This selection minimizes the influence of module-to-module temperature variations and allows ΔT to represent the localized thermal elevation caused by the hotspot. Following the identification of the minimum and maximum temperature, the reference temperature is computed as the mean temperature of the surrounding non-hotspot PV region to establish a consistent thermal baseline for hotspot severity evaluation, as defined in Equation (3.13).

$$T_{ref} = \frac{1}{|\Omega_n|} \sum_{i \in \Omega_n} T_i \tag{3.13}$$

where:
Ω_n: the set of non-hotspot pixels within the same segmented PV module after excluding the hotspot region Ω_h.
T_i : the temperature value of the *i*-th pixel within this region.
$|\Omega_n|$: indicates the total number of pixels within the reference region

Subsequently, the temperature difference ΔT is computed as the difference between the maximum and reference temperature, as expressed in Equation (3.14). The temperature difference is a primary indicator of hotspot severity, reflecting the magnitude of the abnormal thermal elevation within the affected PV cells.

$$\Delta T = T_{max} - T_{ref} \tag{3.14}$$

The minimum and maximum temperature values within the segmented hotspot region are obtained using the implemented MATLAB-based processing pipeline. Cursor-based inspection on the thermal image is subsequently performed to verify the extracted temperature extrema and ensure accurate identification of localized thermal peaks, particularly in regions exhibiting complex hotspot morphology. The validated minimum and maximum temperature values are then utilized to compute the temperature difference ΔT according to Equation (3.14), ensuring consistency and reliability in hotspot severity characterization.

Figure 3.14 illustrates the workflow of thermal-based HC feature extraction, where key temperature parameters (T_{min}, T_{max} and T_{ref}) are extracted from the corresponding thermal image and subsequently used to compute the temperature difference ΔT for hotspot severity characterization. These values quantify the temperature structure within the hotspot region and serve as fundamental indicators of anomaly intensity, as frequently reported in fault analysis studies [127]. The extraction of T_{min}, T_{max}, T_{ref}, and ΔT was performed automatically using the segmented hotspot and non-hotspot regions. Specifically, Tref was computed as the mean temperature of all non-hotspot pixels within the same PV module after excluding the hotspot region. Manual cursor inspection was performed solely for verification and was not used in calculating the thermal descriptors.

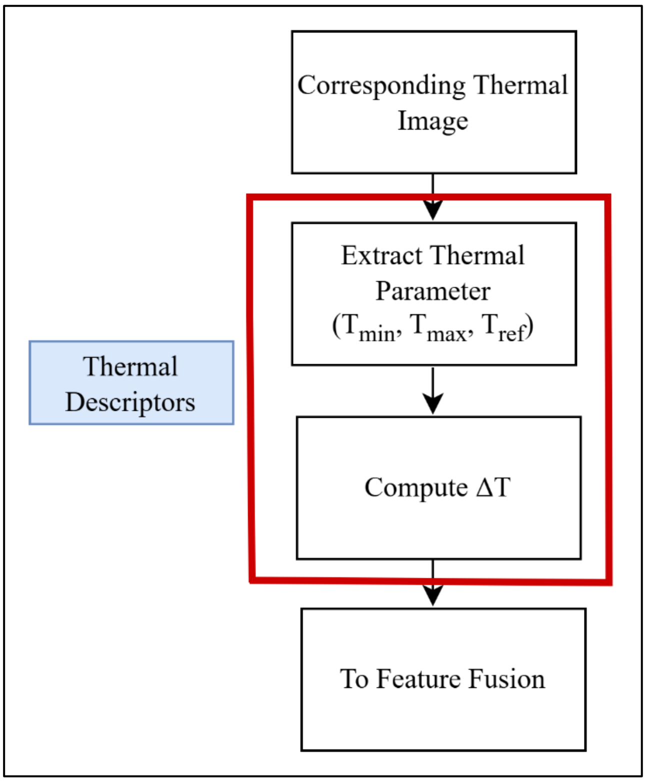


Figure 3.14 Workflow of Thermal-based Handcrafted Feature Extraction for Hotspot Severity Analysis

Figure 3.15 presents an example of a thermal image used to extract temperature-based features for hotspot severity analysis. The T_{min} and T_{max} temperature values within each segmented hotspot region are automatically extracted from the thermal image data using the implemented MATLAB-based processing workflow.

In contrast, the T_{ref} is determined through a cursor-assisted interaction that identifies the appropriate reference and hotspot locations within each thermal image. The corresponding temperature values are then retrieved directly from the thermal data at the selected locations. This semi-automated approach ensures accurate localization of reference and hotspot regions. At the same time, maintain consistency in temperature extraction. The extracted temperature parameters are subsequently used to compute the ΔT, which serves as a key descriptor for hotspot severity characterization across the dataset.

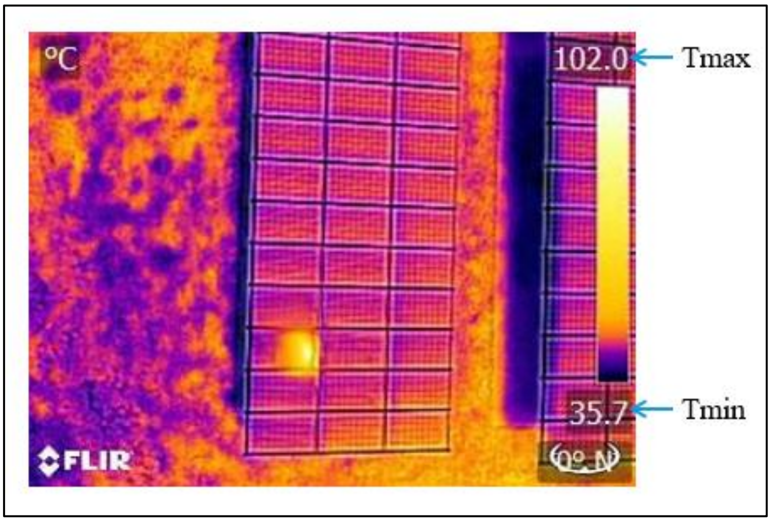


Figure 3.15 Example of a Thermal Image Illustrating Hotspot Regions and the Extraction of Minimum (T_{min}) and Maximum (T_{max}) Temperature Values Used for Temperature-Based Feature Computation

Figure 3.16 illustrates the variation of temperature-based features across the evaluated thermal images, including the T_{ref}, T_{max}, and ΔT. The ΔT value is computed as the difference between T_{max} and T_{ref} and serves as the primary indicator for hotspot severity characterization.

Consistent with the hotspot severity categorization proposed in [56], [122] severity levels were defined using ΔT thresholds to distinguish different hotspot conditions. Specifically, low-severity hotspots correspond to ΔT values of 5°C or below, indicating minor thermal deviations that may arise from normal operating variations or early-stage anomalies. Medium-severity hotspots are characterized by ΔT values between 5°C and 15°C, representing moderate thermal stress that may require closer monitoring. High-severity hotspots are identified when ΔT exceeds 15°C, reflecting pronounced thermal anomalies commonly associated with significant fault conditions and an increased risk of module degradation.

High-severity cases show larger ΔT values and are more clearly distinguishable from the reference temperature, whereas low and medium-severity cases tend to overlap. This makes lower severity levels harder to differentiate. As a result, ΔT serves as a practical and interpretable indicator for classifying hotspot severity.

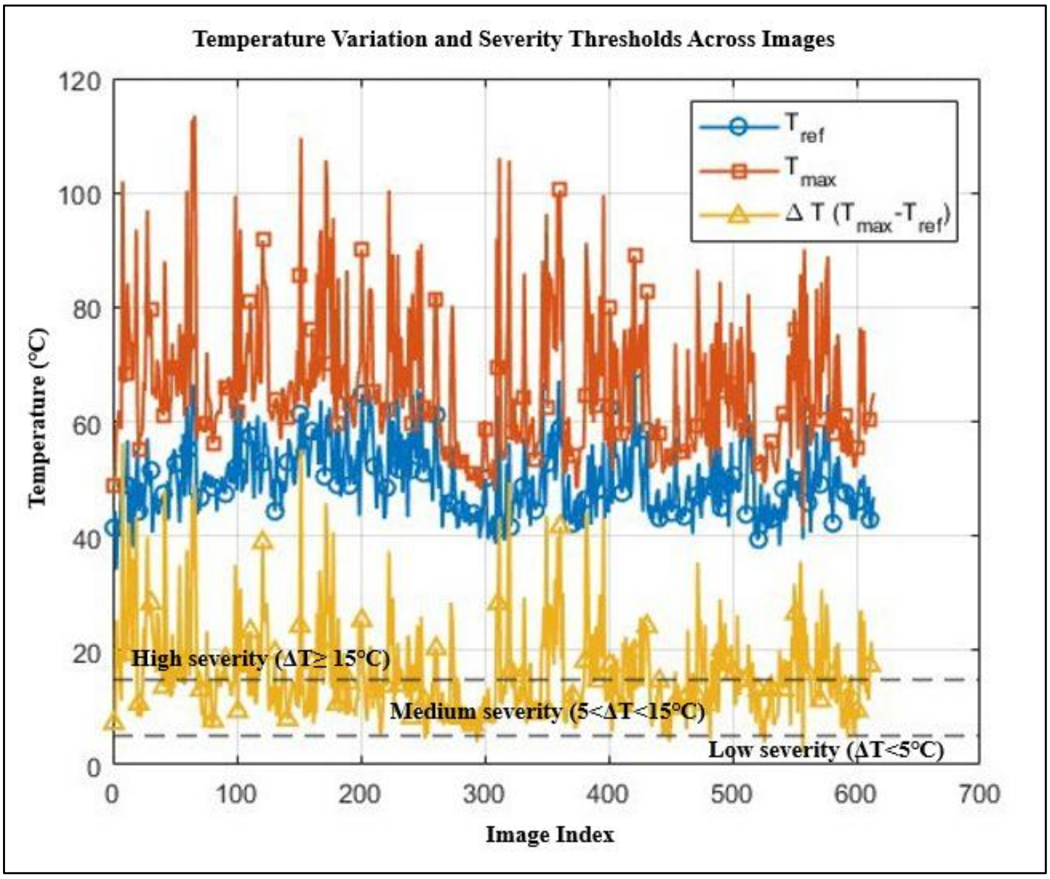


Figure 3.16 Temperature Variation Across Images

The ΔT is a relative measure of the intensity of thermal anomalies and consistent across a range of environmental conditions, including changes in ambient temperature and irradiance [139]. Unlike raw pixel temperatures, ΔT provide a normalized measure of fault intensity, with meaningful thresholds and gradients for distinguishing between different severity levels [54]. Previous research has also associated ΔT with early degradation behaviour and increased thermal stress risks in PV modules, reinforcing its importance as a severity descriptor [140], [141].

The detailed feature fusion techniques for both CNN and HC streams are further described in Subsections 3.6.3. This integration lays a strong foundation for the classification process in the subsequent stage.

3.6.3 Dual-input Feature Fusion Strategy for CNN Classification Model

Once features are extracted from thermal PV imagery, either through CNNs or HC methods, they must be fused and prepared for classification. This process involves structuring the extracted feature vectors and selecting optimal combinations to enhance the classifier's ability to distinguish between different hotspot severity levels. Two strategies were explored in this study: single-input and dual-input feature fusion. The single input strategy was implemented using two independent configurations: image-based features extracted from thermal images or thermal-based features derived from the hotspot region. In contrast, the dual-input feature fusion strategy integrates image-based and thermal-based features into a unified representation to enhance discrimination of hotspot severity. The proposed structure is shown in Figure 3.17, which illustrates how both modalities are combined into a unified representation for classification.

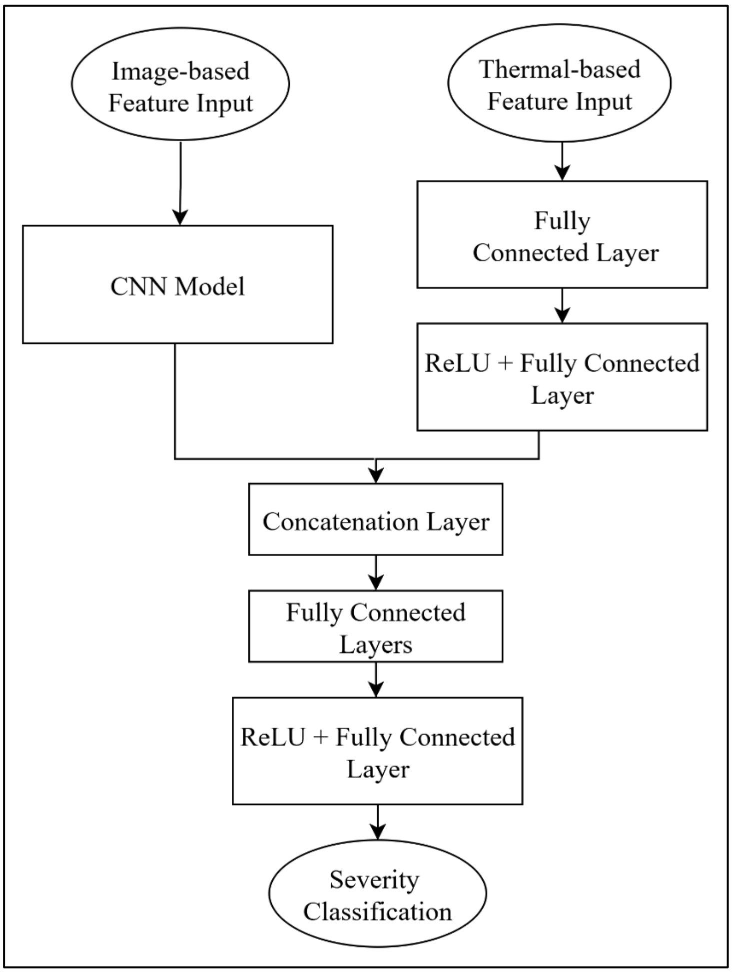


Figure 3.17 Dual-input Architecture Integrating CNN-derived Spatial Features for Hotspot Severity Classification

The CNN branch extracts spatial features from the segmented hotspot image, while the thermal descriptor branch processes temperature-based features derived from the hotspot region. The resulting feature vectors are concatenated to form a fused feature representation, which is subsequently passed to the fully connected classification layers for hotspot severity prediction.

To evaluate the contribution of each input modality to hotspot severity classification, three experimental configurations were investigated. The first configuration utilized image-based CNN features extracted from the segmented hotspot regions. The second configuration employed thermal descriptor features, namely T_{max}, T_{min}, T_{ref}, and ΔT. The third configuration integrated both image-based and thermal descriptor features through the proposed dual-input feature fusion strategy. These experimental configurations enabled a systematic comparison of individual and combined feature representations, thereby establishing a baseline for assessing the effectiveness of the proposed feature fusion approach.

Both inputs' advantages and disadvantages can be easily seen by analyzing each feature type separately. Numerical temperature distributions provide more consistent and accurate thermal thresholds, while visual thermal images offer rich contextual and textural details [142]. This comprehensive evaluation provides a solid foundation for understanding how integrating both inputs into a dual-feature configuration can improve classification accuracy and robustness.

3.7 Stage 3: Enhanced CNN Model for Classifying Hotspot Severity

This section describes Stage 3 of the study, which focuses on improving CNN-based hotspot severity classification performance. Based on the overall workflow presented in Figure 3.1, the detail of proposed method is shown presented in Figure 3.18. Generally, the hotspot regions extracted in the previous stage are processed using an enhanced CNN-based classification model.

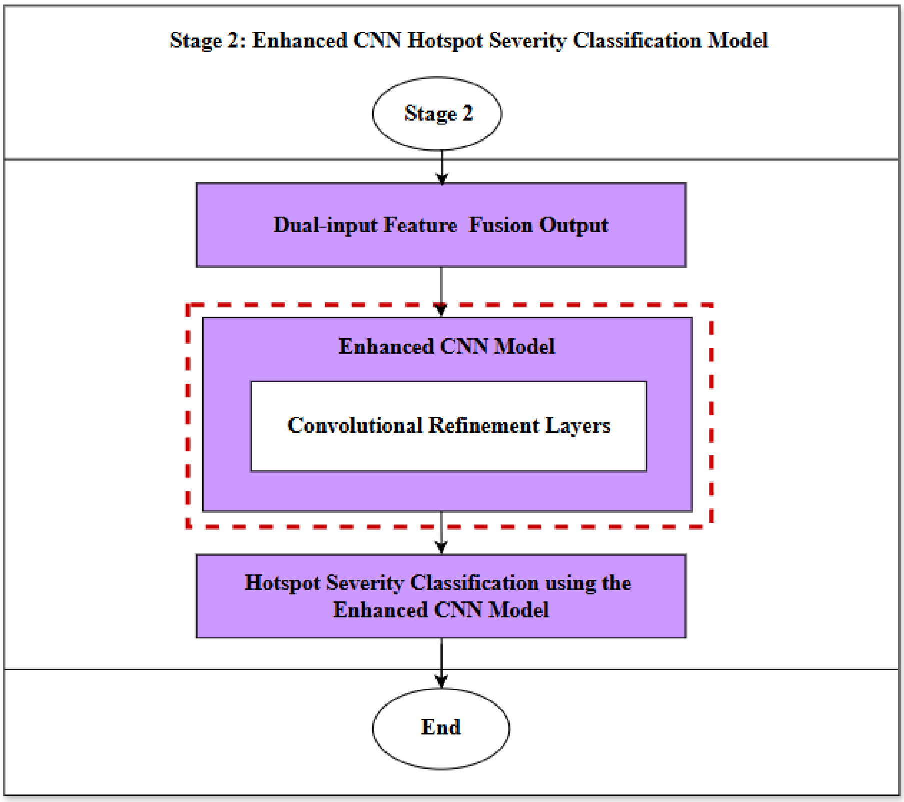


Figure 3.18 Overall Method Flow for Stage 3 (Research Objective 3)

In this stage, hotspot regions extracted from the previous stage are processed using an enhanced CNN-based classification model that integrates refined hotspot information to improve severity discrimination. The workflow demonstrates the transition from hotspot localization to severity-level classification, forming the final decision stage of the proposed framework. The enhanced CNN model in this stage aims to strengthen feature representation and classification stability for accurate hotspot severity categorization.

3.7.1 Proposed Enhanced CNN Model

Conventional CNN architectures such as U-Net, ResNet-18, and ResNet-50 have demonstrated strong capability in feature extraction for image segmentation and classification tasks. However, hotspot severity classification in PV thermal imagery requires more discriminative high-level feature representations, as severity differentiation depends not only on spatial localization but also on fine-grained thermal patterns [31]. To address this requirement, an enhanced CNN model is proposed by introducing additional convolutional refinement layers at the final encoding stage of the baseline CNN architectures.

In CNNs, deeper layers progressively capture higher-level, more abstract feature representations by increasing receptive fields and strengthening feature interactions [143]. While baseline CNN models are effective for spatial learning, their original architectures are not explicitly optimized for severity-aware discrimination. Therefore, additional convolutional layers are incorporated to further refine high-level features before classification, rather than redesigning the entire network architecture. This refinement strategy enhances severity-discriminative features by introducing additional nonlinear transformations, while preserving the original architecture of each CNN model. The added convolutional layers progressively filter out low-activation or non-informative feature responses and emphasize high-intensity thermal patterns associated with hotspot severity [144]. Furthermore, severity classification differs from segmentation in that it requires global discriminative cues rather than solely precise pixel-level delineation [96]. Additional convolutional refinement layers enable the network to suppress non-informative responses and amplify severity-related features, thereby improving classification robustness under varying thermal conditions.

To systematically investigate the impact of architectural refinement, one, two, and three additional convolutional layers were introduced, denoted Conv-1, Conv-2, and Conv-3, respectively. The single-layer configuration represents minimal refinement and evaluates whether limited enhancement is sufficient to improve feature discrimination. Two layers enable deeper feature interaction and increased non-linearity, supporting more complex severity representations. Three layers serve as an upper-bound refinement configuration to examine diminishing performance gains and potential overfitting effects. This progressive augmentation strategy enables controlled evaluation of feature refinement while avoiding excessive architectural complexity.

The detailed implementation settings of the additional convolutional refinement blocks are presented in Table 3.6. Each block comprises a 3×3 convolutional layers with 64 filters, followed by batch normalization, ReLU activation, and dropout regularization. These settings were kept constant across the Conv-1, Conv-2, and Conv-3 configurations to ensure that performance differences were attributed solely to the number of refinement blocks introduced.

Table 3.6
Configuration of Additional Convolutional Refinement Blocks

Layer	Kernel Size	Filters	Padding	Activation	Dropout
Conv Block 1	3×3	64	Same	ReLU	0.3
Conv Block 2	3×3	64	Same	ReLU	0.3
Conv Block 3	3×3	64	Same	ReLU	0.3

The integration of the additional convolutional refinement layers and the subsequent feature fusion process is illustrated in Figure 3.19.

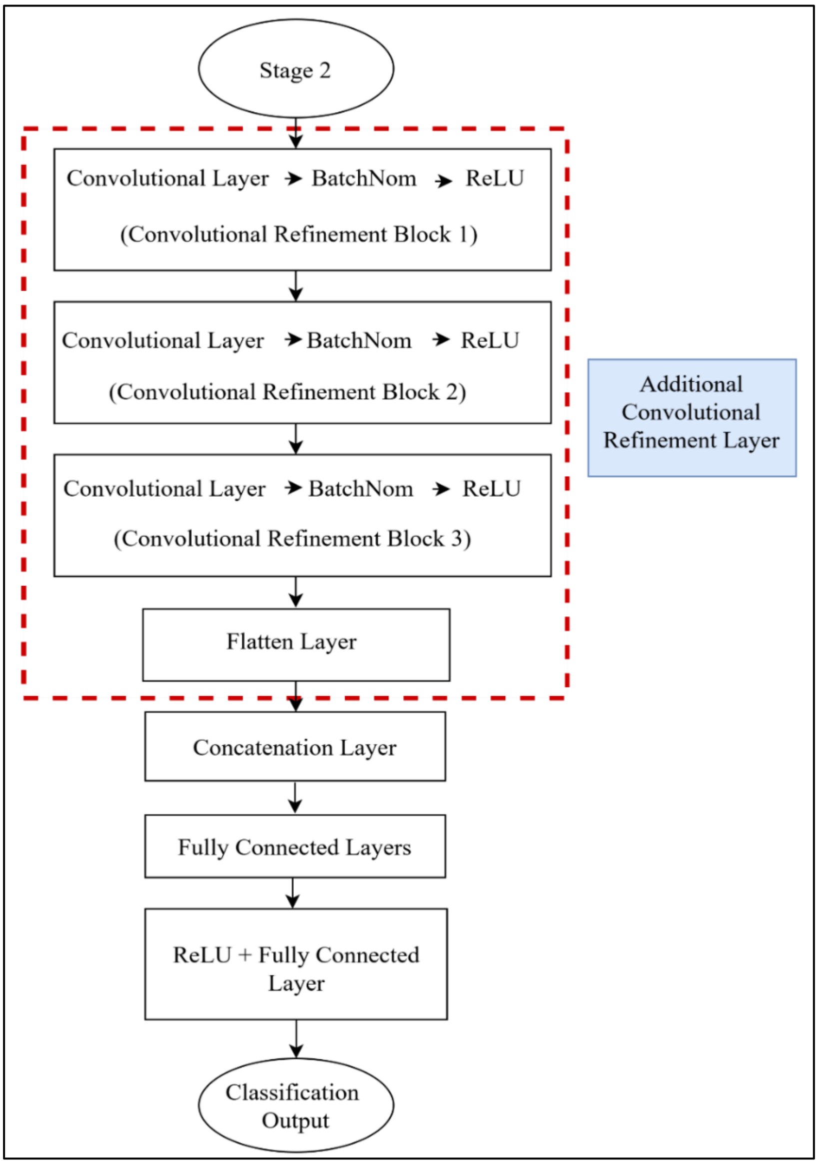


Figure 3.19 Architecture of the Enhanced CNN-based Hotspot Severity Classification Framework, Illustrating the Integration of Additional Convolutional Refinement Layers using Dual-input Feature Fusion Strategy

Figure 3.19 illustrates the architecture of the enhanced CNN-based hotspot severity classification framework. In this stage, a series of convolutional refinement blocks are progressively appended to the baseline CNN architecture to enhance discriminative feature learning. Each refinement block consists of a convolutional layer followed by batch normalization and ReLU activation. The refined feature maps produced by the final refinement block are subsequently flattened to form a compact one-dimensional CNN feature vector, as mathematically expressed in Equation (3.15).

$$F_{CNN} = Flatten(Conv_{Output}) \tag{3.15}$$

To incorporate thermal severity information, the flattened CNN feature vector is concatenated with temperature-based HC descriptors through a feature fusion layer, resulting in a unified representation as defined in Equation (3.16). The fused features are then passed through fully connected layers to generate the final hotspot severity classification output.

$$F_{fusion} = Concatenate[F_{CNN}, \Delta T] \tag{3.16}$$

In this study, ΔT was used as the final handcrafted thermal descriptor for the feature fusion stage. The parameters T_{min}, T_{max}, and T_{ref} were extracted during thermal feature computation to characterize the hotspot temperature profile and to derive ΔT. Therefore, Equation (3.16) represents the fusion between the CNN feature vector and the final temperature-based descriptor, rather than the direct concatenation of all extracted thermal parameters.

The proposed enhanced CNN model incorporates progressive convolutional refinement layers into baseline CNN architectures. Thermal descriptors are further integrated through feature fusion to enhance feature discrimination for hotspot severity classification, while preserving the original network structures. This controlled refinement strategy establishes a robust, interpretable framework for severity-aware classification across different thermal imaging conditions.

3.7.2 Hotspot Severity Classification

Hotspot severity classification in PV thermal imagery is a challenging task due to the interaction between spatial hotspot patterns and temperature variations. While temperature difference (ΔT) is commonly used as a severity indicator, accurate classification requires consideration of both thermal intensity and spatial characteristics [31], [69]. Therefore, an integrated feature representation is necessary for reliable severity assessment. In this study, hotspot severity was categorized into multiple classes based on established thermographic assessment criteria. Table 3.7 summarizes the ΔT-based severity specifications and corresponding recommended maintenance actions adopted in this study, following establish infrared thermography guidelines.

Table 3.7
ΔT Specifications and Recommended Actions Used In This Study for Condition Monitoring of the Hotspot [56]

Condition	Priority level	ΔT between similar components (ºC)	Recommended actions
Low	3	ΔT≤5	Minor overheating; warrants investigation.
Medium	2	5<ΔT<15	Indicates probable deficiency; repair as time permits.
High	1	ΔT ≥15	Major discrepancy; repair immediately.

In this study, hotspot regions obtained from the segmentation stage were subjected to a feature extraction process aimed at enhancing severity discrimination. Image-based features were automatically learned through CNN to encode spatial patterns, texture, and morphological characteristics of the hotspot regions. In addition, thermal descriptors such as maximum temperature and temperature difference (ΔT) were extracted to capture intensity variations within each hotspot. These descriptors provide complementary information to CNN-derived features by explicitly representing temperature behavior associated with different severity levels. The integration of spatial and thermal features enables a more comprehensive representation of hotspot characteristics, supporting robust severity classification.

