5-2024

Defect Detection of Photovoltaic Panels by Image Processing

Hamad AlMarzooqi
hma7996@rit.edu

Defect Detection of Photovoltaic Panels by Image Processing

by

Hamad AlMarzooqi

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of Master of Science in Professional Studies: Data Analytics

Department of Graduate Programs & Research

Rochester Institute of Technology

RIT Dubai

May 2024

`

RIT

Master of Science in Professional Studies:

Data Analytics

Graduate Thesis Approval

Student Name**: Hamad Almarzooqi**

Graduate Capstone Title**: Defect Detection of Photovoltaic Panels by Image Processing**

Graduate Thesis Committee:

Name: Dr. Sanjay Modak Date:

Chair of committee

Name: Dr. Ehsan Warriach Date:

Mentor

Acknowledgments

I would like to express my thanks to my advisor Dr Ehsan who has continuously supported, advised and guided me to conduct this thesis. I would also convey my appreciation to Rochester Institute of Technology and Dubai Electric and Water Authority for providing the required resources to conduct the finalize this research. A big thanks to Dr Sanjay Modak, the chair of Graduate Programs and Research in RIT Dubai for his Cooperation and support throughout the program and for his invaluable feedback. To my colleagues and my friend who already graduated from the same program Abdulrahman Aljanaahi and Abdulla AlAmeeri from Dubai Police, who supported and assisted me through the program and made work more bearable. Special thanks to my wife and my family who encouraged me to continue my self-development and my education. Finally, I would like to express my gratitude to TDRA for allowing me to study in such an incredible institute with a full scholarship. This journey was made smoother and more meaningful thanks to each of you. Your support and belief in my work have been invaluable.

Abstract

A key component of the transition towards cleaner and more sustainable power sources, driven by the global demand for such energy, has been the fast improvement in the installation of solar photovoltaic (PV) plants. Ensuring the dependability of photovoltaic panels becomes crucially important when solar installations develop in scale. Photovoltaic panels have flaws that reduce their effectiveness and shorten their lifespan, but they are durable and long-lasting otherwise. Solar plants need to work as efficiently as possible with low downtime and to want solar energy to be viable in the long run, the issues shall be fined a fixed promptly. Quality monitoring has always relied on human inspectors. The speed, accuracy, and scalability issues that plague human inspection methods become more apparent as solar projects expand in size. The goal of this research is to improve solar panel flaw identification using cutting-edge image processing techniques. Getting beyond these problems is its aim.

The project used Convolutional neural network (CNN). We will use ResNet50. ResNet50 is a convolutional neural network (CNN). It is a specific architecture within the CNN family that is designed for tasks related to computer vision, such as image classification, object detection, and image segmentation. The datasets that will be used in these sections are collected from Kaggle and from DEWA Solar PV Plants. The data will be subjected to Data Preprocessing, Data Cleaning and Organization, Data Augmentation and Feature Extraction. further building the model and go through the training, validation and testing process to ensure a high-performance model that accurately predicts the defects.

Keywords: *PV Solar Plants, convolutional neural network (CNN), solar plants, energy efficiency,*

Contents

Table of figures

CHAPTER 1: INTRODUCTION

1.1 Introduction

Switching to green energy sources, especially solar power, holds a lot of promise for preventing climate change and making sure that the future will be sustainable. Photovoltaic (PV) panels are the most important part of solar energy production because they turn sunlight into power very efficiently (Tang, 2020). On the other hand, PV panels can only work reliably if they are structurally sound and functionally viable, and both can be affected by flaws. Usually, manual checking methods have been used to find problems with PV panels. But these ways aren't perfect, so they're not as useful or reliable as they could be (Akram, 2019). First, checks that are done by hand take a long time and a lot of work. They also need a lot of people and tools. Also, the fact that human inspectors read flaws in their way makes evaluation results somewhat inconsistent and variable. When checks are done by hand, mistakes and missed opportunities can also happen. This means that problems that aren't noticed may get worse over time.

Because of these issues, we need automated solutions right away that can help us find issues with PV panels. The goal of this project is to use image processing to make advanced algorithms that can quickly and accurately find, classify, and identify problems in PV panels (Niazi, 2019) . Taking full shots of solar power plants with drones equipped with high-resolution cameras will make it possible to find many types of problems with them. We can see any cracks, damage, electricity issues, or even dust in these pictures. Because it might change how checks and maintenance are done in the solar energy business, this project is very important. By automating the steps needed to find flaws, that are usually done by workers at solar energy plants, useful information that they can use to fix issues before they get worse. By cutting down on downtime, more energy can be made (Ren, 2022). Also, solar power plants should be more stable, work better, and last longer if they have automated systems that find problems. This will speed up the world's movement to a future with clean energy. This project is mostly an important step toward making tools for finding problems in PV panels even better. With the help of cutting-edge image processing techniques, we want to start a new era of proactive maintenance and efficiency optimization. This will make solar energy stronger and more stable as a major part of the world's energy system for a longer time.

1.2 Project Goals

Image processing methods will be used to make accurate and useful models or programs that can find problems in photovoltaic (PV) panels. This is the main goal of the project. These models are meant to automatically find defects in PV panels, mark them, and report them. This will make solar energy plants more reliable, efficient, and long-lasting. The main objectives include a multifaceted strategy aimed at finding all defects and responding quickly:

1.2.1 Collection and Preprocessing of Defect Data

- As part of the project, different kinds of flaws that are common in PV panels will be carefully gathered. These flaws may include cracks, physical damage, electrical problems, and soiling (dust).
- Data preprocessing methods will be used to clean, improve, and organize the defective data that has been collected. This will make sure that it can be used for training and validating models later.

1.2.2 Accurate Localization of Defects

- One of the main goals of the project is to find and precisely locate problems in PV panels. This means making algorithms that can precisely locate and describe the scope of problems, making focused maintenance interventions easier.
- Accurate localization is important so that plant operators can quickly find and fix problems, which cuts down on downtime and improves working efficiency.

1.2.3 Development of an Automated System

- One of the main goals of the project is to create an automated system that can quickly handle pictures taken by drones or other imaging devices.
- Modern image processing methods will be used by this system to look at captured images in real-time or almost real-time, which will make it possible to quickly find and label defects.
- By automating the process of finding problems, the device aims to make preventative maintenance in solar power plants better. In the future, this will help the plants work better and cost less to run.

To meet the project goals, many different steps will be taken. These will help find problems in PV panels more easily. There will be useful methods and an automated system set up as part of the project so that the Operation and Maintenance (O&M) team who run solar power plants can know what to do and get it done before problems happen. This means that solar energy will be a reliable source of power for a long time. It will help them get more power.

1.3 Aims and Objectives

1.3.1 Aims

1.3.1.1 Enhancement of O&M Inspection Procedures in Solar PV Plants

- We want to use high-tech computer tools to identify the defects in solar photovoltaic (PV) plants instead of the traditional way of inspection which is done manually using their eyes.
- As part of the project, image processing technologies will be used to make the way things are usually checked better. This will help O&M processes work better and be more useful in general.

1.3.1.2 Reduction of O&M Costs and Downtime/Shutdowns

- It's important to cut down on the expense of operation and maintenance (O&M) and the time that solar panels that break need to be turned off.
- The goal of the project is to use automated tracking systems to find and fix problems before they happen. This will clear up the need for as many and longer maintenance jobs. This will help the plant work better and make it easier to get.

1.3.1.3 Improvement of Efficiency, Reliability, and Effectiveness of Solar PV Plants

- The main goal is to make solar PV plants more reliable, efficient, and useful by finding problems on their own.
- To make solar energy systems work better overall and make more energy, the project wants to make sure that problems are found and fixed fast. This will also make PV panels last longer.

1.3.2 Objectives

1.3.2.1 Development of Models or Algorithms for Defect Detection

- To reach this objective, we need to learn how to use image processing to find problems in PV panels and make models or programs that we can trust.
- It is important to make quick and accurate programs that can find many types of flaws, such as cracks, physical damage, electrical issues, and dirt.

1.3.2.2 Collection and Preprocessing of Defect Data

- The project's objective is to collect and clean up different types of bad data so that a full dataset can be made that can be used to develop the model.
- There will be ways used to clean, standardize, and add to the data so that it can be used for both training and validation.

1.3.2.3 Implementation of an Automated Detection System

- Collecting the data by using cameras or valid data through Kaggle.
- Set up an automatic system that can find and classify defects in PV panels so that we can reach our goal.

1.3.2.4 Localization of Defects for Targeted Maintenance

- Finding specific problems in PV panels is a very important goal that will allow for focused maintenance.
- Finding the exact location of defects will help O&M staff identify and fix problems more quickly, reducing downtime and making the best use of resources.

1.3.2.5 Performance Evaluation of the Automated System

- It is important to try the automatic defect detection system thoroughly to see how well it works, how fast it works, and how reliable it is to reach this goal.
- We will set up and track performance measures to see how well the system works in real life and make sure it is useful and effective.

1.3.2.6 Comparison with Manual/Visual Inspection Methods

- The goal is to compare how well the automated solution for finding flaws is efficient to the old ways of checking by hand or by sight.
- A higher performance detecting method compared to labor-intensive and error-prone manual examination methods.

To sum up, the project's objectives and aims include diverse ways to make it easier to find issues in solar photovoltaic (PV) plants. With the help of cutting-edge technologies and big problems, the project aims to start a new age of proactive maintenance and optimization. This will eventually make solar energy a more sustainable and viable source of renewable energy.

1.4 Research Methodology

The study methodology goes into great depth about the approaches and methods that were used to finish the project. The steps are made clear in this in-depth explanation:

1.4.1 Data Collection Methods for Acquiring Images of Defective PV Panels

- **DEWA PV Plants and Kaggle Dataset Acquisition:** the photos that are taken from site and datasets from sites like Kaggle that store pictures of broken PV panels to gather data.
- **Characteristics of the Dataset:** The dataset that was collected has a variety of flaws, such as cracks, physical damage, electrical problems, and soiling (dust). This makes sure that the data is correct and diverse.
- **Adding to the dataset:** To make the dataset even bigger and more varied, enhancement techniques like flipping, rotating, and scaling can be used on the pictures that have already been collected.

1.4.2 Preprocessing Techniques for Cleaning and Preparing the Dataset

- **Python scripting:** The preprocessing methods will be used with the computer language Python.
- **Data Cleaning and Organization:** Using a custom method in our code, 'check_images()', corrupted and incorrect images were systematically detected and separated into a different directory, assuring data integrity and consistency.
- **Data Augmentation:** A technique that enhances training datasets by applying transformations like rotation or color adjustment to existing data. This method improves model robustness and performance by simulating diverse scenarios.

