Master's thesis

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
Faculty of Information Technology and Electrical Engineering
Department of ICT and Natural Sciences

NTNU
Norwegian University of Science and Technology

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
Faculty of Information Technology and Electrical Engineering
Department of ICT and Natural Sciences

NTNU
Norwegian University of Science and Technology

ABSTRACT

Solar Energy is an emerging renewable energy source that has been gaining significant attention in recent years. The increasing demand for sustainable energy solutions has led to the exploration of various solar energy technologies, including land-based and floating photovoltaic (PV) systems. However, with widespread adoption, there is a need to evaluate the performance of these systems under different environmental conditions to determine their efficiency, cost-effectiveness, and environmental impact.

The main objective of this study is to develop deep learning models to detect and classify defects in photovoltaic (PV) systems, specifically focusing on land-based and floating PV systems. The research aims to address the challenges associated with the performance of these systems, particularly in terms of their efficiency, cost-effectiveness, and environmental impact. The study will utilize simulation tools to model the performance of land-based and floating PV systems under various environmental conditions, such as solar irradiance, temperature variations, and the influence of water bodies on the efficiency of floating PV systems.

SAMMENDRAG

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

CHAPTER

ONE

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

. 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 detection 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.

CHAPTER

TWO

LITERATURE REVIEW

This chapter will give a brief overview of the theoretical background about PV systems (Land Based and Floating) behind the methods and algorithms used in this thesis. It will also give an overview of the data used, and how it is structured. The theory will be explained in a way that is easy to understand, and will not go into too much detail.

2.1 Land-based photovoltaic (PV) systems

Land-based photovoltaic (PV) systems are solar energy systems installed on the ground or rooftops, utilizing solar panels to convert sunlight into electricity. These systems harness solar energy for residential, commercial, and utility-scale applications. Key components include solar panels (typically silicon-based), inverters for DC–AC conversion, mounting structures, and electrical connections. Mounting structures ensure panels are securely positioned for optimal sunlight capture. These systems can vary from small residential setups to large-scale solar farms feeding the grid. They play a crucial role in promoting renewable energy adoption, reducing greenhouse gas emissions, and supporting a sustainable energy future.

According to the IEA, land-based PV is one of the fastest-growing sources of renewable energy worldwide, benefiting from low operating costs, scalability, and versatile siting—including rooftops, open fields, and brownfield sites. Advances in solar technology have improved efficiency and lowered costs, enhancing economic viability. Recent studies support these advantages: Belloni et al. (2023) demonstrated how machine-learning optimization can improve O&M efficiency; Bindi et al. (2022) documented dramatic cost reductions, making PV is competitive with fossil fuels, and environmental assessments indicate minimal land-use impact.

2.2 Floating PVs

Floating photovoltaic (PV) systems, or floating solar, are innovative installations deployed on bodies of water such as lakes, reservoirs, and ponds. These systems consist of solar panels mounted on floating platforms or pontoons, allowing them to harness sunlight while floating on the water's surface. ()

Floating PV systems offer several advantages over traditional land-based installations:

- **Reduced land use**: They do not occupy terrestrial land and can be deployed over existing water bodies—including reservoirs, irrigation ponds, and brownfields.
- **Natural cooling and higher efficiency**: Water cooling reduces operating temperatures, increasing electrical output by 5–15%, with some field tests reporting up to 11% improvements.
- **Water evaporation suppression**: Shading from panels reduces reservoir

evaporation by 42–83% in empirical tests—e.g., a 130 kW installation in Brazil cut evaporation by 60%.

- **Higher energy yield and LCA performance**: FPV systems can outperform ground-mounted configurations and feature shorter energy payback times (approximately 1.3 years) and lower lifecycle GHG emissions.

- **Synergy with hydropower**: They can integrate with dams and transmission infrastructure, enhancing grid flexibility.

2.3 Anomalies types in Floating PVs

Like their land-based counterparts, floating photovoltaic (FPV) systems can experience various types of anomalies that affect their performance and efficiency. These anomalies include soiling, algae growth, shading, and structural issues. Detecting these anomalies early is crucial for maintaining an optimal system performance and ensuring the longevity of the floating PV system.

2.4 Anomalies Types in Land-based PVs

PV systems can experience various anomalies affecting their performance and efficiency. These anomalies include hot spots, cell cracks, soiling, shading, and degradation. Detecting these anomalies early is crucial for maintaining optimal system performance.

