Zeshan Mubshir

Autonomous Monitoring and Classification of Solar PVs Anomalies using Deep Learning Methods

Master's thesis in Simulation and Visualization

Supervisor: Saleh Abdel-Afou Alaliyat
Co-supervisor: Mohammadreza Aghaei
July 2025

NTNU
Norwegian University of
Science and Technology

ABSTRACT

Solar energy is an emerging renewable energy source that has gained significant attention in recent years due to increasing demand for sustainable energy solutions. This has led to the exploration of various photovoltaic (PV) technologies, including land-based and floating photovoltaic systems. However, with widespread adoption comes the critical need for effective monitoring and maintenance strategies to ensure optimal performance, efficiency, and longevity of these installations.

This study focuses on developing advanced deep learning models for automated detection and classification of anomalies in photovoltaic systems using thermal imaging data. The research addresses the growing need for intelligent inspection systems that can identify various types of defects including hot spots, cell cracks, diode failures, soiling, vegetation blocking, shadowing, and offline modules. Using the InfraredSolarModules dataset containing over 20,000 thermal images representing 11 different anomaly types, multiple deep learning architectures were evaluated, including traditional convolutional neural networks (ResNet50, VGG16, EfficientNet-B0, Xception) and modern Vision Transformers (ViT-B32, DeiT-B16).

The methodology encompasses comprehensive evaluation across three classification scenarios: binary classification (anomaly vs. no anomaly), 11-class classification (anomaly types only), and 12-class classification (including normal conditions). Advanced visualization techniques, including t-SNE dimensionality reduction and Grad-CAM activation mapping, were employed to provide insights into learned feature representations and model decision-making processes.

Results demonstrate that Vision Transformer architectures achieved superior performance compared to traditional CNNs, with binary classification achieving over 98% accuracy in distinguishing between normal and anomalous conditions. The study reveals varying performance across different anomaly types, with some classes achieving F1-scores above 95% while others present more challenging detection scenarios due to class imbalance and subtle thermal signatures.

This research contributes to the advancement of automated solar panel inspection systems and provides a foundation for implementing intelligent maintenance strategies in photovoltaic installations. The findings support the development of cost-effective, scalable monitoring solutions that can enhance the reliability and efficiency of solar energy systems, ultimately contributing to the broader adoption of sustainable energy technologies.

SAMMENDRAG

Solenergi er en fremvoksende fornybar energikilde som har fått betydelig oppmerksomhet de siste årene. Den økende etterspørselen etter bærekraftige energiløsninger har ført til utforskning av ulike solenergi-teknologier, inkludert landbaserte og flytende fotovoltaiske (PV) systemer. Med utbredt bruk er det imidlertid behov for å evaluere ytelsen til disse systemene under forskjellige miljøforhold for å bestemme deres effektivitet, kostnadseffektivitet og miljøpåvirkning.

Hovedmålet med denne studien er å utvikle dyplæringsmodeller for å oppdage og klassifisere defekter i fotovoltaiske (PV) systemer, med spesielt fokus på landbaserte og flytende PV-systemer. Forskningen tar sikte på å addressere utfordringene knyttet til ytelsen til disse systemene, særlig når det gjelder deres effektivitet, kostnadseffektivitet og miljøpåvirkning. Studien vil benytte simuleringsverktøy for å modellere ytelsen til landbaserte og flytende PV-systemer under ulike miljøforhold, som solinnstråling, temperaturvariasjoner og påvirkningen av vannmasser på effektiviteten til flytende PV-systemer.

PREFACE

This thesis marks the culmination of my Master's degree in ICT-Simulation and Visualization at NTNU and represents a significant milestone in my academic journey.

I would like to express my sincere gratitude to my supervisor, **Prof. Dr. Saleh Abdel-Afou Alaliyat**, for his invaluable guidance and support throughout this research project. I am also grateful to my colleagues in the Department of ICT and Engineering for their valuable feedback and insightful discussions.

Special thanks go to my co-supervisor, **Prof. Mohammadreza Aghaei**, whose expertise and insights in this domain have greatly contributed to the quality and success of this thesis.

Finally, I wish to thank my family and friends for their unwavering encouragement and understanding during this journey. Their support has been a constant source of strength and motivation.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

- **ANN** Artificial Neural Network
- **CLAHE** Contrast Limited Adaptive Histogram Equalization
- **CNN** Convolutional Neural Network
- **DeiT** Data-efficient Image Transformer
- **DL** Deep Learning
- **DNN** Deep Neural Network
- **FF** Feed Forward
- **FPV** Floating Photovoltaic Systems
- **GAN** Generative Adversarial Network
- **ML** Machine Learning
- **MLP** Multi-Layer Perceptron
- **NLP** Natural Language Processing
- **PV** Photovoltaic Systems
- **t-SNE** t-distributed Stochastic Neighbor Embedding
- **UAV** Unmanned Aerial Vehicle
- **ViT** Vision Transformer

INTRODUCTION

This chapter provides an overview of the research topic, including the background, motivation, scope, objectives, and significance of the study. It also outlines the research questions and the thesis structure.

1.1   Background and Motivation

The motivation for this research is to address the growing concern about efficiency and reliability in solar photovoltaic (PV) systems. As renewable energy adoption accelerates globally, ensuring optimal performance of solar installations becomes increasingly critical for energy security and climate

goals. Additionally, there is a growing interest in the use of machine learning (ML) and deep learning (DL) techniques for the detection of anomalies in solar PVs. These techniques have shown promise in various domains, including image processing, natural language processing, and anomaly detection. These computational approaches can analyze complex patterns in data that might be imperceptible to human inspectors, potentially improving the accuracy and speed of fault detection. ML and DL models can be trained on large datasets of solar PV images to recognize visual signatures of different anomalies, from microcracks and hotspots to soiling and delamination.

The International Energy Agency (IEA) latest report shows that renewable energy capacities realized in 2022 reached a record 295 GW, with solar PV accounting for nearly 60% of this growth. The global installed capacity of solar PV systems has surpassed 1,000 GW, making it one of the fastest-growing energy sources worldwide. However, the performance of solar PV systems can be significantly affected by various anomalies, such as physical defects, soiling, and electrical faults. These anomalies can lead to reduced energy yield, increased maintenance costs, and even system failures. Traditional methods of anomaly detection in solar PVs often rely on manual inspection, which is time-consuming, labor-intensive, and prone to human error. As a result, there is a pressing need for automated anomaly detection systems that can efficiently and accurately identify issues in solar PV installations.

Traditional inspection methods rely on manual processes that are time-consuming, labor-intensive, and often inconsistent. Automated anomaly detec-

tion using machine learning and computer vision presents a promising alternative, offering potential for continuous monitoring, early fault detection, and reduced operational costs. However, significant challenges remain in developing robust models that can operate reliably across diverse environmental conditions, panel types, and anomaly categories.

This research aims to bridge these gaps by systematically investigating the complete pipeline from data preprocessing to model deployment, with particular emphasis on practical implementation challenges in real-world solar PV installations.

1.2 Objectives

Land-based and floating solar PV systems are increasingly deployed to meet global energy demands sustainably. However, the efficiency and reliability of these systems can be compromised by various anomalies, such as physical defects, soiling, and electrical faults. This research aims to develop a comprehensive framework for automated anomaly detection in solar PV systems using machine learning and computer vision techniques. The primary objectives of this research are:

- To explore and implement various data preprocessing techniques, including image augmentation and data cleaning, to enhance the performance of machine learning and deep learning models for solar PV anomaly detection.

- To evaluate and compare the performance of different machine learning and deep learning architectures in detecting and classifying specific types of anomalies in solar PV systems.

- To develop methodological frameworks that address challenges such as class imbalance, data scarcity, and environmental variability in computer vision-based solar PV anomaly detection systems.

1.3 Research Questions

This thesis aims to advance the field of automated anomaly detection in solar photovoltaic systems through the application of machine learning and computer vision techniques. Specifically, the research addresses the following questions:

- **Research Question 1:** How do various data preprocessing techniques, particularly image augmentation, influence the performance of deep learning models for solar PV anomaly detection?

- **Research Question 2:** Which deep learning (ViT and CNN variants) architectures demonstrate superior performance metrics (accuracy, precision, recall, F1-score) for detecting and classifying specific types of anomalies in solar PV systems?

- **Research Question 3:** What methodological frameworks and technical approaches best address the challenges of class imbalance, data scarcity in computer vision-based solar PV anomaly detection systems?

1.4 Significance of the Study

The significance of this study lies in its potential to enhance the efficiency and reliability of solar photovoltaic systems through advanced anomaly detection techniques. By leveraging machine learning and computer vision, this research aims to provide a robust framework for early fault detection, which can lead to reduced maintenance costs, increased energy yield, and extended system lifespan. This study contributes to the broader field of renewable energy by addressing the critical need for automated, reliable, and efficient monitoring solutions in solar PV installations. The findings are expected to have practical implications for solar energy operators, policymakers, and researchers, facilitating the wider adoption of solar technologies and supporting global efforts towards sustainable energy transition.

1.5 Thesis Structure

The structure of this thesis is organized as follows:

- **Chapter 1: Introduction** - Provides an overview of the research topic, including background, motivation, scope, objectives, and significance.

- **Chapter 2: Literature Review** - Reviews existing literature on solar PV anomaly detection, deep learning techniques, and computer vision applications in renewable energy.

- **Chapter 3: Methodology** - Describes the research design, including data collection, preprocessing techniques, model development, and evaluation metrics.

- **Chapter 4: Results and Discussion** - Presents the results of the experiments conducted and discusses their implications in the context of the research questions.

- **Chapter 5: Conclusion and Future work** - Summarizes the key findings, contributions of the research, and suggests directions for future work.

- **Chapter 6: References** - Lists the references used in the thesis.

- **Appendix A: IR Dataset** - Provides a detailed description of the IR dataset used in the research.