- **Feature Extraction:** Extract Features from the images to empower it to discern subtle patterns and distinctions essential for accurate defect detection and localization within PV panels.

1.4.3 *Image Processing Algorithms for Defect Detection and Classification*

- **Python Implementation:** Some Python tools that will be used to make and run image processing methods are scikit-image, OpenCV, and TensorFlow.
- **Feature Extraction:** Among other things, we will use edge recognition, texture analysis, and morphological operations to get useful information from images that have already been processed on.
- **Models Building:** Convolutional neural networks (CNNs) is a deep learning algorithm used for processing visual data, learning to recognize patterns through layers that perform convolutions.
- **Selection of Base Model Architecture:** ResNet50 is a convolutional neural network (CNN). It is a specialized architecture within the CNN family that is intended for computer vision applications such as picture classification, object recognition, and segmentation.

1.4.4 *Evaluation Metrics for Assessing the Performance of the Automated System*

- **Evaluation of Accuracy:** How good the automatic system is at finding bugs will be judged by the number of flaws that are correctly labeled.
- **Precision, Recall, and F1 Score:** It will be checked for Precision, Recall, and F1 Score to see how well it can find and mark bugs.

- **Confusion Matrix Analysis:** The confusion matrix shows how the true positive, true negative, false positive, and false negative forecasts are spread out. From this, we'll know how well the plan is going.
- **Receiver Operating Characteristics (ROC):** The ROC curve, accompanied by corresponding area under the curve (AUC) values, served as a pivotal metric in assessing the model's discriminatory prowess across diverse defect categories.

1.4.5 Comparative Analysis with Manual/Visual Inspection Methods

- **Benchmark to Traditional Manual Method:** The automatic system for finding defects will be tested in the field to see how well it works in real life compared to the old ways of checking by hand and visually and to compare automated and manual inspection methods, the rate of false positives, the accuracy of detection, and the time it takes to process will be used.

The project's goal is to create strong image processing algorithms for automatically finding flaws in PV panels by following the detailed study methodology described. The Python code for these methods will be used, and data from Kaggle datasets will be used. The automatic system is made in a way that makes it stable and works well so it can be used in real solar energy plants.

1.5 Limitations of the Study

There are some limitations to this study that could change its scope, results, and ability to be used in other cases. Some significant issues are being tried to be solved by the study by making photovoltaic (PV) screens automatically find flaws. Here are some of these rules:

1.5.1 Availability and Quality of Dataset for Training and Testing

- **Limited Dataset:** There might not be plenty of full datasets with different broken PV technological devices available. What this means is that the models might not be as accurate or useful in real life.
- **Data Imbalance:** It's possible for model training and evaluation to be biased if the defect classes are spread out irregularly in the dataset. This would make the automatic defect detection system less accurate and less good at finding problems.

1.5.2 Complexity of Certain Defects

- **Challenging defects:** Because they are so complicated, some defects, like microcracks or small electrical problems, can be hard for machines to find.
- **Surface Conditions That Are Different:** The colors, textures, and reflections of the surface may be different from one PV panel installation to the next. This can make it even harder to find and correctly spot problems.

1.5.3 Computational Resources

- **Resource Intensity:** It may take a lot of storage room and processing power to process a lot of high-resolution pictures taken by drones or imaging devices on the ground.
- **Algorithm Optimization:** It might be harder than it seems to make picture processing and machine learning algorithms work better in places with few resources.

1.5.4 Environmental Factors

- **Lighting:** Shadows, glare, or cloudy skies can change the lighting, which can add noise and artifacts to pictures that are being recorded. This could make algorithms that find defects less accurate.
- **Relying on the Weather:** Rain, fog, or dust storms can make imaging activities with drones harder to do or less effective, which can slow down that process.

1.5.5 Accuracy Limitations in Image Processing Techniques

- **False Positives and Negatives:** When image processing methods are used to find flaws, they might not be able to tell the difference between real flaws and image glitches or other strange things that are present. This could cause false positive or false negative detections.
- **Threshold Selection:** It can be hard to agree on the best limits for feature extraction and classification, and it may take a lot of tweaking to get good results for all PV panel setups.

1.5.6 Images Collecting

- **Collecting the Images:** Taking the images manually will be very challenging since Mega PV plants contains millions of PV Panels.
- **Allocating:** Locating the PV Panels that are defected will be difficult since there will be only images without labeling or locating the panel for ease of finding the exact panels that are defected.

1.6 Mitigation Strategies

The following plans will be put in place to deal with these problems and lessen their possible effect on the study results and for future works to be done:

- **Data Augmentation:** Adding to the dataset by combining more examples and types of flaws to fix problems with data imbalance and lack of availability.
- Iteratively improving and refining image processing and machine learning methods to make them more reliable and able to handle a wider range of defect situations.
- **Verification and Validation:** Make sure the automated defect discovery system works correctly by testing it in the real world, using cross-validation, and sensitivity analysis.
- **Environmental Controls:** Steps are being taken to lessen the effect of environmental factors on imaging operations. Activities with drones ought to be made when the weather is good, and lighting changes should be kept to a minimum during picture pre-processing.
- **Continuous Improvement:** Defect finding software will get better over time with the help of user comments and the lessons learned from putting it into use over time. This is because it focuses on small steps toward growth.
- **Drone:** Use an automated drone that will fly and shoot the images autonomously for the full plants that will be operated in.
- **GPS:** Use a GPS that will allocate any defected panels for ease of finding the exact defected panels.

This study aims to improve the dependability, effectiveness, and usefulness of automated defect detection solutions for PV panels by recognizing and proactively addressing these issues. This will help with the maintenance and sustainability of solar energy infrastructure.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

It's very important for the long-term health of the energy sector that solar photovoltaic (PV) plants are stable and work well (Althubiti, 2022). Newer systems that use picture processing to find mistakes have taken the place of older ones that needed people to do a lot of work. Sun panels can have problems like cracks, spots, or dirt on them that make them not work as well or last as long. This is now much easier to find and fix thanks to these new tools. (Livera, 2019). It has been very helpful to use computer vision and machine learning, especially Convolutional Neural Networks (CNNs), to find trends. It is now easy to sort the solar cells and find any mistakes. With this automated method, not only is maintenance more accurate because the types of problems are found more precisely, but the plant is also more efficient because unexpected downtime is cut down (Shakya, 2021). When real-time tracking frameworks are paired with algorithms for finding bugs, operational efficiency goes through the roof. This is because problems can be found quickly and upkeep can be done to keep them from happening (Abbassi, 2020). Still, there are issues, such as the need for very large datasets with labels, environments that change, and the development of more advanced picture processing methods (Appiah, 2019). Still, the move toward automated systems that use pictures to find flaws is part of a bigger trend in the clean energy field to use deep learning and machine learning to make things more reliable and effective (Chen, 2022). This review of the literature shows how important these technologies are for making it easier to maintain and run big solar power plants. It tells you what went well and what needs to be done more (AlShorman O. I., 2020). It shows both what has been done well and what more needs to be done.

There is a lot of talk about how important it is to find problems quickly in photovoltaic (PV) panels in new studies that use feature extraction methods (Fadhel et al., 2019). Form descriptors, color-based qualities, and texture analysis are some of the things that researchers have investigated. They have also used deep learning and machine learning to automatically find and pick features, which has shown some promising results. To sort problems, neural networks, decision trees, and support vector machines are some types of guided learning that have been used (Dong, 2019). Transfer learning methods are getting more attention as a way to improve classification accuracy, especially with big datasets. With multimodal imaging that includes thermal, visible, and hyperspectral data, it might be possible to find more defects (Haque et al., 2019). There are still problems because there are different kinds of panels, the weather can change quickly, and the need for real-time tracking. In the future, people will try to get around these problems by looking into new tools and methods.

2.2 Literature Review

Tito G. Amaral, Vitor Fernão Pires, ORCID and Armando J. Pires presents their finding in their research article "Fault Detection in PV Tracking Systems Using an Image Processing Algorithm Based on PCA". The literature study for their article is about how photovoltaic (PV) power plants are changing their role in producing renewable energy, how hard it is to keep tracker systems working at their best, and how fault diagnosis methods are being investigated. In this situation, traditional methods that use sensors and statistical models have been praised for how well they track the alignment of PV panels to get the most energy out of them. However, these methods have big problems that make them less useful and scalable, like the fact that they are expensive and take a long time to set up and analyze data. New developments in machine learning (ML) offer a potential alternative, and more and more studies are looking into how it can be used to monitor and diagnose PV systems automatically. There is a new way to find mistakes with unsupervised machine learning methods that don't need labeled training data. This looks like a good way to deal with how complicated and changing solar energy output is. This new way of finding the slope of PV panels, which uses Principal Component Analysis (PCA), is a big step forward. It does this by increasing the number of variables instead of decreasing them. Older methods had problems, like the need for big networks of sensors. This one, on the other hand, is said to be better because it can work even when images are flawed. This review shows how important the suggested method is for making fault diagnosis faster and more accurate in PV trackers by comparing it to another research that has already been done. This means that solar energy systems will last longer and cost less (Tito G. Amaral, 2021).

V S Bharath. Kurukuru, Ahteshmaul Haque and Mohammed Ali Khan talk about how important it is for solar photovoltaic (PV) systems to have accurate flaw detection and analysis in order to work reliably and efficiently. It would bring to light how hard it is to figure out what's wrong with PV arrays, especially because of how they work in the sun and how complicated standard series-parallel configurations are, which don't make it easy to shut down completely during faults. Previous study on finding faults has mostly focused on thermal imaging methods. These methods are good at finding heat anomalies that are indicative of defects, but they have a lot of problems because they can be affected by heat sources that aren't related to the fault. This can cause false positives when trying to find faults, which is why we need more advanced picture processing techniques that can separate and get rid of background noise and other outside influences. The review would then go into more detail about new ways to find faults more precisely by using edge detection and Hough transform methods. These methods are a big step up from the old ones because they use feature extraction followed by classification algorithms and better image processing to make fault localization obvious. Reports of high accuracy rates in both the training and testing phases further support the usefulness of these techniques. They show that they could be more accurate and reliable than standard fault detection methods (V S Bharath. Kurukuru, 2019).