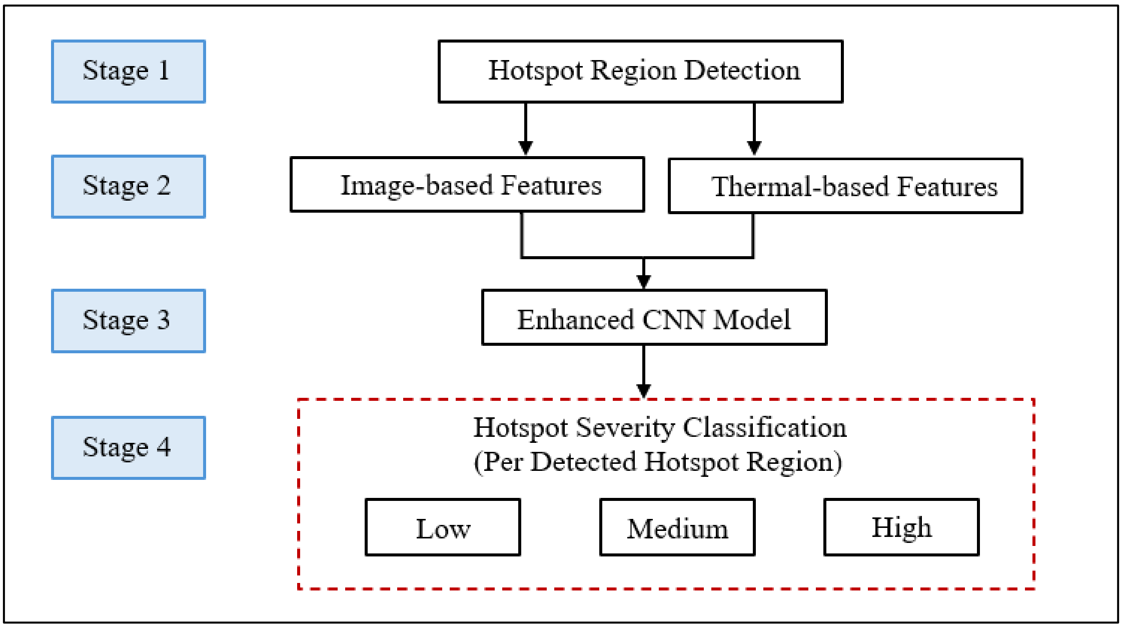


Figure 3.20 Overall Stage for Hotspot Severity Classification

As illustrated in Figure 3.20, hotspot severity classification was performed as an overall stage following hotspot segmentation. The segmented hotspot regions were first extracted as regions of interest to ensure that feature analysis was confined to relevant areas. Two categories of features were subsequently derived, namely image-based features learned through the CNN-based and thermal-based features representing temperature-related information. These complementary features were jointly used to classify hotspot severity into three predefined levels: low, medium, and high, based on ΔT thresholds. This structured classification framework enables systematic differentiation of severity by integrating spatial characteristics and thermal intensity. This feature-based severity classification framework provides a consistent basis for subsequent performance evaluation and expert validation.

3.8 Stage 4: Performance Evaluation of Proposed Hotspot Localization and Severity Classification Model

This stage presents a thorough performance analysis of the proposed hotspot severity classification model using both metric-based evaluations and thermographer validation. The aim is to ensure that the proposed model is technically accurate in detecting and classifying thermal anomalies in PV systems, with supporting evidence from experts to end users. The metrics' performance evaluation emphasizes standard metrics for segmentation and classification, which include Accuracy, mean IoU (mIoU), precision, F1-score, recall and confusion metrics. These metrics are selected due to their broad adoption in semantic segmentation tasks and their effectiveness in assessing model accuracy, particularly in imbalanced or complex image datasets. In contrast, qualitative assessment encompasses human-centric insights by involving domain experts and users in the evaluation process. The system's interpretability, usability, and perceived reliability are validated through the collection of user feedback via structured questionnaires and guided annotation exercises. By combining both forms of evaluation, this stage offers a holistic understanding of the model's performance in both technical and operational contexts.

3.8.1 Performance Evaluation of Stage 1: Two-Tier Semantic Segmentation Method

This subsection explains the two-tier segmentation phase where the main intention of this operation is to localize hotspot region. To quantitatively assess the segmentation performance, several standard evaluation metrics. These metrics provide a comprehensive assessment of segmentation effectiveness in terms of classification correctness, spatial overlap, detection reliability, and computational efficiency. In this study, the model's performance parameters are calculated using the values from confusion matrix. The definition of the term used in the confusion matrix is explained in Table 3.8.

Table 3.8
Interpretation of the Term Used in the Confusion Matrix

Matrix Evaluation	**Interpretation**
True Positive (TP)	The confusion matrix shows the number of correct predictions for positive value.
True Negative (TN)	The confusion matrix shows the number of correct predictions for negative value.
False Positive (FP)	The confusion matrix shows the yes value for actual class is no.
False Negative (FN)	The confusion matrix shows the no value for actual class is yes.

Below are the lists of performance metrics using the confusion matrix components to evaluate the proposed method.

i. Accuracy: accuracy is employed to evaluate the overall correctness of the Stage-1 segmentation by measuring the proportion of pixels correctly classified as hotspot or non-hotspot regions within the PV module. The mathematical expression for accuracy is given by Equation (3.17):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \tag{3.17}$$

ii. Intersection over Union (IoU): the ratio between the pixels that are correctly classified (true positive) divided by the sum of the total number of pixels that have been labelled with that class or that in the ground truth belong to that same class, as in Equation (3.18):

$$IoU = \frac{TP}{TP + FP + FN} \times 100\% \tag{3.18}$$

iii. Mean Intersection over Union (mIoU): the average spatial overlap between the predicted and ground truth segmentations by computing the ratio of the intersection to the union of the two regions across all classes. The mathematical expression for mIoU is given by Equation (3.19):

$$mIoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} = \frac{(A \cap B)}{(A \cup B)} \tag{3.19}$$

where:
C: the number of segmentation classes.
IoU_c: the Intersection over Union value for the c-th class.

iv. Positive Predictive Value (PPV) or Precision: the ratio of positively predicted correct class to the number of predicted positive, as in Equation (3.20):

$$PPV\ or\ Precision = \frac{TP}{TP + FP} \times 100\% \tag{3.20}$$

v. Sensitivity or Recall: the proportion of true hotspot pixels that are correctly detected by the segmentation model and is expressed as Equation (3.21):

$$Sensitivity\ or\ Recall = \frac{TP}{TP + FN} \times 100\% \tag{3.21}$$

Overall, the evaluation results demonstrate the effectiveness of the Stage 1 segmentation in accurately localizing hotspot regions at the pixel level. Thereby providing a reliable input for subsequent severity classification in Stage 2 of the proposed framework.

3.8.2 Performance Evaluation of Stage 2: CNN Classification Model with Dual-input Feature Fusion Technique

This subsection presents the performance evaluation of Stage 2 for the CNN-based classification model, which aims to categorize hotspot severity conditions based on the extracted features. The evaluation is conducted using several standard classification metrics, including accuracy, precision, recall, and F1-score. In addition, k-fold cross-validation is employed to assess the classification reliability and generalization capability of the proposed model. The definitions and formulations of accuracy, precision, and recall employed in this stage follow those presented in Subsection 3.8.1 and are therefore not repeated here for conciseness and consistency. To further evaluate the balance between precision and recall, the F1-score is employed.

F1-score: the ratio of predictive performance based on recall and precision, as given by Equation (3.22):

$$F1 - score = 2 \times \frac{Precision\ \times Recall}{Precision + Recall} \times 100\% \tag{3.22}$$

Table 3.9 illustrates the structure of the confusion matrix and summarizes the performance metrics employed for evaluating the CNN classification models. The figure highlights the relationships among true positive, true negative, false positive, and false negative predictions, which underline the computation of accuracy, precision, recall, and related evaluation metrics.

Table 3.9
Conceptual Representation of the Confusion Matrix and Associated Performance Metrics used for Evaluating the Enhanced CNN Classification Model

		Predicted		
		Low	**Medium**	**High**
Actual	**Low**	TP	FP	FP
	Medium	FN	TP	FP
	High	FN	FN	TP

To further ensure stable model performance across different data distributions, a 5-fold cross-validation procedure is employed. The 5-fold cross-validation was applied to the full dataset, which was randomly partitioned into five equal subsets before cross-validation. The evaluation of the best CNN for the feature extractor was based on test accuracy and k-fold cross-validation. The hold-out validation uses 80% for training, 10% for testing, and 10% for validation, with k-fold cross-validation. Each image used for training, validation, and testing was labelled, and the labels were reused to ensure a non-biased comparison between experiments. Meanwhile, k-fold cross-validation (k=5) was also applied, with the dataset randomly partitioned into 5 subsets. The k-fold (5-fold) cross-validation worked by training the model five times, where each model was trained using four of the partitions, and the remaining partition was used for validation and was then repeated for all folds. The average performance was based on validation accuracy, and the standard deviation across all folds was calculated. Table 3.10 shows the 5-fold cross-validation process for the experiment.

Table 3.10
k-fold Cross-validation Process, k = 5

Folds (k)					Accuracy (%)
1	**2**	**3**	**4**	**5**	
Testing and Validation	Train	Train	Train	Train	Fold 1 Performance
Train	Testing and Validation	Train	Train	Train	Fold 2 Performance
Train	Train	Testing and Validation	Train	Train	Fold 3 Performance
Train	Train	Train	Testing and Validation	Train	Fold 4 Performance
Train	Train	Train	Train	Testing and Validation	Fold 5 Performance
5-folds cross-validation accuracy					Average Performance

Overall, the Stage 2 CNN classification model is evaluated using multiple complementary metrics, including accuracy, precision, recall, and F1-score. These metrics provide a balanced assessment of classification effectiveness from different perspectives. The confusion matrix further supports this evaluation by detailing the distribution of true positive (TP), true negative (TN), false positive (FP), and false negative (FN) for each hotspot severity class. This allows a clearer understanding of class-wise prediction behaviour and potential misclassification patterns. In addition, 5-fold cross-validation enhances the reliability of the results by ensuring consistent performance across different data partitions.

3.8.3 Performance Evaluation of Stage 3 for Enhanced CNN Classification Model

This subsection presents the performance evaluation of the Stage 3 enhanced CNN classification models, which aim to examine the impact of progressive architectural refinement on hotspot severity classification in PV modules. Similar to Stage 2, the evaluation is conducted using standard classification metrics, namely accuracy, precision, recall, and F1-score. The definitions and formulations of these metrics, including the confusion matrix, follow those presented in Subsections 3.8.1 and 3.8.2 and are therefore not repeated here for consistency.

The dataset partitioning strategy remains consistent with Stage 2: 80% for training, 10% for validation, and 10% for testing, to ensure fair performance comparison. The baseline model adopted in this stage corresponds to the best-performing configuration identified in Stage 2.

To investigate the effect of architectural enhancement, additional convolutional layers were progressively integrated into the baseline CNN models. Specifically, one, two, and three convolutional layers were added to evaluate their influence on classification performance. This enhancement strategy was applied consistently across three representative CNN architectures, namely U-Net, ResNet-18, and ResNet-50, enabling a systematic comparison of model behaviour under increasing architectural complexity.

The performance of each enhanced configuration was analyzed based on overall classification metrics and class-wise prediction results derived from the confusion matrix. This evaluation facilitates the identification of performance trends, improvement patterns, and potential misclassifications across different hotspot severity levels. Overall, this evaluation approach enables a structured comparison of the enhanced CNN architectures and evaluates the impact of additional convolutional depth on hotspot severity classification performance.

3.8.4 Performance Evaluation of Stage 4: Overall Performance Evaluation based on Metric and Thermographer Evaluation

This subsection presents an overall performance evaluation of the proposed automated hotspot detection and severity classification framework, integrating quantitative performance metrics with expert-based validation. The earlier stages focused on refining model performance and enhancing the architecture. This stage provides a comprehensive assessment that integrates classification performance analysis, solar thermographer validation, and statistical agreement evaluation.

The objective of this integrated evaluation is to determine whether the enhanced CNN models not only achieve high quantitative performance but also demonstrate reliable agreement with expert judgment. To achieve this, three complementary evaluation components are employed: (i) performance analysis using standard classification metrics, (ii) validation against Solar Thermographer assessments, and (iii) statistical analysis using agreement-based measures. This structured evaluation approach ensures the proposed system is assessed for predictive accuracy, practical reliability, and expert consistency.

a) Performance Analysis

This subsection presents the quantitative performance analysis of the major components of the proposed automated hotspot detection and severity classification framework. The analysis covers the two-tier semantic segmentation module, the CNN classification model with dual-input feature fusion, and the enhanced CNN classification architectures. The evaluation provides a systematic and consistent assessment of model performance at both pixel-level segmentation and severity-level classification stages.

For the two-tier semantic segmentation module, performance evaluation was conducted using accuracy, mean Intersection over Union (mIoU), precision, and recall to comprehensively assess multi-class segmentation quality. The metric definitions and calculation procedures are consistent with those described in the earlier performance evaluation sections.

For the hotspot severity classification stages, including the dual-input feature fusion model and the enhanced CNN architectures, performance evaluation was based on confusion-matrix-derived metrics. The confusion matrix summarizes true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions for each severity class. Based on these components, standard classification metrics such as accuracy, precision, recall, and F1-score were computed to assess overall correctness and class-wise predictive balance.

Comparative analysis was performed across different model configurations to examine performance variations resulting from feature fusion strategies and progressive architectural enhancement. This structured evaluation ensures a comprehensive assessment of segmentation performance and classification consistency within the proposed automated framework.

b) Solar Thermographer Validation

This assessment evaluates the automated system's ability to classify hotspot severity levels. This study employed a questionnaire-based approach to assess the performance and user experience (solar thermographer) of the computerized system. The questionnaire was structured into three parts: gathering personal information, evaluating images to detect abnormalities, and soliciting feedback on preferences regarding the proposed automated model, which was developed based on hotspot severity levels. Informed consent was obtained from all solar thermographer participants before their involvement in the study. Confidentiality and anonymity were implemented to preserve respondents' answers throughout the data collection and analysis phases. The collected data were used solely for research purposes and managed responsibly throughout the study. A sample of the questionnaire is attached to Appendix.

To validate the performance of the proposed automated model, an expert-based questionnaire study was conducted. This procedure consisted of respondent recruitment, image selection and annotation, questionnaire development, and data collection. The solar thermographers evaluated thermal images containing the localized hotspot regions and the corresponding hotspot severity classifications generated by the proposed model. The displayed information enabled the experts to assess both the hotspot localization results and the assigned severity levels against their professional judgement.

i) Questionnaire Development

The questionnaire was designed to obtain structured feedback from domain experts and consisted of three main sections, as summarized in Table 3.11.

Table 3.11
Structure and Components of the Expert Validation Questionnaire

Part	**Task**	**Description**
1	Respondent Background Information	This section gathered demographic and professional details from the respondents, including age, gender, and years of experience in solar thermographic analysis. The purpose was to profile the respondents and enable a comparative study across different levels of expertise in interpreting thermal imagery.

Part	Task	Description
2	Thermal Image Assessment	In this section, respondents were presented with 20 thermal images that had undergone automated hotspot detection and segmentation using the proposed model (see Appendix). The 20 thermal images were selected from the testing dataset to provide representative examples of hotspot conditions encountered in the study. The image selection process included samples from all three severity categories (low, medium, and high) to enable comprehensive expert evaluation across different hotspot conditions. The selected images were randomly and subsequently reviewed to ensure that each severity category was represented in the validation questionnaire. Each displayed image marked the hotspot regions identified by the system. Respondents were asked to assess the severity of the hotspot and classify the thermal condition based on their expert judgment. This evaluation aimed to validate the accuracy and reliability of the model's output by comparing it with expert assessments.
3	System Perception and Preferences	This final section explored respondents' views on the applicability and usability of automated thermal analysis systems. Respondents were asked to indicate their preferences regarding system features, including integration with expert decision-making, suitability for different inspection scenarios, and perceived confidence in the system's hotspot detection capabilities. This section aims to gather insights into how domain experts perceive the effectiveness and practical value of the proposed automated approach in real-world applications.

Overall, the questionnaire's structured design ensured systematic and consistent expert feedback. This feedback was used to validate the proposed hotspot detection and severity classification framework.

ii) Respondents Recruitment

Respondents were recruited from among professional solar thermographers with expertise in thermal imaging. The recruitment process was conducted through professional networks and affiliated institutions. A total of nine (9) independent solar thermographers were selected to participate in the questionnaire. The expert respondents involved in the validation process were drawn from various academic, utility, and industry institutions to ensure a balanced and credible assessment. Table 3.12 summarizes the distribution of expert respondents according to their institutional affiliation.

Table 3.12
Distribution of Expert Respondents by Institutional Affiliation

Institution / Organization	Number of Respondents
Universiti Malaysia Perlis (UniMAP)	3
Tenaga Nasional Berhad (TNB)	1
Universiti Teknologi MARA (UiTM)	1
Centre for Energy Management and Sustainable Campus (COSCEM),	1
Private Electrical Consultancy Firms	3
Total	**9**

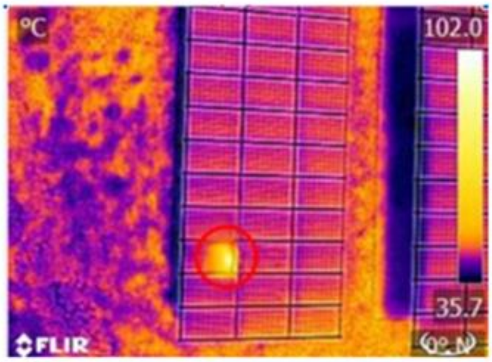
The involvement of qualified solar thermography experts from diverse institutional backgrounds strengthens the credibility of the validation process.

iii) Expert Evaluation and Validation Procedure

This subsection describes the selection of thermal images and the expert-based evaluation procedure employed to validate the hotspot severity assessment performance of the proposed model.

A total of 20 thermal images were selected for inclusion in the questionnaire evaluation. These images were randomly taken and sampled from the testing output results that had been previously evaluated using the proposed automated model. Each image presented to the respondents had undergone hotspot region detection and segmentation during Stage 1 of the methodology. To emphasize the regions of interest, the identified hotspot areas were graphically marked on each image as shown in Figure 3.21. This eliminated the need for hotspot detection, allowing respondents to focus on assessing the severity of the identified hotspots. The use of model-processed images with pre-identified hotspot regions ensured a consistent evaluation basis for all participants. Since the hotspot locations were provided, the expert validation focused on assessing the agreement between the severity classifications assigned by the proposed model and the professional judgement of the solar thermographers. This approach was intended to evaluate the reliability of hotspot severity assessment rather than independent hotspot detection performance.

This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle?



- Low
- Medium
- High

Figure 3.21 Example of Expert Questionnaire Item for Hotspot Severity Classification

To support a structured and unbiased evaluation, respondents were provided with an electronic version of the questionnaire to facilitate viewing thermal images and streamline the evaluation process. Clear instructions accompanied the questionnaire, outlining the objectives of each section and guiding respondents through their assessment tasks step by step. To ensure data integrity and reduce potential bias, all completed questionnaires were collected anonymously. This approach safeguarded respondent confidentiality and promoted impartiality in subsequent data analysis.

Overall, the expert-based image evaluation procedure establishes a structured and unbiased foundation. This foundation supports the subsequent performance analysis and validation of the proposed automated hotspot detection and severity classification model.

c) Statistical Analysis

This subsection outlines the statistical procedures employed to evaluate the agreement between the proposed automated hotspot detection and severity classification model and expert assessments. The objective of this analysis is to quantify the level of consistency between model predictions, expert evaluations, and ground-truth classifications using established statistical measures.

Data collected from the questionnaires in Parts 1 and 3, including demographic information and perceptions regarding automated inspection systems, were used to support descriptive analysis. Meanwhile, responses from Part 2 were statistically analyzed to compare hotspot severity classifications between the Solar Thermographer (expert reference), the proposed automated model, and the validation assessments.

To measure inter-rater agreement and reliability, Cohen's Kappa coefficient (κ) was employed. This coefficient evaluates the degree of agreement between two raters while accounting for chance agreement. The selection of Cohen's Kappa is appropriate for categorical classification problems, particularly in multi-class severity assessment where agreement beyond chance must be quantified. The interpretation of κ values follows the standard classification scale presented in Table 3.13.

Table 3.13
Interpretation Scale of Cohen's Kappa Coefficient

κ Values Range	**Interpretation of Agreement**
$\kappa \leq 0$	No agreement
$0.01 \leq \kappa \leq 0.20$	None to slight agreement
$0.21 \leq \kappa \leq 0.40$	Fair agreement
$0.41 \leq \kappa \leq 0.60$	Moderate agreement
$0.61 \leq \kappa \leq 0.80$	Substantial agreement
$0.81 \leq \kappa \leq 1.00$	Almost perfect agreement

The formula to calculate the correlation between two variables P_o and P_e is given by Cohen Kappa coefficient (κ) is expressed by Equation (3.23).

$$\kappa = \frac{P_o - P_e}{1 - P_e} \tag{3.23}$$

where: P_o = observed agreement between raters.
P_e = expected agreement between raters by chance.

In this study, the Kappa coefficient was used to assess the consistency and reliability of hotspot severity classifications generated by the automated model relative to expert evaluation. This statistical validation ensures that the observed agreement reflects meaningful predictive reliability rather than random coincidence.

3.9 Summary

This chapter presented the complete research methodology developed for automated hotspot detection and severity classification in PV thermal imagery. The proposed methodology was designed as a structured multi-stage framework to address the challenges associated with thermal image quality, background interference, hotspot localization accuracy and severity discrimination under real-world PV inspection conditions.

The chapter began by describing the overall research design, image acquisition process, and data collection strategy. Thermal images of PV modules were acquired under operational conditions and subsequently analyzed. To ensure reliable processing, several pre-processing steps were introduced, including noise suppression and image enhancement. These procedures enhanced thermal contrast and minimized non-uniform intensity variations that could negatively influence hotspot detection and classification performance.

In line with ROI, a new two-tier semantic segmentation strategy was proposed to localize hotspot regions accurately. The first segmentation stage focused on isolating PV modules from the background, while the second stage refined the segmentation to extract hotspot regions within the modules. This hierarchical segmentation approach was designed to minimize background interference and reduce boundary uncertainty. As a result, hotspot localization can be performed more precisely, even under low-contrast and non-uniform irradiance conditions The refined hotspot regions obtained at this stage formed a reliable spatial foundation for subsequent analysis.

Next, the proposed method addressed RO2 by introducing a dual-input feature-extraction strategy to capture both visual characteristics and temperature-based descriptors from the segmented hotspot regions. In this stage, image-based features learned by CNNs were combined with temperature-driven thermal descriptors to form a unified representation of hotspot severity. This dual-input approach allows complementary spatial and thermal information to be jointly exploited. As a result, the discrimination of hotspot severity levels is enhanced, addressing the limitations of purely image-centric analysis.

An enhanced CNN-based hotspot severity classification model was proposed to improve severity-aware discrimination in line with RO3. Progressive convolutional refinement layers were incorporated into baseline CNN architectures to enhance high-level feature representations without altering the original network architecture. In addition, hybrid feature fusion was employed by integrating CNN-derived spatial features with physically meaningful thermal descriptors, enabling the model to learn severity-relevant patterns more effectively. This stage represented the final decision-making component of the proposed framework, producing severity classification outputs for each detected hotspot.

Finally, the proposed stages are evaluated qualitatively using model metrics and expert agreement with assessments provided by solar thermographers. This evaluation addresses RO4 and includes correlation analysis to assess the agreement between automated results and expert judgment. This combined evaluation strategy ensured that the proposed framework was not only technically effective but also practically applicable in real-world PV inspection scenarios. The results and discussion are presented in the subsequent Chapter 4.

CHAPTER 4
RESULTS AND DISCUSSION

4.1 Introduction

A comprehensive methodology for detecting hotspots and classifying their severity in solar photovoltaic (PV) modules was presented in the previous chapter. The proposed framework was designed using deep learning (DL)-based techniques to develop an image segmentation method and classification model. Thus, this section focuses on presenting and discussing the experimental results from the research works, as well as on analysing the findings from the outputs.

The first part of this chapter provides the results and analysis for hotspot localization at Stage 1. The proposed two-tier segmentation method was evaluated to determine the most effective approach for automated hotspot detection in thermal images. A comparative analysis of the proposed techniques is presented by highlighting performance differences, computational efficiency, and segmentation accuracy.

Next, the subsequent section focuses on presenting the results and performance analysis of the classification model by suggesting a new robust feature extraction method. The reliability of the developed model was analyzed through a systematic evaluation of the extracted features and the resulting impact on overall classification performance. The proposed framework was validated using a labelled thermal image dataset to ensure comprehensive analysis across single and dual-input conditions. The improved hotspot categorization accuracy demonstrates the framework's ability to capture complex spatial and thermal patterns. The outcome underscores the importance of domain-specific feature engineering for advancing PV hotspot identification and severity classification.

Furthermore, the next section concentrates on the severity classification process. The segmented thermal images were systematically divided into training and testing datasets to evaluate the generalization capability of the model. Hotspot severity was categorized into predefined class labels to ensure a consistent and structured classification framework across different datasets. The classification structure enhanced both accuracy and reliability during performance evaluation. The analysis also provides valuable insights for further refinement and real-world implementation in solar PV monitoring applications.

The discussion concludes with a comparative analysis between the proposed classification framework and the manual interpretation conducted by a certified solar thermographer. The degree of agreement between the two evaluations was quantified using statistical agreement analysis. Quantitative and qualitative results were examined to validate the performance of the proposed classification approach. The comparative study highlights the alignment between the developed system and expert interpretation, thereby demonstrating the practicality and reliability of the framework. The findings emphasize the potential of the proposed method as a decision-support tool for solar thermographers in diagnosing hotspot severity and ensuring effective PV system maintenance.

4.2 Results and Analysis of Stage 1: Localization of Hotspot Region using A New Two-tier Segmentation Method

This section presents the outcomes of Stage 1, which aims to localize hotspot region using a new two-tier semantic segmentation, as detailed in Subsection 3.5. The results of the first-tier segmentation are discussed in Subsection 4.2.1, whereas the results of image enhancement using Non-Local Means with Contrast-Limited Adaptive Histogram Equalization (NLM-CLAHE) are presented in Subsection 4.2.2, and finally, the second-tier segmentation analysis is provided in Subsection 4.2.3.

4.2.1 Results of First-tier Segmentation

The first-tier segmentation is aimed at separate solar PV modules and the background. The output of this segmentation is important because the hotspot regions within the segmented modules are the subject of the next segmentation stage, which is covered in the next section. Figure 4.1 illustrates qualitative results of the first-tier segmentation outputs of selected samples from the training and testing datasets. The outcomes at this point clearly and distinctly separate the solar panel and the background, demonstrating the efficacy of the separating procedure. The results show that the PV module and background regions in the thermal images were effectively clustered.