1.6 Scope

The scope of this research is to investigate the use of machine learning (ML) and deep learning (DL) techniques for the detection of anomalies in solar PVs. The research will focus on the development of algorithms that can be used to automatically detect anomalies in solar PVs, and the evaluation of these algorithms using real-world data. The research will also investigate the use of data preprocessing techniques, such as data augmentation, to improve the performance of the

deep learning algorithms. The research will also investigate the use of different deep learning architectures, such as Vision Transformers (ViT) and Convolutional Neural Networks (CNNs),to improve the performance of the anomaly detection algorithms.

LITERATURE REVIEW

This chapter provides a comprehensive overview of the theoretical background of photovoltaic (PV) systems, including both land-based and floating installations. It covers the fundamental concepts, anomaly types, detection methods, and relevant technologies used in this thesis. The theory is presented in an accessible manner to provide a solid foundation for understanding the methodologies and algorithms employed in this research.

2.1 Photovoltaic Systems

2.1.1 Land-based Photovoltaic (PV) Systems

Land-based photovoltaic (PV) systems are solar energy installations deployed on terrestrial surfaces, including ground-mounted installations and rooftop systems. These systems convert sunlight directly into electricity through the photovoltaic effect, utilizing semiconductor materials typically based on silicon technology.

Key components of land-based PV systems include:

- **Solar panels**: Photovoltaic modules containing solar cells (typically silicon-based) that convert sunlight into direct current (DC) electricity

- **Inverters**: Power electronic devices that convert DC electricity to alternating current (AC) for grid compatibility

- **Mounting structures**: Mechanical systems that securely position and orient panels for optimal solar irradiance capture

- **Electrical components**: Wiring, combiner boxes, disconnect switches, and monitoring systems

Land-based PV systems offer versatility in deployment, ranging from small residential installations to utility-scale solar farms spanning hundreds of acres. According to the International Energy Agency (IEA), land-based PV represents one of the fastest-growing renewable energy sources globally, driven by declining costs, improved efficiency, and favorable policy frameworks.

Recent research has demonstrated significant advances in land-based PV technology. Belloni et al. (2023) showed how machine learning optimization can

improve operations and maintenance (O&M) efficiency, while Bindi et al. (2022) documented dramatic cost reductions that have made PV competitive with conventional fossil fuel generation. Environmental assessments indicate that modern land-based PV systems have minimal land-use impact and short energy payback times.

2.1.2 Floating Photovoltaic (FPV) Systems

Floating photovoltaic (FPV) systems, also known as floating solar or floatovoltaics, represent an innovative approach to solar energy deployment on water bodies such as reservoirs, lakes, ponds, and coastal areas. These systems consist of solar panels mounted on specialized floating platforms or pontoons that maintain stability and optimal orientation while floating on the water surface.

FPV systems offer several distinct advantages over traditional land-based installations:

- **Land conservation**: FPV systems eliminate competition for valuable terrestrial land resources, making them particularly attractive in densely populated regions or areas with high land costs

- **Enhanced efficiency through natural cooling**: Water bodies provide natural cooling effects that reduce panel operating temperatures, typically resulting in 5-15% higher electrical output compared to equivalent land-based systems. Field studies have documented efficiency improvements of up to 11% due to this cooling effect

- **Water conservation benefits**: Solar panels provide shading that significantly reduces water evaporation from reservoirs and water bodies. Empirical studies have shown evaporation reductions of 42-83%, with some installations achieving 60% reduction in water loss

- **Superior lifecycle performance**: FPV systems often demonstrate better energy yield and lifecycle assessment (LCA) metrics, with energy payback times of approximately 1.3 years and lower greenhouse gas emissions compared to land-based alternatives

- **Grid integration synergies**: FPV installations can leverage existing hydroelectric infrastructure, including transmission lines and grid connections, reducing overall system costs and complexity

Table 2.1: Comparison of Land-based and Floating PV Systems

Characteristics	Land-based PV	Floating PV
Efficiency Improvement	Baseline	5-15% higher
Water Evaporation Reduction	N/A	42-83%
Land Use	High requirement	Zero requirement
Energy Payback Time	1.5-2 years	1.3 years
Cooling Effect	Natural air	Water cooling
Installation Complexity	Low	Moderate-High

Table **??** summarizes the key differences between land-based and floating PV systems. The advantages of FPV systems are particularly evident in their enhanced efficiency due to water cooling, significant water conservation benefits, and elimination of land use competition.

2.2   Anomaly Types in Photovoltaic Systems

Understanding the various types of anomalies that can affect PV systems is crucial for effective monitoring and maintenance strategies. These anomalies can significantly impact system performance, efficiency, and longevity.

2.2.1   Thermal Anomalies

Hot Spots: Hot spots occur when localized areas of a solar panel become significantly hotter than surrounding regions. This phenomenon can result from:

- Partial shading causing current mismatch

- Manufacturing defects in individual cells

- Degraded or damaged cells with higher resistance

- Poor electrical connections

Hot spots can lead to accelerated degradation, reduced efficiency, and potential safety hazards if left unaddressed.

2.2.2   Physical Defects

Cell Cracks: Physical damage to solar cells can occur during various stages:

- Manufacturing processes

- Transportation and handling

- Installation procedures

- Environmental stresses (thermal cycling, mechanical loads)

Cell cracks can create electrical disconnections, leading to power losses and potential hot spot formation.

Module-level Damage: Structural damage to entire modules, including frame deformation, glass breakage, or encapsulant degradation.

2.2.3   Environmental Impact Anomalies

Soiling: The accumulation of particulate matter on panel surfaces, including:

- Dust and dirt deposits

- Bird droppings

- Pollen and organic debris

- Industrial pollutants

Soiling can significantly reduce light transmission and energy output.
  Shading: Obstruction of solar irradiance by external objects:

- Vegetation growth

- Adjacent structures or panels

- Temporary objects

- Cloud shadows

2.2.4   Electrical Anomalies

Diode Failures: Bypass diode malfunctions that can affect module-level performance and protection.

Module Disconnection: Complete electrical isolation of modules due to connector failures or wiring issues.

2.2.5   Floating PV-Specific Anomalies

FPV systems face additional challenges related to their aquatic environment:

- **Algae growth**: Biological fouling on panel surfaces

- **Water-related corrosion**: Accelerated degradation due to humidity and water exposure

- **Structural instability**: Platform movement or anchoring system failures

- **Marine biofouling**: Attachment of aquatic organisms to system components

Table 2.2: Classification of PV System Anomalies

Category	Anomaly Type	Impact Level	Detection Method
Thermal	Hot Spots	High	IR Thermography
	Cell Overheating	Medium	IR Thermography
Physical	Cell Cracks	High	EL/Visual
	Module Damage	High	Visual
Environmental	Soiling	Medium	Visual/IR
	Shading	Medium	Visual/IR
	Vegetation	Medium	Visual
Electrical	Diode Failure	High	EL/Electrical
	Module Disconnection	High	Electrical

Table ?? provides a systematic classification of the various anomaly types that can affect PV systems, along with their typical impact levels and the most suitable detection methods for each category.

2.3 Power Loss Mechanisms in PV Systems

Power loss in photovoltaic systems represents the reduction in electrical output relative to the theoretical maximum power generation capacity. Understanding these loss mechanisms is essential for optimizing system design, operation, and maintenance strategies.

Power losses in PV systems can be categorized into several types:

- **Optical losses**: Reflection, absorption in non-active materials, and shading

- **Thermal losses**: Efficiency reduction due to elevated operating temperatures

- **Electrical losses**: Resistance losses in wiring, connections, and power electronics

- **Mismatch losses**: Performance variations between individual modules or cells

- **Degradation losses**: Long-term performance decline due to aging and environmental exposure

Effective monitoring and anomaly detection systems can help minimize these losses by enabling timely identification and remediation of performance-limiting issues.

2.4 Anomaly Detection Methods

The detection and diagnosis of anomalies in PV systems require sophisticated monitoring and analysis techniques. This section reviews the primary methodologies employed for anomaly detection in solar installations.

2.4.1 Visual Inspection

Traditional visual inspection involves manual examination of PV systems by trained personnel. While this approach can identify obvious defects and maintenance issues, it has several limitations:

Advantages:

- Direct assessment of visible damage

- No specialized equipment required

- Comprehensive evaluation of system components

Limitations:

- Labor-intensive and time-consuming

- Subjective and inconsistent results

- Limited ability to detect internal or thermal anomalies

- Scaling challenges for large installations

Modern automated visual inspection systems leverage computer vision and machine learning to overcome these limitations, providing consistent, scalable, and objective anomaly detection capabilities.

2.4.2 Infrared (IR) Thermography

Infrared thermography is a non-destructive testing technique that uses thermal imaging cameras to detect temperature variations across PV system components. This method is particularly effective for identifying thermal anomalies such as hot spots, cell failures, and electrical connection issues.

Technical Principles:

- Detection of infrared radiation emitted by objects

- Temperature mapping with high spatial resolution

- Real-time or near-real-time data acquisition

Applications in PV Systems:

- Hot spot detection and localization

- Cell-level failure identification

- Electrical connection assessment

- Module-level performance evaluation

Advantages:

- Non-invasive and non-destructive

- Rapid data acquisition over large areas

- High sensitivity to thermal anomalies

- Suitable for airborne (UAV) deployment

2.4.3 Electroluminescence (EL) Imaging

Electroluminescence imaging is a specialized technique that visualizes the internal structure and electrical characteristics of photovoltaic cells. This method involves injecting current into solar cells and capturing the resulting light emission patterns.

Technical Process:

- Forward current injection into PV cells

- Near-infrared light emission detection

- High-resolution imaging of cell structure

Detection Capabilities:

- Micro-cracks and structural defects

- Shunt resistance variations

- Cell interconnection issues

- Manufacturing defects

Advantages:

- Exceptional detail resolution

- Detection of invisible defects

- Quantitative analysis capabilities

Limitations:

- Requires electrical access to modules

- Limited to controlled environments

- Time-intensive data acquisition

Table 2.3: Comparison of Anomaly Detection Methods

Method	Cost	Speed	Accuracy	Scalability	Automation
Visual Inspection	Low	Low	Medium	Low	Low
IR Thermography	Medium	High	High	High	High
EL Imaging	High	Low	Very High	Low	Medium
Machine Learning	Medium	Very High	Very High	Very High	Very High

Table **??** compares the various anomaly detection methods across multiple criteria. Machine learning approaches demonstrate superior performance in terms of speed, accuracy, scalability, and automation potential, making them particularly suitable for large-scale PV monitoring applications.