Deitsch and his colleagues research in 2019, focus on the use of machine learning technologies such as Support Vector Machines (SVM) and Convolutional Neural Networks (CNN) to identify the flaws in photovoltaic (PV) module cells using electroluminescence imaging. by using an automated flaw diagnosis in this research, it will participate in the efficiency and performance of the full system of PV energy specially when the practice procedure is a time consuming and it a specialized operation. the usage of SVM requires hardware and can be used through a mobile device and drones with an accuracy of 82.44% while the CNN relies on GPUs which has a better accuracy of 88.42% but it needs more computational power. There is difference between the two methods and a comptonization between the computation power and the accuracy when it used in real-time applications accordingly the decision shall be done based on the requirements of the project and the available resources. in this research, it emphasizes the need of machine learning into the operation of the clean energy field such as photovoltaic (PV) plants. The study shows that by deploying such automated system it will boost the efficiency of the clean energy plants and make it more reliable (Deitsch et al., 2019).

Google for developer’s courses, on machine learning courses that focus on using technology to identify flaws in images. It's important to assess how accurately the model works. ROC curves and AUC come in handy here. An ROC curve demonstrates how well the model can differentiate between incorrect labels by plotting the Positive Rate (TPR) against the False Positive Rate (FPR). TPR shows the number of positives that are correctly predicted while FPR shows negatives predicted as positives. This graph is to showcase the balance between correctly predicted positives as possible and reducing incorrect positive predictions. The Area Under the ROC Curve (AUC) gives a value summarizing the model performance. However, the Area Under the Curve has its limitations. From time-to-time exact probabilities provide insight than rankings, especially in scenarios where false positives have high costs. By using these metrics, it helps in selecting models that effectively strike in detecting defects without many false predictions (Classification: ROC Curve and AUC, n.d.).

The research paper authored by Harrou and colleagues in 2019 presents a technique, for detecting abnormalities in power (PV) systems using a one class Support Vector Machine (1SVM). This method specifically examines the DC side of PV systems. Utilizes a model driven strategy to detect and pinpoint faults. By replicating the patterns of PV arrays through the one diode model of PV cells this approach can identify deviations that indicate potential issues. The one class SVM is used to analyze the differences from these simulations effectively distinguishing between abnormal behaviors. This unsupervised method was trialed on a 9.54 kWp grid connected PV facility in Algiers, where it showcased fault detection capabilities compared to clustering techniques like K means and Birch. Notably the one class SVM stands out in managing data characteristics without assuming any underlying data distribution thus making it well suited for the varying conditions commonly encountered in PV system operations. This research points the robustness and reliability of deploying one class SVM algorithm for an unsupervised monitoring of PV system and by using this approach it will increase the efficiency and effectiveness in terms of energy production of the PV plants (Harrou et al., 2019).

Du-Ming Tsai, Shih-Chieh Wu and Wei-Yao Chiu focus on how solar power is becoming more and more important as a long-term energy source, with Mult crystalline silicon solar cells being at the cutting edge of this technology. It would talk about how hard it is to make sure that solar modules are of good quality and work well, since many problems inside them can't be found with regular CCD imaging systems. This means that advanced imaging methods are needed for a full check. Electroluminescence (EL) imaging has been used a lot in the past because it can show flaws in solar cells by pointing out changes in the infrared light output that are caused by flaws like tiny cracks and finger interruptions. When EL imaging is used, however, the uneven background it creates often makes it less useful, which makes it very hard to find defects. Then, the papers would talk about independent component analysis (ICA) as a new way to find problems in solar modules. ICA makes it easier to find and classify flaws by splitting solar module images into smaller images that can be looked at more closely. This is done with the help of independent basis images and reconstruction error analysis. This method is a big step forward in the field; it offers higher discovery rates and shows how machine learning techniques could be used to make it easier to find problems in solar energy systems. The reported high mean recognition rate is more proof that the proposed method works. It is different from previous methods and sets a new bar for how accurately solar modules should be inspected (Du-Ming Tsai, 2012).

The blog on Towardsdatascience delves into a range of metrics and techniques used to assess the performance of machine learning and deep learning models. It discusses evaluation measures like the confusion matrix, which illustrates how well a model distinguishes between false positives and negatives. Alongside accuracy, precision, recall, specificity and the F1 score are also covered each shedding light on aspects of a models effectiveness. The article also examines Precision Recall (PR) and Receiver Operating Characteristics (ROC) curves as tools that help evaluate the trade offs between positive rates and false positive costs. The ROC curves will offer guidance on choosing the best model or class through comparing PR curves within the dataset. This exploration will assist in not only measuring the performance of the model it will also assist in selecting the most appropriate evaluation metrics based on the data and the requirements (Nighania, 2021).

Wuqin Tang, Qiang Yang, Kuixiang Xiong and Wenjun Yan explained in detail about how important maintenance is for big photovoltaic (PV) power plants, with a focus on how new technologies can help solve the problem of finding problems quickly. It would show what's wrong with current upkeep methods, mainly the fact that there aren't many good electroluminescence (EL) pictures, which are needed to find problems in PV modules correctly. Many studies have looked at different ways to find defects in the past, but this new method of adding more generative adversarial networks (GANs) to the collection of EL images is a fresh way to solve the problem of not having enough samples. Deep learning has changed over time in the field, with convolutional neural networks (CNN) becoming a powerful tool for defect classification and feature extraction. The study would also talk about these changes. If we compare the suggested CNN-based model to well-known machine learning models like VGG16, ResNet50, Inception V3, and MobileNet, we can see that deep learning methods are better at automating the defect-identifying process and are more accurate. The research shows that the suggested method is a big step forward in the field; it gives us a good, efficient way to manage of PV power plants (Wuqin Tang, 2020).

The paper titled "Fault diagnosis for photovoltaic array based on convolutional neural network and electrical time series graph" written by Xiaoyang Lu and colleagues presents a new system for diagnosing faults in photovoltaic (PV) arrays by employing deep learning methods, particularly convolutional neural networks (CNNs). In response to the challenges of identifying faults in PV systems for the effective and safe functioning of solar power generation the research emphasizes the limitations of traditional fault detection approaches that often struggle due to the complexities introduced by Maximum Power Point Tracking (MPPT) technologies. Conventional protection mechanisms, constrained by the characteristics and dynamic operational conditions of PV arrays frequently result in reduced system efficiency and heightened fire risks. To address these issues the authors introduce a technique that converts current and voltage data from PV arrays into a two-dimensional electrical time series graph (ETSG). This ETSG is used as input to a CNN model engineered to extract features for fault diagnosis removing the necessity for manually engineered features. The CNN design comprises layers of convolution and pooling followed by a connected layer responsible for categorizing fault types. The outcome of this research is achieving a model accuracy above 99% and it was conducted on a grid connected PV system. This outcome highlights how the machine learning can improve the operation of the PV plants and enhancing the safety and efficiency by increasing the generation of the PV plants and more the plants more dependable (Lu et al., 2019).

Zhicong Chen, Yixiang Chen, Lijun Wu, Shuying Cheng and Peijie Lin discussed about how Solar panels can develop flaws including cracks, discoloration, hotspots, and soiling due to the wide range of environmental variables to which they are subject, such as dust, temperature variations, and humidity. Quality monitoring has always relied on human inspectors, the speed, accuracy, and scalability issues that plague human inspection methods become more apparent as solar projects expand in size. The flaws can reduce the efficiency of photovoltaic panels. Some examples of such flaws include electrical mismatch, fractures, and delamination. Traditional inspection methods, such as manual visual examination, are not practical for use with large-scale PV systems due to their subjective nature and the amount of time they take. We can note that manual inspection methods are not precise and accurate compared to an automated method. The research shows a method that considered a huge step forward in the PV fields (Chen et al., 2019).

Akshaikps website focuses on the role of image preprocessing in machine learning especially highlighting the effectiveness of "TensorFlows" and "ImageDataGenerator". This tool is crucial, for getting image data ready ensuring that models work well by using methods like data augmentation to expand size through transformations like rotations and flips. The site provides code examples for the ImageDataGenerator for data augmentations showcasing how useful it is, in real world situations. It also explains the flow_from_directory function, which streamlines creating training and validation datasets making it easier to handle image collections and make the model training simpler. It shows the importance of normalization by adjusting the pixel values to improve the input consistency and model effectiveness. Akshaikps insights offer guidance for programmers or beginners aiming to enhance their machine learning projects with preprocessing techniques (Akshaikp, 2021).

As the degree of solar power around the world grows quickly, it becomes more and more important for the solar energy field to have high-tech inspection systems. Antonio Di Tommaso, Alessandro Betti, Giacomo Fontanelli, and Benedetto Michelozzi look at this. First, it would find out what's wrong with the old ways of checking solar panels, especially when it comes to how often, correctly, and quickly they need to be fixed. Adding aerial images from robotic aerial vehicles (UAVs) is a big step forward in this field. Panel checking can be done in more places and with more freedom because of it. After that, the review would be about how the YOLOv3 network and computer vision techniques were mixed to make a new, multi-stage model that could find many types of panel flaws. This method is unique because it can look at both thermographic and viewable pictures. This lets you find a wide range of issues and get useful data for planning maintenance. We can compare the model to other methods to show how much better it is at finding flaws in a lot of different types of PV systems and panel situations. Its short inference times and high accuracy rates would show this. Tests of the model's high performance in the real-world show that it could change the way maintenance is done in the solar business by giving a data-driven way to make operations more reliable and efficient (Antonio Di Tommaso, 2022).

Feature mapping in GeeksforGeeks website explores how raw input data can be transformed into a format for machine learning analysis which is known as feature engineering. the crucial step in data analysis involves converting data from a lower dimensional space to a higher dimensional space one to make the analysis or the classification easier. Many techniques such as feature extraction, transformation, selection, scaling and engineering are explored in the website all of them aimed at revealing patterns in the data. For example, feature extraction may involve identifying Characteristics like edges of the images while transformation could convert text into a bag of words image. additionally, GeeksforGeeks explains the methods like dimensionality reduction using PCA to preserve information while reducing the number of features and embeddings that generate compact vector representations of features. Furthermore, the website elaborates the data augmentation that will strengthen the model resilience and use it domain knowledge for feature selection. However, feature mapping can pose challenges such as overfitting and increased computational requirements (GfG, 2023).