Hot Spots: Hot spots occur when a portion of a solar panel becomes significantly hotter than the surrounding areas. This can happen due to partial shading, cell mismatch, or manufacturing defects. Hot spots can lead to reduced efficiency and even permanent damage to the solar cells if not addressed promptly.

Cell Cracks: Cell cracks are physical damage to the solar cells that can occur during manufacturing, transportation, or installation. These cracks can reduce the electrical output of the affected cells and may lead to further degradation over time. Detecting cell cracks early is essential to prevent further damage and ensure the longevity of the solar panel. **Soiling:** Soiling refers to the accumulation of dirt, dust, and other debris on the surface of solar panels. This layer of dirt can block sunlight from reaching the solar cells, significantly reducing their efficiency. Regular cleaning and maintenance are necessary to prevent soiling from impacting the performance of the PV system. **Shading:** Shading occurs when objects such as trees, buildings, or other structures block sunlight from reaching the solar panels. Even partial shading can cause significant drops in energy production, as solar panels are designed to operate optimally in direct sunlight. Identifying and mitigating shading issues is crucial for maintaining the efficiency of PV systems. **Degradation:** Over time, solar panels can experience degradation due to exposure to environmental factors such as UV radiation, temperature fluctuations, and moisture. This degradation can lead to a gradual decline in the performance of the solar panels. Regular monitoring and maintenance are necessary to assess the condition of the panels and address any degradation issues promptly.

2.5 Power loss in PV systems

Power loss in photovoltaic (PV) systems refers to the reduction in electrical output due to various factors, including shading, soiling, temperature effects, and anomalies such as hot spots and cell cracks. These losses can significantly impact the overall efficiency and performance of the PV system, leading to lower energy production and increased operational costs. Understanding the causes of power loss is crucial for optimizing the design and operation of PV systems, ensuring that they operate at their maximum potential. By implementing effective monitoring and maintenance strategies, operators can minimize power loss and enhance the overall efficiency of photovoltaic systems.

2.6 Anomalies Detection Methods

Anomaly detection methods in photovoltaic (PV) systems are essential for identifying and addressing issues that can impact the performance and efficiency of solar panels. These methods leverage various techniques, including statistical analysis, machine learning, and computer vision, to automatically detect anomalies such as hot spots, cell cracks, soiling, shading, and degradation. By employing these methods, operators can ensure timely maintenance and repairs, ultimately enhancing the reliability and longevity of PV systems.

2.6.1 Visual Inspection

Visual inspection is a traditional method for detecting anomalies in photovoltaic(PV) systems. It involves manually examining solar panels for visible defects such as cracks, discoloration, and soiling. While visual inspection can be effective for identifying obvious issues, it is often time-consuming and labor-intensive, especially for large-scale PV installations. Additionally, human inspectors may overlook subtle anomalies that could affect the performance of the solar panels. To address these limitations, automated visual inspection techniques have been developed, leveraging computer vision and machine learning algorithms to analyze images of solar panels and detect anomalies with higher accuracy and efficiency. These automated systems can process large volumes of data quickly, reducing the need for manual labor and increasing the speed of anomaly detection in PV systems.

2.6.2 IR Thermography

IR thermography, or infrared thermography, is a non-destructive testing technique that uses infrared cameras to detect and visualize temperature variations on the surface of objects. In the context of photovoltaic (PV) systems, IR thermography is employed to identify anomalies such as hot spots, cell cracks, and other thermal issues that can affect the performance of solar panels. By capturing thermal images, operators can quickly assess the condition of PV systems, enabling them to detect potential problems early and take corrective actions to maintain optimal performance.

2.6.3 EL Imaging

Electroluminescence (EL) imaging is a technique used to visualize the internal structure of photovoltaic (PV) cells by capturing the light emitted when an electric current is passed through the cells. This method is particularly useful for detecting anomalies such as micro-cracks, shunts, and other defects that may not be visible through traditional visual inspection methods. EL imaging provides a detailed view of the PV cells, allowing operators to identify and address issues that can affect the performance and efficiency of solar panels. By employing EL imaging, operators can ensure the reliability and longevity of PV systems, ultimately enhancing their energy production capabilities.