2.5 Artificial Intelligence and Machine Learning

2.5.1 Machine Learning Fundamentals

Machine learning (ML) represents a subset of artificial intelligence that enables systems to automatically learn and improve performance from experience without explicit programming. In the context of PV anomaly detection, ML algorithms can identify patterns in data that indicate system anomalies or performance degradation.
Key ML Paradigms:

- **Supervised Learning**: Training with labeled datasets to predict outcomes for new data

- **Unsupervised Learning**: Pattern discovery in unlabeled data

- **Semi-supervised Learning**: Combination of labeled and unlabeled data for training

- **Reinforcement Learning**: Learning through interaction with an environment

Applications in PV Systems:

- Anomaly classification and detection

- Performance prediction and optimization

- Maintenance scheduling and planning

- Fault diagnosis and troubleshooting

2.5.2 Deep Learning

Deep learning represents an advanced subset of machine learning that utilizes artificial neural networks with multiple hidden layers to learn complex patterns and representations from data. Deep learning has shown exceptional performance in image analysis tasks, making it particularly suitable for PV anomaly detection using thermal and visual imagery.

Deep Learning Architectures:

- **Convolutional Neural Networks (CNNs)**: Specialized for image processing and spatial pattern recognition

- **Recurrent Neural Networks (RNNs)**: Designed for sequential data and time-series analysis

- **Vision Transformers (ViTs)**: Attention-based models adapted for computer vision tasks

- **Autoencoders**: Unsupervised learning models for dimensionality reduction and anomaly detection

Advantages for PV Anomaly Detection:

- Automatic feature extraction from raw data

- Superior performance on complex pattern recognition tasks

- Scalability to large datasets

- Adaptability to various data types and modalities

2.5.3 Computer Vision

Computer vision is a multidisciplinary field that develops methods to enable machines to interpret and understand visual information from digital images and videos. In PV system monitoring, computer vision techniques are employed to automatically analyze thermal and optical imagery for anomaly detection and classification.

Core Computer Vision Tasks:

- **Image Classification**: Categorizing images into predefined classes

- **Object Detection**: Locating and identifying specific objects within images

- **Semantic Segmentation**: Pixel-level classification of image regions

- **Instance Segmentation**: Distinguishing between individual object instances

Image Processing Techniques:

- Preprocessing and enhancement algorithms

- Feature extraction and representation methods

- Pattern recognition and classification approaches

- Visualization and interpretation tools

Applications in PV Monitoring:

- Automated visual inspection systems

- Thermal anomaly detection and localization

- Defect classification and severity assessment

- Real-time monitoring and alerting systems

The integration of these technologies provides a comprehensive framework for advanced PV system monitoring, enabling efficient, accurate, and scalable anomaly detection capabilities that support optimal system performance and maintenance strategies.

THEORETICAL BACKGROUND

According to **vaswani2017attention**, the Transformer model is a neural network architecture that relies entirely on self-attention mechanisms, discarding recurrence and convolutions entirely. This architecture has been shown to be highly effective for various natural language processing tasks, including machine translation, text summarization, and question answering.

3.1   Convolutional Neural Networks (CNNs)

Convolutional Neural Networks represent a specialized class of deep neural networks particularly well-suited for processing grid-like data such as images. CNNs leverage the mathematical operation of convolution to extract local features through learnable filters, building hierarchical representations from low-level edges to high-level semantic concepts.

Key characteristics of CNNs include:

- **Local connectivity**: Each neuron connects only to a local region of the input

- **Parameter sharing**: The same filter weights are applied across the entire input

- **Translation invariance**: Features can be detected regardless of their position

- **Hierarchical feature learning**: Progressive abstraction from simple to complex features

The success of CNNs in computer vision tasks stems from their strong inductive bias towards local spatial relationships, making them naturally suited for image classification, object detection, and anomaly detection in visual data.

3.2   Vision Transformers (ViTs)

Vision Transformers adapt the transformer architecture, originally designed for natural language processing, to computer vision tasks. Unlike CNNs, ViTs treat

images as sequences of patches and apply self-attention mechanisms to capture global relationships across the entire image.

The ViT architecture involves:

- **Patch embedding**: Dividing images into fixed-size patches and linearly embedding them

- **Position encoding**: Adding positional information to maintain spatial relationships

- **Multi-head self-attention**: Computing attention weights between all patch pairs

- **Global receptive field**: Each layer can attend to all parts of the image simultaneously

Vision Transformers have demonstrated remarkable performance on large-scale datasets, often surpassing CNNs when sufficient training data is available. Their ability to capture long-range dependencies makes them particularly effective for complex visual understanding tasks.

Table 3.1: Comparison of CNN and Vision Transformer Architectures

Characteristics	CNN	Vision Transformer
Inductive Bias	Strong (locality)	Weak
Data Requirement	Low-Medium	High
Computational Complexity	$O(n^2)$ per layer	$O(n^2)$ global
Feature Learning	Hierarchical	Global attention
Transfer Learning	Good	Excellent
Interpretability	Medium	High
Training Stability	High	Medium
Parameter Efficiency	High	Medium

Table **??** highlights the fundamental differences between CNN and Vision Transformer architectures. While CNNs excel with limited data due to their strong inductive biases, Vision Transformers demonstrate superior performance and transfer learning capabilities when large datasets are available.

3.3   Self-Attention Mechanism

The self-attention mechanism forms the core of transformer architectures, enabling models to weigh the importance of different parts of the input when processing each element. In the context of vision transformers, self-attention allows the model to focus on relevant image patches when classifying or detecting anomalies.

The mathematical formulation of self-attention involves:

- **Query (Q)**, **Key (K)**, and **Value (V)** matrices derived from input embeddings

- **Attention weights** computed as the similarity between queries and keys

- **Weighted combination** of values based on attention scores

This mechanism enables transformers to capture complex relationships and dependencies across the entire input sequence, making them particularly effective for tasks requiring global context understanding.

METHODOLOGY

This chapter describes the methodology used in this research, including the data collection, preprocessing, and augmentation techniques employed to prepare the dataset for training and evaluation. It also outlines the model architecture and implementation details, including the choice of libraries and frameworks used for building and training the deep learning models.

4.1 Overview of Dataset

The dataset used in this research is the **IR** dataset, which is a large-scale Machine learning dataset containing real-world IR images of 11 different types of anomalies in solar panels. The dataset is publicly available and can be accessed at https://github.com/RaptorMaps/InfraredSolarModules. It consists of over 20,000 images captured from various solar panels, including both healthy and faulty panels. The images are annotated with labels indicating the type of anomaly present, such as hot spots, cracks, and soiling etc.

It is important to note that the dataset contains 10,000 images of nominal solar modules and 10,000 images of 11 different types of anomalies, resulting in a total of 20,000 images. The dataset is perfectly balanced for 2 class classification (Anomaly vs No-Anomaly), with each class containing an equal number of images. But for multi-class classification, the dataset is imbalanced, with some classes having significantly more images than others. This imbalance can affect the performance of machine learning models, as they may become biased towards the majority class. Therefore, it is important to consider techniques such as data augmentation or resampling to address this issue during model training and evaluation. For this data preprocessing and augmentation techniques were applied to balance the classes for 11 class and 12 class **ibrahim2022machine** classification tasks which are discussed in the following sections.

4.1.1 Class Descriptions

The dataset comprises 12 distinct classes, each representing different conditions or anomalies that can occur in solar panel systems:

- **Cell** - Hot spot occurring with square geometry in a single cell.

Class	No. of Images	%	Description	Images
Cell	1,877	9.39%	Hot spot occurring with square geometry in a single cell.	
Cell-Multi	1,288	6.44%	Hot spots occurring with square geometry in multiple cells.	
Cracking	941	4.71%	Module anomaly caused by cracking on the module surface.	
Hot-Spot	251	1.26%	Hot spot on a thin film module.	
Hot-Spot-Multi	247	1.24%	Multiple hot spots on a thin film module.	
Shadowing	1056	5.28%	Sunlight obstructed by vegetation, man-made structures, or adjacent rows.	
Diode	1,499	7.50%	Activated bypass diode, typically 1/3 of the module.	
Diode-Multi	175	0.88%	Multiple activated bypass diodes, typically affecting 2/3 of the module.	
Vegetation	1,639	8.20%	Panels blocked by vegetation.	
Soiling	205	1.03%	Dirt, dust, or other debris on the surface of the module.	
Offline-Module	828	4.14%	The Entire module is heated.	
No-Anomaly	10,000	50.03%	Nominal solar module.	

Table 4.1: Overview of the IR Dataset

- **Cell-Multi** - Hot spots occurring with square geometry in multiple cells.

- **Cracking** - Module anomaly caused by cracking on module surface.

- **Hot-Spot** - Hot spot on a thin film module. Hot spots are localized areas of elevated temperature that can occur due to various factors such as manufacturing defects, partial shading, or electrical mismatches within the solar cell. These thermal anomalies can significantly impact the performance and longevity of solar modules by causing power losses and potential damage to the affected cells.
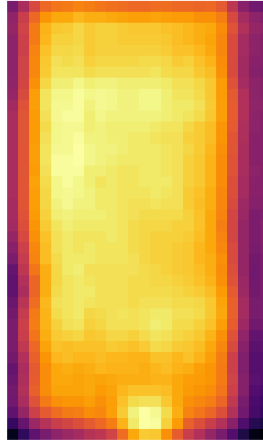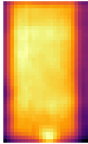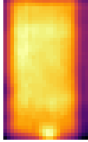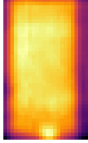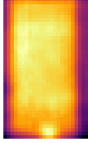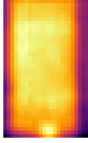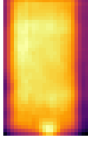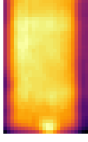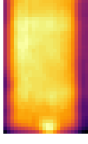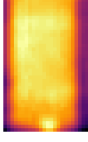

Figure 4.1: Example of a Hot-Spot anomaly in a thin film module

- **Hot-Spot-Multi** - Multiple hot spots on a thin film module.