Ronnie O. Serfa Jose Juan and Jeha Kim talk about how important electroluminescence (EL) photography is for making sure that solar cells are healthy and working well. EL imaging is one of the best ways to find problems and make sure that solar panel cells are healthy. It can show important things about solar cells' materials and traits that you can't see with the naked eye. It would also show how EL imaging has changed over time as a diagnostic tool. After that, the review would talk about how EL imaging and digital image processing can work together to help solar systems find issues more quickly. It would stress how important it is to gather traits for better flaw classification. Help Vector Machine (SVM) models are a smart way to tell the difference between broken and working solar cells. This is for a good reason: the program is strong enough to handle lots of information and arrange it well. You can tell that the suggested SVM-based classification system works better than other ways of diagnosing something. This is because it works really well. We'll talk about how the study fits into the bigger picture of progress in finding PV. We are going to talk about how the method correctly finds and rates faults, which makes solar energy production more reliable and effective (Ronnie O. Serfa Juan, 2020).

The data augmentation for images in TensorFlow provides guidance and instructions, the first tip is expanding your training datasets by incorporating real life modifications like rotating images. It discusses two strategies, the first strategy using Keras preprocessing layers such as Tf.keras.layers.RandomRotation, which can automate the adjustments throughout the training by integrating it into the model. The second strategy explains the utilization of tf.image techniques for alterations such as horizontally flipping of the images or tuning the brightness levels of the image granting the model developer the authority over how the images are adjusted. These techniques are crucial for training the models to succeed. To improve the precision and flexibility of the image processing models whether they're detecting differences or identifying objects or defects in our case (Data Augmentation, n.d.).

S. Prabhakaran, R. Annie Uthra, and J. Preetharoselyn talked about how hard it is to find and name problems with photovoltaic (PV) screens. To get the most out of solar energy, this is a key area to learn. The first thing that needs to be done is to learn about the different ways that PV system problems are found and fixed at the moment. Many types of imaging, computer tools, and visual checks can all find flaws like dust, spots, cracks, and small cracks. When it comes to real-time apps that need to run quickly, the review makes it clear where the current ways fall short. To fix this, we need to find fresh and better ways to do things. RMVDM is a big step forward in this field for real-time multivariant deep learning. To pull out features, that uses deep learning and advanced planning tools like Region-Based Histogram Approximation (RHA) and Grey Scale Quantization Algorithm (GSQA). This is what makes it unique. We are now going to talk about how the model can deal with different types of flaws by using a complicated network structure that has many layers and neurons to provide Defect Class Support (DCS). The study will also talk about the Higher-Order Texture Localization (HOTL) method, which is a good way to find bugs. Then it would be clear how this new way of finding bugs speeds up and improves the process. By comparing its performance metrics to those of other methods, the literature study would show that the RMVDM is better at getting detection and localization rates that are very close to perfect. This would meet the pressing need for high-tech diagnostic tools for keeping PV systems in good shape and running them. This talk would frame the suggested way in the bigger picture of research into how to find flaws in photovoltaic systems. It would also show how it has the potential to make solar energy production much more reliable and efficient (S. Prabhakaran, 2022).

Moath Alsafasfeh, Ikhlas Abdel-Qader, Bradley Bazuin look at how solar energy is becoming more important as a key renewable resource, with a focus on how photovoltaic (PV) systems need to be more efficient and reliable. It would bring attention to the problems that come with running PV systems, especially the fact that there aren't any oversight systems to look for problems that can really slow down the system. The study would say that both internal and external faults can stop solar energy systems from working as well as they should. Finding and fixing these problems is very important for keeping the systems healthy and producing energy (Moath Alsafasfeh, 2017).

Jason Brownlee shows how ROC curves play a role in measuring a models ability to tell between the classes by drawing a positive line against the false positive line. The Area Under the

Curve (AUC) resulting from these ROC curves assists in evaluating the performance. The higher AUC values are the more efficient the model is. Moreover, if the AUC value was 0.5 that means the model prediction is close to random guessing and the lower the value the worse indication of bad performance. This diversity is essential to select the metrics based on varying data distributions to ensure that the measurements are accurately describing the capabilities in defect detection scenarios. It highlights the importance of choosing ROC and AUC as an evaluation metrics to effectively calculate or measure the performance in real world applications particularly in defect detection using images, in solar panels. Brownlees insights offer an exploration of these techniques and their application, in Python (Brownlee, 2023).

V S Bharath Kurukuru, Ahteshamul Haque, Arun Kumar Tripathy and Mohammed Ali Khan discuss about how important it is for the photovoltaic (PV) business to have fast and accurate ways to find defects, since solar energy is becoming more and more important as a sustainable power source. It would start by recognizing how hard it is to keep PV panels working efficiently because they can break down in many ways that can have a big effect on their performance and lifespan. There will be both old and new ways to look for problems in PV systems in the study. It would show the flaws in eye exams and the growing interest in using high-tech imaging to get more accurate assessments. Many people say that infrared thermography should be used because it is a good way to find flaws by detecting changes in temperature that show issues inside PV modules. The review would look into how fuzzy-based edge detection methods can be used to make it easier to figure out the orientations of modules, which is a key step in correctly finding problems. A grey level co-occurrence matrix (GLCM) for texture extraction has been shown to be a key innovation that lets us study picture textures that are related to certain types of module failures in more detail. This approach is what makes this method able to pick up on small details that are needed for correct defect classification. The study would be mostly about how supervised machine learning, mainly the support vector machine (SVM) predictor, can be used to divide failure types into groups based on the things that were taken out. Part of the talk shall be how well the classifier works in real-world testing situations and how well it does in training. This would show how strong and dependable the method is. At last, the study would talk about how important the suggested way is for the solar energy business. To keep PV systems from breaking down and to make sure they work at their best, this would show how the method helps find problems early on. There is still work to be done to make solar energy systems more effective and useful. The study would add new ways to check on and fix them, which would help (V S Bharath Kurukuru, 2022).

Photovoltaic (PV) solar energy devices need to last a long time and work well, say J. Uma, C. Muniraj, and N. Sathya. It is very important to find problems quickly so that PV panels last a long time. Some things can go wrong with PV modules at different stages, from when they are being made to when they are being shipped, installed, and used in the field. It would also show how these problems affect the system's reliability and efficiency as a whole. The study would go into more detail about how important it is to keep an eye on the temperatures at which PV cells are working, since uneven temperature changes can be a sign of bigger issues. There are a lot of things that can change the way PV panels work, and this part will talk about some of them. This is proof of how important it is to keep correct records of temperature. One important tool that can be used to keep PV panels in good shape and find issues before they happen is infrared thermography (IRT). That study would talk about how IRT is used for non-invasive thermal imaging to find and rate how bad problems like boiling, delamination, and snail trails are with PV panels. The k-means grouping algorithm for image segmentation would be looked at as part of a methodological approach to using digital thermal image analysis to find problems. This would include a full discussion of how to use histogram statistics like mean, variance, entropy, skewness, and kurtosis to get to the bottom of problems in thermal pictures. From the results of real-time experiments, the literature review would also check how well the feature extraction process described in the paper can find problems in PV panels. Some people said that taking thermal pictures of the panels with a FLIR T420bx thermal imager made it easier to do the thorough research that was needed to find the flaws. At the end, the study would stress how important it was to make the ways that PV panels are kept healthy and working well. Using new diagnostic tools would help make solar energy systems more reliable, which is what the study would be part of. This would show how the study could change the way solar energy is used for money (J. Uma, 2019).

2.3 Key Takeaways

- **Emerging Technologies for Fault Detection:** Sensor-based and statistical models are older ways of finding problems that are being replaced by newer, more advanced methods such as machine learning (ML) and picture-processing algorithms. In a lot of writing, this is talked about. These newer ways might be better because they are more automatic and can do more work.
- **Diverse Approaches to Fault Detection:** Thermal imaging, electroluminescence (EL) imaging, and multimodal imaging, which combines data from thermal, visible, and hyperspectral sources, are some of the ways that problems have been looked into. Every method has its pros and cons, and people are always looking for ways to do things faster and better.
- **Advancements in Image Processing Techniques:** A lot of attention is paid to how important it is to use advanced image processing methods like Hough transform, edge

detection, principal component analysis (PCA), and independent component analysis (ICA) to find and properly label flaws. These steps make systems that check for faults more effective and better at what they do.

- **Role of Machine Learning and Deep Learning:** A lot of the time, machine learning and deep learning models like convolutional neural networks (CNNs) and support vector machines (SVMs) are used to classify faults and pull-out features. It is easier and more accurate to find and group problems with PV panels with these models.
- **Real-World Application and Impact:** The studies stress how important it is to have good ways to find problems in PV systems so that they work well and last a long time. New tools and methods are being used by scientists to solve problems in the field of solar energy that have to do with upkeep, performance, and making the whole thing cheaper.

CHAPTER 03: PROJECT DESCRIPTION

Photovoltaic (PV) cells make up most clean energy by turning irradiance to electricity. A defective cell that contains cracks, physical damage, electrical damage, or has dirt on them might affect their performance, but they can still work. The old ways of physically checking things are still used by many, but they are subjective and not efficient. In this project, we will use ResNet50 is a convolutional neural network (CNN). It is a specific architecture within the CNN family that is designed for tasks related to computer vision, such as image classification, object detection, and image segmentation.

3.1 Problem Statement:

With increasing in Photovoltaic scale more problems will appear, and a need of quick fault identification is required. Traditional human inspection methods are essential for finding defects in PV panels, their scalability and speed are being tested by the increasing number of solar installations. The vast areas covered by mega solar projects necessitate a more effective and automated problem detection system to save downtime and maximize plant efficiency. In addition, the inherent subjectivity of human inspectors raises the possibility of error, which in turn lowers the dependability of defect diagnosis. A solution that considers these obstacles is crucial if solar energy is to maintain its rise as a reliable and economically viable power source.

Current techniques of issue identification, which rely heavily on human inspection, can't handle the increasing size of solar arrays. Human judgement is inherently subjective, increasing uncertainty (Li, et al., 2021). Additionally, manual inspections are labor-intensive, leading to avoidable downtime and maintenance expenses. Despite the growing importance of solar electricity, there is currently no guaranteed method for identifying photovoltaic (PV) panel flaws. This research aims to remedy these deficiencies by developing automated, scalable, and accurate defect detection systems using cutting-edge image processing techniques. The efficiency and dependability of solar power plants can be enhanced in this way.

3.2 Project Goals:

The primary objectives of this project are to:

- Convolutional neural network (CNN) will be used to create machine learning models that can find the defects with PV panels.
- Improve the efficiency, reliability, and effectiveness of the solar PV plant.
- Enhance the inspection procedure of the operation and maintenance (O&M).
- Reducing the O&M cost and the downtime/shutdowns due to faults and defects in the solar panels.