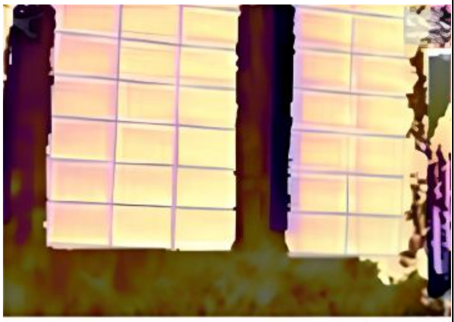
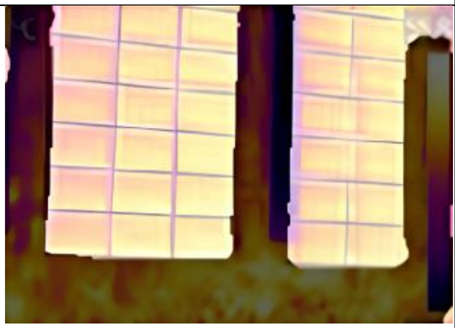
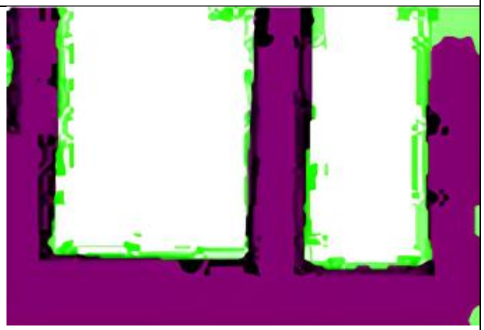
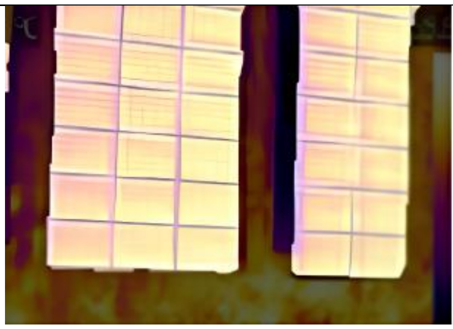
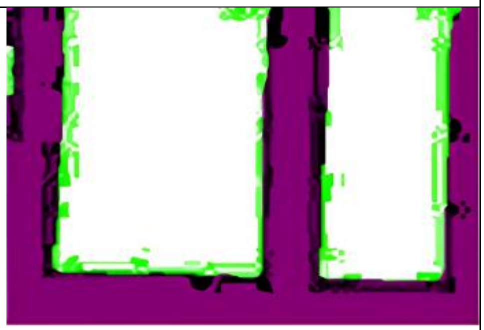
Deep Learning Model	**Thermal Images**	**First-tier Segmentation Result**
a) U-Net		
b) ResNet -18		
c) ResNet -50		
PV Panel	Background or non-panel	GT versus Segmented

Figure 4.1 Sample of Result for the First-tier Segmentation for Separating Solar Panel and Background using Deep Learning (a) U-Net (b) ResNet-18 (c) ResNet-50

The qualitative results of the first-tier segmentation are generated by three DL architectures, namely U-Net, ResNet-18, and ResNet-50. All models effectively distinguished PV panels from the background, with white areas denoting PV panels and purple areas denoting non-panel areas. Green outlines highlight segmentation boundaries and residual noise.

The U-Net model, as shown in Figure 4.1(a), achieved acceptable segmentation accuracy but exhibited noticeable edge noise, as reflected by the green artefacts around the module boundaries. Meanwhile, ResNet-18, as shown in Figure 4.1(b), produced cleaner segmentation by minimizing boundary noise while maintaining some inconsistencies at the edges. The result of ResNet-50, as shown in Figure 4.1(c) achieved the most refined segmentation results, characterized by smoother boundaries and minimal noise, demonstrating a superior ability to capture fine spatial details. This finding can be attributed to the deeper convolutional layers of ResNet-50, which enhance feature extraction and increase the receptive field, enabling better differentiation between similar thermal intensities within PV modules.

The results of first-tier segmentation are further quantitively analyzed for all models, as shown in Table 4.1. Among all, ResNet-50 achieved the highest accuracy of 98.33% and a mean IoU of 96.41%, confirming strong agreement with the ground truth. ResNet-18 followed closely with 97.55% accuracy, exhibiting the highest precision (99.75%) and recall (98.80%). These results demonstrate the network's ability to balance sensitivity and specificity, effectively reducing both false positives and false negatives. In contrast, U-Net recorded the lowest performance, with an accuracy of 95.03% and a mean IoU of 91.54%, suggesting comparatively weaker boundary representation and feature extraction capability than the ResNet-based architectures evaluated in this study. These findings are consistent with recent studies [31], [145], which reported that deeper architectures such as ResNet-50 outperform U-Net for PV fault detection due to enhanced gradient propagation and improved feature representation. In this study, ResNet-50 (50 layers) and ResNet-18 (18 layers) provide deeper hierarchical feature extraction than the comparatively shallower U-Net encoder. Moreover, both ResNet models leverage pretrained weights, enabling better feature generalization and contributing to their superior segmentation performance.

Table 4.1
The First-tier Segmentation Result

Semantic Model	**Accuracy (%)**	**Mean IoU (%)**	**Precision (%)**	**Recall (%)**
U-Net	95.03	91.54	98.28	94.97
ResNet-18	97.55	95.69	99.75	98.80
ResNet-50	**98.33**	**96.41**	**96.87**	**98.79**

Although the reported accuracy values are high, pixel-level accuracy in semantic segmentation may be influenced by class imbalance between PV module and background regions. Therefore, mean IoU, precision, and recall were also considered to provide a more comprehensive assessment of segmentation performance. The high mean IoU achieved by ResNet-50 (96.41%) indicates substantial overlap between the predicted segmentation masks and the ground-truth annotations, suggesting that the superior performance was not solely influenced by the dominance of background pixels.

While ResNet-18 achieved the highest precision (99.75%), ResNet-50 produced higher accuracy and mean IoU values. This indicates that ResNet-50 generated slightly more false-positive pixels while providing more complete PV module coverage and improved boundary localization. Since the first-tier segmentation output is subsequently used for hotspot localization and severity classification, preserving relevant module regions is more critical than adopting an overly conservative segmentation strategy. Consequently, the higher IoU and comparable recall achieved by ResNet-50 provide a stronger foundation for the subsequent stages of the proposed framework.

In contrast, U-Net recorded comparatively lower performance, suggesting weaker feature representation and boundary preservation capability than the ResNet-based architectures evaluated in this study. It should be noted that ResNet-18 and ResNet-50 were implemented as backbone networks within the DeepLabV3+ semantic segmentation framework. Therefore, the observed performance improvement reflects the combined contribution of the DeepLabV3+ decoder and the enhanced feature extraction capability of the ResNet backbones. Furthermore, the use of pretrained weights enabled better feature generalization, contributing to more accurate PV module localization and boundary delineation.

The analysis reveals that ResNet-18 offers a practical balance between segmentation performance and computational efficiency, making it suitable for real-time UAV-based PV inspections. In contrast, ResNet-50 achieved the best overall segmentation performance, particularly in terms of accuracy and mean IoU, indicating superior PV module coverage and boundary localization. These observations are consistent with recent advancements in solar PV inspection research, where residual learning architectures have demonstrated strong feature extraction capabilities while maintaining reliable segmentation performance. Overall, the findings confirm that both ResNet-based architectures provide effective solutions for first-tier PV module segmentation, with ResNet-18 favouring computational efficiency and ResNet-50 prioritizing segmentation accuracy and boundary preservation.

4.2.2 Results of Image Enhancement using Non-Local Means with Contrast-Limited Adaptive Histogram Equalization Algorithm

In Stage 1, the next process after first-tier segmentation is image enhancement using the NLM with the CLAHE algorithm. This section presents the results and output using qualitative and quantitative evaluations. The qualitative is evaluated based on visual validation. Meanwhile, the quantitative evaluation used standard image quality metrics, including Mean Squared Error (MSE), Peak Signal-to-Noise Ratio (PSNR), and Signal-to-Noise Ratio (SNR). These evaluations were used to determine the level of improvement achieved using the NLM-CLAHE enhancement algorithm in Stage 1.

Figure 4.2 illustrates a side-by-side comparison of the original thermal images (Figure 4.2 (i)) and the enhanced images (Figure 4.2 (ii)) produced using NLM and CLAHE. The technique integrates NLM denoising with CLAHE. The enhanced images show a more precise thermal pattern and increased contrast. Based on the findings, this enhancement method enables the identification of previously obscured repetitive cell textures in PV modules. Furthermore, enhanced visibility of such features is clearly improved, which is essential for downstream processes, including segmentation, damage localization, and severity classification in PV inspection systems.

i) Thermal Image (Original)	ii)Enhanced Image

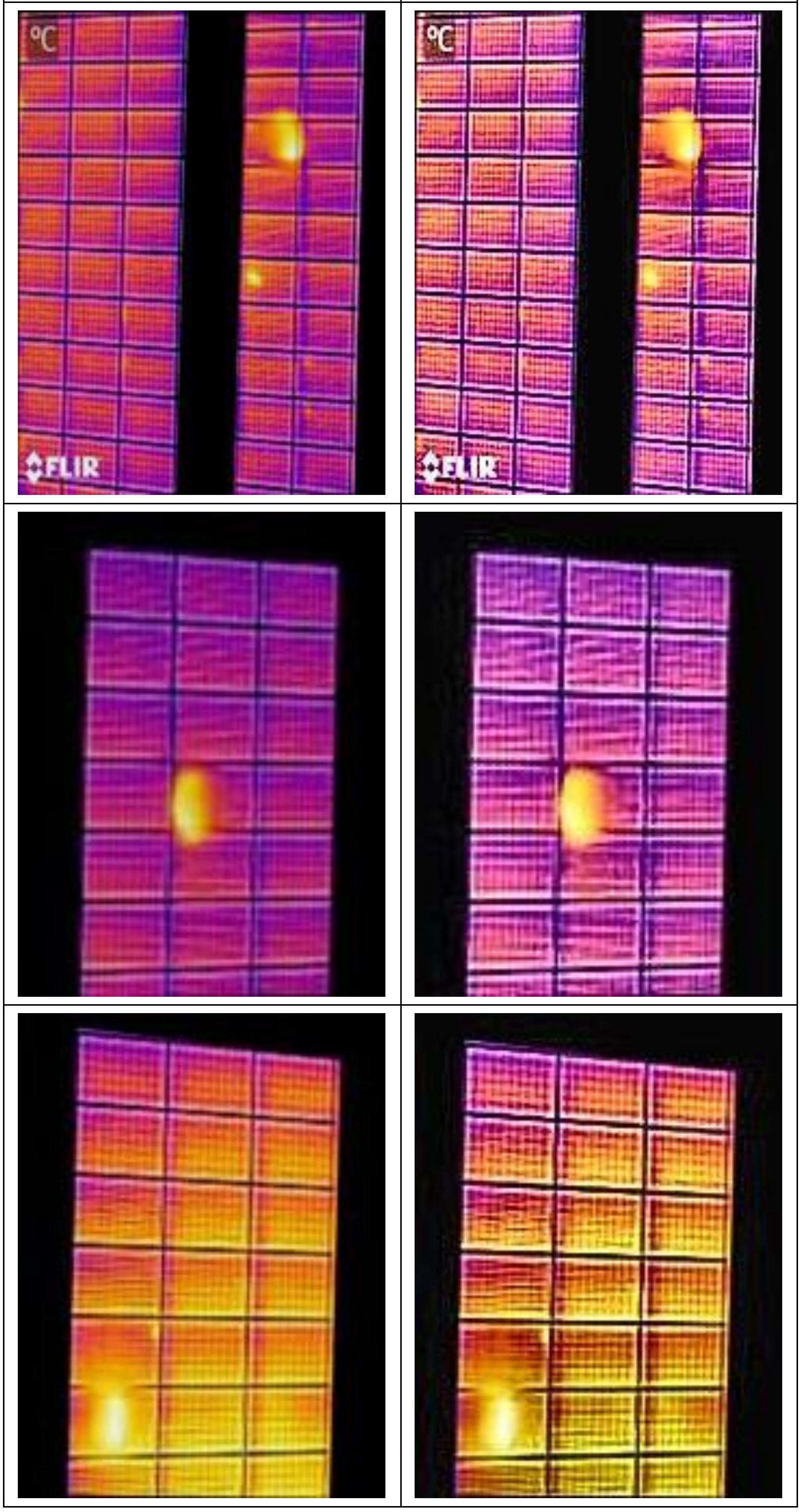


Figure 4.2 Qualitative Results Compared Between the Original (i) and Enhanced Image (ii) using the Proposed Hybrid NLM and CLAHE Method

Table 4.2 presents the quantitative comparison of image quality metrics for the NLM, CLAHE, and combined NLM-CLAHE methods. Based on the results, the NLM-CLAHE method achieves the best overall performance, with the lowest MSE (0.0002), indicating minimal pixel-level error between the enhanced and reference images. This is further supported by the highest PSNR value of 36.4690 dB, demonstrating superior reconstruction quality with low distortion. In contrast, the individual NLM and CLAHE methods produce higher MSE values (0.0138 and 0.0125, respectively) and lower PSNR values (18.7456 dB and 19.1084 dB), indicating comparatively greater pixel variation after enhancement. Similarly, the SNR values indicate that NLM-CLAHE (29.8610 dB) preserves signal strength better than NLM (12.1476 dB) and CLAHE (12.9820 dB). Overall, the results confirm that the combined NLM-CLAHE approach significantly improves image quality while preserving spatial and boundary characteristics, outperforming individual enhancement techniques. Furthermore, MSE, PSNR, and SNR were used as quantitative metrics to assess the preservation of thermal information after image enhancement. Although contrast enhancement modifies image intensity distributions, these metrics provide useful measures of structural preservation relative to the original thermal image, which served as the reference image in this study.

Table 4.2
Quantitative Comparison of Image Quality Metrics for NLM, CLAHE, and NLM-CLAHE Methods

Parameter	MSE	PSNR	SNR
NLM	0.0138	18.7456	12.1476
CLAHE	0.0125	19.1084	12.9820
NLM-CLAHE	**0.0002**	**36.4690**	**29.8610**

The findings confirm that the proposed hybrid NLM-CLAHE enhancement in Stage 1 effectively balances contrast amplification and noise reduction. Additionally, enhanced visibility improves the identification of module boundaries, cell textures, and localized hotspots. This improvement is crucial for enabling precise segmentation and reliable fault analysis in subsequent processing stages. More importantly, the PSNR and SNR results show that the enhancement procedure effectively preserves the images' essential thermal properties, and the overall patterns remain similar to those of the original thermal images. This is consistent with the work of [146] and [147], who noted that PSNR values exceeding 30 dB are generally sufficient to ensure reliable feature preservation for segmentation tasks.

In conclusion, the results of this section highlight the development of a reliable pre-processing stage that supports precise module segmentation and subsequent classification of hotspot severity.

4.2.3 Results of Second-tier Segmentation

This section presents the findings of the second-tier segmentation phase, which is used to detect hotspot regions in the solar PV modules. Figure 4.3 presents a qualitative comparison of the predicted hotspot regions with the corresponding ground-truth annotations for the three evaluated architectures. In each subfigure, the white region represents the correctly segmented hotspot area that overlaps with the ground truth mask, while the pink region denotes the predicted hotspot region. The black area corresponds to non-hotspot regions.

As shown in Figure 4.3(a), U-Net produces a predicted region with noticeable boundary irregularities and partial misalignment relative to the ground truth. The overlap area is comparatively smaller, indicating limitations in capturing fine-grained hotspot morphology. In contrast, ResNet-18 (Figure 4.3(b)) demonstrates improved spatial consistency, with a greater overlap between the predicted and ground truth regions and more stable boundary delineation.

ResNet-50 (Figure 4.3(c)) exhibits the highest degree of correspondence with the ground truth mask. The segmented hotspot region exhibits finer contours and reduced false segmentation, indicating superior feature representation and spatial discrimination. The increased overlap between predicted and reference regions reflects enhanced segmentation precision and structural preservation. Overall, the qualitative observations align with the quantitative performance metrics, confirming that deeper residual architectures improve hotspot boundary localization and robustness in second-tier segmentation.

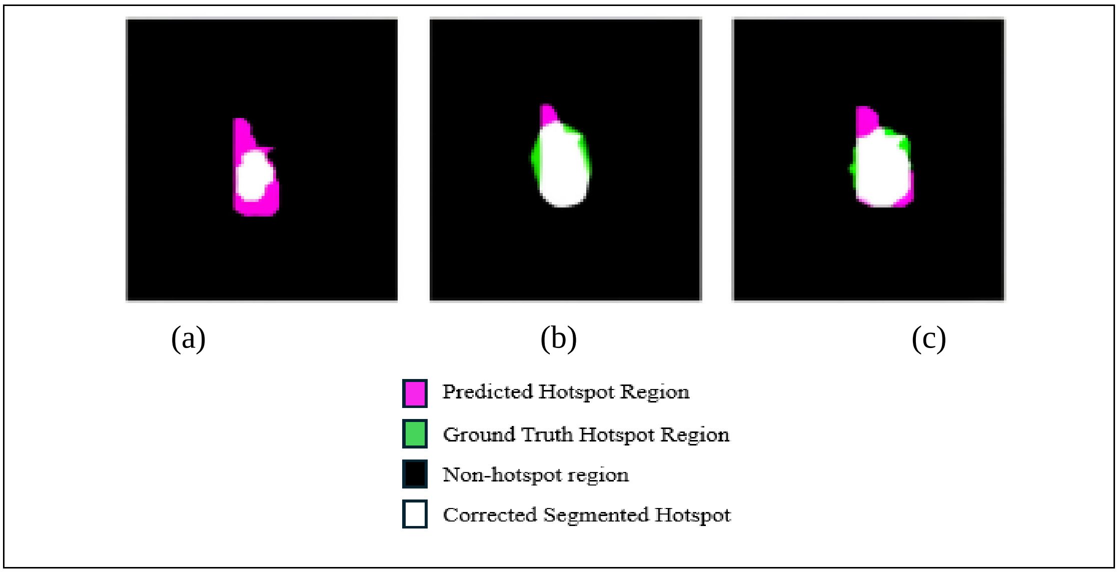


Figure 4.3 Sample of Result Segmented Hotspot Region using the Second-tier Segmentation for (a) U-Net (b) ResNet-18 (c) ResNet-50 Architecture

Next, the quantitative evaluation presents the results of second-tier segmentation with and without NLM-CLAHE enhancement, as shown in Tables 4.3 and 4.4. Without enhancement, ResNet-50 achieved the highest accuracy (85.56%) and mean IoU (76.98%), outperforming U-Net (64.78%, 63.83%) and ResNet-18 (72.45%, 55.57%). Following the application of the proposed NLM-CLAHE enhancement, all evaluated models demonstrated improved segmentation performance. ResNet-18 exhibited the largest performance gain, with accuracy increasing from 72.45% to 87.10% and mean IoU improving from 55.57% to 78.61%. Similarly, ResNet-50 achieved further improvements, with accuracy increasing from 85.56% to 87.82% and mean IoU from 76.98% to 78.95%, while maintaining the best overall segmentation performance. U-Net also benefited from the enhancement, although the improvement was comparatively modest, with accuracy increasing from 64.78% to 66.50% and mean IoU from 63.83% to 65.70%.

The results indicate that the proposed NLM-CLAHE enhancement improves thermal contrast and hotspot visibility, thereby facilitating more accurate hotspot localization. The substantial improvement observed for ResNet-18 suggests that the enhanced thermal information provided additional discriminative features that were effectively utilized by the residual learning architecture. Although ResNet-50 achieved the highest overall performance, the relative improvement was more pronounced for ResNet-18, indicating a greater sensitivity to the enhanced thermal features. These findings are consistent with previous studies [139], [140], which demonstrated that residual learning architectures are effective in improving feature representation and segmentation performance. Overall, the results confirm that the proposed NLM-CLAHE enhancement improves second-tier hotspot segmentation, particularly when integrated with ResNet-based architectures.

Table 4.3
The Second-tier Segmentation Result (Without NLM-CLAHE)

Semantic Model	Accuracy (%)	Mean IoU (%)	Precision (%)	Recall (%)
U-Net	64.78	63.83	66.68	65.45
ResNet-18	72.45	55.57	74.75	70.79
ResNet-50	**85.56**	**76.98**	**78.86**	**74.85**

Table 4.4
The Second-tier Segmentation Result (With NLM-CLAHE)

Semantic Model	Accuracy (%)	Mean IoU (%)	Precision (%)	Recall (%)
U-Net	66.50	65.70	67.69	66.98
ResNet-18	87.10	78.61	80.89	76.83
ResNet-50	**87.82**	**78.95**	**85.89**	**81.86**

Figures 4.4 to 4.7 collectively summarize the quantitative performance of the second-tier segmentation models with and without NLM-CLAHE enhancement. The evaluation metrics include accuracy, mean IoU, precision, and recall. The results demonstrate consistent performance improvements across all models after applying the proposed enhancement method. Among the evaluated architectures, ResNet-50 achieved the highest overall performance, recording an accuracy of 87.82%, a mean IoU of 78.95%, a precision of 85.89%, and a recall of 81.86% after enhancement. ResNet-18 exhibited the largest relative improvement, with accuracy increasing from 72.45% to 87.10% and mean IoU improving from 55.57% to 78.61%, indicating that the enhanced thermal contrast provided more discriminative hotspot features for segmentation. Although U-Net also benefited from the enhancement process, the performance gains were comparatively smaller, with accuracy increasing from 64.78% to 66.50% and mean IoU from 63.83% to 65.70%.

The improvement observed in the ResNet-based models suggests that the proposed NLM-CLAHE enhancement effectively increases hotspot visibility and boundary contrast, facilitating more accurate hotspot localization. Furthermore, the corresponding increases in precision and recall indicate reduced segmentation errors and improved identification of hotspot regions. These findings are consistent with previous studies [148], [149], which reported that residual learning architectures are effective in improving feature representation and segmentation performance in thermal image analysis. Overall, the results confirm that the proposed enhancement strategy improves second-tier hotspot segmentation performance, particularly when integrated with ResNet-based architectures.

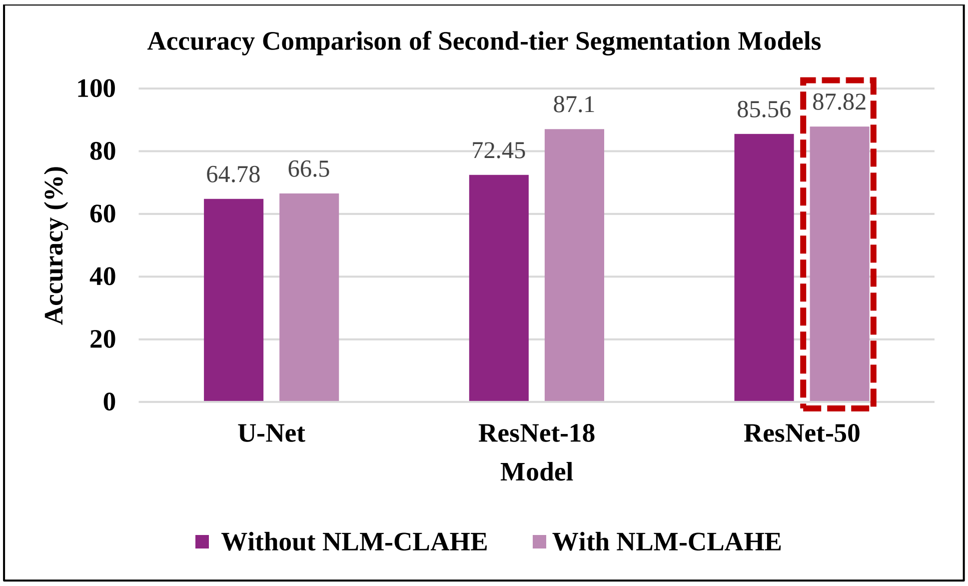


Figure 4.4 Accuracy Comparison of Second-tier Hotspot Segmentation Models with and without NLM-CLAHE Enhancement

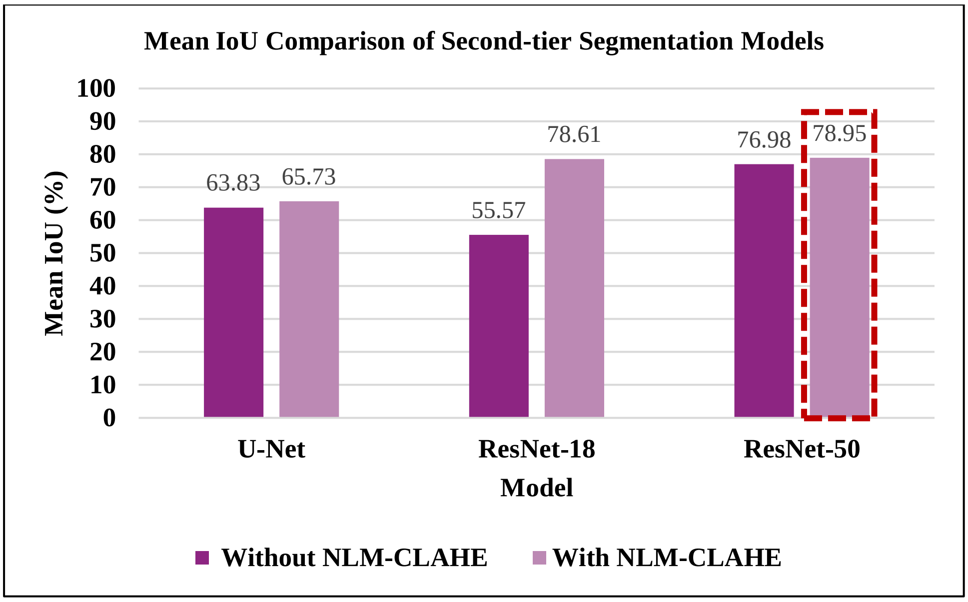


Figure 4.5 Mean IoU Comparison of Second-tier Hotspot Segmentation Models with and without NLM-CLAHE Enhancement

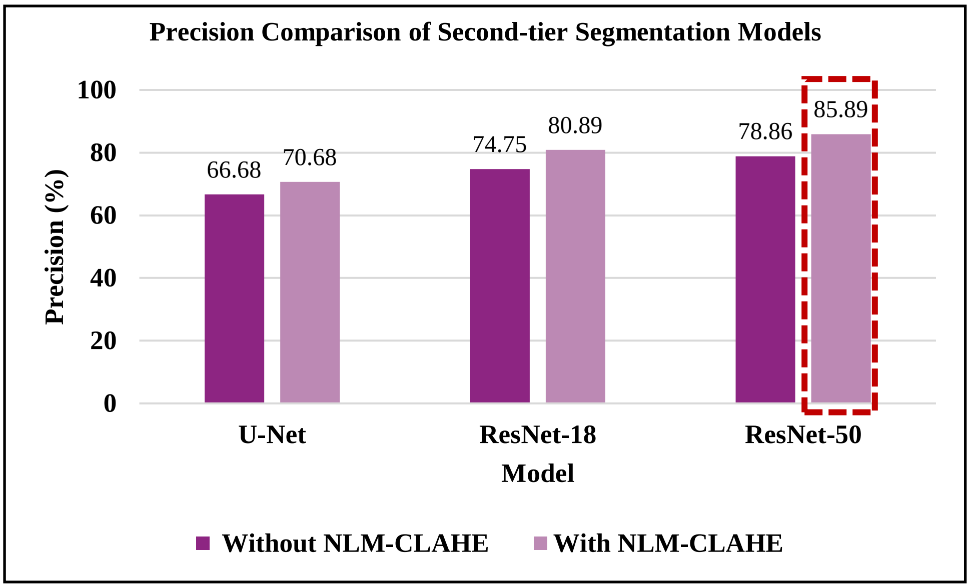


Figure 4.6 Precision Comparison of Second-tier Hotspot Segmentation Models with and without NLM-CLAHE Enhancement

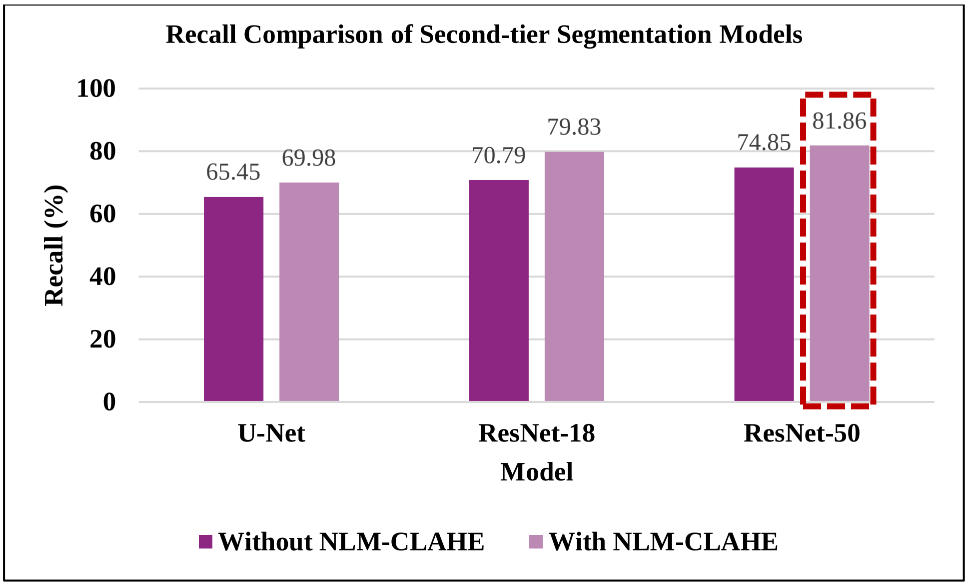


Figure 4.7 Recall Comparison of Second-tier Hotspot Segmentation Models with and without NLM-CLAHE Enhancement

Overall, the results confirm that the proposed two-tier segmentation framework incorporating NLM-CLAHE enhancement improves hotspot segmentation performance across all evaluated models. The enhanced thermal images provide clearer hotspot boundaries and improved feature contrast, contributing to more accurate hotspot localization. ResNet-50 achieved the best overall segmentation performance, while ResNet-18 demonstrated the greatest performance improvement following enhancement. These findings are consistent with previous studies [149] [150], which highlighted the effectiveness of contrast-adaptive pre-processing in supporting CNN-based thermal image analysis. Consequently, the proposed framework provides a reliable solution for hotspot localization and severity assessment in photovoltaic monitoring applications.