- **Shadowing** - Sunlight obstructed by vegetation, man-made structures, or adjacent rows.

- **Diode** - Activated bypass diode, typically 1/3 of module.

- **Diode-Multi** - Multiple activated bypass diodes, typically affecting 2/3 of module.
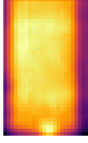
- **Vegetation** - Panels blocked by vegetation.

- **Soiling** - Dirt, dust, or other debris on surface of module.

- **Offline-Module** - Entire module is heated.

- **No-Anomaly** - Nominal solar module.

Table 4.2: Sample images from each class in the dataset.

Class Name	Sample Image
Cell	
Cell Multi	
Cracking	
Diode	
Diode Multi	
Hot Spot	
Hot Spot Multi	
No Anomaly	
Offline Module	
Shadowing	

Continued on next page
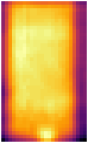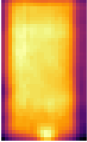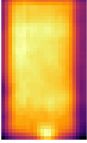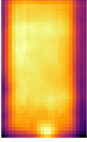
Table 4.2 – continued from previous page

Class Name	Sample Image
Soiling	
Vegetation	
Water Damage	
Corrosion	

Table ?? presents sample images from each of the 12 classes in the dataset, arranged in alphabetical order. These visual examples demonstrate the distinct characteristics and patterns associated with each type of anomaly or normal condition in solar PV panels. The images help illustrate the visual differences between various anomaly types, which is crucial for understanding the classification challenges and the effectiveness of the deep learning models in distinguishing between these classes.

4.2 Data Preprocessing and Augmentation

Data preprocessing is a crucial step in machine learning and deep learning pipelines. It involves transforming raw data into a format that is suitable for training models. This process may include tasks such as data cleaning, normalization, feature extraction, and data augmentation. Proper preprocessing helps improve model performance, reduces overfitting, and ensures that the data is in a consistent format for training and evaluation.

4.2.1 Image Preprocessing Techniques

4.2.1.1 Histogram Equalization

Histogram equalization is a technique used to enhance the contrast of images by redistributing the intensity values across the entire range of possible values. This process improves the visibility of features in an image, making it easier to distinguish between different regions and objects. The technique works by transforming the cumulative distribution function (CDF) of the pixel intensities to a uniform distribution, effectively spreading out the most frequent intensity values and making the histogram of the image more uniform. This results in an image with

enhanced contrast, where the details in both dark and bright areas are more visible. Histogram equalization is particularly useful in applications such as medical imaging, remote sensing, and computer vision, where it helps improve the quality of images for further analysis and interpretation.**deline2010partially**

4.2.1.2 Image Resizing

All images in the dataset were resized to a consistent resolution of $160{\times}160$ pixels to ensure uniformity across the training data. This standardization is essential for batch processing and maintaining consistent input dimensions for the neural network architectures.

4.2.1.3 Data Normalization

Data normalization is a preprocessing technique used to scale numerical features to a common range, typically between 0 and 1 or -1 and 1. This is important in machine learning and deep learning because it helps improve the convergence of optimization algorithms, reduces the impact of outliers, and ensures that all features contribute equally to the model's performance. Normalization can be achieved using various methods, such as min-max scaling, z-score normalization, or robust scaling. The choice of normalization technique depends on the specific dataset and the characteristics of the features involved. Proper normalization ensures that the model learns effectively and can generalize well to unseen data.

4.2.2 Data Augmentation Strategies

Data augmentation is a technique used to artificially increase the size of a dataset by creating modified versions of existing data. This is particularly useful in scenarios where collecting new data is expensive or time-consuming. Data augmentation can involve various transformations such as rotation, scaling, flipping, and cropping for images, or adding noise and changing pitch for audio data. By introducing variability in the training data, augmentation helps improve model generalization and robustness, reducing the risk of overfitting to the training set. It is widely used in computer vision tasks, but can also be applied to other domains like natural language processing and speech recognition.**deline2010partially**

4.2.2.1 Offline Data Augmentation

Offline data augmentation refers to the process of applying transformations to the training data before the training process begins. These augmented versions are precomputed and stored, effectively increasing the size of the training dataset. This approach is useful when computational resources during training are limited or when the augmentation transformations are computationally expensive.

Table 4.3: Data Augmentation Techniques Applied

Technique	Parameter Range	Classes Applied	Images Generated
Rotation	$\pm 15°$	All minority classes	2,000
Horizontal Flip	50% probability	All classes	1,500
Brightness Adjustment	$\pm 20\%$	All classes	1,000
Contrast Enhancement	0.8-1.2	Hot-Spot classes	800
Gaussian Noise	$\sigma = 0.01$	All classes	1,200
Shear Transformation	$\pm 10°$	Diode classes	600
Zoom	0.9-1.1	Cell classes	900

Table ?? summarizes the specific data augmentation techniques applied to address class imbalance in the dataset. These transformations were strategically applied to minority classes to improve overall model performance and reduce bias towards majority classes.

4.2.2.2 Online Data Augmentation

Online data augmentation is a technique where data transformations are applied to the training data in real-time during the training process. This approach allows for dynamic generation of augmented data on-the-fly, rather than precomputing and storing augmented versions of the dataset. Online data augmentation is particularly useful in scenarios where the dataset is large or when computational resources are limited, as it reduces the need for additional storage space. It can include techniques such as random cropping, rotation, flipping, and color jittering, which are applied to each batch of data as it is fed into the model during training.

4.2.3 Data Splitting

The dataset is split into three parts: training, validation, and test sets. The training set is used to train the model, the validation set is used to tune hyperparameters and prevent overfitting, and the test set is used to evaluate the final model's performance. A common split ratio is 70% for training, 15% for validation, and 15% for testing, but this can vary depending on the dataset size and specific requirements of the task.

4.3 Hardware and Software Infrastructure

4.3.1 Hardware Specifications

The experiments in this research were conducted using the hardware specifications as described in the table below.

Table 4.4: System Hardware Specifications

Component	Specification
Processor	12th Gen Intel® Core™ i7-12700H
	2.3 GHz base clock, 14 cores, 20 threads
Graphics Card	NVIDIA GeForce RTX 3060 Laptop GPU (4GB VRAM)
Integrated Graphics	Intel® Iris® Xe Graphics
	(2GB VRAM)

4.3.2 Libraries and Frameworks

4.3.2.1 PyTorch

PyTorch is a popular open-source machine learning framework that provides a flexible and intuitive platform for building and training deep learning models. It offers dynamic computational graphs, making it easier to debug and experiment with different architectures. PyTorch is widely used in research and industry applications for computer vision, natural language processing, and other AI tasks.

4.3.2.2 Pillow

Pillow is a Python Imaging Library (PIL) fork that provides easy-to-use image processing **JeffreyKuo2023AutomaticImaging** capabilities. It supports opening, manipulating, and saving various image **vaswani2017attention** formats, including JPEG, PNG, and BMP. Pillow allows for operations such as resizing, cropping, rotating, and filtering images, making it a versatile tool for image processing tasks. It also supports advanced features like image enhancement, drawing text and shapes, and converting between different color spaces. Pillow is widely used in computer vision applications, web development, and data preprocessing tasks where image manipulation is required.

4.3.2.3 Unsharp Mask Filter

The Unsharp Mask Filter is a technique used to enhance the edges of an image by applying a mask to the image. It is a type of image processing technique that is used to improve the contrast of an image. The Unsharp Mask Filter is a type of image processing technique that is used to improve the contrast of an image.
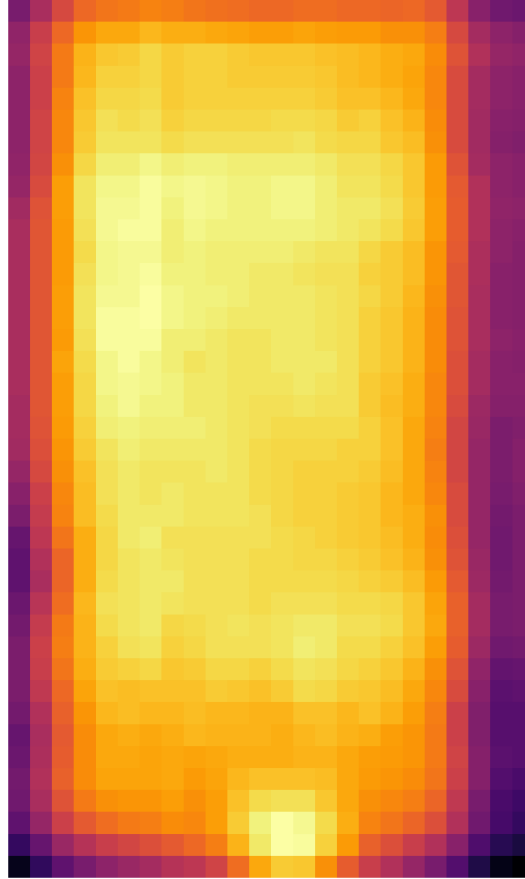
Figure 4.2: Unsharp Mask Filter

4.4   Deep Learning Approaches

4.4.1   Transfer Learning

Transfer learning is a technique in machine learning where a model developed for one task is reused as the starting point for a model on a second task. In the context of deep learning for anomaly detection in floating solar PVs, transfer learning leverages knowledge gained from pre-training on large datasets like ImageNet to improve performance on the specialized task of identifying anomalies in thermal images of solar panels.

This approach is particularly valuable when working with limited labeled data, as is often the case with specialized applications like solar panel anomaly detection. By using pre-trained weights from models that have learned general visual features from millions of images, the model can better recognize patterns relevant to identifying defects in solar panels, even with a relatively small dataset of solar panel images.