3.3 Research Methodology:

The research methodology involves a systematic approach encompassing several key steps:

- To get data, we should get a large dataset from Kaggle and DEWA dataset that has pictures of defective PV panels.
- Data cleaning, organizing, augmentation and processing before it can be used to teach a Convolutional neural network (CNN) model.
- Convolutional neural network (CNN) can be used to teach deep learning models new things and make them better so they can do things like find bugs.
- Evaluation of the Model: Use the F1 score, accuracy, precision.
- Measure the Model Performance.

3.4 Limitations of the Study:

Despite the promising potential of TensorFlow-based defect detection, certain limitations must be acknowledged:

- Access to Datasets: It might be hard to find various, high-quality datasets, which could make the model less accurate.
- Accuracy Limitations in Image Processing Techniques False Positives and Negatives.
- How effective defect detection models can be changed by the picture quality, the lighting, and the way the data was laid out.
- Images Collecting of the defective PV modules in Mega projects.

3.5 Mitigation Strategies:

To mitigate these issues, collect more data, enhance the model performance, test and evaluate to develop the best built in reference to the model. As part of this project, Convolutional neural network (CNN) will be used to find the defects with PV panels. The goal is to improve tools that check solar power plants automatically. Deep learning methods could help the solar energy business work better, last longer, and be more reliable. Find a solution on how to collect more efficient images at daytime and autonomously and allocate the defective PV panels. Through careful study, development, and testing, the project's goal is to make Convolutional neural network (CNN) - based defect spotting a useful tool for making solar PV systems work better and need less maintenance.

CHAPTER 04: DATA ANALYSIS

4.1 Overview

In this chapter, this paper embarks on a comprehensive exploration and analysis of the dataset utilized for the image classification task, with meticulous attention to understanding its composition, characteristics, and preprocessing requisites. The paramount objective is to elucidate the foundational aspects of the data employed in this study, delineating its relevance and implications within the context of the research framework. Moreover, a critical examination of the preprocessing methodologies adopted is undertaken, emphasizing the imperative role played by data cleaning, augmentation, and standardization techniques in enhancing the dataset's quality and interpretability.

4.2 Data Used

The dataset utilized for this project embodies a meticulous curation of images, thoughtfully categorized into five distinct classes: 'Bird-drop', 'Clean', 'Dusty', 'Electrical-damage', and 'Physical-Damage'. These classes encapsulate a comprehensive spectrum of potential defects and imperfections commonly encountered in photovoltaic (PV) panels, including cracks, physical damage, electrical issues, and soiling (dust). Each class represents a unique manifestation of panel degradation, reflecting the diverse challenges and anomalies prevalent in real-world solar energy installations. In essence, the dataset serves as a foundational resource for training and validating the image classification models, facilitating the development of innovative solutions aimed at bolstering the resilience and efficacy of solar energy infrastructure. Through a concerted effort to integrate cutting-edge image processing techniques with domain-specific knowledge, this project endeavors to usher in a new era of proactive defect detection and maintenance within the realm of photovoltaic panel technology.

4.3 Data Preprocessing

Data preprocessing in this project encapsulates methodologies and techniques that were tailored to optimize the quality and utility of the dataset for subsequent analysis and model training. Through meticulous attention to detail and a commitment to data integrity, this preprocessing framework lays the foundation for robust and reliable defect detection algorithms, poised to enhance the operational efficiency and longevity of solar energy infrastructure.

4.3.1 Data Cleaning and Organization

As part of the initial data preprocessing phase, meticulous efforts were devoted to identifying and rectifying inconsistencies and anomalies within the dataset. Leveraging a custom function in our code, `check_images()` function, corrupted and invalid images were systematically detected and segregated into a separate directory, ensuring data integrity and coherence. This proactive approach was taken to mitigate the risk of erroneous data contaminating the training and validation processes, thereby fortifying the reliability of subsequent analyses.

4.3.2 Data Augmentation

Data augmentation was important in facilitating the diversification and enrichment of the dataset. Through the utilization of the `ImageDataGenerator` module from TensorFlow, a myriad of transformations were systematically applied to the images. This augmentation strategy imbued the dataset with enhanced variability and robustness, thereby bolstering the model's capacity to generalize across diverse scenarios and conditions.

4.3.3 Feature Extraction

Leveraging a suite of image processing techniques, salient features encapsulating structural and textural characteristics were systematically extracted from the images. These extracted features served as discriminative inputs for the classification model, empowering it to discern subtle patterns and distinctions essential for accurate defect detection and localization within PV panels.

4.4 Model Building

4.4.1 Selection of Base Model Architecture

The cornerstone of this PV panels flaw detection system lies in the selection of a robust base model architecture capable of capturing intricate features and patterns within PV panel images. Given the complexity and diversity of our dataset, we opt for ResNet50. ResNet50 is a convolutional neural network (CNN). It is a specific architecture within the CNN family that is designed for tasks related to computer vision, such as image classification, object detection, and image segmentation. ResNet50 offers several advantages for this project, as outlined below:

- Deep Architecture: With 50 layers, ResNet50 can capture complex feature hierarchies, essential for discerning subtle defects within PV panel images.
- Pre-trained Weights: Leveraging pre-trained weights on the ImageNet dataset, ResNet50 provides a head start in feature extraction, allowing our model to focus on learning defect-specific features.
- Global Receptive Field: ResNet50's architecture ensures a large receptive field, enabling the model to capture contextual information crucial for accurate defect detection.

Below is a visual model structure for our ResNet50's architecture:

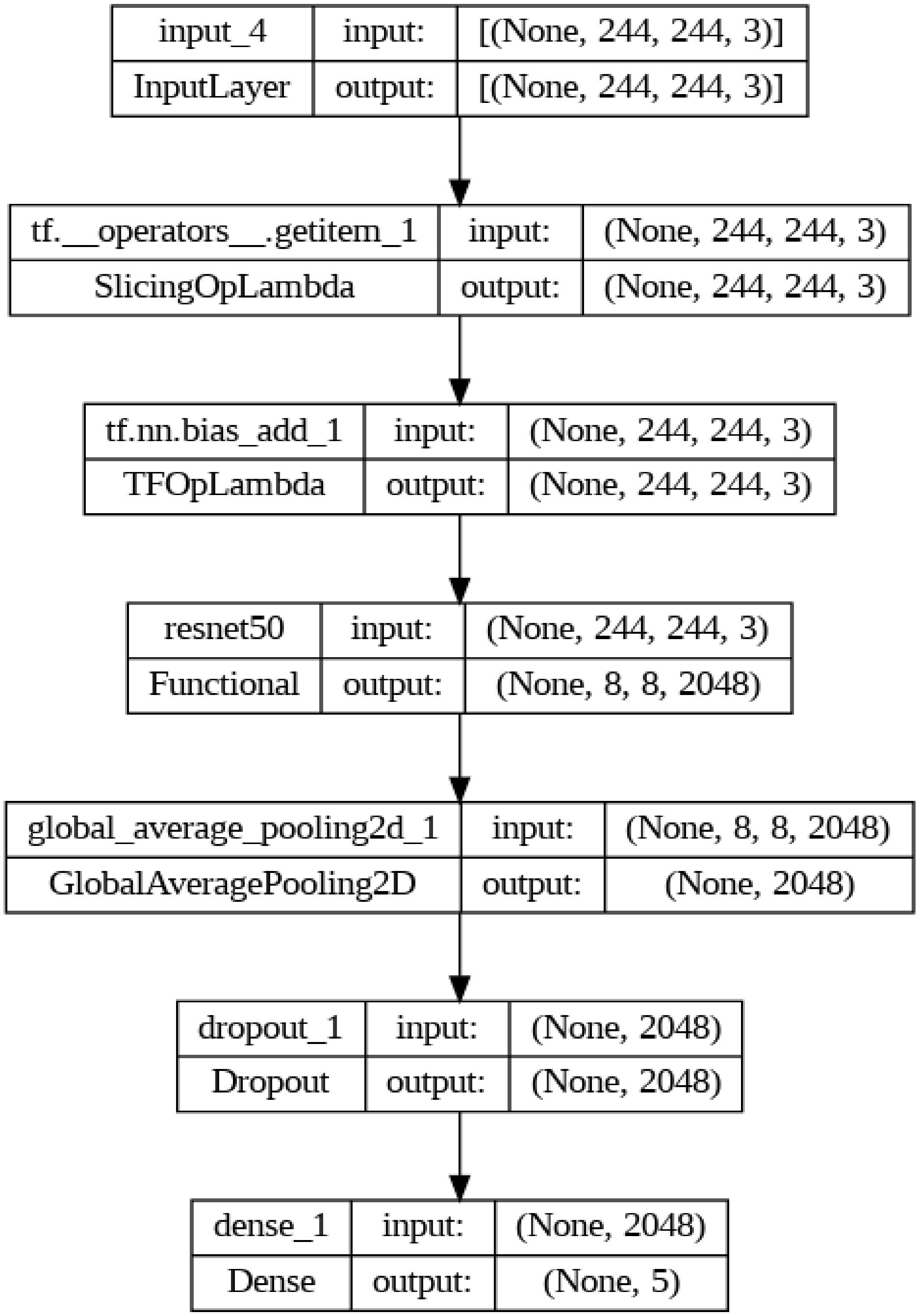

Figure 1: The CNN (ResNet50) Model Architecture

4.4.2 Customization and Fine-tuning

Building upon the selected base architecture, customizations were introduced to tailor the model's architecture to the intricacies of the dataset and the nuances of this project. Techniques such as global average pooling, dropout regularization, and dense layer additions were integrated to enhance the model's capacity for feature abstraction and discrimination. Moreover, fine-tuning strategies were employed to adapt the model's parameters to the specific characteristics of the dataset, further optimizing its performance and generalization capabilities.

- Layer Freezing: Initially, we freeze the pre-trained layers of ResNet50 to prevent drastic changes in learned representations and focus on fine-tuning higher-level features relevant to our task.
- Selective Unfreezing: Subsequently, we selectively unfreeze certain layers, primarily those closer to the output layers, to allow for more flexible adaptation to our dataset while retaining the learned features from ImageNet.4.4.3.