4.3 Results and Analysis of Stage 2: CNN Classification Model using Dual-input Features Extraction Technique

This section presents the results and analysis of a classification model using a feature extraction technique. Two distinct feature extraction approaches were investigated. The first approach employs a single-input strategy, in which image and temperature data are trained separately using CNN. The second approach combines both image and temperature features derived from thermal image characteristics trained in the CNN model.

To ensure a comprehensive evaluation, both feature extraction methods were assessed using balanced datasets. This comparison provides insight into how data distribution influences model performance, particularly in real-world PV monitoring scenarios where class imbalance is common. Through this evaluation, the analysis aims to determine the most effective feature extraction strategy for improving hotspot localization and severity discrimination in PV systems.

4.3.1 Results of CNN Classification Model using Single Input Feature Extraction (Image Input)

This subsection evaluates the CNN classification performance of U-Net, ResNet-18, and ResNet-50 using image-only features for hotspot severity classification. The analysis aims to determine whether visual image characteristics alone, without temperature information, provide sufficient discriminatory capability for severity recognition.

The results in Table 4.5 summarize the classification, while Figure 4.8 illustrates the classification performance of CNN models using a single image feature input across multiple evaluation metrics. Among the three CNN models, ResNet-18 achieved the highest accuracy of 76.53%, marginally outperforming ResNet-50 at 73.60% and considerably surpassing U-Net at 62.40%. However, ResNet-18 maintained superior precision (76.77%), recall (76.53%), and F1-score (76.51%), confirming balanced classification performance and reliable detection.

Table 4.5
The Result for Single-Feature Extraction is Based on Image Input

Metrics	U-Net (%)	ResNet-18 (%)	ResNet-50 (%)
Accuracy	62.40	76.53	73.60
Precision	62.36	76.77	73.60
Recall	62.40	76.53	73.60
F1-score	62.37	76.51	73.55

Meanwhile, Figure 4.8 shows that ResNet-18 achieves superior precision and recall compared to both U-Net and ResNet-50. These results demonstrate that ResNet-18 can derive reliable image-level representations. While ResNet-50 achieves competitive accuracy, its slightly lower performance compared to ResNet-18 indicates that increased model complexity does not always translate to superior generalization under the current dataset size. Observations were reported by [143], [145], who found that deeper CNN architectures require significantly larger datasets to achieve stable optimization. U-Net achieved lower precision and recall (≈62%) than the ResNet models during the classification stage.

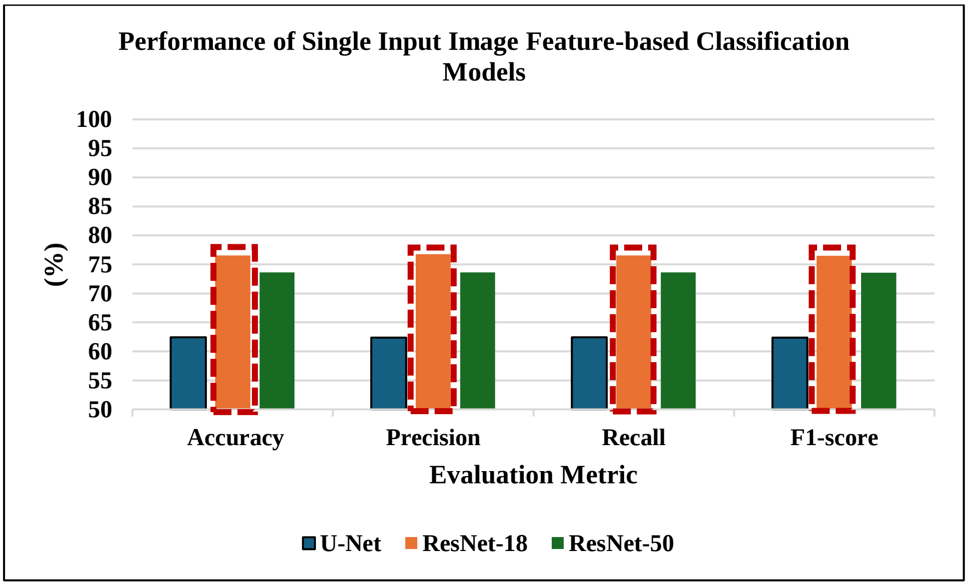


Figure 4.8 Classification of CNN Models Using Single-Image Features

The results convey three key implications. First, although image input alone achieves satisfactory classification accuracy, it remains limited in capturing the intrinsic thermal variations that define hotspot severity. This limitation underscores the need for complementary temperature-based information to enhance the classification robustness. This observation aligns with findings in [17] and [151], which emphasized the role of elevated module temperature and thermal anomaly patterns as reliable indicators for hotspot severity classification. Second, the consistent performance of ResNet-18 and ResNet-50 indicates that both architectures can generalize reasonably well under visual-only conditions when the dataset provides sufficient variability. The comparatively stronger stability of ResNet-18 further suggests that deeper architectures require proportionally larger datasets to fully leverage their representational capacity, consistent with [152], which reported that increasing network depth is associated with higher data requirements. In contrast, the comparatively weaker classification performance of U-Net highlights limitations of its encoder–decoder structure in capturing discriminative thermal patterns compared to the deeper residual models.

In summary, both ResNet-18 and ResNet-50 outperform U-Net when using only image features, with ResNet-18 demonstrating slightly superior discriminative performance across all evaluation metrics. These findings confirm the effectiveness of residual architectures in extracting meaningful spatial patterns from visual inputs. However, despite satisfactory performance, the results indicate limitations when image-based information is used in isolation for hotspot severity classification. This suggests that spatial features alone may not fully capture the thermal characteristics associated with PV defects, thereby motivating further exploration of complementary feature representations in subsequent analyses.

4.3.2 Results of CNN Classification Models Using Single Input Feature Extraction (Thermal Input)

This subsection evaluates the performance of U-Net, ResNet-18, and ResNet-50 in CNN classification when temperature values are used as the sole input feature. Table 4.6 presents quantitative results, while Figure 4.9 illustrates the classification performance of CNN models using a single thermal feature input across multiple evaluation metrics.

Table 4.6
The Result of CNN Classification Single-feature Extraction based on Thermal Input

Metrics	U-Net (%)	ResNet-18 (%)	ResNet-50
Accuracy	74.01	85.37	66.49
Precision	58.11	85.24	66.48
Recall	74.91	85.38	66.49
F1-score	57.56	85.25	66.43

From the result, ResNet-18 achieved the best overall performance with an accuracy of 85.37%. It also demonstrated the highest balance between F1-score (85.25%), recall (85.38%), and precision (85.24%). U-Net recorded a moderate accuracy of 74.01%, indicating reduced generalization capability. The relatively low precision (58.11%) and F1-score (57.56%) compared with the accuracy suggest that the model produced more false-positive predictions for certain severity classes. This indicates that although the U-Net identified a substantial proportion of the target samples, its class discrimination capability was less effective than the ResNet-based models. ResNet-50 showed comparatively weaker results, with an accuracy of 66.49 and precision, recall, and F1 scores of approximately 66%.

The superior performance of ResNet-18 is attributed to its residual skip connections and lightweight design, which enhance gradient stability and feature preservation during training. Despite having fewer layers than ResNet-50, ResNet-18 effectively captures discriminative thermal patterns while avoiding overfitting. The comparatively lower performance of U-Net indicates limitations in capturing discriminative temperature patterns from the thermal feature input. This behaviour is reflected by the lower precision and F1-score values, suggesting greater confusion among hotspot severity classes compared with the ResNet-based architectures. While deeper architecture generally offers greater representational capacity, the present results indicate that performance gains depend on the balance between model complexity and available training data. This observation is consistent with studies in [71], [152], [153], which reported that deeper CNN architectures require larger datasets for convergence.

Figure 4.9 confirms the dominance of ResNet-18 across all metrics, particularly in validation accuracy and F1-score, reflecting strong generalization capability. U-Net's relatively high accuracy. Meanwhile, ResNet-50 exhibits lower overall performance across metrics, suggesting that its deeper architecture requires stronger data augmentation or transfer learning to improve stability.

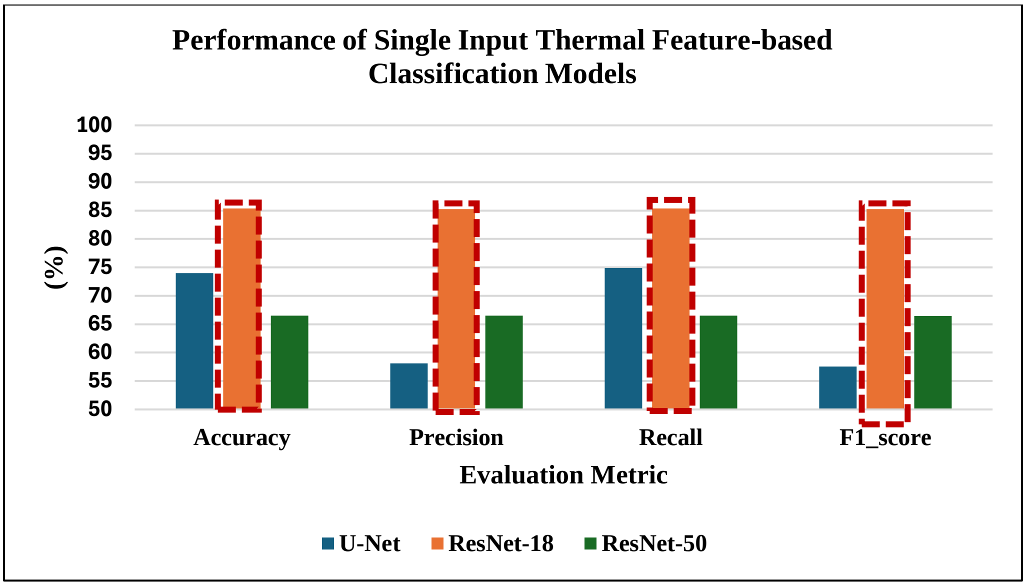


Figure 4.9 Classification of CNN Models Using Single-Thermal Features

These findings highlight several significant consequences. The strong performance of ResNet-18 demonstrates that temperature-based features alone can reliably distinguish between hotspot severity levels in PV systems. In line with [17], this observation suggests that temperature differential is a useful indicator of hotspot behaviour. However, the reliability of thermal measurements may be influenced by irradiance, ambient temperature, emissivity, and wind conditions. Consequently, temperature-based features should be complemented with additional image-derived information to improve classification robustness.

Secondly, the weak performance of U-Net and ResNet-50 indicates that the model's complexity must be addressed with the dataset's size and feature richness. The study in [143] emphasizes that deeper networks require larger and more diverse datasets to avoid instability and redundancy. The ResNet-18 architecture satisfies the scaling principles outlined in [83] by providing a balanced trade-off between computational efficiency and representational capacity. Finally, relying solely on temperature inputs can make the system exposed to fluctuations in irradiance and ambient conditions, as [31] highlights. This reinforces the importance of integrating hybrid feature sets to achieve robust and more robust severity classification.

In summary, the results demonstrate that ResNet-18 achieves the best performance with temperature-only input, consistently attaining high accuracy, precision, recall, and F1 scores. The findings highlight the strong discriminative power of thermal features for identifying hotspot severity levels, particularly when temperature variation is the dominant distinguishing factor. Compared to image-only input, thermal data provides more direct defect-related information, thereby improving classification robustness. Nevertheless, while thermal features exhibit strong predictive capability, relying solely on a single modality may still limit the model's overall representational capacity. This observation supports the need to investigate a dual-input feature integration strategy to further enhance classification performance.

4.3.3 Results of Dual-Input Feature Extraction (Thermal and Image)

This subsection presents the classification performance of U-Net, ResNet-18, and ResNet-50 on dual-input features combining temperature and image data. The experiment aimed to assess whether the fusion of complementary modalities enhances model robustness, accuracy, and generalization beyond single-input configurations. The results are summarized in Table 4.7. At the same time, Figure 4.10 presents the classification performance of CNN models using dual-input across multiple evaluation metrics.

The results show that ResNet-18 achieved the highest overall performance, with an accuracy of 98.67%. It also achieved exceptionally consistent precision, recall, and F1-score values of 98.67%, indicating highly reliable classification performance across all evaluation metrics. The reported performance was obtained using independent training, validation, and testing datasets as described in Chapter 3. The testing dataset was not used during model training or parameter optimization, thereby providing an unbiased evaluation of classification performance. Similarly, ResNet-50 demonstrated strong performance, achieving an accuracy of 96.00%, precision of 96.08%, recall of 96.00%, and an F1-score of 96.01%. In contrast, U-Net performed poorly, achieving only 33.33% accuracy, with precision, recall, and F1-score values of 11.11%, 33.33%, and 16.67%, respectively. This result suggests that the implemented U-Net configuration was less effective in learning discriminative multimodal feature representations than the ResNet-based architectures. The poor performance may be attributed to limitations in integrating complementary thermal image and temperature features within the adopted classification framework, resulting in reduced class discrimination capability.

Table 4.7
The Result of Dual-input Feature Extraction

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

As shown in Figure 4.10, the dual-input feature extraction approach consistently improved the classification performance of the ResNet-based models across all evaluation metrics. The most notable improvement was achieved by ResNet-18, with classification accuracy increasing from 85.37% to 98.67%, a gain of 13.30%. Similar improvements were observed for precision, recall, and F1-score. These findings suggest that thermal images and temperature descriptors provide complementary information for hotspot severity classification. While temperature-based features capture the thermal characteristics and severity-related temperature variations of hotspot regions, image-based features contribute additional spatial and contextual information that facilitates better discrimination between severity levels. Consequently, integrating both modalities yields a more informative and discriminative feature representation than either input alone.

ResNet-50 also benefited from the dual-input feature representation, achieving of 96.00% accuracy and maintaining consistently high precision, recall, and F1-score. In contrast, U-Net recorded substantially lower performance across all evaluation metrics, suggesting that the implemented U-Net configuration was less effective in learning discriminative multimodal feature representations than the ResNet-based architectures. The poor performance may be attributed to limitations in integrating complementary thermal image and temperature features within the adopted classification framework, resulting in reduced class discrimination capability. These observations are consistent with the findings reported in [154] and [155], which demonstrated that multimodal feature fusion enhances feature representation and improves classification performance by leveraging complementary information from multiple input modalities. Overall, the results demonstrate the effectiveness of the proposed dual-input feature extraction strategy, with ResNet-18 achieving the highest overall performance among the evaluated models.

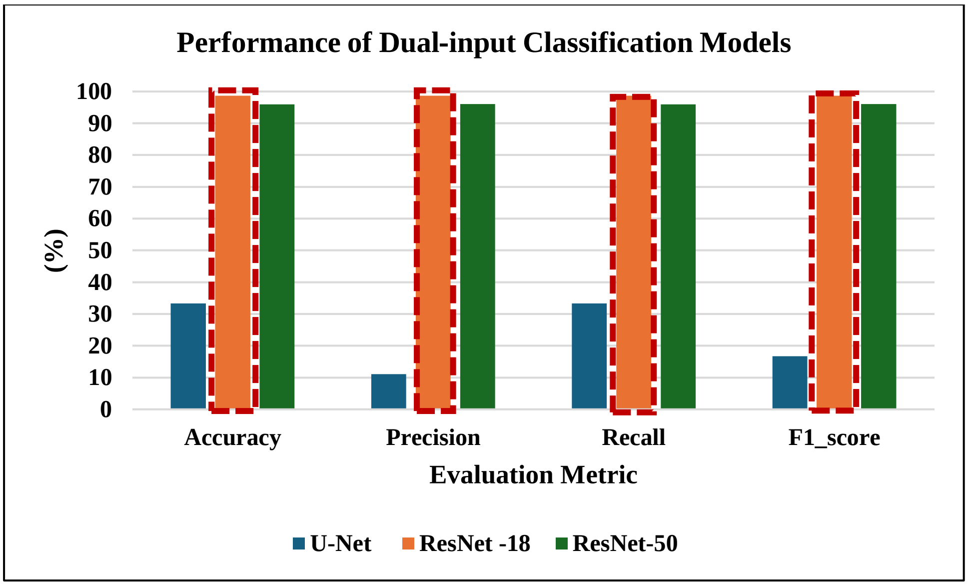


Figure 4.10 Classification of CNN Models Using Dual-input Features

The findings support several essential conclusions. First, the accuracy and robustness are greatly improved by combining temperature and detailed image features. This observed improvement validates the finding in [154], [156] that multimodal fusion improves decision-making capacities in hotspot severity classification. Second, the strong performance of ResNet-18 demonstrates that shallower residual architectures provide a positive balance between computational efficiency and classification accuracy. This efficiency is achieved without compromising discriminative capability when hybrid features are utilized. Third, U-Net showed lower classification performance in this experimental setup, reflecting architectural differences in handling image-level features.

Overall, ResNet-18 achieved the best balance between classification accuracy and computational efficiency under the dual-input configuration. Therefore, ResNet-18 was selected for detailed confusion matrix analysis as it demonstrated the most balanced and stable performance among all evaluated architectures in the dual-input configuration. Figures 4.11-4.13 present the confusion matrices for ResNet-18 with single-image, single-thermal, and dual-input configurations, respectively.

Figure 4.11 presents the confusion matrix for the ResNet-18 model using image-only input. The model demonstrates reasonable classification performance, particularly for the medium class, which records the highest number of correct predictions. However, noticeable misclassification occurs between the low and high-severity levels, indicating overlapping visual characteristics between the two categories. This pattern suggests that spatial image features alone may not adequately capture the thermal intensity variations required for precise discrimination of hotspot severity. The findings highlight the inherent limitation of relying solely on visual information for robust classification performance.

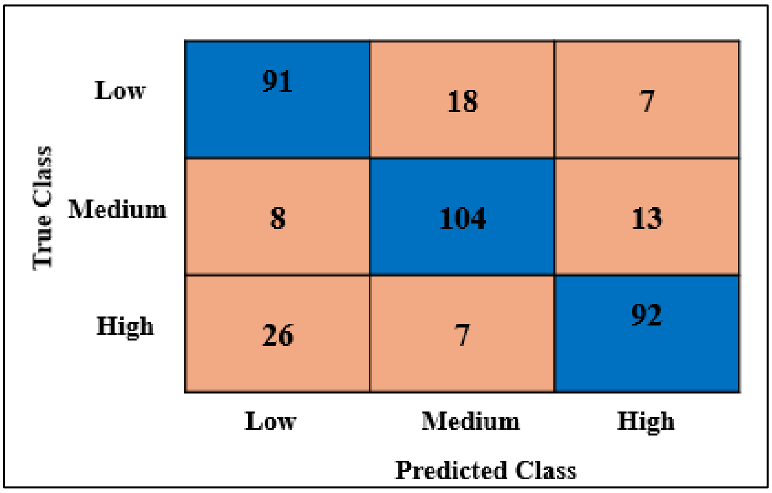


Figure 4.11 Confusion Matrix of Hotspot Severity Classification Using the ResNet-18 Architecture for Single Input Feature Extraction (Image Input)

Figure 4.12 illustrates the confusion matrix for the ResNet-18 model using thermal-only input. Compared to image-based input, the model achieves improved classification consistency across all severity levels, particularly in distinguishing the medium class with minimal misclassification. Although some confusion remains between the low and high categories, the overall prediction accuracy is more balanced and stable. These results indicate that thermal features provide stronger discriminative cues for assessing hotspot severity. Nevertheless, minor classification overlaps suggest that integrating complementary spatial information may further enhance model robustness.

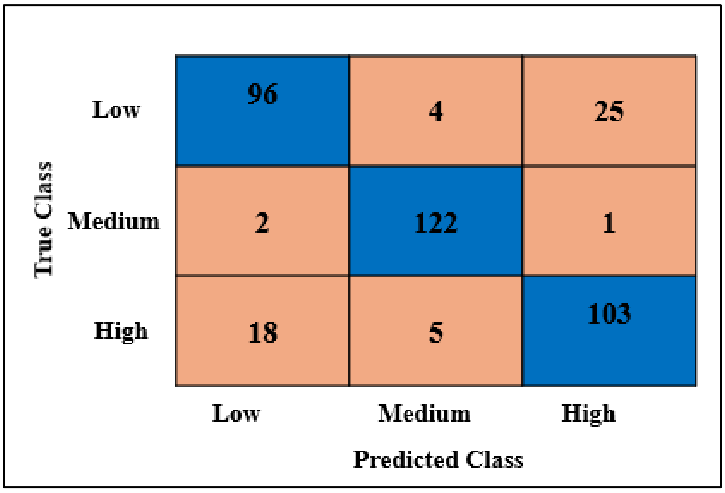


Figure 4.12 Confusion Matrix of Hotspot Severity Classification Using the ResNet-18 Architecture for Single-input Feature Extraction (Thermal Input)

Figure 4.13 demonstrates the confusion matrix for dual-input feature extraction. The classification results show near-perfect performance, with minimal misclassification across all severity levels. This confirms that integrating thermal and visual image features significantly enhances discriminative capability and class separability.

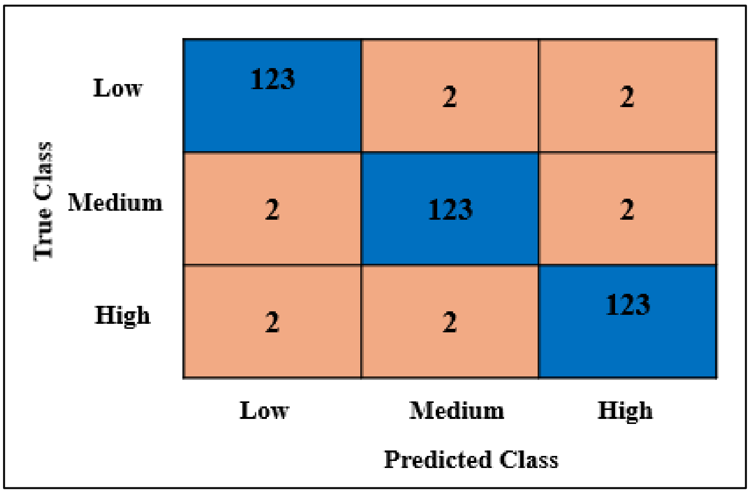


Figure 4.13 Confusion Matrix of Hotspot Severity Classification Using the ResNet-18 Architecture for Dual-input Feature Extraction

In conclusion, dual-input feature extraction offers a substantial advantage in PV hotspot severity analysis by merging image and temperature features. ResNet-18 is identified as the most reliable and efficient model, achieving near-perfect accuracy and consistency across all metrics. The results demonstrate that dual-input fusion significantly improves hotspot assessment performance compared with single-input approaches, further validating the effectiveness of hybrid DL frameworks for PV fault detection and severity classification.

4.3.4 k-fold Cross-validation for Model Generalization

To further assess the robustness and generalization capability of the proposed models, a k-fold cross-validation procedure was implemented. In this study, a 5-fold cross-validation scheme was adopted, focusing on the ResNet-18 architecture, which had demonstrated superior performance in prior experiments. Tables 4.8-4.10 present the cross-validation performance for U-Net, ResNet-18, and ResNet-50.

Table 4.8 presents the cross-validation performance of ResNet-18. Accuracy, precision, recall, and F1-score remain consistently high across all folds, with accuracy ranging from 98.21% to 99.40% and an average of 99.00%. Precision and recall both averaged 99.01% and 99.00%, respectively, confirming the model's stability in detecting hotspot regions with minimal false positives or false negatives. The F1-score also remained constant at 99.00%, signifying well-balanced precision–recall trade-offs.

Table 4.8
k-folds Cross-validation Process for ResNet-18, k=5

Fold	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
1	98.60	98.62	98.60	98.61
2	98.21	98.21	98.21	98.21
3	99.40	99.41	99.40	99.40
4	99.40	99.41	99.40	99.40
5	99.40	99.40	99.40	99.40
Average	**99.00**	**99.01**	**99.00**	**99.00**
Standard Deviation	**0.56**	**0.56**	**0.56**	**0.56**

Table 4.9 summarises the results for ResNet-50, which achieved slightly lower but still competitive performance, with an average accuracy of 97.17 % and precision, recall, and F1-score values of 97.19%, 97.17%, and 97.17%, respectively. Minor

fluctuations between 95.82% and 98.41% indicate excellent consistency and robustness, despite its deeper network configuration.