2.7 Machine Learning

2.8 Deep Learning

2.9 Computer Vision

Computer vision is a field of artificial intelligence that focuses on enabling machines to interpret and understand visual information from the world. It involves the development of algorithms and models that can process, analyze, and extract meaningful information from images and videos. In the context of solar PV systems, computer vision techniques can be employed to automatically detect and classify anomalies in solar panels, such as cracks, soiling, and shading. By leveraging computer vision, it is possible to automate the inspection process, reducing the need for manual labor and increasing the efficiency of anomaly detection in solar PV systems.

CHAPTER

THREE

THEORETICAL BACKGROUND

According to [1], 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.

CHAPTER

FOUR

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 from radar chart **??**, 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

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

Table 4.1: Overview of the IR Dataset

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 preproceessing and augmentation techniques were applied to balance the classes for 11 class and 12 class [2] classification tasks which| are discussed in the following sections.

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

4.1.1 Data Preprocessing

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.

Common techniques include scaling numerical features, handling missing values, and encoding categorical variables.

4.2 Image Filter

4.2.1 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.[3]

4.2.2 Image Resizing

4.2.3 Data Augmentation

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 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.[3]

4.2.4 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.

4.2.5 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.

4.2.6 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.

4.2.7 Data Splitting

According to, 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.

4.2.8 Hardware and Software Specifications

4.3 Libraties and Frameworks

Pillow is a Python Imaging Library (PIL) fork that provides easy-to-use image processing [4] capabilities. It supports opening, manipulating, and saving various image [1] 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.

Pytorch:

Pillow: Pillow is a Python Imaging Library (PIL) fork that provides easy-to-use image

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

Table 4.2: 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)

4.4 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.

4.5 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.

4.6 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 (DeiT) Data-efficient Image Transformer architectures.

4.6.1 CNN Models

4.6.2 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 [5]. 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. 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.

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

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

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).

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

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:

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

Hyperparameter	**Vision Transformer (ViT)**
Learning Rate	*1e-4*
Batch Size	32
Epochs	100
Optimizer	AdamW
Weight Decay	0.01
Image Size	160×160
Patch Size	16×16
Number of Heads	12
Number of Layers	12
Hidden Dimension	768
MLP Dimension	3072
Dropout Rate	0.1
Output layer Activation Function	GELU
Learning Rate Scheduler	CosineAnnealingLR
Early Stopping Patience	10

Table 4.3: Hyperparameters for Transformer-Based Models

4.7 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. 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.

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.

One of the reason to choose 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. By leveraging the distillation token and other techniques, DeiT can effectively learn from smaller datasets while maintaining high accuracy and generalization capabilities.

DeiT employs a distillation loss during training, which encourages the model to learn from a teacher model (often a pre-trained 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.

4.8 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:

4.8.1 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.

4.8.2 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.

4.8.3 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.

4.8.4 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.

4.8.5 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.

4.8.6 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.

4.8.7 Distillation Token Interaction

The distillation token learns through attention with other tokens:

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

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

4.8.8 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.

4.9 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 (A UC-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.

4.9.1 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.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4.14}$$

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.

4.9.2 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.

4.9.3 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.

4.9.4 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.

4.9.5 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.

4.10 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 this challenge 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.

Grad-Cam (Gradient-weighted Class Activation Mapping) is a technique used to visualize the regions of an image that contribute most to a model's prediction. It is particularly useful for understanding the decision-making process of deep learning models, especially convolutional neural networks (CNNs) and Vision Transformers (ViTs). Grad Cam provides insights into which parts of an image are most relevant to a specific class prediction, helping to interpret the model's behavior and identify potential areas of interest or concern.

CHAPTER

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 5.1 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 5.2), 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 5.3 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 5.5 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 5.4 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 5.5, 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 5.6, 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 5.7 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]	2	92.50	92.00	92.00	92.00
	11	78.85	—	—	—
	12	66.43	—	—	—
Residual Ensemble	2	94.40	—	—	—
	12	85.90	—	—	—
Alex-Net Multi-scale [21]	2	97.32	97.63	97.00	97.32
	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

CHAPTER

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

What will be the impact of the results? How is Deit better than ViT for detecting solar PV anomalies?

The results of our experiments demonstrate the effectiveness of [Your Model, e.g., the fine-tuned ViT model] in detecting anomalies in infrared images of solar panels. The high F1-score of [Your Score] for the [Best Class] class suggests that [2] the model is particularly adept at identifying [Type of Anomaly].