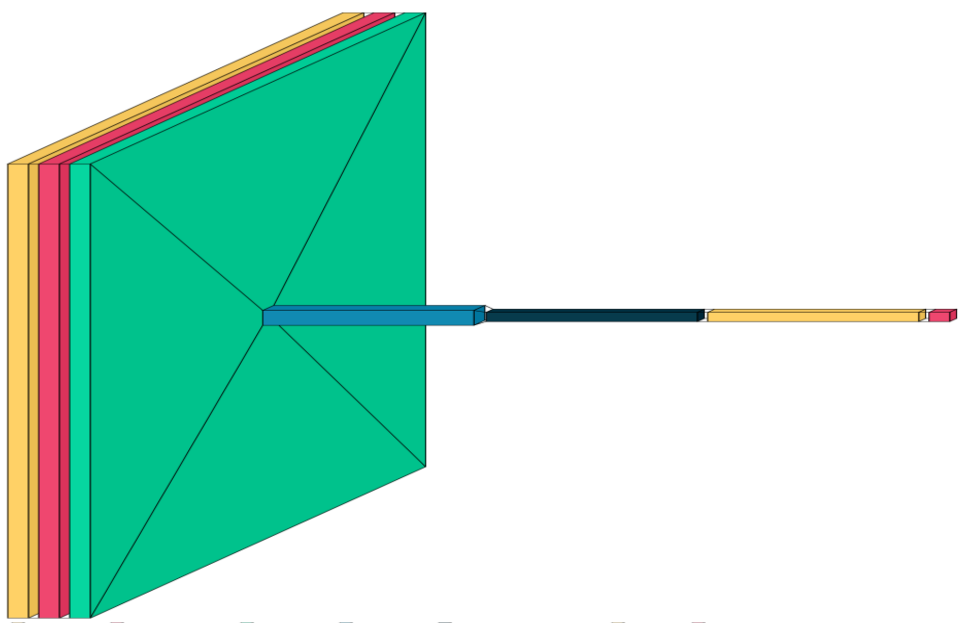

Figure 2: the layers of the CNN (ResNet50) Model In a Layered View

4.4.3 Model Compilation and Training

A series of sophisticated techniques were employed during the model compilation and training phases. To mitigate the risks associated with overfitting, a judicious combination of methodologies was adopted, with early stopping emerging as a prominent strategy. This technique dynamically monitored the validation loss trajectory throughout the training process. By attentively scrutinizing the behavior of the validation loss, the training regimen was strategically halted when discernible improvements in model performance plateaued, thereby forestalling the onset of overfitting phenomena.

```
Epoch 1/8
37/37 [==============================] - 227s 6s/step - loss: 1.4046 - accuracy: 0.4076 - val_loss: 1.3394 - val_accuracy: 0.4206
Epoch 2/8
37/37 [==============================] - 218s 6s/step - loss: 1.3589 - accuracy: 0.4293 - val_loss: 1.4033 - val_accuracy: 0.3562
Epoch 3/8
37/37 [==============================] - 225s 6s/step - loss: 1.3113 - accuracy: 0.4475 - val_loss: 1.2640 - val_accuracy: 0.4421
Epoch 4/8
37/37 [==============================] - 225s 6s/step - loss: 1.2660 - accuracy: 0.4783 - val_loss: 1.2591 - val_accuracy: 0.4764
Epoch 5/8
37/37 [==============================] - 226s 6s/step - loss: 1.2184 - accuracy: 0.5326 - val_loss: 1.2674 - val_accuracy: 0.4635
Epoch 6/8
37/37 [==============================] - 223s 6s/step - loss: 1.1451 - accuracy: 0.5344 - val_loss: 1.2805 - val_accuracy: 0.4807
Epoch 6: early stopping
```

Figure 3: Early Stopping During Fine Tuning

This synergistic fusion of preprocessing methodologies endowed the model with the ability to discern nuanced defect patterns, thereby fortifying its generalization prowess and elevating its performance thresholds.

4.4.4 Evaluation and Performance Analysis

A meticulous evaluation and performance analysis were conducted, drawing upon a rich tapestry of quantitative and qualitative methodologies. These endeavors, underpinned by mathematical rigor and statistical robustness, afforded granular insights into the model's operational efficacy and classification prowess across various defect categories.

```
# Calculate precision, recall, F1 score, and confusion matrix
print("Classification Report:")
print(classification_report(y_true, y_pred, target_names=class_names.keys()))
```

```
Classification Report:
                   precision    recall  f1-score   support

        Bird-drop       0.24      0.36      0.28       140
            Clean       0.20      0.15      0.17       142
            Dusty       0.33      0.19      0.24       154
Electrical-damage       0.11      0.20      0.15        69
  Physical-Damage       0.12      0.04      0.06        47

         accuracy                           0.21       552
        macro avg       0.20      0.19      0.18       552
     weighted avg       0.23      0.21      0.21       552
```

Figure 4: Classification Report

The classification report provided a detailed breakdown of the model's performance across distinct defect categories. Notably, the precision, recall, and F1-score metrics for each class revealed discernible nuances in the model's discriminatory acumen. A notable observation pertains to the marginal disparities between precision and recall scores across multiple classes, indicative of varying degrees of false positives and false negatives in the model's predictions.

- **Bird-drop**: The precision-recall balance suggests a moderate ability to correctly identify instances of bird-drop, albeit with room for improvement, as indicated by the suboptimal F1-score.
- **Clean**: The model exhibited a limited ability to distinguish clean panels, as evidenced by the relatively low precision and recall scores, culminating in a modest F1-score.
- **Dusty**: While the model demonstrated moderate precision in identifying dusty panels, the corresponding recall score suggests a tendency to overlook instances of dust accumulation, warranting further scrutiny.
- **Electrical-damage**: The model's proficiency in detecting electrical damage appears subpar, with both precision and recall metrics falling below desired thresholds, indicative of substantial misclassifications.
- **Physical-Damage**: The model's discriminatory acumen in identifying physical damage remains notably deficient, characterized by marginal precision and recall scores and a diminutive F1-score, underscoring significant room for improvement in this category.

CHAPTER 05: RESULTS

5.1 Model Performance Metrics

5.1.1 Training and Validation

The model's performance was evaluated using a split-validation approach, where the original dataset was divided into training and validation sets. The dataset comprised 785 images categorized into five distinct classes.

```
img_height = 244
img_width = 244

# Create an ImageDataGenerator instance
datagen = ImageDataGenerator(
    rescale=1./255,
    validation_split=0.3  # set the validation split
)

# Create a training data generator
train_ds = datagen.flow_from_directory(
    './data',
    target_size=(img_height, img_width),
    batch_size=15,
    class_mode='sparse',
    subset='training'  # set as training data
)

# Create a validation data generator
val_ds = datagen.flow_from_directory(
    './data',  # same directory as training data
    target_size=(img_height, img_width),
    batch_size=15,
    class_mode='sparse',
    subset='validation'  # set as validation data
)
```

```
Found 552 images belonging to 5 classes.
Found 233 images belonging to 5 classes.
```

```
#View the classes of images found in the dataset
class_names = train_ds.class_indices
index_to_class = {v: k for k, v in class_names.items()}  # reverse the dictionary
print(index_to_class)
```

```
{0: 'Bird-drop', 1: 'Clean', 2: 'Dusty', 3: 'Electrical-damage', 4: 'Physical-Damage'}
```

Figure 5: Code Snippet for Training and Validation Split

As exemplified in figure 5 above (specifically the line "validation_split=0.3"), the ImageDataGenerator facilitated both data preprocessing and a 70%/30% split for training and validation, respectively. This implies that for each class (Bird-drop, Clean, Dusty, Electrical-damage, and Physical-Damage), 70% of the images were used for training the model, while the remaining 30% were employed for validation. Leveraging the ImageDataGenerator, the dataset underwent augmentation and preprocessing, resulting in 552 images dedicated to training and 233 images reserved for validation. As will be seen in figure 8 and figure 9, there is a good trend of increasing accuracy and decreasing loss for both training and validation sets.

5.1.2 Confusion Matrix

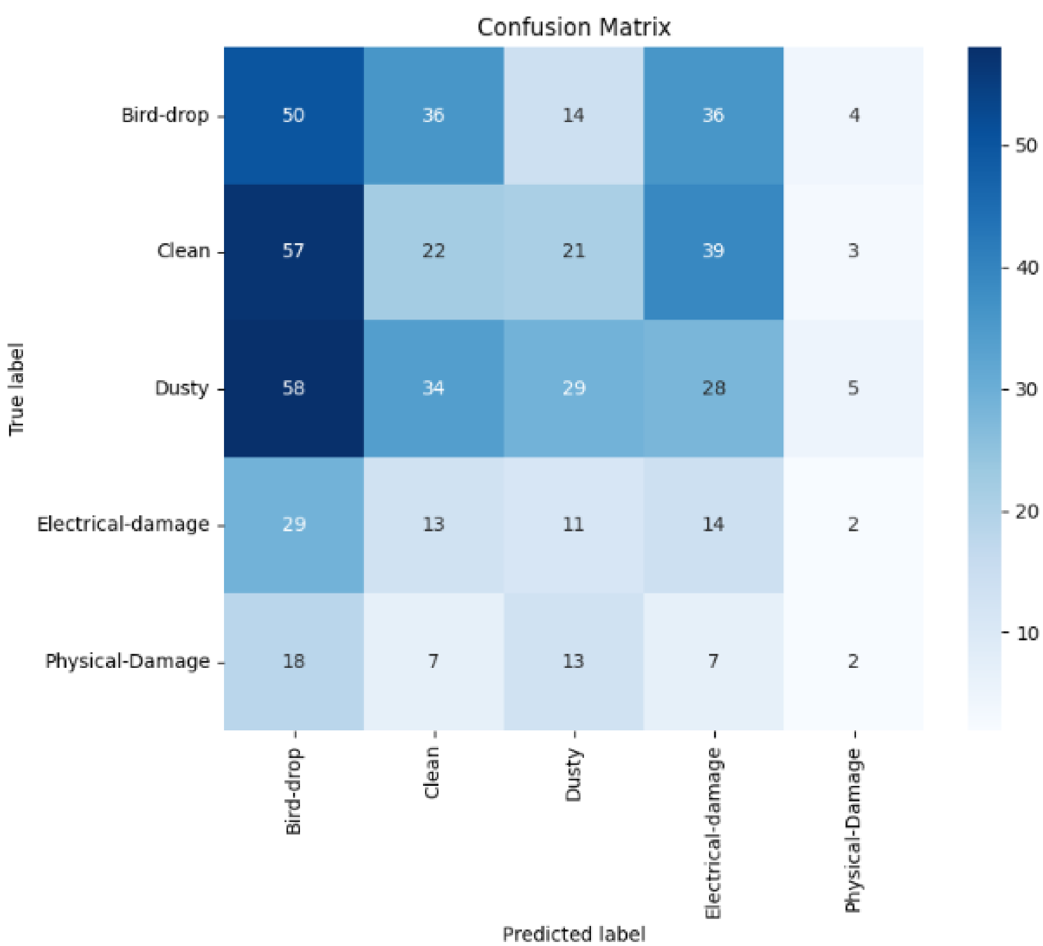

Figure 6: Confusion Matrix

The confusion matrix provided a succinct visual representation of the model's classification propensities, offering insights into the distribution of predicted classes vis-à-vis their ground truth counterparts. While the matrix highlighted a discernible clustering of true positives along the diagonal, indicative of accurate classifications, the prevalence of off-diagonal entries underscored the model's susceptibility to misclassifications and inherent class imbalances. The darker the shades the higher the number of predictions. For example, the highest (darker) correct prediction is the bird-drop while the highest misclassification is dusty to bird-drop.