Table 4.9
k-folds Cross-validation Process for ResNet-50, k=5

Fold	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
1	96.21	96.23	96.21	96.21
2	98.41	98.41	98.41	98.41
3	98.21	98.23	98.21	98.21
4	95.82	95.82	95.81	95.82
5	97.21	97.25	97.21	97.21
Average	**97.17**	**97.19**	**97.17**	**97.17**
Standard Deviation	**1.16**	**1.16**	**1.16**	**1.16**

In contrast, Table 4.10 shows that U-Net recorded highly inconsistent results, with accuracies varying from 33% to 89% and an overall average of 44.53%. Precision (26.86%), recall (44.47%), and F1-score (31.17%) were significantly lower, confirming limited generalization capability under dual-input conditions.

Table 4.10
k-folds Cross-validation Process for U-Net, k=5

Fold	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-score (%)**
1	33.33	11.11	33.33	16.67
2	33.47	11.16	33.33	16.72
3	89.04	89.77	89.04	89.09
4	33.47	11.16	33.33	16.72
5	33.33	11.11	33.33	16.67
Average	**44.53**	**26.86**	**44.47**	**31.17**
Standard Deviation	**24.88**	**35.17**	**24.91**	**32.38**

The comparison of all three models demonstrates a clear performance hierarchy. ResNet-18 achieved the most stable and accurate outcomes, followed closely by ResNet-50, while U-Net underperformed severely. The minimal variation across ResNet folds indicates strong model regularization and balanced learning between training and validation sets, consistent with the trends reported in [152] and [157], which showed that residual architectures maintain high reliability across repeated validation cycles.

Conversely, U-Net's inconsistent results highlight its architectural limitations in classification tasks requiring multimodal fusion, reaffirming observations in [158] that encoder–decoder networks optimized for segmentation fail to retain stability under non-pixel-level learning. The weak generalization of U-Net suggests that its spatial feature encoder cannot effectively align dual-input modalities, leading to high variance and unreliable predictions.

The standard deviation values further support the stability analysis of the evaluated models. ResNet-18 achieved the lowest standard deviation across all evaluation metrics (0.56%), indicating highly consistent performance across the five folds. ResNet-50 also demonstrated good stability, with standard deviation values of approximately 1.16%. In contrast, U-Net recorded substantially higher standard deviation values, particularly for precision (35.17%) and F1-score (32.38%), confirming unstable classification behaviour and limited generalization capability under the dual-input configuration.

Overall, the results establish that residual networks, particularly ResNet-18, offer superior stability, balanced precision and recall behaviour, and high predictive reliability across multiple data folds. These findings confirm ResNet-18 as the optimal model and validate the generalization strength of the proposed hybrid DL framework for PV hotspot localization and severity classification.

4.3.5 Comparison and Analysis of CNN Classification Results Using Single and Dual-Input Features

A comparative analysis across all input configurations and model architecture provides a comprehensive overview of their relative effectiveness. Table 4.11 summarizes the classification performance of ResNet-18 across single thermal, single image, and dual-input feature configurations. Under all evaluated configurations, ResNet-18 consistently outperforms both ResNet-50 and U-Net, demonstrating superior versatility and computational efficiency.

Table 4.11
Performance Comparison of ResNet-18 Using Single and Dual-input Feature Configurations

Evaluation Metric (%)	Single Input		Dual Input
	Image -based	Thermal-based	
Accuracy	76.53	85.37	98.67
Precision	76.77	85.24	98.67
Recall	76.53	85.38	98.67
F1-score	76.51	85.25	98.67

In single-input experiments, ResNet-18 demonstrates stable and balanced performance across all evaluation metrics. ResNet-18 achieved balanced, stable performance, with F1-scores of 76.51% for the image input and 85.25% for the thermal input. This difference indicates that temperature information provides stronger discriminative capability for hotspot severity classification than visual features alone.

A substantial performance improvement is observed when dual-input features are employed. The dual-input configuration demonstrated substantial improvement across all evaluation metrics. Compared with the thermal-based configuration, the proposed dual-input feature extraction approach improved classification accuracy by 13.30%, precision by 13.43%, recall by 13.29%, and F1-score by 13.42%. When compared with the image-based configuration, the improvements were even more pronounced, with gains of 22.14%, 21.90%, 22.14%, and 22.16% in accuracy, precision, recall, and F1-score, respectively. These results quantitatively demonstrate the effectiveness of integrating thermal and image information for hotspot severity classification.

These findings confirm that ResNet-18 effectively leverages feature fusion to improve classification robustness without increasing architectural complexity. The results further indicate that compact CNN architectures can achieve high performance when supported by well-designed multi-modal input representations. This characteristic is particularly advantageous for scalable, real-time PV fault-monitoring systems, where both accuracy and computational efficiency are critical.

Figure 4.14 compares the proposed dual-feature fusion approach (ResNet-18 + dual-input) with state-of-the-art hybrid feature extraction techniques. The proposed method achieved a testing accuracy of 98.67%, outperforming hybrid local features with infrared and clustering (96.80%) [159], hybrid RGB-texture-HOG-LBP with SVM (92.00%) [160], and global plus local features using binary Harris Hawks Optimization with SVM (98.41%) [161]. The superior accuracy demonstrates that integrating DL-based feature extraction with hybrid input modalities offers a more robust, automated alternative to traditional HC features.

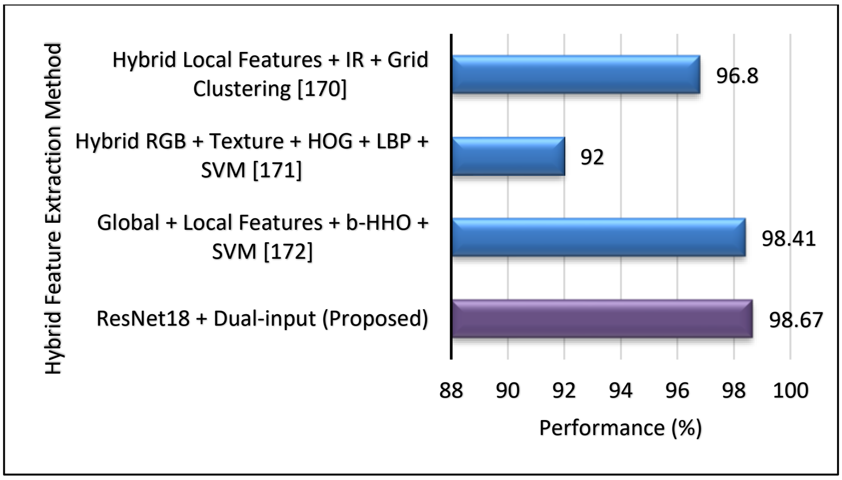


Figure 4.14 Accuracy Comparison of the Proposed Hybrid Feature Fusion Method (ResNet18 + Dual-input) with Selected State-of-the-art Hybrid Feature Extraction Approaches for Hotspot Detection in Solar PV Systems

This comparison serves as a contextual benchmark rather than a direct performance ranking. The reported results were obtained from different datasets, experimental settings, and evaluation protocols. As a result, differences in reported accuracy may be influenced by variations in dataset characteristics and experimental procedures. Nevertheless, the comparison provides a useful indication of the competitiveness of the proposed dual-input feature fusion framework relative to existing hybrid feature extraction approaches reported in the literature.

Within the context of the evaluated dataset and experimental protocol, the observed improvement can be attributed to the ability of the deep network to automatically extract high-level discriminative representations from fused thermal and visual domains, consistent with findings in [162] and [163]. These results indicate that the hybrid fusion process enhances both feature complementarity and model generalization, contributing to improved performance under varied operational conditions.

The outcomes validate the potential of hybrid feature fusion strategies in enhancing PV hotspot severity detection. ResNet-18 offers the best balance between model depth, computational efficiency, and classification accuracy, making it highly suitable for large-scale deployment in PV monitoring systems. The results further demonstrate that integrating temperature and image-based features improve feature richness and noise robustness, enabling reliable hotspot localization and severity classification.

In summary, the proposed hybrid ResNet-18 framework demonstrates competitive performance compared with selected hybrid feature extraction approaches reported in the literature. The findings highlight the potential of multimodal feature fusion to improve hotspot severity classification and support the integration of complementary thermal and image data in PV monitoring applications.

4.4 Result and Analysis of Stage 3: Enhanced CNN Classification Models

This section presents the outcomes and analysis of the CNN-based classification models utilized for hotspot severity classification in PV systems. The evaluation focuses on quantitative performance metrics derived from the classification outputs of different DL architectures. This step evaluates multiple performance metrics, including accuracy, precision, recall, and F1-score, to clarify the strengths and weaknesses of each model in detecting thermal abnormalities of varying severity. A comprehensive comparison is performed to provide insight into the architectural performance, computational efficiency, and layer-specific behaviour of the utilized models.

4.4.1 Comparison Results of CNN Layers Depth

This subsection presents the results of enhanced CNN models namely U-Net, ResNet-18 and ResNet-50. All models are trained and tested on thermal image datasets, with enhanced layers added to the model architecture. Each architecture is evaluated based on its ability to accurately classify hotspot severity categories when CNN layers are added to the model. The results are analyzed with reference to the confusion matrix output to determine which model offers the best trade-off in performance.

The comparative results across architectures and CNN layer depths provide important insights into the classification performance of the tested models. Tables 4.12–4.14 illustrate the overall performance of ResNet-18, ResNet-50, and U-Net with one, two, and three additional CNN layers. When comparing architectures, both ResNet-18 and ResNet-50 consistently outperformed U-Net in all metrics. ResNet-18 demonstrated the strongest performance, achieving up to 99.47% accuracy in the three-layer configuration, while ResNet-50 achieved slightly lower results at 99.20%. U-Net, by contrast, maintained 33.3% accuracy across all layer configurations, indicating that the implemented architecture was unable to effectively learn discriminative representations for hotspot severity classification.

Table 4.12
Performance of a CNN with an Additional One (1) Convolutional Layer

Metric	Accuracy	Precision	Recall	F1-score
U-Net (%)	33.33	11.11	33.33	16.67
ResNet-18 (%)	97.87	97.88	97.87	97.87
ResNet-50 (%)	96.53	96.56	96.53	96.52

Table 4.13
Performance of a CNN with an Additional Two (2) Convolutional Layers

Metric	Accuracy	Precision	Recall	F1-score
U-Net (%)	33.33	11.11	33.33	16.67
ResNet-18 (%)	98.13	98.17	98.13	98.14
ResNet-50 (%)	98.93	98.93	98.93	98.93

Table 4.14
Performance of a CNN with an Additional Three (3) Convolutional Layer

Metric	Accuracy	Precision	Recall	F1-score
U-Net (%)	33.33	11.11	33.33	16.67
ResNet-18 (%)	99.47	99.48	99.47	99.47
ResNet-50 (%)	99.20	99.21	99.20	99.20

When comparing CNN depth, both ResNet architectures showed notable improvements as the number of layers increased. ResNet-50 exhibited a sharper performance gain from one to two layers, while ResNet-18 displayed steady improvements, ultimately surpassing ResNet-50 in the three-layer configuration. These findings demonstrate that performance improvements are associated with the incremental addition of convolutional layers in the ResNet-based models, rather than CNN depth in general. U-Net did not show similar benefits, highlighting task-specific architectural differences.

The consistently low performance of U-Net across all layer configurations suggests that the model may have converged on a dominant-class prediction strategy during training. This behaviour is reflected by the unchanged accuracy (33.33%), precision (11.11%), recall (33.33%), and F1-score (16.67%) values despite the addition of convolutional layers. The results indicate that the implemented U-Net configuration was unable to effectively learn discriminative multimodal feature representations for hotspot severity classification. Consequently, increasing network depth alone was insufficient to improve classification performance. The limitation appears to be related to feature representation and multimodal feature integration rather than network depth.

Overall, the comparative analyses across architecture and CNN layer depths confirm that ResNet-18 with three additional CNN layers offers the most effective and robust configuration for hotspot severity classification in PV systems.

4.4.2 Results of Hotspot Severity Classification

This section presents a detailed analysis of hotspot severity classification results based on confusion matrices. The analysis focuses on class-wise prediction behavior and misclassification patterns across different CNN architectures. Figures 4.15–4.17 illustrate the confusion matrices obtained for hotspot severity classification using U-Net, ResNet-18, and ResNet-50 architectures, respectively.

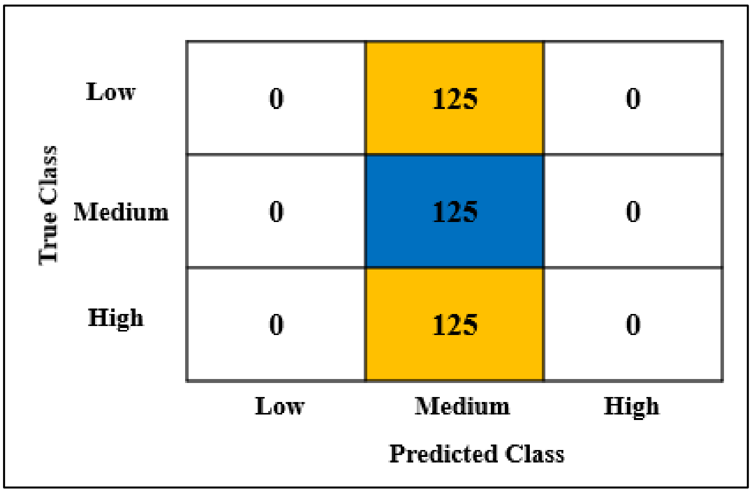


Figure 4.15 Confusion Matrix of Hotspot Severity Classification Using the U-Net Architecture

As shown in Figure 4.15, the U-Net model exhibits substantial misclassification across the three severity categories. The confusion matrix indicates that predictions are predominantly concentrated in the medium class, with low and high-severity samples largely misclassified. This pattern indicates limited class separability and suggests that the U-Net struggles to distinguish fine-grained severity within differences in thermal hotspot data. The dispersion of off-diagonal errors further reflects reduced discriminative capability for multi-class severity classification.

The quantitative performance metrics presented in Table 4.15 reinforce these observations. While the model achieves 100% recall for the medium class, it fails to correctly identify low and high-severity samples, resulting in zero precision and recall for these categories. Consequently, the overall F1-score remains low, and the overall classification accuracy is 33.33%. Given that the dataset is balanced across the three classes, this accuracy corresponds to random-level performance. These findings confirm that although U-Net is effective for pixel-level segmentation tasks, it is not well-suited to multi-class hotspot severity classification without additional feature refinement.

Table 4.15
Precision, Recall and F1-score of U-Net for Three-Class Hotspot Severity Classification

Metric	Low	Medium	High	Overall
Precision (%)	0.00	33.33	0.00	11.11
Recall (%)	0.00	100.00	0.00	33.33
F1-score (%)	0.00	50.00	0.00	16.67
Accuracy (%)	-	-	-	33.33

In contrast, the ResNet-18 model demonstrates a highly concentrated diagonal pattern in the confusion matrix, indicating strong class separability across all severity levels. As shown in Figure 4.16, almost all samples are correctly classified, with only two medium-severity cases misclassified as high severity. Importantly, no low or high severity samples are incorrectly predicted as normal or lower-risk categories.

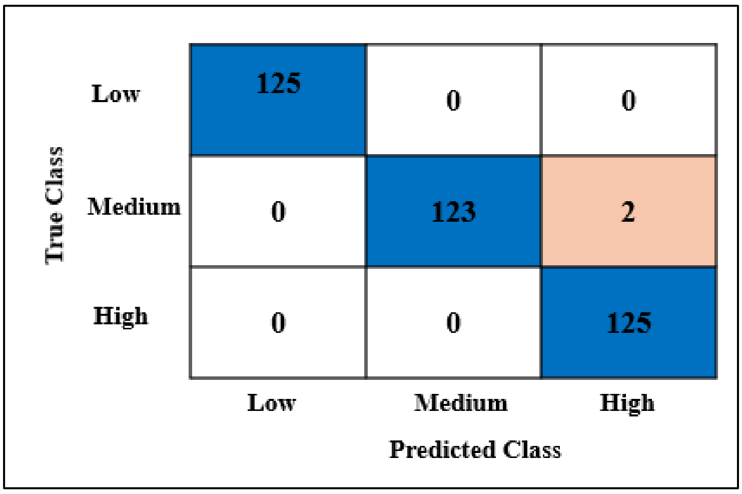


Figure 4.16 Confusion Matrix of Hotspot Severity Classification Using the ResNet-18 Architecture

Table 4.16 further confirms the strong classification performance of ResNet-18, achieving an overall accuracy of 99.47%. Precision, recall, and F1-score values remain above 98% across all hotspot severity classes, indicating highly reliable classification performance. The slightly lower precision for the high-severity class (98.43%) is attributed to two medium-severity samples being classified as high severity, as observed in Figure 4.16. Nevertheless, the overall classification results demonstrate excellent class separability and robust discrimination of severity.

Table 4.16
Precision, Recall and F1-score of ResNet-18 for Three-Class Hotspot Severity Classification

Metric	Low	Medium	High	Overall
Precision (%)	98.43	100.00	98.43	98.95
Recall (%)	100.00	98.40	100.00	99.47
F1-score (%)	99.21	99.19	99.21	99.20
Accuracy (%)	-	-	-	99.47

These results demonstrate that ResNet-18 effectively captures discriminative thermal patterns and severity-related feature representations. The residual learning mechanism likely contributes to this robustness by facilitating stable gradient propagation and deeper feature abstraction, enabling reliable classification across varying hotspot intensities. The consistency between precision and recall values suggests a balanced classification behaviour without bias toward specific severity categories.

Although ResNet-50 achieves high overall classification accuracy, its confusion matrix in Figure 4.17 shows only a few misclassifications across adjacent severity classes. Specifically, minor confusion is observed between low and medium categories, indicating slight overlap in learned feature representations.

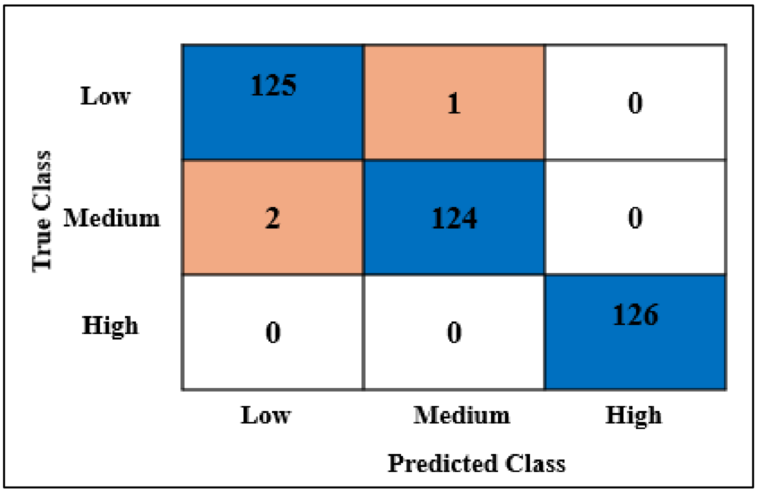


Figure 4.17 Confusion Matrix of Hotspot Severity Classification Using the ResNet-50 Architecture

These errors may be attributed to the increased architectural complexity of ResNet-50, which, while enhancing representational capacity, may introduce reduced class-wise stability when trained on moderately sized datasets. This observation suggests that deeper architectures do not necessarily guarantee superior classification robustness, particularly when severity discrimination is primarily driven by differences in temperature intensity rather than highly complex spatial patterns.

As summarized in Table 4.17, ResNet-50 achieves strong performance across all evaluation metrics, with an overall accuracy of 99.20%. Class-wise precision and recall values remain above 97% for all severity levels, demonstrating stable discriminative capability. The overall precision, recall, and F1-score values were calculated as the average of the corresponding metrics across the three hotspot severity classes. However, compared with ResNet-18, the marginal performance gains indicate that increasing network depth does not yield substantial improvements for this task. This reinforces the suitability of moderately deep residual networks for hotspot severity classification in thermal PV imagery.

Table 4.17
Precision, Recall and F1-score of ResNet-50 for Three-Class Hotspot Severity Classification

Metric	**Low**	**Medium**	**High**	**Overall**
Precision (%)	99.21	100.00	98.43	99.21
Recall (%)	100.00	97.60	100.00	99.20
F1-score (%)	99.60	98.79	99.21	99.20
Accuracy (%)	-	-	-	99.20

Overall, the comparative analysis shows that residual-based architectures significantly outperform U-Net for hotspot severity classification. While U-Net exhibits limited class separability and substantial misclassification across adjacent severity levels, both ResNet-18 and ResNet-50 achieve highly concentrated diagonal structures in their confusion matrices, indicating strong discriminative capability. Among the evaluated models, ResNet-18 delivers the best overall performance, with superior class balance and minimal misclassification, and slightly outperforms the deeper ResNet-50 architecture. These findings suggest that moderate-depth residual networks provide an optimal balance between representational capacity and generalization for thermal-based hotspot severity classification. The results further confirm that increasing architectural depth does not necessarily improve performance, particularly when severity differentiation is driven primarily by temperature-intensity variations rather than highly complex spatial features.

4.5 Results and Analysis for Stage 4: Overall Performance of Hotspot Localization and Severity Classification

This section presents the comprehensive evaluation of the proposed hotspot localization and severity classification framework under Stage 4. Unlike previous stages that examined individual components independently, this stage assesses the integrated system to determine its overall robustness, reliability, and practical applicability. The evaluation is conducted using two complementary approaches: quantitative performance metrics and expert-based thermographer validation. By combining objective statistical measurements with expert interpretation, this analysis provides a holistic assessment of the system's operational effectiveness.

4.5.1 Results and Analysis based on Performance Metrics Evaluation

Overall, the experimental findings demonstrate progressive performance improvements across all stages of the proposed framework. The two-tier segmentation approach achieved reliable hotspot localization, with ResNet-50 providing the best segmentation performance. In the classification phase, the dual-input severity classification strategy significantly improved discriminative capability compared to single-modality configurations. The integration of spatial and thermal features enabled more refined class separation, particularly between closely related severity levels. Furthermore, strategies that incorporate additional convolutional refinement strengthen classification robustness and stability. Among all evaluated configurations, the three-layer ResNet-18 model achieved the highest overall performance, attaining 99.47% accuracy and outperforming deeper architectures while maintaining improved computational efficiency.

In addition to accuracy, consistent improvements were observed across precision, recall, and F1-score metrics, indicating balanced class prediction and reduced misclassification between severity levels. These findings confirm that integrating spatial and thermal information enhances feature representation and strengthens decision stability. Agreement analysis with certified solar thermographers further validated the reliability and practical applicability of the proposed framework.

4.5.2 Results and Analysis based on Solar Thermographer Evaluation

This subsection presents an expert-based evaluation conducted by certified solar thermographers to assess the consistency between human interpretation and the proposed hotspot severity classification framework. A total of 20 representative thermal images were independently reviewed by nine certified thermographers, who assigned severity labels across three predefined hotspot categories. The expert annotations were subsequently compared with the established ground truth to evaluate inter-rater agreement and alignment with the reference classification. This evaluation provides an external validation of the system's practical reliability and its consistency with professional inspection standards.

Table 4.18 presents the hotspot severity classifications assigned by nine professional solar thermographers across 20 thermal images, along with the corresponding ground truth (GT) labels used for expert agreement analysis. In addition, the table includes the classification results generated by the proposed model (PM), enabling a direct comparison between human expert assessments and the automated system. This comparison facilitates the evaluation of the proposed model's consistency and agreement with expert judgments.

As observed in the evaluation results, a high level of agreement is evident among thermographers for images associated with severe hotspot conditions, with most expert assessments matching the ground truth labels. This indicates that marked thermal anomalies are generally easier to interpret and classify due to their distinct temperature contrasts and spatial patterns. In contrast, moderate and low-severity cases exhibit relatively higher variability in expert judgments. Such variation is expected, as fine-grained temperature differences and borderline hotspot patterns often require subjective interpretation, even among experienced thermographers.

The observed trends from the solar thermographer evaluation are consistent with the statistical agreement analysis reported in Subsection 4.5.3. While Cohen's Kappa provides a quantitative measure of model reliability, the expert-based evaluation offers complementary qualitative validation. It further confirms that the proposed framework produces severity classifications that align well with professional judgement, particularly for critical hotspot conditions. This dual validation strengthens the practical applicability of the proposed hotspot localization and severity classification approach for real-world PV inspection tasks.

Table 4.18
Solar Thermographer Evaluation of Hotspot Severity Classification Results

Image	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
ST 1	M	H	H	H	L	M	H	H	H	H	H	H	L	M	M	M	L	L	H	L
ST 2	H	H	H	H	H	H	H	H	H	H	H	H	M	H	H	H	M	M	L	L
ST 3	H	H	H	H	M	M	H	H	H	H	H	H	M	M	M	H	L	L	H	L
ST 4	L	M	M	H	L	L	H	H	H	H	H	H	L	M	M	M	L	L	H	L
ST 5	L	H	H	M	L	M	H	H	H	H	H	H	L	M	M	L	M	L	M	L
ST 6	L	H	H	H	L	M	H	H	H	H	H	H	L	H	L	H	L	L	H	L
ST 7	M	M	M	M	L	L	H	H	H	H	H	H	L	M	L	H	L	L	H	L
ST 8	L	M	M	L	L	M	H	H	H	H	H	H	L	M	M	M	L	L	H	L
ST 9	L	L	L	M	L	L	H	H	H	H	H	H	L	L	L	M	L	L	M	L
GT	M	M	M	M	M	L	H	H	H	H	H	H	H	M	L	M	L	L	L	L
Proposed Model	M	M	M	M	M	L	H	H	H	H	H	H	M	M	M	M	L	M	M	L

ST: Solar Thermographer
GT: Ground Truth
Hotspot Severity Class:
L: Low
M: Medium
H: High

4.5.3 Results based on Statistical Analysis

This subsection quantitatively evaluates the consistency between the automated system and certified solar thermographer interpretations using Cohen's Kappa (κ) coefficient. Cohen's Kappa was selected due to its suitability for categorical classification tasks, as it measures inter-rater agreement while adjusting for agreement occurring by chance. This provides a more reliable representation of the consistency of classification between raters.

Figure 4.18 presents the κ values for each solar thermographer (ST1–ST9) across two comparisons: Solar thermographer versus ground truth (ST vs GT) and proposed model versus solar thermographer (PM vs ST). The ST vs GT results demonstrate noticeable variability among thermographers, with κ values ranging from 0.11 (ST6) to 0.57 (ST7). This spread indicates differences in individual interpretation when assigning hotspot severity levels, particularly for borderline cases where visual distinctions between low and medium classes are less pronounced. The observed variability confirms that manual severity assessment is inherently influenced by subjective judgment.