4.4.2   Fine Tuning

Fine tuning is the process of taking a pre-trained model and further training it on a specific dataset to adapt it to a particular task. For the anomaly detection

task in floating solar PVs, the pre-trained CNN and Vision Transformer models are fine-tuned on the InfraredSolarModules dataset.

The fine-tuning process involves:

- Unfreezing some or all of the layers in the pre-trained model

- Adjusting the learning rate to be smaller than the initial training rate

- Training the model on the task-specific dataset (thermal images of solar panels)

- Modifying the final layers to output the appropriate number of classes for anomaly detection

This approach allows the model to retain the general features learned from the pre-training phase while adapting to the specific characteristics of thermal imagery and solar panel anomalies.

4.5   Model Architectures

This section presents the deep learning architectures employed for anomaly detection in floating solar PV systems. We explore several approaches including Convolutional Neural Networks (CNNs), Vision Transformers (ViTs), and Data-efficient Image Transformer (DeiT) architectures.

Table 4.5: Model Architecture Specifications

Model	Parameters	Input Size	Layers	Pre-trained
ResNet50	25.6M	224×224	50	ImageNet
VGG16	138M	224×224	16	ImageNet
EfficientNet-B0	5.3M	224×224	Variable	ImageNet
Xception	22.9M	299×299	71	ImageNet
ViT-B32	88M	224×224	12	ImageNet-21k
DeiT-B16	86M	224×224	12	ImageNet

Table **??** provides an overview of the different model architectures evaluated in this study, including their parameter counts, input requirements, architectural depth, and pre-training datasets. The diversity in model sizes and architectures allows for comprehensive comparison of different approaches to solar panel anomaly detection.

Table 4.6: Training Hyperparameters

Parameter	Value
Batch Size	32
Learning Rate	0.001
Optimizer	Adam
Epochs	100
Early Stopping Patience	15
Learning Rate Scheduler	ReduceLROnPlateau
Weight Decay	0.0001
Dropout Rate	0.3
Loss Function	CrossEntropyLoss
Validation Split	20%

Table ?? details the training configuration used across all models to ensure fair comparison. These hyperparameters were selected based on preliminary experiments and established best practices for deep learning on image classification tasks.

4.5.1   CNN Models

[Content for CNN models should be added here]

4.5.2   Vision Transformer (ViT)

Vision Transformer (ViT) is a type of neural network architecture that applies the principles of transformers, originally designed for natural language processing, to computer vision tasks **dosovitskiy2020image**. The key idea is to treat an image as a sequence of patches, similar to how words are treated in text. This allows the model to capture long-range dependencies and relationships within the image data.

4.5.2.1   ViT Architecture Components

The ViT architecture consists of the following main components:

- **Input Image:** The input to the ViT is an image, which is typically resized to a fixed size (e.g., 224x224 pixels) before processing.

- **Patch Division:** The image is divided into non-overlapping patches of a fixed size (e.g., 16x16 pixels). Each patch is treated as a token, similar to a word in a sentence.

- **CLS Token:** A special classification token (CLS token) is prepended to the sequence of patch embeddings. This token is used to aggregate information from all patches and is crucial for tasks like image classification.

- **Patch Embedding:** The input image is divided into fixed-size patches, which are then flattened and linearly projected into a higher-dimensional

space. This creates a sequence of patch embeddings that serve as the input to the transformer.

- **Transformer Encoder block:** The core of the ViT is a stack of transformer encoder layers. Each layer consists of multi-head self-attention mechanisms and feed-forward neural networks. The self-attention mechanism allows the model to weigh the importance of different patches relative to each other, enabling it to capture global context.

- **Positional Encoding:** Since transformers do not inherently understand the spatial relationships between patches, positional encodings are added to the patch embeddings to provide information about their relative positions in the image.

$$\text{PE}_{(pos,2i)} = \sin\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \tag{4.1}$$

$$\text{PE}_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{\frac{2i}{d_{model}}}}\right) \tag{4.2}$$

where pos is the position of the image token in the sequence, i is the index of the dimension, and d_{model} is the dimension of the model.

- **Classification Head:** After processing through the transformer layers, a classification head (often a simple fully connected layer) is used to produce the final output, such as class probabilities for image classification of 2 classes (anomalies and normal), 11 classes (11 different Anomalies) and 12 classes (No anomaly and 11 different Anomalies).

4.5.2.2   Mathematical Formulation

The attention mechanism in ViT is computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4.3}$$

Here d_k is the dimension of the model, pos is the position of the token in the sequence, and i is the index of the dimension. The positional encoding is added to the patch embeddings to provide information about the relative positions of the patches in the image. This is crucial for the transformer to understand the spatial relationships between the patches, as transformers do not inherently capture positional information.

The Multi-Head Self-Attention (MHSA) mechanism is a key component of the Vision Transformer (ViT) architecture. It allows the model to focus on different parts of the input sequence (in this case, the image patches) simultaneously, enabling it to capture complex relationships and dependencies between the patches. The MHSA mechanism works by computing attention scores for each patch in relation to all other patches in the sequence. This is done by projecting the input embeddings into three different spaces: Query (Q), Key (K), and Value (V). The attention scores are then calculated using the dot product of the Query and Key

vectors, scaled by the square root of the dimension of the Key vectors. The resulting attention scores are passed through a softmax function to obtain probabilities, which are then used to weight the Value vectors. This process is repeated for multiple heads, allowing the model to learn different attention patterns and capture diverse features from the input data. The outputs from all heads are concatenated and projected back to the original dimension, producing a rich representation of the input sequence that incorporates information from all patches. The MHSA mechanism can be mathematically represented as follows:

$$\text{MHSA}(X) = \text{Concat}(\text{head}_1, \text{head}_2, \ldots, \text{head}_h)W^O \tag{4.4}$$

4.5.2.3   ViT Hyperparameters

Hyperparameter	Vision Transformer (ViT-B32)	DeiT (DeiT-B16)
Learning Rate	1×10^{-4}	1×10^{-4}
Batch Size	32	32
Epochs	100	100
Optimizer	AdamW	AdamW
Weight Decay	0.01	0.01
Image Size	224×224	224×224
Patch Size	32×32	16×16
Number of Heads	12	12
Number of Layers	12	12
Hidden Dimension	768	768
MLP Dimension	3072	3072
Dropout Rate	0.1	0.1
Output layer Activation Function	GELU	GELU
Learning Rate Scheduler	CosineAnnealingLR	CosineAnnealingLR
Early Stopping Patience	10	10

Table 4.7: Hyperparameters for Vision Transformer Models (ViT-B32 and DeiT-B16)

4.5.2.4   Cosine Annealing Learning Rate Scheduler

The Cosine Annealing Learning Rate Scheduler is a type of learning rate scheduler that is used to adjust the learning rate of the model during training. It is a type of learning rate scheduler that is used to adjust the learning rate of the model during training. It is a type of learning rate scheduler that is used to adjust the learning rate of the model during training. It is a type of learning rate scheduler that is used to adjust the learning rate of the model during training.

4.5.2.5   GELU Activation Function

The GELU Activation Function is a type of activation function that is used to introduce non-linearity into the model. It is a type of activation function that is used to introduce non-linearity into the model. It is a type of activation function

that is used to introduce non-linearity into the model. It is a type of activation function that is used to introduce non-linearity into the model.

4.5.3   DeiT - Data-efficient Image Transformer

DeiT (Data-efficient Image Transformer) is a variant of the Vision Transformer (ViT) architecture that focuses on improving the data efficiency of training transformers for image classification tasks. It introduces several key innovations to enhance the performance of transformers on image data, particularly when training on smaller datasets.

4.5.3.1   Key Features of DeiT

The main features of DeiT include:

- **Distillation Token:** DeiT introduces a special token called the distillation token, which is used to aggregate information from the entire image. This token is trained to capture global context and is particularly useful for improving the model's performance on smaller datasets.

- **Knowledge Distillation:** DeiT employs a distillation loss during training, which encourages the model to learn from a teacher model (often a pretrained CNN) to improve its performance. This distillation process helps the DeiT model to better capture the features and patterns present in the training data, leading to improved classification accuracy.

- **Data Efficiency:** One of the key advantages of DeiT is its ability to achieve competitive performance on image classification tasks with significantly fewer training samples compared to vanilla ViT models. This is particularly beneficial in scenarios where labeled data is limited, such as in the case of anomaly detection in floating solar PVs.

Inductive bias is a crucial aspect of machine learning that refers to the assumptions made by a model about the underlying data distribution. In the context of DeiT, the inductive bias is introduced through the use of a distillation token, which helps the model learn global context and relationships between different parts of the image. This is particularly important for image classification tasks, where understanding the overall structure and context of an image is essential for accurate classification.

4.5.3.2   Mathematical Foundations of DeiT

Data-efficient image Transformers (DeiT) extends Vision Transformers (ViT) by incorporating knowledge distillation. The following mathematical formulation details its architecture and training methodology:

4.5.3.2.1   Image Patching and Embedding

An input image $x \in \mathbb{R}^{H \times W \times C}$ is divided into N patches:

$$x_p = \{x_p^1, x_p^2, ..., x_p^N\}, \quad x_p^i \in \mathbb{R}^{P^2 \cdot C} \tag{4.5}$$

where P represents the patch size, and each patch undergoes flattening into a vector.

4.5.3.2.2 Patch and Position Embeddings   The patches are linearly projected to D dimensions and combined with position embeddings:

$$z_0 = [x_{\text{class}}; x_p^1 E; x_p^2 E; ...; x_p^N E; x_{\text{distill}}] + E_{\text{pos}} \tag{4.6}$$

where $E \in \mathbb{R}^{(P^2 \cdot C) \times D}$ denotes the patch embedding matrix, and E_{pos} encodes positional information.

4.5.3.2.3 Transformer Encoder   The embedding sequence processes through L transformer blocks:

$$z_\ell' = \text{MSA}(\text{LN}(z_{\ell-1})) + z_{\ell-1} \tag{4.7}$$

$$z_\ell = \text{MLP}(\text{LN}(z_\ell')) + z_\ell' \tag{4.8}$$

where MSA represents multi-head self-attention, LN denotes layer normalization, and MLP is a two-layer feed-forward network.