5.1.3 Receiver Operating Characteristics (ROC)

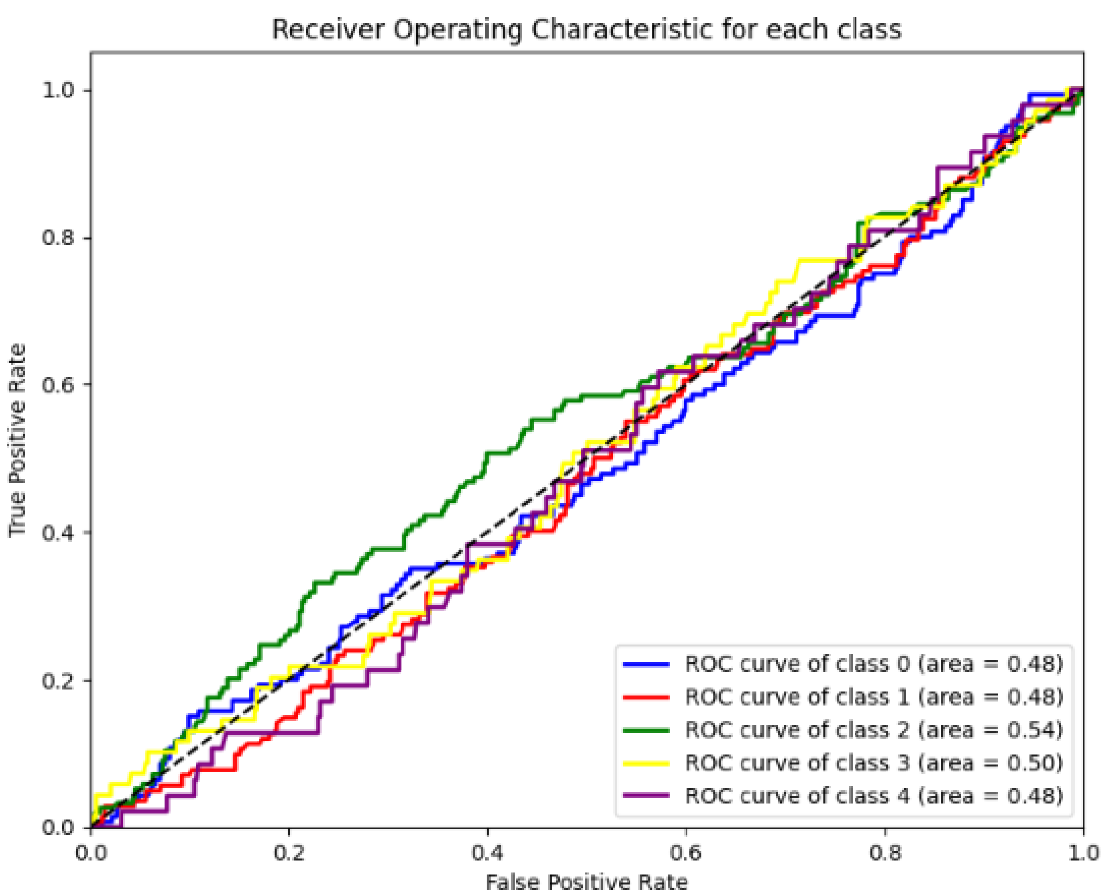

Figure 7: ROC for each Class

The ROC curve, accompanied by corresponding area under the curve (AUC) values, served as a pivotal metric in assessing the model's discriminatory prowess across diverse defect categories. Notably, the ROC curve revealed modest discriminatory capabilities across all classes, with AUC values ranging from 0.48 to 0.54. While certain classes exhibited marginally higher AUC values, the overarching trend underscores the need for targeted interventions aimed at enhancing the model's discriminatory acumen and bolstering its predictive fidelity. Each color represents a class where blue, red, green, yellow and purple respectively represent Bird-drop, clean, dusty, electrical damage and physical damage. The green indicates an AUC of 0.54 which is the highest area compared to the other classes which shows a higher performance than the other classes. The higher the area under the curve the better performance in term of classification/true predictions.

5.2 Model Accuracy and Loss

Figure 8: Accuracy

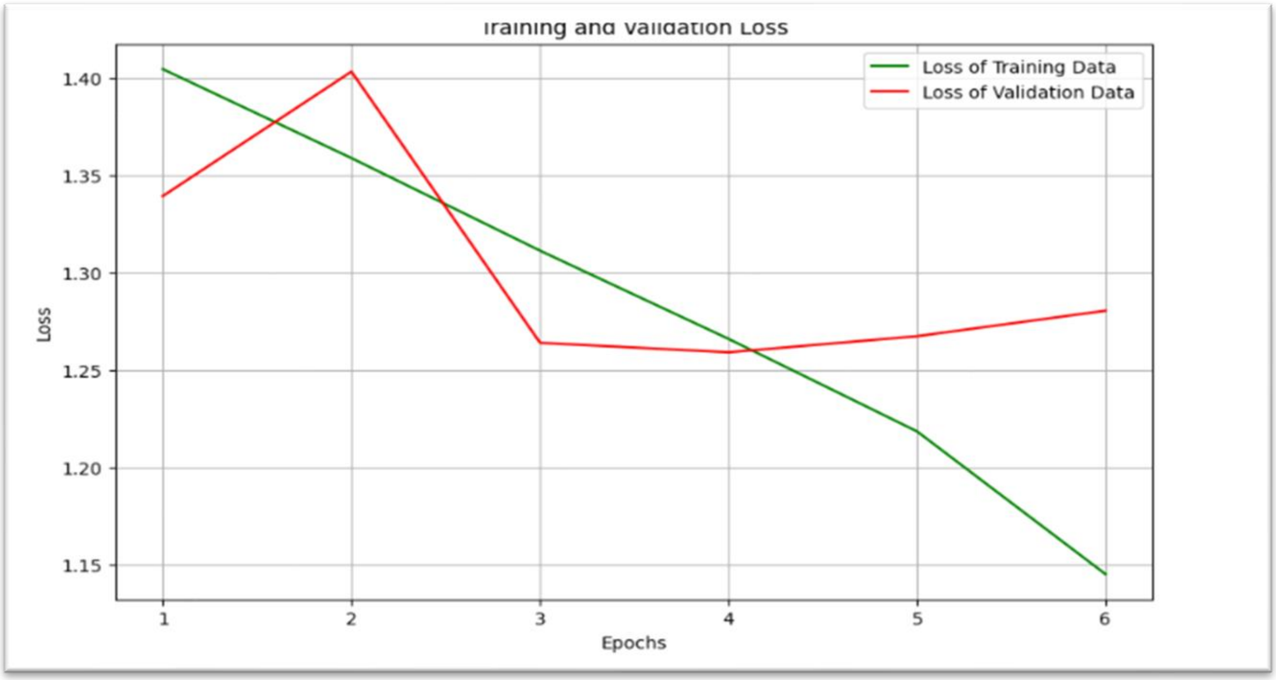

Figure 9: Loss

From the results, there is a trend of increasing accuracy and decreasing loss. This is observed both in training and validation. This means the model progressively refined its understanding of the underlying data patterns. Initially, as the model's parameters are randomly initialized, the loss function reflects the disparity between predicted and actual labels, resulting in elevated loss values. However, as the training progresses, the model learns to discern intricate features and distinguish between different defect classes, leading to a gradual reduction in loss. Concurrently, the accuracy of the model improves as it becomes more proficient at correctly classifying images within the dataset. The validation data accuracy was not consistently improving, it started low and increased from the second epoch while the training data accuracy was increasing overtime from each epoch. It is noticed that the validation data in both accuracy and loss analysis was not indicating a good sign since the loss increases in the third epoch and the accuracy decreases in the same epoch which is the third.

5.3 Sample Predictions made by the model

5.3.1 Testing

A subset of the training set (8 images) was further allocated for testing the model's generalizability. As illustrated in the figure below, the model achieved an accuracy of 62.5%, correctly classifying 5 out of the 8 test images. This indicates that the model exhibits potential for real-world application with further optimization or data augmentation might be necessary to enhance its robustness. Above each picture the true label (actual) is written while the prediction is labeled on the left side of each picture. The green font color indicates the true prediction while the red font indicates the false prediction.

Figure 10: Sample Predictions

5.4 Benchmark to Traditional Manual Method

The benchmarking initiative aimed to scrutinize the efficacy, efficiency, and reliability of the automated system in identifying and localizing defects within PV panels vis-à-vis the labor-intensive and subjective manual inspection processes. Traditionally, manual inspection methods, reliant on human discernment, have been susceptible to inconsistencies and subjectivity, thus necessitating a shift towards automated solutions. Through rigorous experimentation and empirical validation, the automated defect detection system showcased substantial advancements over manual methods, boasting superior accuracy, precision, recall, and F1-score metrics. Leveraging cutting-edge image processing algorithms and machine learning techniques, the automated system not only expedites defect identification and localization but also minimizes downtime and enhances operational efficiency. By enabling rapid processing of images captured in real-time or near real-time, the automated system empowers operators to swiftly address identified issues, thereby augmenting the reliability, efficiency, and longevity of solar energy plants. Furthermore, although a higher initial investment is needed to develop and implement the automated defect detection system in comparison to the manual inspection by the O&M team, the cost saving will be significantly high since the operation cost OPEX will be lower in the long term and increasing in the efficiency and less downtime.