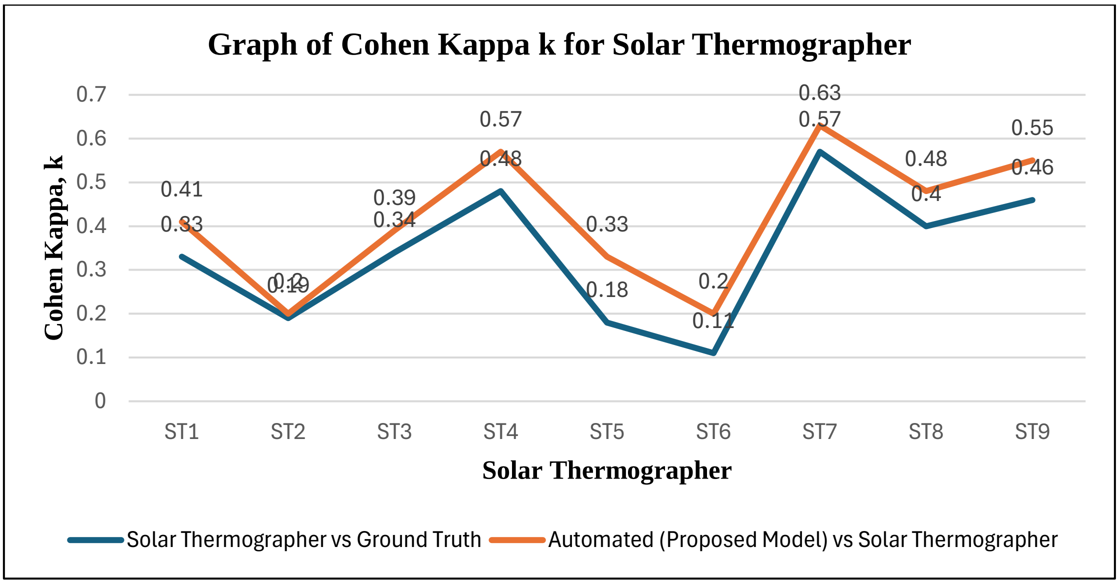


Figure 4.18 Cohen's Kappa Agreement Between Solar Thermographers and Ground Truth, and Between Automated (Proposed Model) and Solar Thermographers

A comparable trend is observed in the PM vs ST comparison. The κ values indicate fair to moderate overall agreement, with stronger agreement observed among thermographers who also showed greater consistency with the ground truth. This pattern suggests that the model's predictions align closely with expert interpretation and reflect similar decision boundaries in severity classification.

Table 4.19 summarizes the average κ values across all thermographers. The mean agreement between ST and GT is 0.36 ± 0.164, indicating fair-to-moderate inter-expert consistency. In comparison, the agreement between the proposed model and thermographers (PM vs ST) is slightly higher at 0.40 ± 0.131, suggesting that the automated system achieves a level of consistency comparable to, and marginally more stable than, human inter-rater agreement.

Table 4.19
Cohen's Kappa For The Evaluation Of The System Developed Against The Ground Truth (GT) and Solar Thermographer (ST)

Test	Item / Dataset	Average score ± Standard Deviation
Cohen's Kappa, κ	ST vs GT	0.36 ± 0.164
	PM vs ST	0.40 ± 0.131

Overall, while variability exists among human thermographers, the proposed framework demonstrates performance comparable to that of experts. The system achieves agreement levels comparable to expert judgment, supporting its applicability as a decision-support tool in real-world PV thermal inspection. Cohen's Kappa analysis confirms the effectiveness and stability of the proposed model, particularly in scenarios where human agreement is inconsistent.

4.6 Comparative Analysis of Hotspot Localization and Severity Classification with Existing Methods

In this section, the performance of the proposed hotspot localization and severity classification framework is benchmarked against several recent studies that utilize infrared thermography and DL approaches for PV fault analysis. The comparison focuses on both hotspot detection or localization and severity classification capabilities to provide a structured evaluation of methodological differences. Table 4.20 presents a comparison between the proposed framework and recent studies in terms of hotspot localization and severity classification performance. While prior studies demonstrate strong performance in either detection or classification tasks, the two components are often examined independently. As a result, integrated evaluation of both localization and severity classification remains limited in the existing literature. In contrast, the proposed approach integrates these components into a single, fully automated framework, enabling end-to-end processing from segmentation to severity grading.

Table 4.20
Comparative Analysis of Hotspot Localization and Severity Classification with Existing Methods

Study	Pixel-Level Localization	Severity Classification	Classes	Accuracy (%)
Ali et al. [161]	-	Yes	3-class	98.41
Acıkgoz et al.[67]	-	Binary	2-class	98.65
Sandeep et al.[164]	Partial	Binary	2-class	~98
Proposed Framework	IoU 98.33% PV, 78.95% Hotspot	Dual Input	3-class	99.47

Recent studies have reported performance in hotspot detection and classification using both HC feature-based techniques and DL architectures. For instance, Ali et al. [161] proposed a hybrid global–local feature extraction approach combined with optimization-based feature selection for three-class PV fault classification, achieving 98.41% accuracy. Similarly, Acıkgoz et al. [67] evaluated multiple pre-trained CNNs and reported a highest accuracy of 98.65% for binary hotspot classification. Sandeep et al. [164] introduced a CNN-based model for hotspot detection with approximately 98% classification accuracy. While these studies demonstrate strong classification performance, explicit pixel-level localization metrics are not consistently reported. In contrast, the proposed framework evaluates localization performance using pixel-level IoU, achieving 98.33% IoU for solar PV region segmentation and 78.95% IoU for hotspot region segmentation. Since IoU is computed at the pixel scale, it provides a more detailed assessment of segmentation overlap, supporting precise spatial delineation of hotspot regions.

In terms of hotspot severity classification, existing studies largely emphasize binary categorization rather than structured multi-level grading. For example, Acıkgoz et al. [67] and Sandeep et al. [164] primarily distinguish between hotspot and non-hotspot conditions, without further differentiating severity levels. Although Ali et al. [161] extended the problem into a three-class PV fault classification setting, the approach relied on manually engineered feature extraction and did not integrate segmentation-driven severity assessment. In contrast, the proposed framework performs three-class hotspot severity classification (Low, Medium, and High) using a dual-input feature extraction strategy that combines spatial image information with thermal characteristics. The enhanced dual-input ResNet-18 configuration achieved an overall classification accuracy of 99.47%, demonstrating balanced performance across precision, recall, and F1-score metrics. By coupling pixel-level hotspot segmentation with structured severity grading, the framework ensures that classification decisions are based on accurately delineated hotspot regions rather than on global image representations alone.

Overall, the comparative evaluation indicates that while existing studies achieve strong performance in either hotspot detection or classification, fully integrated solutions remain limited. The proposed framework addresses this gap by combining two-tier pixel-level segmentation with structured three-class severity grading within a single integrated and fully automated framework. The incorporation of dual-input feature fusion further enhances classification robustness by leveraging both spatial and thermal characteristics. Although direct numerical comparisons may vary due to differences in datasets and evaluation methods, the proposed framework demonstrates competitive localization performance and strong severity classification accuracy. Overall, the results indicate that the approach provides a practical and comprehensive solution for PV hotspot inspection.

4.7 Summary

This chapter presents and discusses the results of the proposed methodology for classifying hotspot severity in thermal images. The results are structured into four main components: two-tier segmentation, CNN classification performance with dual-input feature extraction, enhanced CNN model comparison, and overall performance evaluation.

In the first part, a two-tier semantic segmentation approach is proposed. The resulting segmentation output provides a reliable basis for the subsequent hotspot identification stage. The analysis demonstrates that the proposed two-tier framework consistently achieves higher localization accuracy than a single-stage approach. This improvement is particularly evident in thermal images affected by low contrast and non-uniform irradiance conditions. The integration of hybrid image enhancement techniques, including NLM denoising and CLAHE, effectively improves thermal contrast while suppressing noise. Consequently, hotspot visibility and boundary definition are enhanced. As a result, the refined hotspot regions yield more discriminative spatial and thermal features, thereby improving robustness and reliability in subsequent severity classification.

Next, in terms of feature extraction, the study found that integrating spatial image-based features with temperature-driven thermal descriptors yields more discriminative representations for hotspot severity classification. While single-feature inputs capture only partial characteristics of hotspot behaviour, the dual-feature fusion strategy effectively preserves both geometric hotspot patterns and underlying thermal intensity variations. The use of refined hotspot regions derived from the two-tier segmentation further enhances feature reliability by minimizing background interference and boundary imprecision. Consequently, this dual-feature approach produces a more comprehensive and informative image representation. Thus, it enhances the model’s ability to improve class separability and achieve more stable severity classification performance.

Furthermore, the performance of the enhanced CNN classification models with varying convolutional layer depths is analyzed. The experimental results indicate that progressively increasing the number of convolutional layers leads to consistent performance improvements, as deeper architectures enable more effective hierarchical feature learning. Specifically, models with additional convolutional layers demonstrate greater ability to capture complex spatial and thermal patterns associated with hotspot severity. Among the evaluated configurations, the model with three convolutional layers achieves the best overall classification performance, indicating an optimal balance between representational capacity and model complexity.

Moreover, qualitative validation by solar thermographers supported the effectiveness of the proposed framework. The fair-to-moderate between the predicted severity classes and expert assessments indicates high interpretative reliability. Minor disagreements were primarily observed at boundary hotspot regions, reflecting inherent variations in expert judgment when interpreting slight thermal differences. Overall, these findings demonstrate the potential of the proposed system to support consistent and repeatable hotspot severity assessments within operational PV inspection workflows.

In conclusion, the proposed methodology for classifying hotspot severity from thermal images demonstrates improved classification accuracy and robustness. The integration of a two-tier segmentation architecture provides precise and reliable hotspot localization, even under challenging thermal conditions. By combining spatial image-based features with temperature-driven thermal features, the proposed framework enables more discriminative and stable severity representations. Furthermore, enhancing the CNN architecture through deeper convolutional layers improves hierarchical feature learning, with the three-layer configuration achieving optimal classification performance. Comparative analysis with existing approaches confirms that the proposed hybrid framework demonstrates competitive performance relative to existing hotspot localization and severity classification approaches for PV thermal imagery. Qualitative validation by solar thermographers further supports the reliability and practical applicability of the proposed system. Overall, these results highlight the potential of the framework to facilitate consistent hotspot severity assessment, support preventive maintenance strategies, reduce performance degradation, and enable effective management of PV assets.

Despite the promising results, several limitations should be acknowledged. The developed framework was evaluated using thermal images collected from a single PV installation site, which may limit its applicability to different environmental and operational conditions. In addition, the solar thermographer validation involved a limited number of experts, and no independent external dataset was available for further verification. Future work should evaluate the proposed framework using larger multi-site datasets and broader expert participation to further assess its applicability under diverse operating conditions. Nevertheless, the findings demonstrate the potential of the proposed framework as a practical and reliable tool for automated hotspot severity assessment in PV monitoring applications.

CHAPTER 5
CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusion

This study presented a comprehensive methodology for hotspot localization and severity classification in solar photovoltaic (PV) modules using aerial thermal imagery. The proposed approach integrates image processing and deep learning (DL) techniques into a systematic pipeline comprising pre-processing, segmentation, and severity classification stages. The results demonstrate that the proposed methodology significantly improves the accuracy and reliability of hotspot detection. This improvement provides valuable support for effective solar PV system monitoring and maintenance.

The proposed framework begins with a two-tier semantic segmentation strategy for hotspot localization, addressing Objective (1). In the first stage, the model segments solar PV modules from the background, followed by fine-grained segmentation of hotspot regions within the identified PV modules. The segmentation architecture incorporates U-Net, ResNet-18, and ResNet-50 to evaluate multi-level feature learning capabilities. Experimental results indicate that the two-tier segmentation approach achieves strong performance, with ResNet-50 attaining the highest Intersection over Union (IoU) of 78.95%, followed by ResNet-18 (78.61%) and U-Net (63.70%). Both ResNet-18 and ResNet-50 demonstrate superior feature extraction capability compared to the baseline U-Net model. These results validate the effectiveness of the proposed framework in improving hotspot localization accuracy and support reliable severity assessment in thermal PV imagery.

The second objective of this study (Objective 2) focuses on enhancing hotspot severity classification by effectively extracting features from segmented hotspot regions. To achieve this, a dual-input feature extraction strategy was developed by integrating CNN-based deep features derived from thermal images with temperature-based numerical features. Comparative analysis shows that the single-input CNN model and the temperature-only feature model achieved classification accuracy of 76.53% and 85.37%, respectively. By contrast, the proposed dual-input feature fusion approach significantly improves classification performance, achieving an outstanding accuracy of 98.67%. These results demonstrate that combining complementary thermal image features and explicit temperature-based descriptors substantially enhances the discriminative capability of the classification model. The findings validate the effectiveness of the proposed dual-input strategy in improving hotspot severity classification in thermal PV imagery.

Building upon the effectiveness of the proposed dual-input feature extraction strategy, Objective (3) evaluates the effectiveness of enhancing CNN-based classification models for hotspot severity assessment by extending convolutional depth. A series of classification experiments was conducted using the same baseline architectures, namely U-Net, ResNet-18, and ResNet-50, under identical experimental settings. One to three additional convolutional layers were added to each model to assess the impact of architectural depth on classification performance. The results indicate that both ResNet-18 and ResNet-50 benefit from deeper configurations, with ResNet-18 demonstrating the most consistent improvement. In particular, the enhanced ResNet-18 model with three additional convolutional layers achieved the highest performance among the evaluated models, attaining an accuracy and F1-score of 99.47%. Although ResNet-50 also showed performance gains, these improvements were less consistent, while U-Net exhibited limited suitability for classification tasks. These findings confirm that increasing convolutional depth in residual network architectures, particularly ResNet-18, is an effective strategy for improving performance in hotspot severity classification.

Following the demonstrated effectiveness of the enhanced CNN-based classification model, Objective (4) focuses on validating the reliability of the proposed hotspot severity classification framework through expert-based assessment. Cohen's κ analysis was conducted to evaluate the level of agreement between the proposed classification results and assessments provided by nine independent solar thermographers. The results indicate a fair to moderate level of agreement between the proposed framework and expert evaluations, indicating reasonable alignment with expert judgement. In contrast, comparing individual solar thermographers with expert-derived ground truth reveals greater variability, particularly for low and moderate-severity hotspot cases. Such variation is expected among expert assessments. This is primarily due to minor temperature differences and borderline thermal patterns, which often require subjective interpretation even among experienced thermographers.

Overall, these findings confirm that the proposed framework delivers reliable classification outcomes comparable to inter-expert variability. This supports its practical applicability for thermal inspection and decision-making in solar PV system monitoring.

5.2 Research Contribution

This research contributes to the advancement of automated hotspot localization and severity classification in solar PV systems using aerial thermal imagery and deep learning techniques. In alignment with the research objectives, the key contributions of this study are summarized as follows:

i. A novel two-tier semantic segmentation framework for automated hotspot localization is proposed. The first segmentation stage isolates PV module regions from the background, while the second stage extracts fine-grained hotspots within individual modules. This hierarchical segmentation strategy effectively reduces background interference and boundary uncertainty, enabling more accurate hotspot localization under varying thermal contrast and non-uniform irradiance conditions.

ii. An enhanced thermal image scheme integrating Non-Local Means (NLM) denoising with Contrast Limited Adaptive Histogram Equalization (CLAHE) is introduced. This hybrid (NLM-CLAHE) enhancement technique improves thermal contrast while suppressing noise, resulting in clearer hotspot boundaries and more reliable extraction of the hotspot region for subsequent segmentation and analysis.

iii. A new model of dual-input feature extraction strategy is proposed for hotspot severity classification by integrating CNN-based image features with temperature-driven numerical descriptors. This fusion approach enables the joint exploitation of complementary spatial and thermal information. As a result, the limitations of purely image-centric analysis are addressed, and the discrimination of hotspot severity levels is improved.

iv. An enhanced CNN-based hotspot severity classification model is developed by incorporating progressive convolutional refinement layers into baseline

CNN architectures. This controlled refinement strategy strengthens high-level feature representations without altering the original network architecture. As a result, classification robustness and stability are improved across diverse thermal imaging conditions.

v. A comprehensive performance evaluation and expert-based validation framework is established to assess the reliability and practical applicability of the proposed method. Quantitative performance metrics are complemented by expert-agreement analysis based on evaluations by certified solar thermographers. This demonstrates improved consistency and alignment between automated severity classification and expert judgement.

Overall, the proposed contributions provide a structured and robust solution for automated hotspot detection and severity classification in PV thermal imagery. The integration of refined segmentation, hybrid feature representation, enhanced CNN classification, and expert validation supports the development of reliable, scalable, and practically applicable PV inspection systems.

5.3 Recommendations for Future Work

The findings of this study highlight the importance of accurate segmentation and effective image enhancement for hotspot detection. Furthermore, hybrid learning approaches enhance the accuracy of PV system severity assessment. The proposed multi-layered approach provides a solid framework for automated thermal analysis. However, several opportunities remain to improve further the reliability, adaptability, and scalability of future PV assessment systems.

Future research should focus on designing learning architectures that are optimized for limited computational environments. This is important for automating hotspot detection in large-scale PV installations. A key challenge is the inability to implement complex models on technologies with limited computational capacity. Addressing this issue requires adopting optimization methods that reduce computational load without compromising segmentation accuracy or severity classification. Improvements of this nature would support the use of the system on edgeoriented platforms, including internet of thing (IoT)-based monitoring devices and compact processing units in field applications. Ultimately, such optimization will support wider adoption of real-time fault detection, making automated hotspot assessment more practical, affordable, and scalable across PV installations of varying sizes.

Furthermore, distributed edge processing architectures are a high-priority research area. Large PV installations will be able to rapidly, in real time, execute advanced fault-detection algorithms using this method. Besides, this approach will promote more adaptable decision-making across diverse environmental conditions. This direction will improve operational monitoring and enable continuous performance assessment of PV installations.

Further improvement in segmentation performance also presents an essential direction for development. Introducing similar indices across all segmented regions within an image may increase confidence in identifying repeated hotspot patterns, especially when hotspot boundaries vary due to inconsistent temperature distribution. Investigating alternative segmentation strategies that address shape variability and thermal dispersion characteristics could also strengthen region delineation and reduce misclassification risks.

The development and standardization of thermal imagery databases is another essential objective for further research. Implementing a thorough benchmark dataset with uniform assessment standards and annotation guidelines would increase repeatability and make cross-study comparisons more insightful. A larger and more diverse data repository would further support robust model validation and reflect real-world variations in PV operating conditions.

In summary, this work develops a thorough automated framework to enhance PV system severity categorization, hotspot localization, and image terminology. Furthermore, this study offers several opportunities for future advancement. Future PV inspection systems can be made more reliable and scalable by incorporating real-time edge-based processing, expanding standardized datasets, and optimizing learning architectures for resource-constrained environments. Further improvements in segmentation techniques will support the development of more adaptive and intelligent monitoring systems. These advancements can reduce operational risks, facilitate predictive maintenance, and enhance the long-term reliability of large-scale solar infrastructure.

REFERENCES

[1] O. A. Marzouk, “Summary of the 2023 Report of TCEP (Tracking Clean Energy Progress) by the International Energy Agency (IEA), and Proposed Process for Computing a Single Aggregate Rating,” *E3S Web of Conferences*, vol. 601, pp. 1–31, 2025, doi: 10.1051/e3sconf/202560100048.

[2] I. ENERGY and AGENCY, “Renewables 2023,” 2023. doi: 10.1002/peng.20026.

[3] Ministry of Economy Malaysia, “National Energy Transition Roadmap (NETR) 1,” 2023.

[4] I. R. Ibrahim, M. N. Bin Azman, N. A. Kamarzaman, and R. Darus, “Maximum Demand Peak Shave Approach Utilising a Hybrid Solar PV and Battery Energy Storage System,” *Proceedings - 13th IEEE International Conference on Control System, Computing and Engineering, ICCSCE 2023*, pp. 243–248, 2023, doi: 10.1109/ICCSCE58721.2023.10237141.

[5] M. Vlaminck, R. Heidbuchel, W. Philips, and H. Luong, “Region-Based CNN for Anomaly Detection in PV Power Plants Using Aerial Imagery,” *Sensors*, vol. 22, no. 3, Feb. 2022, doi: 10.3390/s22031244.

[6] C. Buerhop, L. Bommes, J. Schlipf, T. Pickel, and A. Fladung, “Infrared imaging of photovoltaic modules: a review of the state of the art and future challenges facing gigawatt photovoltaic power stations,” *Progress in Energy*, vol. 4, no. 4, 2022, doi: 10.1088/2516-1083/ac890b.

[7] I. E. Agency, *World Energy Outlook 2023*. 2023. [Online]. Available: www.iea.org.terms

[8] 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, doi: 10.3390/rs15071914.

[9] T. Gupta, G. Tiwari, and K. Joshi, “A Survey of Artificial Intelligence Integrating Image Processing Techniques for Predicting Solar Energy Production: Challenges and Innovations,” *Proceedings of the 2024 International Conference on Artificial Intelligence and Emerging Technology, Global AI Summit 2024*, pp. 282–287, 2024, doi: 10.1109/GlobalAISummit62156.2024.10947776.

[10] A. Ivanda, L. Seric, and M. Braovic, “Exploring Applications of Convolutional Neural Networks in Analyzing Multispectral Satellite Imagery: A Systematic Review,” *Big Data Mining and Analytics*, vol. 8, no. 2, pp. 407–429, 2025, doi: 10.26599/BDMA.2024.9020086.

[11] B. Emek Soylu, M. S. Guzel, G. E. Bostanci, F. Ekinci, T. Asuroglu, and K. Acici, “Deep-Learning-Based Approaches for Semantic Segmentation of Natural Scene Images: A Review,” *Electronics (Switzerland)*, vol. 12, no. 12, 2023, doi: 10.3390/electronics12122730.

[12] C. Xiaoni, D. Wei, Y. He, and M. Shen, “A fault diagnosis method of hot spots for photovoltaic clusters based on model parameters,” *Energy Sci. Eng.*, vol. 12, no. 8, pp. 3453–3464, 2024, doi: 10.1002/ese3.1829.

[13] M. I. Islam, M. S. Bin Jadin, A. Al Mansur, and T. Alharbi, “Electrical Performance and Degradation Analysis of Field-Aged PV Modules in Tropical Climates: A Comparative Experimental Study,” *Energy Conversion and Management: X*, vol. 24, no. October, p. 100719, 2024, doi: 10.1016/j.ecmx.2024.100719.

[14] M. Jalal, I. U. Khalil, and A. ul Haq, “Deep learning approaches for visual faults diagnosis of photovoltaic systems: State-of-the-Art review,” *Results in Engineering*, vol. 23, no. July, p. 102622, 2024, doi: 10.1016/j.rineng.2024.102622.

[15] W. Jianxun, C. Xin, J. Weicheng, H. Li, L. Junyi, and S. Haigang, “PVNet: A novel semantic segmentation model for extracting high-quality photovoltaic panels in large-scale systems from high-resolution remote sensing imagery,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 119, no. January, p. 103309, 2023, doi: 10.1016/j.jag.2023.103309.

[16] L. Chen, B. Liu, K. Sun, and J. Zhao, “A Hot Spot Detection Method of Photovoltaic Module Based on Image Semantic Segmentation,” *2023 IEEE 6th International Electrical and Energy Conference, CIEEC 2023*, pp. 837–841, 2023, doi: 10.1109/CIEEC58067.2023.10167018.

[17] 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, doi: 10.1016/j.ijleo.2024.171627.

[18] L. S. Pinho, T. D. Sousa, C. D. Pereira, and A. M. Pinto, “Anomaly Detection for PV Modules Using Multi-Modal Data Fusion in Aerial Inspections,” *IEEE*

Access, vol. 13, pp. 88762–88779, 2025, doi: 10.1109/ACCESS.2025.3570815.

[19] G. M. El-Banby, N. M. Moawad, B. A. Abouzalm, W. F. Abouzaid, and E. A. Ramadan, “Photovoltaic system fault detection techniques: a review,” *Neural Comput. Appl.*, vol. 35, no. 35, pp. 24829–24842, 2023, doi: 10.1007/s00521-023-09041-7.

[20] N. N. Nyangiwe, A. D. K. Kenfack, N. Thantsha, and M. Msimanga, “Performance Monitoring of Photovoltaic Modules Using Machine-Learning-Based Solutions: A Survey of Current Trends,” *Energy Sci. Eng.*, 2025, doi: 10.1002/ese3.70423.

[21] E. K. Ukiwe, S. A. Adeshina, T. Jacob, and B. B. Adetokun, “Deep learning model for detection of hotspots using infrared thermographic images of electrical installations,” *Journal of Electrical Systems and Information Technology*, vol. 11, no. 1, 2024, doi: 10.1186/s43067-024-00148-y.

[22] C. Yang, F. Sun, Y. Zou, Z. Lv, and L. Xue, “A Survey of Photovoltaic Panel Overlay and Fault Detection Methods,” *Energies (Basel).*, vol. 17, no. 4, pp. 1–37, 2024, doi: 10.3390/en17040837.

[23] R. Duan, Y. Wang, X. Chen, and S. Li, “An enhanced algorithm for cell-level anomaly segmentation in photovoltaic solar panels using electroluminescence imaging,” *Energy*, vol. 331, no. April, p. 136711, 2025, doi: 10.1016/j.energy.2025.136711.

[24] Y. Jiang and C. Zhao, “Attention classification-and-segmentation network for micro-crack anomaly detection of photovoltaic module cells,” *Solar Energy*, vol. 238, no. February 2021, pp. 291–304, 2022, doi: 10.1016/j.solener.2022.04.012.

[25] H. Zhang, H., Lu, Q., Huang, S. C., Bai, J., and Kang, “Research on Hot Spot Defect Detection in Photovoltaic Modules Based on UAV Remote Sensing Technology and Deep Learning,” *Unmanned Systems*, 2025, doi: 10.1142/S2301385027500439.

[26] A. Khatri, S. Khadka, N. Lamichhane, and R. Shrestha, “A comprehensive review of infrared thermography and deep learning applications for solar photovoltaic systems,” *Infrared Phys. Technol.*, vol. 148, no. April, p. 105878, 2025, doi: 10.1016/j.infrared.2025.105878.

[27] A. Teta, B. Korich, D. Bakria, N. Hadroug, and A. Rabehi, “Fault detection

and diagnosis of grid-connected photovoltaic systems using energy valley optimizer based lightweight CNN and wavelet transform," *Sci. Rep.*, vol. 14, no. 1, pp. 1–22, 2024, doi: 10.1038/s41598-024-69890-7.

[28] 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, doi: 10.1016/j.seja.2025.100090.

[29] W. S. Ebhota and P. Y. Tabakov, "Influence of photovoltaic cell technologies and elevated temperature on photovoltaic system performance," *Ain Shams Engineering Journal*, vol. 14, no. 7, p. 101984, 2023, doi: 10.1016/j.asej.2022.101984.

[30] I. Kaaya, J. Ascencio-Vasquez, K. A. Weiss, and M. Topic, "Assessment of uncertainties and variations in PV modules degradation rates and lifetime predictions using physical models," *Solar Energy*, vol. 218, no. January, pp. 354–367, 2021, doi: 10.1016/j.solener.2021.01.071.

[31] N. Fernando, L. Seneviratne, N. Weerasinghe, N. Rathnayake, and Y. Hoshino, "Efficient Hotspot Detection in Solar Panels via Computer Vision and Machine Learning," *Information (Switzerland)*, vol. 16, no. 7, pp. 1–50, 2025, doi: 10.3390/info16070608.

[32] U. B. Hakan Acikgoz, Deniz Korkmaz, "Photovoltaic cell defect classification based on integration of residual-inception network and spatial pyramid pooling in electroluminescence images," *Expert Syst. Appl.*, vol. 229, 2023, doi: https://doi.org/10.1016/j.eswa.2023.120546.

[33] A. Thakfan and Y. bin Salamah, "Artificial-Intelligence-Based Detection of Defects and Faults in Photovoltaic Systems: A Survey," *Energies (Basel).*, vol. 17, no. 19, pp. 1–22, 2024, doi: 10.3390/en17194807.