This performance can be attributed to [Reason, e.g., the Vision Transformer's ability to capture global context within the image...]. When compared to the baseline [e.g., ResNet50] model, our proposed architecture achieved a [e.g., 5

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 [Your Technology] 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.
- **Increased Energy Yield:** Early detection of issues like **<empty citation>** soiling, cracking, or faulty diodes prevent long-term power loss 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.

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

6.2 Limitations

While this study provides valuable insights, it is important to acknowledge its limitations. Gitstash is a good model for detecting solar PV anomalies.

Now we need to compare the results of Gitstash with the results of ViT and ResNet50. we need to use the same dataset and the same evaluation metrics.

- **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.

- **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.

6.3 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 and Vision Transformers, on the publicly available *IR* dataset.

CHAPTER

SEVEN

CONCLUSIONS

Give a concise summary of your research and finding here, and include a short summary of any future work as well.

- Summary of research
- Key findings
- Future work

Research Question 1

Research Question 2

Research Question 3

Background information on the research topic: The research topic is centered around image processing, specifically focusing on noise reduction techniques and their impact on image quality. The study explores various algorithms and their effectiveness in enhancing images while preserving important details.

7.1 Future Work and Recommendations

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

- **Exploring More Architectures:** Future work could expand the analysis to include other advanced architectures, such as hybrid CNN-Transformer models or more recent state-of-the-art object detection models like YOLO or DETR, to not only classify but also localize the anomalies.

- **Real-World Deployment and Testing:** The next logical step is to test the trained model on a live video feed from a drone-mounted thermal camera to evaluate its real-time performance and robustness in a dynamic field environment.

- **Addressing Class Imbalance:** Further investigation into advanced techniques for handling class imbalance, such as cost-sensitive learning or generative adversarial networks (GANs) for synthetic data generation, could improve performance on rare anomaly classes.[6]

- **Transfer Learning from Other Domains:** Exploring transfer learning from related domains, such as medical imaging or satellite imagery analysis, could provide additional insights for improving anomaly detection performance in solar panel inspection tasks.

- **Multi-Modal Data Fusion:** Future research could explore fusing thermal data with corresponding visual-spectrum (RGB) images to provide richer information to the model, potentially improving classification accuracy for ambiguous cases like shadowing versus soiling.

- **Longitudinal Studies:** Conducting longitudinal studies to monitor the performance of solar panels over time, using the developed models to detect changes in performance and predict future anomalies, could provide valuable insights into the long-term reliability of solar energy systems.

- **Integration with IoT Systems:** Future work could focus on integrating the anomaly detection system with IoT platforms for real-time monitoring and alerting, enabling proactive maintenance strategies in solar energy systems.

- **User Interface Development:** Developing a user-friendly interface for maintenance teams to visualize detected anomalies and access detailed reports could enhance the practical utility of the developed system, making it easier to implement in real-world scenarios.

  Future work could also explore multi-modal approaches that combine thermal [3] imaging with other data sources, such as visual spectrum images or electrolumsinescence images to improve the model performance and diversity to captures a wider range of anomalies.

REFERENCES

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin, S. Bengio, and Y. Yang, "Attention is all you need," *Advances in Neural Information Processing Systems (NeurIPS)*, 2017. [Online]. Available: https://arxiv.org/abs/1706.03762.

[2] M. Ibrahim, A. Alsheikh, F. M. Awaysheh, and M. D. Alshehri, "Machine learning schemes for anomaly detection in solar power plants," *Energies*, vol. 15, no. 3, p. 1082, 2022.

[3] C. Deline, "Partially shaded operation of multi-string photovoltaic systems," in *2010 35th IEEE Photovoltaic Specialists Conference*, IEEE, 2010, pp. 000 394–000 399.

[4] C. F. Jeffrey Kuo, S. H. Chen, and C. Y. Huang, "Automatic detection, classification and localization of defects in large photovoltaic plants using unmanned aerial vehicles (UAV) based infrared (IR) and RGB imaging," *Energy Conversion and Management*, vol. 276, p. 116 495, Jan. 2023. DOI: 10.1016/J.ENCONMAN.2022.116495. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0196890422012730.

[5] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, et al., "An image is worth 16x16 words: Transformers for image recognition at scale," *International Conference on Learning Representations (ICLR)*, 2020. [Online]. Available: https://arxiv.org/abs/2010.11929.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. [Online]. Available: https://arxiv.org/abs/1512.03385.

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-filereadme fileepository link

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

NTNU

Norwegian University of
Science and Technology