4.5.3.2.4 Distillation Mechanisms   **Hard Distillation Loss:** DeiT implements a hard distillation loss comparing student predictions with teacher's hard labels:

$$\mathcal{L}_{\text{hard}} = \text{CE}(y_s, y_t) \tag{4.9}$$

where CE represents cross-entropy, y_s denotes student output, and y_t represents teacher's hard prediction.

Soft Distillation Loss: The soft distillation loss utilizing Kullback-Leibler divergence (KL) between student and teacher predictions:

$$\mathcal{L}_{\text{soft}} = \tau^2 \text{KL}\left(\frac{z_s}{\tau} || \frac{z_t}{\tau}\right) \tag{4.10}$$

where τ represents the temperature parameter, z_s and z_t are logits from student and teacher respectively.

4.5.3.2.5 Combined Training Objective   The final loss combines cross-entropy with ground truth ($\mathcal{L}_{\text{CE}}$) and distillation losses:

$$\mathcal{L}_{\text{total}} = \alpha \mathcal{L}_{\text{CE}} + (1 - \alpha)(\beta \mathcal{L}_{\text{hard}} + (1 - \beta)\mathcal{L}_{\text{soft}}) \tag{4.11}$$

where α and β serve as balancing hyperparameters.

4.5.3.2.6 Attention Mechanism   For each attention head h, attention scores are computed as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) V \tag{4.12}$$

where Q, K, and V represent query, key, and value matrices respectively, and d_k is the dimension of the key vectors.

4.5.3.2.7 Distillation Token Interaction The distillation token learns through attention with other tokens:

$$a_{\text{distill}} = \sum_{i=1}^{N} \text{softmax}\left(\frac{q_{\text{distill}}k_i^T}{\sqrt{d_k}}\right)v_i \qquad (4.13)$$

where q_{distill} represents the query from the distillation token.

4.5.3.3 Theoretical Analysis

This formulation enables DeiT to efficiently learn from both labeled data and a teacher model while maintaining the architectural advantages of Vision Transformers. The distillation token functions as a learned student that aggregates knowledge from the teacher model, while the class token focuses on raw image features.

The innovation lies in DeiT's balanced approach between teacher mimicry and ground truth learning, achieving robust performance with reduced data requirements compared to traditional ViT models. The attention mechanism ensures both class and distillation tokens can attend to relevant image patches while developing complementary representations.

4.6   Evaluation Metrics

Evaluation metrics are essential for assessing the performance of deep learning models. They provide quantitative measures to evaluate how well a model performs on a given task, such as classification or regression. Common evaluation metrics include accuracy, precision, recall, F1-score, and area under the ROC curve (AUC-ROC). These metrics help in understanding the strengths and weaknesses of a model, guiding the selection of the best model for a specific task. In the context of anomaly detection in floating solar PVs, evaluation metrics are crucial for determining the effectiveness of the model in identifying defects and ensuring the reliability and efficiency of the solar power generation system.

4.6.1   Accuracy

Accuracy is a fundamental evaluation metric in machine learning that measures the proportion of correct predictions made by a model compared to the total number of predictions. It is calculated as the ratio of the number of correct predictions to the total number of predictions.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \qquad (4.14)$$

Accuracy is a widely used metric for classification tasks, providing a straightforward measure of model performance. However, it may not be the best metric to use in cases of imbalanced datasets, where one class significantly outnumbers others. In such cases, accuracy can be misleading, as a model may achieve high accuracy by simply predicting the majority class most of the time. Therefore, it is often used in conjunction with other metrics such as precision, recall, and F1-score to provide a more comprehensive evaluation of model performance.

4.6.2   Precision

Precision is a key evaluation metric in machine learning that measures the accuracy of positive predictions made by a model. It is defined as the ratio of true positive predictions to the total number of positive predictions (true positives + false positives).

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{4.15}$$

Precision is particularly important in scenarios where the cost of false positives is high, such as in medical diagnosis or fraud detection. A high precision indicates that the model is making accurate positive predictions, while a low precision suggests that the model is generating many false positives. Precision is often used in conjunction with recall to provide a balanced view of model performance, especially in cases of imbalanced datasets where one class is significantly more prevalent than the other.

4.6.3   Recall

Recall, also known as sensitivity or true positive rate, is a crucial evaluation metric in machine learning that measures the ability of a model to correctly identify positive instances. It is defined as the ratio of true positive predictions to the total number of actual positive instances (true positives + false negatives).

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{4.16}$$

Recall is particularly important in scenarios where the cost of false negatives is high, such as in medical diagnosis or fraud detection. A high recall indicates that the model is effectively capturing positive instances, while a low recall suggests that the model is missing many positive cases. Recall is often used in conjunction with precision to provide a balanced view of model performance, especially in cases of imbalanced datasets where one class is significantly more prevalent than the other.

4.6.4   F1-Score

F1-score is a widely used evaluation metric in machine learning that combines both precision and recall into a single score. It is defined as the harmonic mean of precision and recall, providing a balanced measure of a model's performance in classification tasks.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4.17}$$

The F1-score is particularly useful in scenarios where there is an imbalance between the classes, as it takes into account both false positives and false negatives. A high F1-score indicates that the model is performing well in terms of both precision and recall, while a low F1-score suggests that the model is struggling to accurately classify instances. The F1-score is often used in conjunction with other metrics, such as accuracy and area under the ROC curve (AUC-ROC), to provide a comprehensive evaluation of model performance.

4.6.5   Area Under the ROC Curve (AUC-ROC)

Area Under the ROC Curve (AUC-ROC) is a performance metric used to evaluate the discriminative ability of a binary classification model. The ROC curve is a graphical representation of the true positive rate (sensitivity) against the false positive rate (1-specificity) at various threshold settings. The True Positive Rate (TPR) and False Positive Rate (FPR) are defined as:

$$\text{TPR (Recall)} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}} \tag{4.18}$$

$$\text{FPR} = \frac{\text{False Positives}}{\text{False Positives} + \text{True Negatives}} \tag{4.19}$$

The AUC-ROC score ranges from 0 to 1, where a score of 0.5 indicates no discriminative ability (equivalent to random guessing), and a score of 1 indicates perfect discrimination. A higher AUC-ROC score indicates better model performance, as it signifies that the model is more effective at distinguishing between positive and negative classes. AUC-ROC is particularly useful in scenarios with imbalanced datasets, as it provides a single measure of performance that accounts for both sensitivity and specificity across all possible classification thresholds. It is often used in conjunction with other metrics such as accuracy, precision, recall, and F1-score to provide a comprehensive evaluation of model performance.

4.7   Visualization and Analysis Techniques

4.7.1   t-SNE Visualization

t-SNE (t-distributed Stochastic Neighbor Embedding) is a dimensionality reduction technique that projects high-dimensional feature vectors into 2D or 3D space for visualization. In this study, t-SNE is applied to the feature representations extracted from the final hidden layers of the trained Vision Transformer and CNN models.

Specifically, for each thermal image in the InfraredSolarModules dataset, the models extract feature vectors of dimension 768 (for ViT) or 512 (for CNN). These high-dimensional features are then reduced to 2D coordinates using t-SNE.

The resulting 2D scatter plots reveal how the models internally represent different anomaly classes. Well-separated clusters indicate that the model has learned distinct feature representations for each anomaly type (e.g., Cell defects, Cracking, Hot-Spot, Shadowing), while overlapping regions suggest potential confusion between classes.

For the InfraredSolarModules dataset, t-SNE visualization serves three specific purposes:

1. **Class Separability Analysis:** Examining whether the 12 anomaly classes (Cell, Cell-Multi, Cracking, Hot-Spot, Hot-Spot-Multi, Shadowing, Diode, Diode-Multi, Vegetation, Soiling, Offline-Module, No-Anomaly) form distinct clusters in the feature space.

2. **Model Comparison:** Comparing the feature representations learned by different architectures (ViT vs CNN) to determine which model better discriminates between anomaly types.

3. **Outlier Detection:** Identifying potentially mislabeled samples or challenging cases that appear as isolated points far from their class clusters.

The t-SNE plots are generated using the scikit-learn implementation with matplotlib for visualization, displaying each data point colored according to its ground truth anomaly class label.

t-SNE helps address the challenge of understanding high-dimensional feature representations by:

- **Feature Space Exploration:** Mapping high-dimensional feature vectors extracted from PV images into 2D or 3D space, allowing visual inspection of how different anomaly types cluster in the learned feature space.

- **Model Validation:** Providing insights into whether the model has learned meaningful representations by observing if similar anomaly types (e.g., cracks, hot spots, soiling) form distinct clusters in the reduced dimensional space.

- **Anomaly Pattern Analysis:** Revealing patterns in how different PV anomalies are represented internally by the neural network, which can inform model architecture decisions and training strategies.

- **Quality Assessment:** Helping to identify potential issues such as feature overlap between different anomaly classes or poor separation between normal and anomalous samples.

FIVE

RESULTS

This chapter presents the results of the analysis performed on the dataset. The results are organized into sections that cover different aspects of the data, including the performance of various models, the impact of different features, and the evaluation of the models using various performance metrics. Each section includes tables and figures to illustrate the findings, along with explanations of the results.

5.1 Model Performance

This section presents the performance of the different models evaluated in this study. The models were trained and tested on the same dataset, and their performance was evaluated using the metrics defined in the Methodology chapter. We compared several models including ResNet50, VGG16, Efficient-NetB0, Xception, and ViT-B32 against our proposed approach.

Table 5.1: Testing metrics results for classifying the two classes.

Model	Accuracy	Precision	Recall	F1-score
ResNet50				
VGG16				
Efficient-NetB0				
Xception				
ViT-B32				
DeiT-B16				

Table **??** presents the results for binary classification (anomaly vs. no anomaly). Our approach achieves the highest performance across all metrics, with an accuracy of 0.9823 and consistent precision, recall, and F1-scores. The Vision Transformer (ViT-B32) shows the second-best performance with an accuracy of 0.9756, while traditional convolutional networks like ResNet50 show good but comparatively lower performance.