[34] S. Pimenow, O. Pimenowa, and P. Prus, "Challenges of Artificial Intelligence Development in the Context of Energy Consumption and Impact on Climate Change," *Energies (Basel).*, vol. 17, no. 23, pp. 1–34, 2024, doi: 10.3390/en17235965.

[35] U. Hijjawi, S. Lakshminarayana, T. Xu, G. Piero Malfense Fierro, and M. Rahman, "A review of automated solar photovoltaic defect detection systems: Approaches, challenges, and future orientations," *Solar Energy*, vol. 266, no. October, p. 112186, 2023, doi: 10.1016/j.solener.2023.112186.

[36] M. A. Ebied, A. Munshi, S. A. Alhuzali, M. M. El-sotouhy, A. I. Shehta, and

M. S. Elborlsy, “Advanced deep learning modeling to enhance detection of defective photovoltaic cells in electroluminescence images,” *Sci. Rep.*, vol. 15, no. 1, pp. 1–24, 2025, doi: 10.1038/s41598-025-14478-y.

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

[38] O. Bamisile, C. Acen, D. Cai, Q. Huang, and I. Staffell, “The environmental factors affecting solar photovoltaic output,” *Renewable and Sustainable Energy Reviews*, vol. 208, no. October 2024, p. 115073, 2025, doi: 10.1016/j.rser.2024.115073.

[39] T. Nakamura, M. Imaizumi, K. Murai, R. Kawamura, and I. Nanno, “Mechanism of thermal runaway in solar cells,” *Jpn. J. Appl. Phys.*, 2023.

[40] T. Sun, H. Xing, S. Cao, Y. Zhang, S. Fan, and P. Liu, “A novel detection method for hot spots of photovoltaic (PV) panels using improved anchors and prediction heads of YOLOv5 network,” *Energy Reports*, vol. 8, pp. 1219–1229, 2022, doi: 10.1016/j.egyr.2022.08.130.

[41] 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, doi: 10.1016/j.uncres.2025.100156.

[42] M. L. Adekanbi, E. S. Alaba, T. J. John, T. D. Tundealao, and T. I. Banji, “Soiling loss in solar systems: A review of its effect on solar energy efficiency and mitigation techniques,” *Cleaner Energy Systems*, vol. 7, no. October 2023, p. 100094, 2024, doi: 10.1016/j.cles.2023.100094.

[43] P. C. Okonkwo, S. C. Nwokolo, S. O. Udo, A. U. Obiwulu, and U. N. Onnoghen, “Solar PV systems under weather extremes: Case studies, classification, vulnerability assessment, and adaptation pathways,” *Energy Reports*, vol. 13, no. October 2024, pp. 929–959, 2025, doi: 10.1016/j.egyr.2024.12.067.

[44] H. Al Mahdi, P. G. Leahy, M. Alghoul, and A. P. Morrison, “A Review of Photovoltaic Module Failure and Degradation Mechanisms: Causes and Detection Techniques,” *Solar*, vol. 4, no. 1, pp. 43–82, 2024, doi: 10.3390/solar4010003.

[45] N. Kumari, S. K. Singh, S. Kumar, and V. K. Jadoun, “Analysis of single-cell

shading impact on polycrystalline PERC module under hot-dry conditions," *Energy Conversion and Management: X*, vol. 27, no. July, p. 101143, 2025, doi: 10.1016/j.ecmx.2025.101143.

[46] D. Bhor, R. T. Pote, N. Chavan, and R. Pote, "Comparative Thermal Analysis of Different Solar Panel Materials using ANSYS SEE PROFILE Comparative Thermal Analysis of Different Solar Panel Materials using ANSYS," *Article in International Journal of Engineering Research*, vol. 11, no. 06, pp. 125–128, 2022, [Online]. Available: www.ijert.org

[47] T. Rahman, A. Al Mansur, M. S. Hossain Lipu, M. S. Rahman, and R. H. Ashique, "Investigation of Degradation of Solar Photovoltaics: A Review of Aging Factors, Impacts, and Future Directions toward Sustainable Energy Management," *Energies (Basel).*, vol. 16, no. 9, 2023, doi: 10.3390/en16093706.

[48] D. Marangis, A. Livera, G. Tziolis, G. Makrides, A. Kyprianou, and G. E. Georghiou, "Trend-Based Predictive Maintenance and Fault Detection Analytics for Photovoltaic Power Plants," vol. 2400473, pp. 1–10, 2024, doi: 10.1002/solr.202400473.

[49] A. Livera, M. Theristis, A. Charalambous, J. S. Stein, and G. E. Georghiou, "Decision support system for corrective maintenance in large-scale photovoltaic systems," *Conference Record of the IEEE Photovoltaic Specialists Conference*, pp. 306–311, 2021, doi: 10.1109/PVSC43889.2021.9518796.

[50] L. M. Shaker, A. A. Al-Amiery, M. M. Hanoon, W. K. Al-Azzawi, and A. A. H. Kadhum, "Examining the influence of thermal effects on solar cells: a comprehensive review," *Sustainable Energy Research*, vol. 11, no. 1, 2024, doi: 10.1186/s40807-024-00100-8.

[51] E. E. van D. M. Vumbugwa, F.J. Vorster, J.L. Crozier McCleland, "Common misinterpretations of thermal signatures on polycrystalline PV modules under different operational conditions," *Solar Energy*, vol. Volume 263, 2023, doi: https://doi.org/10.1016/j.solener.2023.111957.

[52] W. M. Xinfeng Zhao, Yangjing Zhao, Shunchang Hu, Hongyan Wang, Yuyan Zhang, "Progress in Active Infrared Imaging for Defect Detection in the Renewable and Electronic Industries," 2023, doi: https://doi.org/10.3390/s23218780.

[53] V. Ramani, P. Arjunan, K. Poolla, and C. Miller, “Semantic segmentation of longitudinal thermal images for identification of hot and cool spots in urban areas,” *Build. Environ.*, vol. 249, no. October 2023, p. 111112, 2024, doi: 10.1016/j.buildenv.2023.111112.

[54] V. T. Gomathy Balasubramani, “Thermal Image Analysis of Photovoltaic Panel for Condition Monitoring Using Hybrid Thermal Pixel Counting Algorithm and XGBoost Classifier,” *Electric Power Components and Systems*, 2023, doi: https://doi.org/10.1080/15325008.2023.2276828.

[55] A. Wang and Y. Xuan, “Close examination of localized hot spots within photovoltaic modules,” *Energy Convers. Manag.*, vol. 234, no. February, p. 113959, 2021, doi: 10.1016/j.enconman.2021.113959.

[56] G. K. Balakrishnan, C. T. Yaw, S. P. Koh, T. Abedin, and A. A. Raj, “A Review of Infrared Thermography for Condition-Based Monitoring in Electrical Energy: Applications and Recommendations,” *Energies (Basel).*, vol. 15, no. 16, 2022, doi: 10.3390/en15166000.

[57] A. M. M. Kamarulzaman, W. S. W. M. Jaafar, N. A. A. Bakar, K. Chew, and N. I. K. Anuar, “Comprehensive Analysis of Solar Photovoltaic System Defects and Inspection Techniques in Tropical Environments Using Thermal UAV: Study Case in Marang, Terengganu,” *E3S Web of Conferences*, vol. 599, 2025, doi: 10.1051/e3sconf/202459903003.

[58] M. Dhimish and P. I. Lazaridis, “An empirical investigation on the correlation between solar cell cracks and hotspots,” *Sci. Rep.*, vol. 11, no. 1, pp. 1–11, 2021, doi: 10.1038/s41598-021-03498-z.

[59] R. M. Gawande, S. Nambiar, S. Shinde, S. S. Banait, A. V. Sonawane, and H. B. Vanjari, “Machine Learning Approaches for Fault Detection and Diagnosis in Electrical Machines: A Comparative Study of Deep Learning and Classical Methods,” *Panamerican Mathematical Journal*, vol. 34, no. 2, pp. 121–137, 2024, doi: 10.52783/pmj.v34.i2.930.

[60] 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, doi: 10.1002/pip.3859.

[61] S. Boubaker, S. Kamel, N. Ghazouani, and A. Mellit, “Assessment of Machine and Deep Learning Approaches for Fault Diagnosis in Photovoltaic

Systems Using Infrared Thermography," *Remote Sens. (Basel).*, vol. 15, no. 6, 2023, doi: 10.3390/rs15061686.

[62] Y. Zefri, A. Elkettani, I. Sebari, and S. A. Lamallam, "Thermal Infrared and Visual Inspection of Photovoltaic Installations by UAV Photogrammetry — Application Case: Morocco," pp. 1–24, 2018, doi: 10.3390/drones2040041.

[63] Z. Xu, Y. Shen, K. Zhang, and H. Wei, "A Segmentation Method for PV Modules in Infrared Thermography Images," *Asia-Pacific Power and Energy Engineering Conference, APPEEC*, vol. 2021-Novem, pp. 1–5, 2021, doi: 10.1109/APPEEC50844.2021.9687630.

[64] S. Kovalskyi and V. Koval, "Comparison of image processing techniques for defect detection," *CEUR Workshop Proc.*, vol. 3716, pp. 158–167, 2024.

[65] Q. Wan, M. Smigaj, B. Brede, and L. Kooistra, "Optimizing UAV-based uncooled thermal cameras in field conditions for precision agriculture," *International Journal of Applied Earth Observation and Geoinformation*, vol. 134, no. February 2024, doi: 10.1016/j.jag.2024.104184.

[66] M. Akteruzzaman, "Development of a Predictive Simulation Model for Solar Photovoltaic System Performance Analysis Considering Environmental, Technical, and Economic Efficiency Factors," *Review of Applied Science and Technology*, vol. 04, no. 02, pp. 639–668, 2025, doi: 10.63125/8ftjw526.

[67] H. Acikgoz, D. Korkmaz, and C. Dandil, "Classification of Hotspots in Photovoltaic Modules with Deep Learning Methods," *Turkish Journal of Science and Technology*, vol. 17, no. 2, pp. 211–221, 2022, doi: 10.55525/tjst.1158854.

[68] 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, doi: 10.24191/esteem.v20iSeptember.18.

[69] W. Ahmed, "Thermal and chromatic analysis for scalable photovoltaic hotspot detection," *Solar Energy*, vol. 306, no. January 2026, doi: 10.1016/j.solener.2025.114227.

[70] A. Patil and M. Rane, "Convolutional Neural Networks: An Overview and Its Applications in Pattern Recognition," *Smart Innovation, Systems and Technologies*, vol. 195, pp. 21–30, 2021, doi: 10.1007/978-981-15-7078-0_3.

[71] 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, doi: 10.1186/s40537-021-00444-8.

[72] J. B. S. Carlos Quiterio Gómez Muñoz, Fausto Pedro García Márquez, “Hybrid CNN Architecture for Hot Spot Detection in Photovoltaic Panels Using Fast R-CNN and GoogleNet,” *CMES - Computer Modeling in Engineering and Sciences*, 2025, doi: https://doi.org/10.32604/cmes.2025.069225.

[73] 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, doi: 10.1016/j.solener.2025.113965.

[74] P. Goel, M. Elmeligy, S. Shapsough, I. Zualkernan, R. Elfakharany, and H. Saleem, “Performance Comparison of Deep Learning Models for Photovoltaic Hotspot Detection Using Thermal Imaging,” *Proceedings of the 3rd International Conference on Machine Learning and Pattern Recognition, MLPR 2025*, no. February, pp. 1–8, 2025, doi: 10.1145/3760678.3760679.

[75] X. W. and J. Z. B. Liu, L. Chen, K. Sun, “A Hot Spot Identification Approach for Photovoltaic Module Based on Enhanced U-Net With Squeeze-and-Excitation and VGG19,” *IEEE Trans. Instrum. Meas.*, vol. 73, 2024, doi: 10.1109/TIM.2024.3379412.

[76] M. Z. Ab Hamid, K. Daud, Z. H. Che Soh, M. K. Osman, I. S. Isa, and N. H. Ishak, “Enhanced Solar Panel Segmentation and Hotspot Recognition Using U-Net: A Multiclass Semantic Segmentation Approach,” *Journal of Electrical & Electronic Systems Research*, vol. 26, no. APR2025, pp. 27–33, 2025, doi: 10.24191/jeesr.v26i1.004.

[77] 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 Biomedical Sciences (ICIIBMS)*, vol. 8, pp. 295–299, 2023, doi: 10.1109/iciibms60103.2023.10347663.

[78] W. Jiangtao, N. I. R. Ruhaiyem, and F. Panpan, “A Comprehensive Review of U-Net and Its Variants: Advances and Applications in Medical Image Segmentation,” *IET Image Process.*, vol. 19, no. 1, 2025, doi: 10.1049/ipr2.70019.

[79] M. L. Sherozbek Jumaboev, Dadajon Jurakuziev, “Photovoltaics Plant Fault Detection Using Deep Learning Techniques,” pp. 1–16, 2022.

[80] N. A. Mazlan, K. A. Othman, S. Shahbudin, and M. Kassim, “Convolution Neural Network (CNN) Architectures Analysis for Photovoltaic (PV) Module Defect Images Classification,” *Proceeding of the International Conference on Computer Engineering, Network and Intelligent Multimedia, CENIM 2022*, pp. 390–395, 2022, doi: 10.1109/CENIM56801.2022.10037564.

[81] B. Mahaur, K. K. Mishra, and N. Singh, “Improved Residual Network based on norm-preservation for visual recognition,” *Neural Networks*, vol. 157, pp. 305–322, 2023, doi: 10.1016/j.neunet.2022.10.023.

[82] I. Trinchieri, A. Tridello, and M. E. E, “Semantic Scene Segmentation for Indoor Robot Navigation,” 2023.

[83] G. K. Pandey and S. Srivastava, “ResNet-18 comparative analysis of various activation functions for image classification,” *6th International Conference on Inventive Computation Technologies, ICICT 2023 - Proceedings*, no. April, pp. 595–601, 2023, doi: 10.1109/ICICT57646.2023.10134464.

[84] 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, doi: 10.1038/s41598-025-09101-z.

[85] R. Aman, M. Rizwan, and A. Kumar, “Comparative Analysis of Transfer Learning Techniques for Enhanced Fault Detection in Solar Photovoltaic Modules,” *2024 IEEE 4th International Conference on Sustainable Energy and Future Electric Transportation, SEFET 2024*, pp. 1–6, 2024, doi: 10.1109/SEFET61574.2024.10718126.

[86] L. Koshy, M. V. Vaishnav, S. Sunil, S. V. Abraham, and S. Vidhyadharan, “A Thermal Image-based Fault Detection System for Solar Panels,” *2024 International Conference on Cybernation and Computation, CYBERCOM 2024*, pp. 246–250, 2024, doi: 10.1109/CYBERCOM63683.2024.10803145.

[87] E. Tsalera, A. Papadakis, M. Samarakou, and I. Voyiatzis, “Feature Extraction with Handcrafted Methods and Convolutional Neural Networks for Facial Emotion Recognition,” *Applied Sciences (Switzerland)*, vol. 12, no. 17, 2022, doi: 10.3390/app12178455.

[88] Z. Y. Shen, H. De Shen, and F. He, “A hybrid deep learning-based model for enhanced feature representation in image retrieval systems,” *Egyptian Informatics Journal*, vol. 31, no. February 2025, doi: 10.1016/j.eij.2025.100717.

[89] J. Ma, X. Jiang, A. Fan, J. Jiang, and J. Yan, “Image Matching from Handcrafted to Deep Features: A Survey,” *Int. J. Comput. Vis.*, vol. 129, no. 1, pp. 23–79, 2021, doi: 10.1007/s11263-020-01359-2.

[90] A. Oulefki, Y. Himeur, T. Trongtirakul, K. Amara, and S. Agaian, “Detection and analysis of deteriorated areas in solar PV modules using unsupervised sensing algorithms and 3D augmented reality,” *Heliyon*, vol. 10, no. 6, p. e27973, 2024, doi: 10.1016/j.heliyon. 2024.e27973.

[91] R. Wang, Z. Zhou, S. Li, and Z. Zhang, “Advances and challenges in infrared-visible image fusion: a comprehensive review of techniques and applications,” *Artif. Intell. Rev.*, vol. 59, no. 1, 2026, doi: 10.1007/s10462-025-11426-0.

[92] P. Gonidakis, A. Sonora-Mengana, B. Jansen, and J. Vandemeulebroucke, “Handcrafted Features Can Boost Performance and Data-Efficiency for Deep Detection of Lung Nodules from CT Imaging,” *IEEE Access*, vol. 11, no. October, pp. 126221–126231, 2023, doi: 10.1109/ACCESS.2023.3331315.

[93] P. Manon, L. Beuque, D. Ing, M. B. I. Lobbes, Y. Van Wijk, and Y. Widaatalla, “Combining Deep Learning and Handcrafted Radiomics for Classification of Suspicious Lesions on Contrast-enhanced Mammograms,” no. Ml, 2023.

[94] S. Tang, Y. Xing, L. Chen, X. Song, and F. Yao, “Review and a novel strategy for mitigating hot spot of PV panels,” *Solar Energy*, vol. 214, no. November 2020, pp. 51–61, 2021, doi: 10.1016/j.solener.2020.11.047.

[95] Almas, S. Sundaram, and U. D. Dwivedi, “Predictive analysis of power degradation rate in solar PV systems emphasizing hot spots and visual effects-based failure modes,” *Renew. Energy*, vol. 228, no. November 2023, p. 120684, 2024, doi: 10.1016/j.renene.2024.120684.

[96] G. Goudelis, P. I. Lazaridis, and M. Dhimish, “A Review of Models for Photovoltaic Crack and Hotspot Prediction,” *Energies (Basel).*, vol. 15, no. 12, 2022, doi: 10.3390/en15124303.

[97] S. Kanthalakshmi, S. Maalathy, and P. S. Kumar, “Image Processing and CNN based Manufacturing Defect Detection and Classification of Faults in Photovoltaic Cells,” *Ictact Journal on Image and Video Processing*, vol. 9102, no. February, p. 3, 2024, doi: 10.21917/ijivp.2024.0460.

[98] Y. Wang, Y. Shen, C. Li, K. Zhang, and H. Wei, “Hotspot detection of photovoltaic modules in infrared thermal image based on saliency analysis,”

Proceedings of the 34th Chinese Control and Decision Conference, CCDC 2022, no. 2018, pp. 1479–1484, 2022, doi: 10.1109/CCDC55256.2022.10033497.

[99] G. B. Balachandran, M. Devisridhivyadharshini, M. E. Ramachandran, and R. Santhiya, "Comparative investigation of imaging techniques, pre-processing and visual fault diagnosis using artificial intelligence models for solar photovoltaic system – A comprehensive review," *Measurement (Lond).*, vol. 232, no. January 2024, doi: 10.1016/j.measurement.2024.114683.

[100] C. Buerhop, L. Bommes, J. Schlipf, T. Pickel, and A. Fladung, "Infrared imaging of photovoltaic modules: a review of the state of the art and future challenges facing gigawatt photovoltaic power stations," *Progress in Energy*, vol. 4, no. 4, pp. 0–24, 2022, doi: 10.1088/2516-1083/ac890b.

[101] A. Najiah Nurul Afifah, Indrabayu, A. Suyuti, and Syafaruddin, "A New Approach for Hot Spot Solar Cell Detection based on Multi-level Otsu Algorithm," *Proceedings - 2021 International Seminar on Intelligent Technology and Its Application: Intelligent Systems for the New Normal Era, ISITIA 2021*, pp. 278–282, 2021, doi: 10.1109/ISITIA52817.2021.9502239.

[102] H. Ling, M. Liu, and Y. Fang, "Deep Edge-Based Fault Detection for Solar Panels," *Sensors*, vol. 24, no. 16, pp. 1–15, 2024, doi: 10.3390/s24165348.

[103] V. Nancy and G. Balakrishnan, "Thermal image-based object detection and identification on varying environmental conditions," *Journal of Environmental Protection and Ecology*, vol. 21, no. 4, pp. 1426–1436, 2020.

[104] J. Beck, "Quality aspects of annotated data," *AStA Wirtschafts- und Sozialstatistisches Archiv*, vol. 17, no. 3–4, pp. 331–353, 2023, doi: 10.1007/s11943-023-00332-y.

[105] I. H. Sarker, "Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 6, pp. 1–20, 2021, doi: 10.1007/s42979-021-00815-1.

[106] R. Archana and P. S. E. Jeevaraj, *Deep learning models for digital image processing: a review*, vol. 57, no. 1. Springer Netherlands, 2024. doi: 10.1007/s10462-023-10631-z.

[107] M. Benghanem, A. Mellit, and C. Moussaoui, "Embedded Hybrid Model (CNN–ML) for Fault Diagnosis of Photovoltaic Modules Using Thermographic Images," *Sustainability (Switzerland)*, vol. 15, no. 10, 2023,

doi: 10.3390/su15107811.

[108] 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, doi: 10.1109/M2GARSS52314.2022.9840254.

[109] L. Pratt, J. Mattheus, and R. Klein, “A benchmark dataset for defect detection and classification in electroluminescence images of PV modules using semantic segmentation,” *Systems and Soft Computing*, vol. 5, no. November 2022, p. 200048, 2023, doi: 10.1016/j.sasc.2023.200048.

[110] F. Mazen Ali Mazen, Y. O. Shaker, and R. Ahmed Abul Seoud, “Attention-Based SegNet: Toward Refined Semantic Segmentation of PV Modules Defects,” *IEEE Access*, vol. 12, no. July, pp. 100792–100804, 2024, doi: 10.1109/ACCESS.2024.3431098.

[111] J. Swami and V. Kumar, “Solar Panel Problem of Hotspot and Detection and Their Possible Solutions,” *International Journal of Creative Research Thoughts*, vol. 11, no. 5, pp. 2320–2882, 2023, [Online]. Available: www.ijcrt.org

[112] U. Rashid, “Predictive Maintenance for Solar Photovoltaic Systems: Deep Learning-based Anomaly Identification using Infrared Thermography,” 2025.

[113] A. W. Kandeal, M. R. Elkadeem, A. Kumar Thakur, G. B. Abdelaziz, and R. Sathyamurthy, “Infrared thermography-based condition monitoring of solar photovoltaic systems: A mini review of recent advances,” *Solar Energy*, vol. 223, no. May, pp. 33–43, 2021, doi: 10.1016/j.solener.2021.05.032.

[114] Y. Zefri, I. Sebari, H. Hajji, and G. Aniba, “Developing a deep learning-based layer-3 solution for thermal infrared large-scale photovoltaic module inspection from orthorectified big UAV imagery data,” *International Journal of Applied Earth Observation and Geoinformation*, vol. 106, 2022, doi: 10.1016/j.jag.2021.102652.

[115] G. Tanda and M. Migliazzi, “Infrared thermography monitoring of solar photovoltaic systems: A comparison between UAV and aircraft remote sensing platforms,” *Thermal Science and Engineering Progress*, vol. 48, no. October 2023, p. 102379, 2024, doi: 10.1016/j.tsep.2023.102379.

[116] A. Haeruman, S. Ul Haq, M. Mohandes, S. Rehman, and S. S. I. Mitu, “AI-Based PV Panels Inspection using an Advanced YOLO Algorithm,” *Materials Research Proceedings*, vol. 43, pp. 230–237, 2024, doi: 10.21741/9781644903216-30.

[117] P. Haidari, A. Hajiahmad, A. Jafari, and A. Nasiri, “Deep learning-based model for fault classification in solar modules using infrared images,” *Sustainable Energy Technologies and Assessments*, vol. 52, no. PA, p. 102110, 2022, doi: 10.1016/j.seta.2022.102110.

[118] S. F. Bello, I. U. Wada, O. B. Ige, and E. C. Chianumba, “AI-driven predictive maintenance and optimization of renewable energy systems for enhanced operational efficiency and longevity,” 2024.

[119] A. Venckauskas, J. Toldinas, and N. Morkevicius, “Improving Multi-Class Classification for Recognition of the Prioritized Classes Using the Analytic Hierarchy Process,” *Applied Sciences (Switzerland)*, vol. 15, no. 13, pp. 1–18, 2025, doi: 10.3390/app15137071.

[120] D. B. Olawade, O. Z. Wada, A. O. Ige, B. I. Egbewole, A. Olojo, and B. I. Oladapo, “Artificial intelligence in environmental monitoring: Advancements, challenges, and future directions,” *Hygiene and Environmental Health Advances*, vol. 12, no. August, p. 100114, 2024, doi: 10.1016/j.heha.2024.100114.

[121] A. Thakfan and Y. bin Salamah, “Development and Performance Evaluation of a Hybrid AI-Based Method for Defects Detection in Photovoltaic Systems,” *Energies (Basel).*, vol. 18, no. 4, pp. 1–19, 2025, doi: 10.3390/en18040812.

[122] S. Gallardo-Saavedra, L. Hernández-Callejo, and Ó. Duque-Pérez, “Analysis and characterization of PV module defects by thermographic inspection,” *Revista Facultad de Ingenieria*, no. 93, pp. 92–104, 2019, doi: 10.17533/udea.redin.20190517.

[123] N. R. E. Laboratory, S. A. Sandia National Laboratory, and the S. N. L. M. Partnership, and (SuNLaMP) PV O&M Best Practices Working Group, “Best Practices for Operation and Maintenance of Photovoltaic and Energy Storage Systems,” *Nrel*, no. December, p. 153, 2018, [Online]. Available: https://www.nrel.gov/research/publications.html

[124] R. Iea-pvps, *Review on Infrared and Electroluminescence Imaging for PV*

Field Applications. 2018.

[125] Y. Lei, X. Wang, A. An, and H. Guan, "Deeplab-YOLO: a method for detecting hot-spot defects in infrared image PV panels by combining segmentation and detection," *J. Real. Time. Image Process.*, vol. Volume 21, no. Issue 2, 2024, doi: 10.1007/s11554-024-01415-x.

[126] S. Mislav, T. Mirko, and M. Leo, "Using CNNs for Photovoltaic Panel Defect Detection via Infrared Thermography to Support Industry 4.0," *Business Systems Research*, vol. 15, no. 1, pp. 45–66, 2024, doi: 10.2478/bsrj-2024-0003.

[127] S. F. Ahmed, M. S. Bin Alam, M. Hassan, M. R. Rozbu, and T. Ishtiak, *Deep learning modelling techniques: current progress, applications, advantages, and challenges*, vol. 56, no. 11. Springer Netherlands, 2023. doi: 10.1007/s10462-023-10466-8.

[128] S. Kumar, P. Asiamah, O. Jolaoso, and U. Esiowu, "Enhancing Image Classification with Augmentation: Data Augmentation Techniques for Improved Image Classification," 2025, [Online]. Available: http://arxiv.org/abs/2502.18691

[129] 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, doi: 10.1109/ACCESS.2024.3505754.

[130] D. Prattic, F. Lagan, M. Versaci, D. Frankovi, and A. Jakopli, "Efficient Attentive U-Net for Fault Diagnosis and Predictive Maintenance of Photovoltaic Panels Through Infrared Thermography," pp. 1–28, 2025.

[131] A. Hamza, Z. Ali, S. Dudley, K. Saleem, M. Uneeb, and N. Christofides, "A multi-stage review framework for AI-driven predictive maintenance and fault diagnosis in photovoltaic systems," *Appl. Energy*, vol. 393, no. March, p. 126108, 2025, doi: 10.1016/j.apenergy.2025.126108.

[132] Y. Hong, R. Pan, J. Su, and M. Li, "Infrared image detection of defects in lightweight solar panels based on improved MSRCR and YOLOv8n," *Infrared Phys. Technol.*, vol. 141, no. April, p. 105473, 2024, doi: 10.1016/j.infrared.2024.105473.

[133] U. R. Qureshi, A. Rashid, N. Altini, V. Bevilacqua, and M. La Scala,

"Radiometric Infrared Thermography of Solar Photovoltaic Systems: An Explainable Predictive Maintenance Approach for Remote Aerial Diagnostic Monitoring," *Smart Cities*, vol. 7, no. 3, pp. 1261–1288, 2024, doi: 10.3390/smartcities7030053.

[134] X. Zhang, H. Wu, K. Qi, Y. Qian, and Y. Zhang, "Detailed PV Monitor: A Highly Generalized Photovoltaic Panels Segmentation Network Integrating Context-Aware and Deep Feature Reconstruction," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 18, pp. 10131–10143, 2025, doi: 10.1109/JSTARS.2025.3558471.

[135] O. A. Rozak, M. Z. Daud, S. Bakhri, and M. Z. Ibrahim, "Review of the Impact of Tropical Climate on Solar Photovoltaic Modules Performances Using the Combination of Cooling Systems," *J. Sustain. Sci. Manag.*, vol. 18, no. 1, pp. 228–236, 2023, doi: 10.46754/jssm.2023.01.014.

[136] Z. Feng, M. L. Tan, M. A. Mahamud, J. Chuah, and F. Zhang, "Impacts of solar radiation modification on temperature extremes and heatwaves in Southeast Asia," *Weather Clim. Extrem.*, vol. 49, no. December 2024, p. 100789, 2025, doi: 10.1016/j.wace.2025.100789.

[137] M. Mustafa and M. O. F. Malik, "Factors Hindering Solar Photovoltaic System Implementation in Buildings and Infrastructure Projects: Analysis through a Multiple Linear Regression Model and Rule-Based Decision Support System," *Buildings*, vol. 13, no. 7, 2023, doi: 10.3390/buildings13071786.

[138] M. Dhimish, "HOTSPOT-YOLO: A Lightweight Deep Learning Attention-Driven Model for Detecting Thermal Anomalies in Drone- Based Solar Photovoltaic Inspections," pp. 1–29.

[139] H. Ayunts, S. Agaian, and A. Grigoryan, "SlantNet: A Lightweight Neural Network for Thermal Fault Classification in Solar PV Systems," *Electronics (Switzerland)*, vol. 14, no. 7, pp. 1–20, 2025, doi: 10.3390/electronics14071388.

[140] M. Afridi, A. Kumar, F. ibne Mahmood, and G. S. TamizhMani, "Comparative Analysis of Hotspot Stress Endurance in Pristine and Thermal Cycled Prestressed Glass–Glass Photovoltaic Modules," *Sustainability (Switzerland)*, vol. 15, no. 16, 2023, doi: 10.3390/su151612131.

[141] M. Dhimish and G. Badran, "Investigating defects and annual degradation in

UK solar PV installations through thermographic and electroluminescent surveys," *Npj Mater. Degrad.*, vol. 7, no. 1, 2023, doi: 10.1038/s41529-023-00331-y.

[142] M. Jamali, P. Davidsson, R. Khoshkangini, M. G. Ljungqvist, and R. C. Mihailescu, "Context in object detection: a systematic literature review," *Artif. Intell. Rev.*, vol. 58, no. 6, 2025, doi: 10.1007/s10462-025-11186-x.

[143] D. Bhatt, C. Patel, H. Talsania, J. Patel, R. Vaghela, and S. Pandya, "CNN variants for computer vision: History, architecture, application, challenges and future scope," *Electronics (Switzerland)*, vol. 10, no. 20, pp. 1–28, 2021, doi: 10.3390/electronics10202470.

[144] F. Ghribi and F. Hamdaoui, "Innovative deep learning architectures for medical image diagnosis: a comprehensive review of convolutional, recurrent, and transformer models," *Visual Computer*, vol. 41, no. 13, pp. 11603–11628, 2025, doi: 10.1007/s00371-025-04122-1.

[145] F. Wang, Z. Wang, Z. Chen, D. Zhu, X. Gong, and W. Cong, "An Edge-Guided Deep Learning Solar Panel Hotspot Thermal Image Segmentation Algorithm," *Applied Sciences (Switzerland)*, vol. 13, no. 19, 2023, doi: 10.3390/app131911031.

[146] K. Zhang, M. He, L. Dong, and C. Ou, "The Application of Tsallis Entropy Based Self-Adaptive Algorithm for Multi-Threshold Image Segmentation," *Entropy*, vol. 26, no. 9, 2024, doi: 10.3390/e26090777.

[147] A. J. P. Hamadi, Halim, Ni Mas Zahra Zhafira, "Quality improvement and segmentation of mammography images using Python," in *AIP Conference Proceeding*, 2023.

[148] D. Matusz-Kalász, I. Bodnár, and M. Jobbágy, "An Overview of CNN-Based Image Analysis in Solar Cells, Photovoltaic Modules, and Power Plants," *Applied Sciences (Switzerland)*, vol. 15, no. 10, 2025, doi: 10.3390/app15105511.

[149] K. Awedat, G. Comert, M. Ayad, and A. Mrebit, "Advanced fault detection in photovoltaic panels using enhanced U-Net architectures," *Machine Learning with Applications*, vol. 20, no. February, p. 100636, 2025, doi: 10.1016/j.mlwa.2025.100636.

[150] 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, doi: 10.1038/s41598-024-81101-x.

[151] R. Ashwini, M. Dhiraj, and M. Yogita, "Performance of fault classification on Photovoltaic modules using Thermographic images.," 2022, pp. 1–6.

[152] M. Krichen, "Convolutional Neural Networks: A Survey," *Computers*, vol. 12, no. 8, pp. 1–41, 2023, doi: 10.3390/computers12080151.

[153] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image Segmentation Using Deep Learning: A Survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, 2022, doi: 10.1109/TPAMI.2021.3059968.

[154] B. Li, R. Xia, J. Li, J. Zhang, and Z. Zhang, "Multimodal deep learning with hyperspectral imaging for accurate origin classification of wolfberries," *Food Chem. X*, vol. 31, no. October 2025, doi: 10.1016/j.fochx.2025.103166.

[155] B. Wang, Y. Zou, L. Zhang, Y. Li, Q. Chen, and C. Zuo, "Multimodal super-resolution reconstruction of infrared and visible images via deep learning," *Opt. Lasers Eng.*, vol. 156, no. March 2022, doi: 10.1016/j.optlaseng.2022.107078.

[156] M. Hussain, M. O'Nils, J. Lundgren, and S. J. Mousavirad, "A Comprehensive Review on Deep Learning-Based Data Fusion," *IEEE Access*, vol. 12, no. December, pp. 180093–180124, 2024, doi: 10.1109/ACCESS.2024.3508271.

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

[158] R. Kv, K. Prasad, and P. Peralam Yegneswaran, "Segmentation and Classification Approaches of Clinically Relevant Curvilinear Structures: A Review," *J. Med. Syst.*, vol. 47, no. 1, 2023, doi: 10.1007/s10916-023-01927-2.

[159] U. Mamodiya, I. Kishor, R. Garine, P. Ganguly, and N. Naik, "Artificial intelligence-based hybrid solar energy systems with smart materials and adaptive photovoltaics for sustainable power generation," *Sci. Rep.*, vol. 15, no. 1, 2025, doi: 10.1038/s41598-025-01788-4.

[160] A. Z. Muhammad Umair Ali, Hafiz Farhaj Khan, Manzar Masud, Karam Dad Kallu, "A machine learning framework to identify the hotspot in photovoltaic

module using infrared thermography," *Solar Energy*, pp. 643–651, 2020, doi: https://doi.org/10.1016/j.solener.2020.08.027.

[161] 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, 2024, doi: 10.1155/2024/5586605.

[162] Y. L. Yang, Kaixuan, Wei Xiang, Zhenshuai Chen, Jian Zhang, "A review on infrared and visible image fusion algorithms based on neural networks☆," *J. Vis. Commun. Image Represent.*, vol. 101, no. March 2024.

[163] J. Meng, J. Zou, Z. Xiang, C. Wang, and S. Wang, "Visible and thermal image fusion network with diffusion models for high-level visual tasks," *Applied Intelligence*, vol. 55, no. 4, 2025, doi: 10.1007/s10489-024-06210-6.

[164] B. Sandeep, D. Saiteja Reddy, R. Aswin, and R. Mahalakshmi, "Monitoring of PV Modules and Hotspot Detection Using Convolution Neural Network Based Approach," *IFIP Adv. Inf. Commun. Technol.*, vol. 654 IFIP, no. Icears, pp. 311–323, 2022, doi: 10.1007/978-3-031-16364-7_24.

APPENDICES

APPENDIX 1

Expert Evaluation Form for Classification of Solar PV Hotspots in Thermal Images

Section 1 of 5

Expert Evaluation Form for Classification of Solar PV Hotspots in Thermal Images

Assalamualaikum wbt and Hello, Prof/Dr./Sir/Madam.

My name is Nurul Huda binti Ishak, a doctorate candidate of Doctor of Philosophy (Electrical Engineering) student at Universiti Teknologi MARA (UiTM), Permatang Pauh Campus. I am hoping for your assistance as a thermal image expert in responding to an expert evaluation form for the classification of hotspots in thermal images on classifying hotspot severity conditions.

This evaluation form includes 2 sections. The first section is related to assessor background information and opinions on the computer-aided diagnosis (CAD) system. The second section contains PV aerial thermal imagery evaluation on whole images.

This Expert Evaluation Form is designed to assess the evaluation of expert/manual/visual diagnosis in segmented hotspot regions in thermal images based on the expert selection. Your responses will be used as the qualitative assessment of the CAD classification model developed in this study.

This feedback is extremely beneficial to this research in terms of developing new techniques for developing a CAD system for hotspot severity classification. Thank you for taking the time to fill out this evaluation form.

Please take your time (within the stipulated time given) to evaluate each image carefully.

Rest assured that your response is strictly private and confidential. It will be utilised exclusively for this study and will not be publicised among the assessors.

Thank you for your time and feedback.

This form is automatically collecting emails from all respondents. Change settings

Email *

Short answer text

Section 2 of 5

Section 1: Assessor Details

Please answer based on your expertise and background in assessing thermal images.

Description (optional)

Name: *

Short answer text

Occupation and position. *

Short answer text

Current Institution. *

Short answer text

Are you previously/currently involved in solar PV activity in your institution? *

- ◯ Yes
- ◯ No

How long is your experience in assessing thermal images (year)? *

- ◯ less than 1 year
- ◯ 1-2 years
- ◯ 2-3 years
- ◯ 4-5 years
- ◯ more than 5 years

Section 3 of 5

Section 2: Computer-Aided Diagnosis (CAD) System for Hotspot Analysis

Description (optional)

1) Do you think that the implementation of a CAD system is important for maintaining PV system performance? *

- ◯ Yes
- ◯ No

2) Generally, are you satisfied with your existing CAD system? *

- ◯ Yes
- ◯ No

3) From your point of view, is it crucial to propose improvements to your existing CAD system? *

- ◯ Strongly support – The improvements are necessary and will enhance system performance.
- ◯ Support – The improvements are beneficial, though some areas could be further refined.
- ◯ Neutral – I neither support nor oppose the proposed improvements.
- ◯ Oppose – The improvements may not add significant value or may cause complications.
- ◯ Strongly oppose – The proposed changes are unnecessary or potentially harmful to the current system.

4) How does the CAD used to assist in maintaining the PV system process, and what are its key capabilities? *

- Increases efficiency and reduces processing time.
- Enhances accuracy and reduces errors.
- Provides advanced features (e.g. 3D modeling, automation).
- Improves collaboration and data sharing.
- Limited impact – does not significantly assist the process.
- Other: ____________

5) Have you used any infrared thermography and/or CAD systems in your institution/company (current and past)? *

- Yes
- No

6) What challenges have you encountered when using CAD in thermal image interpretation?(if any) *

- Difficulty in integrating thermal image data with CAD models.
- Limited accuracy in detecting thermal anomalies within the CAD environment.
- Lack of user-friendly features or interface for thermal analysis.
- Software limitations or compatibility issues.
- No significant challenges encountered.
- Other: ____________

7) Have you received any training or continuing education related to CAD in thermal interpretation? *

- Yes
- No

Section 4 of 5

Section 2 (Continue): Computer-Aided Diagnosis (CAD) System for Hotspot Analysis

Please rate the following questions based on your satisfaction levels (1=Strongly Disagree, 2=Moderately Disagree, 3=Neutral, 4=Moderately Agree, 5=Strongly Agree)

Do you think that if a CAD thermal system focusing on hotspot severity conditions is developed, it will contribute to new knowledge findings? *

	1	2	3	4	5
Strongly Disagr...	☐	☐	☐	☐	☐

Do you think that if a CAD thermal system focusing on hotspot severity conditions is developed, it will reduce the maintenance team's workload? *

	1	2	3	4	5
Strongly Disagr...	☐	☐	☐	☐	☐

Do you think that developing a CAD thermal system that focuses on hotspot severity conditions will induce collaboration with current artificial intelligence technology? *

	1	2	3	4	5
Strongly Disagr...	☐	☐	☐	☐	☐

Do you think that if a CAD thermal system focusing on hotspot severity conditions is developed, it will complement or enhance your current workflow? *

	1	2	3	4	5
Strongly Disagr...	☐	☐	☐	☐	☐

If the CAD thermal system focusing on hotspot severity conditions is developed, as an expert, I definitely will use the system. *

	1	2	3	4	5
Strongly Disagr...	☐	☐	☐	☐	☐

What are your expectations when/if using thermal CAD during inspection assessment? *

- ☐ Improved Detection Accuracy
- ☐ Second Opinion and Support
- ☐ Efficiency and Time Savings
- ☐ Reducing False Negatives
- ☐ Standardization of Interpretation
- ☐ Other: ....

Please provide additional comments/suggestions: *

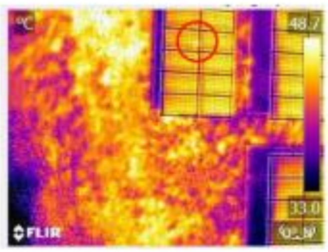
Short answer text

Section 5 of 5

Section 3: Hotspot Assessment

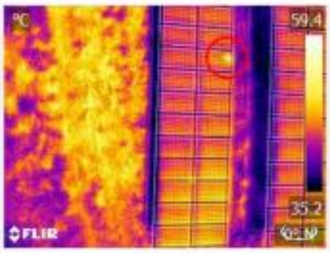
Please refer to the attached thermal images below for the classification of hotspot severity conditions from the solar farm dataset.

1) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



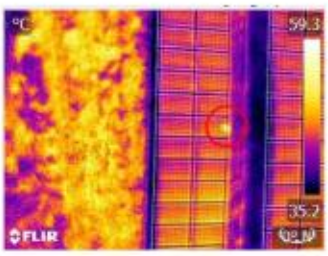
- Low
- Medium
- High

2) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



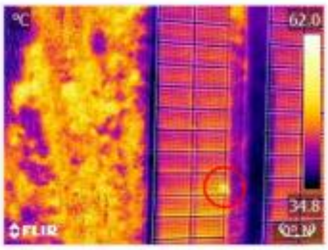
- Low
- Medium
- High

3) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



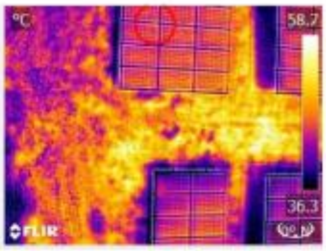
- Low
- Medium
- High

4) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



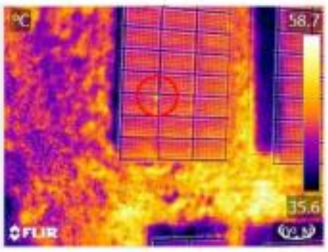
- Low
- Medium
- High

5) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



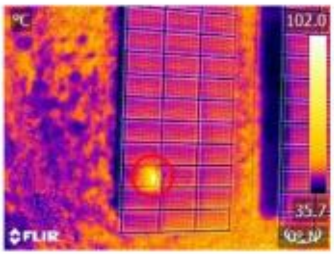
- Low
- Medium
- High

6) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



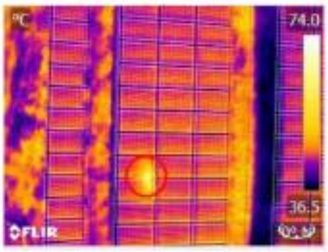
- Low
- Medium
- High

7) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



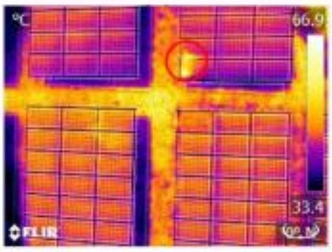
- Low
- Medium
- High

8) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



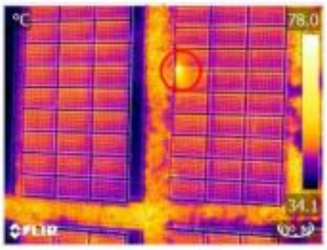
- Low
- Medium
- High

9) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



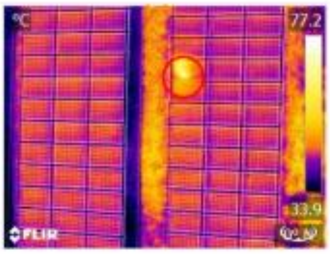
- Low
- Medium
- High

10) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



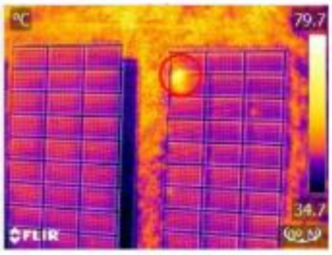
- Low
- Medium
- High

11) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



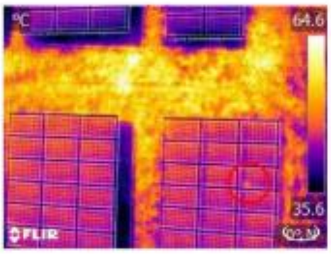
- Low
- Medium
- High

12) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



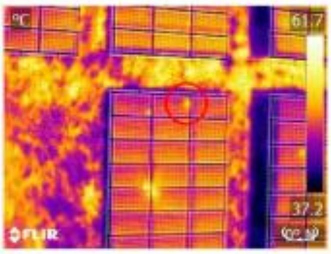
- Low
- Medium
- High

13) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



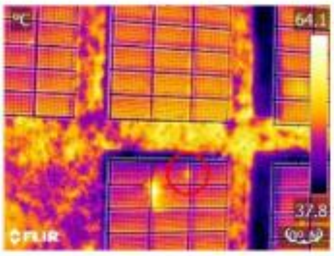
- Low
- Medium
- High

14) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



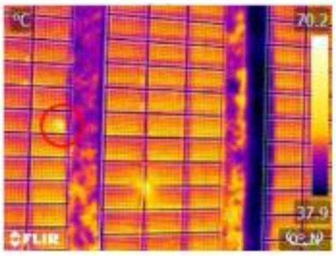
- Low
- Medium
- High

15) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



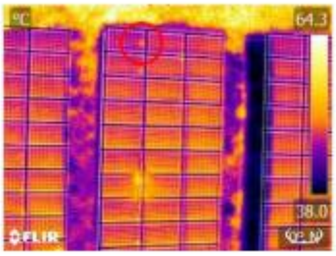
- Low
- Medium
- High

16) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



- Low
- Medium
- High

17) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *

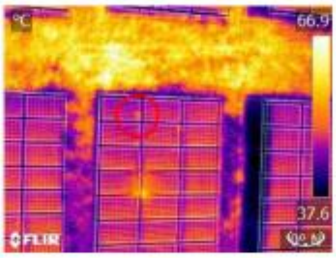


◯ Low

◯ Medium

◯ High

18) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *

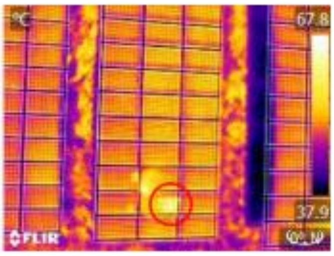


◯ Low

◯ Medium

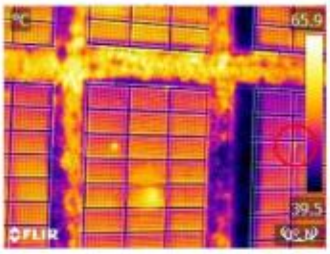
◯ High

19) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



- Low
- Medium
- High

20) This thermal image has been captured by a thermal camera. From your view, what is your assumption on hotspot severity as per the red circle? *



- Low
- Medium
- High

AUTHOR'S PROFILE

Nurul Huda binti Ishak completed her Bachelor of Electrical Engineering at the Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), Shah Alam, in 2006. She obtained her Master of Science in Electronic System Design Engineering from the School of Electrical and Electronics Engineering, Universiti Sains Malaysia, in 2009. She is currently a lecturer at Universiti Teknologi MARA (UiTM), Permatang Pauh Campus, Pulau Pinang, under the Department of Power, Faculty of Electrical Engineering. Her main research interests include power systems, thermal imaging, and artificial intelligence.

LIST OF PUBLICATION

Grant

1. Title: A Formulation of Cascaded Non-Subsampled Contourlet Transform (NSCT) Algorithm Based on Gradient Temperature Thermographic Image Features for Extracting Characteristics of Thermal Gradient in Fused Images.
   Code: FRGS/1/2023/ICT02/UITM/02/12
   Amount: RM 100,296.00
   From: Fundamental Research Grant (FRGS), Ministry of Higher Education, Malaysia.
   Members: **Nurul Huda Ishak**, Iza Sazanita Isa, Muhammad Khusairi Osman, Mohd Shawal Jadin, Kamarulazhar Daud and Nur Ashida Salim.
   Duration: 1 October 2023-30 September 2026

Referred Journal

Nurul Huda Ishak, Iza Sazanita Isa, Muhammad Khusairi Osman, Mohd Shawal Jadin, Kamarulazhar Daud and Mohd Zulhamdy Ab Hamid, "2TSS: Two-Tier Semantic Segmentation Framework With Enhancement for Hotspot Detection of Solar Photovoltaic Thermal Images," in IEEE Access, vol. 13, pp. 88888-88904, 2025, doi: 10.1109/ACCESS.2025.3570974. **WoS and Scopus Indexed, Status: Published.**

Mohd Zulhamdy Ab Hamid ,Kamarulazhar Daud, Zainal Hisham Che Soh, Muhammad Khusairi Osman, Iza Sazanita Isa and **Nurul Huda Ishak**, “ Enhanced Solar Panel Segmentation and Hotspot Recognition Using U-Net: A Multiclass Semantic Segmentation Approach”, in Journal of Electrical & Electronic System Research, 2025. **Scopus Indexed, Status: Published.**

Nurul Huda Ishak, Iza Sazanita Isa, Muhammad Khusairi Osman, Mohd Shawal Jadin, Kamarulazhar Daud and Noor Fadzilah Razali, "Performance Improvement of CNN-Based Model for Multiclass Hotspot Severity Classification in Photovoltaic Thermal Imagery” in ESTEEM Academic Journal, vol.22, doi: 10.24191/esteem.v22March.10388. **Mycite.**

Proceedings

1. **Nurul Huda Ishak**, Iza Sazanita Isa, Muhammad Khusairi Osman, Kamarulazhar Daud and Mohd Zulhamdy Ab Hamid, "Hotspot Detection of Solar Photovoltaic System: A Perspective from Image Processing," in 2023 IEEE 3rd International Conference in Power Engineering Applications (ICPEA), Putrajaya, Malaysia, 2023, **Scopus Indexed, Status: Published.**
2. Mohd Zulhamdy Ab Hamid, Kamarulazhar Daud, Zainal Hisham Che Soh, Muhammad Khusairi Osman, Iza Sazanita Isa and **Nurul Huda Ishak**, "Review of Deep Learning-Based Hotspot Detection in Solar Photovoltaic Arrays," in 2024 IEEE 4th International Conference in Power Engineering Applications (ICPEA), Pulau Pinang, Malaysia, 2024, **Scopus Indexed, Status: Published.**
3. **Nurul Huda Ishak**, Iza Sazanita Isa, Muhammad Khusairi Osman, Kamarulazhar Daud, Mohd Zulhamdy Ab Hamid and Mohd Shawal Jadin, "Automated CNN-based Semantic Segmentation for Thermal Image of Solar Photovoltaic (PV) Panel," in 2024 IEEE 14th International Conference on

Control System, Computing and Engineering (ICCSCE), Penang, Malaysia, 2024, **Scopus Indexed, Status: Published.**

4. Nuzul Lokmanhakim Zulkifli, Iza Sazanita Isa, **Nurul Huda Ishak**, Muhammad Khusairi Osman, Kamarulazhar Daud,Rosheila Darus and Normasni Ad Fauzi, “ Optimizing Thermal Image Segmentation Using K-Means Clustering and Morphological Operations for Photovoltaic”, in 2025 IEEE 15th International Conference on Control System, Computing and Engineering (ICCSCE), Penang, Malaysia,

Book Chapter

1. **Nurul Huda Ishak**, Muhammad Ameer Izzat Mohammad Halim and Iza Sazanita Isa, “Detection of Power Distribution Fault in Thermal Images Using CNN” for Intelligent Multimedia Signal Processing for Smart Ecosystems, Springer (2023). **Scopus Indexed, Status: Published.**

Copyright/Patent

1. Intelligence Hotspot Segmentation (IHS) for Solar Panels Based on Deep Learning.
- No. CRLY2024P04898, (2024). Intellectual Property Corporation Malaysia (MyIPO).
2. HDL-Hotspot: Hybrid Deep-Learning Tool for PV Hotspot Severity.
- No. CRDV2025P04262, (2025). Intellectual Property Corporation Malaysia (MyIPO).

Award

1. **Nurul Huda Ishak,** Iza Sazanita Isa, Muhammad Khusairi Osman, Mohd Shawal Jadin, Kamarulazhar Daud and Faizal Mohamad Twon Tawi, “Intelligence Hotspot Segmentation (IHS) for Solar Panels Based on Deep Learning” in AI & IOT International Innovation Expo 2024, **Gold Award**.
2. **Nurul Huda Ishak,** Iza Sazanita Isa, Muhammad Khusairi Osman, Kamarulazhar Daud and Mohd Shawal Jadin, “Postgraduate Research Image Competition 2025”. **3rd place**.
3. Nurul Huda Ishak, Iza Sazanita Isa, Muhammad Khusairi Osman, Mohd Shawal Jadin, Kamarulazhar Daud and Faizal Mohamad Twon Tawi, “HDL-

Hotspot: An Intelligent Hybrid Deep Learning Tool for Hotspot Severity Classification in Solar Photovoltaic Systems” in Penang International Invention, Innovation and Design (PIID2025), **Gold Award**.

APPENDIX

NO	TITTLE	HARDCOPY	SOFTCOPY
1	HOTSPOT LOCALIZATION AND SEVERITY CLASSIFICATION IN PHOTOVOLTAIC AERIAL THERMAL IMAGERY USING SEMANTIC SEGMENTATION AND CONVOLUTIONAL NEURAL NETWORK.		/