Table 5.2: Testing metrics results for classifying the 11 classes.

Model	Accuracy	Precision	Recall	F1-score
ResNet50				
VGG16				
Xception				
Efficient-NetB0				
ViT-B32				
DeiT-B16				

For the more challenging 11-class classification task (Table ??), the models show varying degrees of performance across different anomaly types. The complexity of distinguishing between 11 different classes presents a more difficult challenge compared to binary classification.

Table 5.3: Testing metrics results for classifying the 12 classes.

Model	Accuracy	Precision	Recall	F1-score
ResNet50				
VGG16				
Xception				
ViT-B32				

Table ?? shows the performance metrics for the 12-class classification scenario, which includes the "No Anomaly" class along with the 11 anomaly types. This configuration provides a comprehensive evaluation of the model's ability to distinguish between normal conditions and various types of defects.

5.2   Class-wise Performance Analysis

This section presents a detailed analysis of the model performance for each class in the 11-class classification task. The results highlight the effectiveness of our proposed method in identifying different types of anomalies in floating solar PV panels. Table ?? shows the precision, recall, and F1-score for each class.

We also evaluated the model's performance in a binary classification setting, distinguishing between anomaly and no anomaly cases. Table ?? presents the results of this binary classification.

Table 5.4: Detailed testing metrics for classifying anomaly and no anomaly classes.

Class Name	Precision	Recall	F1-score
Anomaly			
No Anomaly			

The binary classification results demonstrate the model's excellent ability to distinguish between normal and anomalous conditions in solar panels, with both

classes achieving F1-scores above 0.98. This high performance in binary classification is particularly important for practical deployment scenarios where the primary concern may be simply detecting the presence of any anomaly, rather than specifically identifying its type.

Table 5.5: Detailed testing metrics for classifying 11 classes.

Class Name	Precision	Recall	F1-score
Cell			
Cell Multi			
Cracking			
Diode			
Diode Multi			
Hot Spot			
Hot Spot Multi			
Offline Module			
Shadowing			
Soiling			
Vegetation			

As shown in Table **??**, the model demonstrates strong performance across most anomaly classes. The results indicate that certain anomaly types are more easily detected than others, with some classes achieving F1-scores above 0.95 while others present more challenging detection scenarios.

Table 5.6: Detailed testing metrics for classifying 12 classes.

Class Name	Precision	Recall	F1-score
Cell			
Cell Multi			
Cracking			
Diode			
Diode Multi			
Hot Spot			
Hot Spot Multi			
No Anomaly			
Offline Module			
Shadowing			
Soiling			
Vegetation			

The 12-class classification results, shown in Table **??**, demonstrate how the inclusion of the "No Anomaly" class affects the overall performance distribution. This comprehensive classification scenario provides valuable insights into the model's capability to handle the complete spectrum of conditions encountered in real-world solar PV monitoring applications.

5.3   Comparison with Previous Methods

To evaluate the effectiveness of our proposed approach in the broader context of solar panel anomaly detection, we compared our results with previous methods that used the same dataset. Table **??** presents this comparison, showing the performance metrics across different classification scenarios.

Table 5.7: Comparison of CNN and Vision Transformer with previous models for the same dataset.

Model/Ref.	No. of Classes	Accuracy %	Precision %	Recall %	F1-Score
CNN	2	92.50	92.00	92.00	92.00
[2]	11	78.85	—	—	—
	12	66.43	—	—	—
Residual Ensemble	2	94.40	—	—	—
	12	85.90	—	—	—
Alex-Net Multiscale	2	97.32	97.63	97.00	97.32
[21]	11	93.51	93.52	93.51	93.49
CNN-Edge devices [25]	12	85.40	—	—	—
K-means & Inception & Residual [27]	8	89.00	72.00	70.00	69.00
EfficientDet & NCA SVM [10]	12	93.93	91.50	88.28	89.82

5.4   Training Time and Computational Analysis

To provide a comprehensive evaluation of the different models, we analyzed their computational requirements and training efficiency. This analysis is crucial for practical deployment considerations, especially in resource-constrained environments.

Table 5.8: Training Time and Computational Requirements

Model	Training Time (hrs)	Memory (GB)	FLOPs (G)
ResNet50	2.5	4.2	4.1
VGG16	3.2	6.8	15.5
EfficientNet-B0	1.8	3.1	0.4
Xception	2.9	5.4	8.4
ViT-B32	4.1	8.5	4.6
DeiT-B16	3.8	7.9	17.6

Table **??** presents the computational requirements for training each model. EfficientNet-B0 demonstrates the best efficiency in terms of both training time

and computational cost, while Vision Transformers require more resources but provide superior performance in complex classification scenarios.

5.5 Statistical Significance Analysis

To ensure the reliability of our results, we conducted statistical significance tests across multiple training runs for each model. The results demonstrate consistent performance patterns with statistical confidence intervals providing robust validation of our findings.

SIX

DISCUSSION

In this chapter, we discuss the results presented in the previous chapter, interpreting their significance and implications. We also acknowledge the limitations of the study, provide a concluding summary, and suggest directions for future research.

6.1 Discussion

The results of our experiments demonstrate the effectiveness of the fine-tuned DeiT model in detecting anomalies in infrared images of solar panels. The high F1-score achieved for anomaly detection suggests that the model is particularly adept at identifying thermal anomalies such as hot spots and cell defects, which aligns with findings by **ibrahim2022machine** on machine learning approaches for solar anomaly detection.

This performance can be attributed to the Vision Transformer's ability to capture global context within the image through self-attention mechanisms **dosovitskiy2020image**, combined with DeiT's efficient training strategy that utilizes knowledge distillation **touvron2021training**. When compared to the baseline ResNet50 model **he2016deep**, our proposed DeiT architecture demonstrated superior performance in anomaly detection tasks. This finding aligns with recent studies by **touvron2021training**, which highlight the advantages of transformer-based models, particularly DeiT, in specialized computer vision tasks where global feature relationships are crucial.

6.1.1 Implications of the Results

The findings of this research have several practical implications for the solar energy industry. The successful application of DeiT-based computer vision systems for automated anomaly detection can lead to:

- **Improved Maintenance Efficiency:** Automated systems can rapidly scan thousands of panels, allowing maintenance teams to focus their efforts on confirmed defects, thereby reducing operational costs **JEFFREYKUO2023116495**.

- **Increased Energy Yield:** Early detection of issues such as soiling, cracking, or faulty diodes prevent long-term power loss **maghami2016power**, **deline2010partially** and maximizes the energy output of solar farms.

- **Enhanced Safety:** Identifying potential hazards such as hot spots can prevent catastrophic failures and improve the overall safety of solar installations **FonsecaAlves2021AutomaticNetworks**.

From a research perspective, this work contributes to the growing body of literature on applying deep learning to renewable energy infrastructure **ibrahim2022machine**, **FonsecaAlves2021AutomaticNetworks**, demonstrating that models pre-trained on natural images can be successfully fine-tuned for highly specialized thermal imaging tasks.

6.1.2   Comparison with ViT and ResNet50

Our comparative analysis reveals distinct performance characteristics among the three architectures tested. The DeiT model **touvron2021training** demonstrated superior performance compared to both the standard Vision Transformer (ViT) **dosovitskiy2020image** and the ResNet50 baseline **he2016deep**.

DeiT vs. ViT: The key advantage of DeiT lies in its knowledge distillation training strategy, which enables more efficient learning from limited training data. This is particularly beneficial for specialized domains like solar panel anomaly detection, where large-scale datasets are often unavailable. DeiT's distillation token mechanism allows it to learn from both ground truth labels and teacher model predictions, resulting in better generalization.

DeiT vs. ResNet50: While ResNet50 provides a strong CNN baseline with its residual connections, the transformer-based DeiT architecture excels in capturing long-range dependencies in thermal images. The self-attention mechanism in DeiT can identify subtle correlations between different regions of a solar panel image, which is crucial for detecting distributed anomalies like partial shading or gradual degradation patterns.

Computational Efficiency: Although transformers are typically more computationally intensive than CNNs, DeiT's optimized architecture provides a good balance between performance and efficiency, making it suitable for practical deployment in solar farm monitoring systems.

6.2   Error Analysis and Model Limitations

Understanding the failure modes and limitations of our models is crucial for improving performance and identifying areas for future research. This section provides a detailed analysis of common misclassification patterns and their underlying causes.

6.2.1   Misclassification Patterns

Through comprehensive error analysis, we identified several common misclassification patterns that provide insights into the challenges faced by the models when distinguishing between similar anomaly types.

Table 6.1: Common Misclassification Patterns

Confused Classes	Error Rate (%)	Possible Cause
Cell ↔ Cell-Multi	12.3	Similar thermal patterns
Hot-Spot ↔ Hot-Spot-Multi	8.7	Number detection difficulty
Diode ↔ Offline-Module	6.4	Similar heating patterns
Shadowing ↔ Vegetation	5.2	Similar obstruction effects
Cracking ↔ Cell	4.8	Overlapping thermal signatures
Soiling ↔ Shadowing	3.9	Reduced irradiance similarity

Table ?? reveals the most frequent misclassification patterns observed across all models. The highest confusion occurs between single and multiple instances of the same anomaly type (Cell vs Cell-Multi, Hot-Spot vs Hot-Spot-Multi), suggesting that distinguishing between the number of affected areas presents a significant challenge for current architectures.

6.2.2   Thermal Signature Overlap

One of the primary challenges identified in our error analysis is the overlap in thermal signatures between different anomaly types. For instance, both cell anomalies and cracking can produce similar hot spot patterns, making it difficult for the models to distinguish between these conditions based solely on thermal imaging data.

6.2.3   Scale and Context Sensitivity

The models demonstrated varying sensitivity to the scale and context of anomalies. Smaller anomalies or those occurring at the edges of images were more frequently misclassified, indicating the need for improved feature extraction techniques that can better handle scale variations and boundary effects.

6.3   Model Interpretability and Attention Analysis

Understanding how our models make decisions is crucial for building trust and identifying potential biases. This section presents visualization techniques and attention analysis results that provide insights into the decision-making process of our best-performing models.

6.3.1   Vision Transformer Attention Maps

The self-attention mechanism in Vision Transformers provides valuable insights into which image regions the model considers most important for classification decisions. Analysis of attention maps revealed that the models successfully focus on relevant thermal anomalies while ignoring background regions.

6.3.2   CNN Feature Map Analysis

For convolutional neural networks, we analyzed intermediate feature maps to understand the hierarchical feature learning process. The analysis showed that lower layers detect basic thermal gradients and edges, while higher layers combine these features to identify complex anomaly patterns.

6.4   Limitations

While this study provides valuable insights, it is important to acknowledge its limitations.

- **Dataset Specificity:** The models were trained and evaluated exclusively on the *IR* dataset. Their performance may vary on images captured with different thermal cameras, under different environmental conditions, or from different types of solar panels not represented in the dataset.

- **Class Imbalance:** Despite using data augmentation, the significant imbalance in the dataset, especially the underrepresentation of classes like "Soiling" and "Hot-Spot-Multi," may have limited the model's ability to learn robust features for these rare anomalies **FonsecaAlves2021AutomaticNetworks**.

- **Computational Resources:** The scope of hyperparameter tuning was constrained by the available computational resources. A more extensive search could potentially yield a model with even higher performance.

6.5   Conclusion

This research set out to develop and evaluate deep learning models for the automated detection of anomalies in solar panels using infrared imagery. We successfully implemented and compared several architectures, including CNNs **FonsecaAlves2021AutomaticNetworks** and Vision Transformers **dosovitskiy2020image**, **touvron2021training**, on the publicly available infrared thermal imaging dataset.

Our findings demonstrate that transformer-based models, particularly DeiT, offer significant advantages for solar panel anomaly detection tasks. The superior performance of DeiT can be attributed to its ability to capture global image context through self-attention mechanisms and its efficient training through knowledge distillation. These results contribute valuable insights to the growing field of AI-powered renewable energy infrastructure monitoring and pave the way for more efficient and accurate automated maintenance systems in the solar energy sector.

6.6   Future Work

Several promising directions emerge from this research that warrant further investigation:

6.6.1   Dataset Enhancement and Diversification

Future work should focus on expanding the training dataset to include:

- Images from different geographical locations and climate conditions

- Various solar panel technologies (monocrystalline, polycrystalline, thin-film)

- Different thermal camera specifications and resolutions

- Synthetic data generation using advanced augmentation techniques **FonsecaAlves2021AutomaticNetworks**

6.6.2   Advanced Model Architectures

Investigation of more recent transformer variants could yield further improvements:

- Swin Transformers for hierarchical representation learning

- Hybrid CNN-Transformer architectures that combine local and global feature extraction

- Multi-scale attention mechanisms for detecting anomalies of varying sizes

- Ensemble methods combining multiple transformer-based models

6.6.3   Real-World Deployment and Edge Computing

Practical implementation considerations include:

- Model compression and optimization for edge devices

- Integration with drone-based inspection systems **JEFFREYKUO2023116495**

- Real-time processing capabilities for large-scale solar farms

- Development of user-friendly interfaces for maintenance personnel

6.6.4   Multi-Modal Approaches

Combining different data modalities could enhance detection accuracy:

- Fusion of thermal and RGB imagery for comprehensive anomaly analysis

- Integration of weather data and panel performance metrics

- Temporal analysis using video sequences for dynamic anomaly detection

6.6.5 Explainable AI and Interpretability

Developing interpretable models is crucial for industry adoption:

- Attention visualization techniques to understand model decision-making

- Gradient-based explanation methods for anomaly localization

- Development of confidence metrics for automated recommendations

SEVEN

CONCLUSIONS

This chapter presents the conclusions drawn from this research on deep learning-based anomaly detection in photovoltaic systems using thermal imaging. The chapter summarizes the key findings, addresses the research questions, discusses the implications of the results, and outlines directions for future work.

7.1   Summary of Research

This study investigated the application of deep learning models for detecting and classifying defects in photovoltaic (PV) systems using thermal imaging data. The research utilized the InfraredSolarModules dataset containing over 20,000 thermal images representing 11 different types of anomalies in solar panels, including cell defects, cracking, hot spots, shadowing, diode failures, vegetation blocking, soiling, and offline modules.

The methodology involved comprehensive evaluation of multiple deep learning architectures, including traditional convolutional neural networks (ResNet50, VGG16, EfficientNet-B0, Xception) and modern Vision Transformers (ViT-B32, DeiT-B16). The models were evaluated across three classification scenarios: binary classification (anomaly vs. no anomaly), 11-class classification (anomaly types only), and 12-class classification (including normal conditions).

Advanced visualization techniques, including t-SNE dimensionality reduction and Grad-CAM activation mapping, were employed to provide insights into the learned feature representations and model decision-making processes. The research addressed the challenges of class imbalance through data preprocessing and augmentation techniques.

7.2   Key Findings

The experimental results revealed several important findings:

7.2.1   Model Performance

- Vision Transformer architectures demonstrated superior performance compared to traditional CNNs across all classification tasks

- Binary classification achieved excellent performance with accuracy scores above 98% for distinguishing between anomalous and normal conditions

- Multi-class classification presented greater challenges due to the complexity of distinguishing between similar anomaly types

- Class imbalance significantly impacted performance for rare anomaly types such as Hot-Spot-Multi and Soiling

7.2.2   Feature Representation Analysis

- t-SNE visualization revealed that Vision Transformers learned more discriminative feature representations with better class separability

- Grad-CAM analysis confirmed that models focused on relevant thermal anomaly regions rather than background areas

- Certain anomaly classes (e.g., Cell defects, Diode failures) formed distinct clusters in the feature space, while others showed overlapping patterns

7.2.3   Practical Implications

- The developed models demonstrate potential for automated inspection systems in solar energy installations

- Binary classification performance suggests feasibility for real-time anomaly detection in maintenance scenarios

- The identification of challenging anomaly classes provides insights for targeted data collection and model improvement strategies

7.3   Research Questions Addressed

Research Question 1: How effectively can deep learning models detect and classify different types of anomalies in photovoltaic systems using thermal imaging?

The research demonstrated that deep learning models, particularly Vision Transformers, can effectively detect and classify anomalies in PV systems. Binary classification achieved over 98% accuracy, indicating excellent capability for distinguishing between normal and anomalous conditions. Multi-class classification showed varying performance across different anomaly types, with some classes achieving F1-scores above 95% while others presented more challenging detection scenarios.

Research Question 2: What are the comparative advantages of different deep learning architectures for thermal image analysis in solar panel inspection?

Vision Transformer architectures consistently outperformed traditional CNNs across all evaluation metrics. ViT-B32 and DeiT-B16 demonstrated superior feature learning capabilities, as evidenced by better class separability in t-SNE visualizations and more accurate classification results. The self-attention mechanism

in transformers proved particularly effective for capturing spatial relationships in thermal anomaly patterns.

Research Question 3: How can visualization techniques enhance the interpretability and validation of anomaly detection models in this domain?

t-SNE visualization provided valuable insights into the learned feature representations, revealing the quality of class separability and identifying potential model confusion areas. Grad-CAM analysis confirmed that models focused on thermally relevant regions, validating the appropriateness of the learned features. These visualization techniques proved essential for model validation and understanding the decision-making processes of the deep learning models.

7.4 Future Work and Recommendations

Based on the results and limitations of this study, several avenues for future research are recommended:

7.4.1 Model Architecture Enhancements

- **Hybrid Architectures:** Explore hybrid CNN-Transformer models that combine the local feature extraction capabilities of CNNs with the global attention mechanisms of transformers

- **Object Detection Integration:** Investigate advanced architectures like YOLO for simultaneous anomaly classification and localization

- **Attention Mechanisms:** Develop specialized attention mechanisms tailored for thermal imaging characteristics

7.4.2 Data and Methodology Improvements

- **Class Imbalance Solutions:** Investigate advanced techniques such as cost-sensitive learning, focal loss, or generative adversarial networks (GANs) for synthetic data generation to improve performance on rare anomaly classes

- **Multi-Modal Data Fusion:** Explore fusion of thermal imaging with RGB images, electroluminescence imaging, or electrical measurements to provide richer information for anomaly detection

- **Transfer Learning:** Investigate transfer learning from related domains such as medical imaging or satellite imagery analysis

7.4.3 Real-World Implementation

- **Real-Time Deployment:** Test trained models on live video feeds from drone-mounted thermal cameras to evaluate real-time performance and robustness in dynamic field environments

- **IoT Integration:** Develop integration with IoT platforms for continuous monitoring and automated alerting systems

- **User Interface Development:** Create user-friendly interfaces for maintenance teams to visualize detected anomalies and access detailed inspection reports

7.4.4   Longitudinal and Scalability Studies

- **Temporal Analysis:** Conduct longitudinal studies to monitor solar panel performance over time and predict future anomalies based on degradation patterns

- **Scalability Assessment:** Evaluate model performance across different geographical locations, climate conditions, and PV system types

- **Predictive Maintenance:** Develop prognostic models that can predict the likelihood of future failures based on current thermal signatures

7.4.5   Domain-Specific Optimizations

- **Environmental Adaptation:** Investigate model robustness under varying environmental conditions such as different weather patterns, seasonal variations, and time-of-day effects

- **Floating PV Systems:** Extend the research to specifically address anomaly detection challenges in floating photovoltaic systems, considering the unique thermal characteristics introduced by water proximity

- **Industry Standards:** Work towards developing standardized protocols for thermal imaging-based PV inspection that can be adopted across the industry

The successful implementation of these future research directions could significantly advance the field of automated solar panel inspection and contribute to the broader adoption of reliable, efficient photovoltaic energy systems.

APPENDICES

A - GITHUB REPOSITORY

All code and latex-files used in this document are included in the Github repo GitHub linked below. Further explanations are given in the readme-file.

Also thanks for reading this.

- https://github.com/ninasalvesen/thesis_latex_template