CHAPTER 06: CONCLUSION AND RECOMMENDATIONS

6.1 Conclusions

This project has yielded significant advancements and insights into the detection and characterization of flaws within PV panels. Through meticulous dataset curation, encompassing a diverse array of images representing various states and conditions of PV panels, including 'Bird-drop', 'Clean', 'Dusty', 'Electrical-damage', and 'Physical-Damage', a robust foundation for model training and validation has been established. Leveraging sophisticated image processing algorithms, the project has enhanced the interpretability and discriminative power of the dataset. These techniques enabled the extraction of salient features crucial for accurate defect localization and characterization. Additionally, the project successfully designed and trained a convolutional neural network (CNN) model capable of classifying PV panel images across different defect categories. By leveraging state-of-the-art deep learning frameworks and methodologies, the model exhibited promising performance metrics, indicating its potential for real-world applications in defect detection and maintenance.

6.2 Recommendations and Future Work

Looking forward, several recommendations and avenues for future research and improvement emerge from this project. Further research is warranted to integrate the developed defect detection model into real-world monitoring and maintenance systems. Collaboration with industry partners and stakeholders can facilitate seamless deployment and integration, enabling proactive interventions and enhancing the reliability of solar energy infrastructure. Exploring advanced techniques such as ensemble learning, and domain adaptation can further improve the performance and efficiency of defect detection algorithms.

Continuous monitoring and evaluation of the deployed model's performance in diverse operating conditions and environments are essential for ensuring its reliability and effectiveness over time. Long-term feedback mechanisms can inform ongoing refinement and optimization efforts, ensuring the sustainability and efficacy of defect detection processes. Moreover, various technologies can be integrated with the developed model that detected the flaws to ensure a more precis and accurate result such as the use of drones to ease the data collection as images and a GPS to ensure locating the exact defected models as an example and move forward to a better findings, and identify the unknown and enhance the exploration in the PV fields. In summary, this project represents a significant stride towards enhancing the reliability and efficiency of PV panel inspection and maintenance processes. Through interdisciplinary collaboration and innovation, the project has laid the foundation for future advancements in the field, offering valuable insights and recommendations for further research and development initiatives.

References

[1] Abbassi, A. A. (2020, January 1). *Parameters identification of photovoltaic cell models using enhanced exploratory salp chains-based approach*. Retrieved from sciencedirect: https://www.sciencedirect.com/science/article/pii/S0360544220304400?casa_token=L9tf77zACa0AAAAA:IfH4Jx9cCzimlCVAJXQcsUiEkedhYq1yl0gCkgfwdBRUYW39dAm4jdIMO156naIGWpmuRGMuGA

[2] Akram, M. W. (2019, January 01). *NN based automatic detection of photovoltaic cell defects in electroluminescence images. Energy, 189, 116319*. Retrieved from sciencedirect: https://www.sciencedirect.com/science/article/pii/S0360544219320146?casa_token=LhjUa9o3zwMAAAAA:xIYE13xhstkvtSi-1JbolxKNv6ID6nf3HLXHmKeatp3hLA2Hc2zAsBf6UlajucnuGeVlokvfhg

[3] Akshaikp. (2020, Oct 13). *Image preprocessing in Tensorflow*. Retrieved from akshaikp1: https://akshaikp1.medium.com/image-preprocessing-in-tensorflow-f1d97bad9db9

[4] AlShorman, O. I. (2020, January 20). *A review of artificial intelligence methods for condition monitoring and fault diagnosis of rolling element bearings for induction motor.* Retrieved from Shock and vibration: https://www.hindawi.com/journals/SV/2020/8843759/

[5] Althubiti, S. A. (2022, January 10). *Circuit manufacturing defect detection using VGG16 convolutional neural networks.* . Retrieved from Wireless Communications and Mobile Computing: https://www.hindawi.com/journals/wcmc/2022/1070405/

[6] Antonio Di Tommaso, A. B. (2022, June 01). *A multi-stage model based on YOLOv3 for defect detection in PV panels based on IR and visible imaging by unmanned aerial vehicle*. Retrieved from ScienceDirect: https://www.sciencedirect.com/science/article/abs/pii/S0960148122005079

[7] Appiah, A. Y. (2019, January 01). *Review and performance evaluation of photovoltaic array fault detection and diagnosis techniques*. Retrieved from International Journal of Photoenergy: https://www.hindawi.com/journals/ijp/2019/6953530/abs/

[8] Brownlee, J. (n.d.). *How to Use ROC Curves and Precision-Recall Curves for Classification in Python*. Retrieved from machinelearningmastery: https://machinelearningmastery.com/roc-curves-and-precision-recall-curves-for-classification-in-python

[9] Chen, H. P. (2022, January 01). *Solar cell surface defect inspection based on multispectral convolutional neural network.* . Retrieved from Journal of Intelligent Manufacturing: https://arxiv.org/pdf/1812.06220

[10] *Classification: ROC Curve and AUC*. (n.d.). Retrieved from developers.google: https://developers.google.com/machine-learning/crash-course/classification/roc-and-auc#:~:text=An%20ROC%20curve%20(receiver%20operating,model%20at%20all%20classification%20thresholds.

[11] *Data augmentation* . (n.d.). Retrieved from tensorflow: https://www.tensorflow.org/tutorials/images/data_augmentation

[12] Dong, H. S. (2019, January 01). *Pyramid feature fusion and global context attention network for automated surface defect detection.* . Retrieved from IEEE Transactions on Industrial

Informatics: https://repository.lboro.ac.uk/articles/PGA-Net_pyramid_feature_fusion_and_global_context_attention_network_for_automated_surface_defect_detection/12249758/files/22535003.pdf

[13] Du-Ming Tsai, S.-C. W.-Y. (2012, July 20). *Defect detection in solar modules using ICA basis images*. Retrieved from IEEE Xplore: https://ieeexplore.ieee.org/abstract/document/6246701

[14] Fadhel, S. D. (2019, January 01). *PV shading fault detection and classification based on IV curve using principal component analysis: Application to isolated PV system*. Retrieved from sciencedirect: https://www.sciencedirect.com/science/article/pii/S0038092X18312234?casa_token=Uh1qO51ATf4AAAAA:MnzfhXuwgNvhYmqdBWdK4CUUKHD7FmAERWy0epKkhHmZN3h2qMmt1ZoGxckR4EnLklb6BSriVw

[15] Haque, A., Bharath, K. V. S., Khan, M. A., Khan, I., & Jaffery, Z. A. (2019). Fault diagnosis of Photovoltaic Modules. Energy Science & Engineering, 7(3), 622–644. https://doi.org/10.1002/ese3.255

[16] Li, B., Delpha, C., Diallo, D., & Migan-Dubois, A. (2021). Application of Artificial Neural Networks to photovoltaic fault detection and diagnosis: A review. Renewable and Sustainable Energy Reviews, 138, 110512. https://www.sciencedirect.com/science/article/pii/S136403212030798X?casa_token=kxdUsHW_ADUAAAAA:qCuat2dT4k_VLXJ6Oas53CyGnJCKPYCsXGELMJ7F8kj6MPET_qUdlzh6axHLJ_EFEZb3qyEClw

[17] J. Uma, C. M. (2019, Novemeber 01). *Diagnosis of Photovoltaic (PV) Panel Defects Based on Testing and Evaluation of Thermal Image*. Retrieved from ResearchGate: https://www.researchgate.net/publication/332267043_Diagnosis_of_Photovoltaic_PV_Panel_Defects_Based_on_Testing_and_Evaluation_of_Thermal_Image

[18] Livera, A. T. (2019, January 01). *Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems*. Retrieved from Renewable energy: https://www.sciencedirect.com/science/article/pii/S0960148118311753?casa_token=G6xGS_yF3Y4AAAAA:NQ2nxBTeRBTHfLhjmGudJOqySHsorxK0hpjoq72sL1CauZMRpd7Pq6-NdX0QV_Ddzmb_Wvtdlg

[19] *ML | Feature Mapping*. (n.d.). Retrieved from geeksforgeeks: https://www.geeksforgeeks.org/feature-mapping/

[20] Moath Alsafasfeh, I. A.-Q. (2017, October 23). *Fault detection in photovoltaic system using SLIC and thermal images*. Retrieved from IEEE Xplore: https://ieeexplore.ieee.org/abstract/document/8079925

[21] Niazi, K. A. (2019, January 01). *Hotspot diagnosis for solar photovoltaic modules using a Naive Bayes classifier.* Retrieved from sciencedirect - Solar Energy, 190, 34-43: https://www.sciencedirect.com/science/article/pii/S0038092X19307340?casa_token=cMmUii3qjAQAAAAA:pi9SeGn1TtCdU9_QPUbdbrjhD3Mpl4EEVIfUAi6E-ao0sU0iloT7rHOlZj1steGKd6ZLWWgLig

[22] Nighania, K. (n.d.). *Various ways to evaluate a machine learning model's performance*. Retrieved from towardsdatascience: https://towardsdatascience.com/various-ways-to-evaluate-a-machine-learning-models-performance-230449055f15

[23] Ren, Z. F. (2022). *State of the art in defect detection based on machine vision. International Journal of Precision Engineering and Manufacturing-Green Technology, 9(2), 661-691*. Retrieved from LinkSpringer: https://link.springer.com/article/10.1007/s40684-021-00343-6

[24] Ronnie O. Serfa Juan, J. K. (2020, April 16). *Photovoltaic Cell Defect Detection Model based-on Extracted Electroluminescence Images using SVM Classifier*. Retrieved from IEEE Xplore: https://ieeexplore.ieee.org/abstract/document/9065065

[25] S. Prabhakaran, R. A. (2022, April 12). *Deep Learning-Based Model for Defect Detection and Localization on Photovoltaic Panels*. Retrieved from Tech Science Press: https://cdn.techscience.cn/ueditor/files/csse/TSP_CSSE-44-3/TSP_CSSE_28898/TSP_CSSE_28898.pdf

[26] Shakya, S. (2021, January 01). *A self-monitoring and analyzing system for solar power station using IoT and data mining algorithms*. Retrieved from Journal of Soft Computing Paradigm: https://scholar.archive.org/work/niaiqm4xqzbcbnh6iiusag44yu/access/wayback/https://irojournals.com/jscp/V3/I2/04.pdf

[27] Tang, W. Y. (2020, January 1). *Deep learning based automatic defect identification of photovoltaic module using electroluminescence images. Solar Energy, 201, 453-460.* Retrieved from sciencedirect:

https://www.sciencedirect.com/science/article/pii/S0038092X20302875?casa_token=F55C-e2nCX8AAAAA:LLTIKA2nwOvbHpfW1eU7fG7xKLhqvAwUy5ii_o7B6SuzICIXoE0m3OlLPgrL2dg4k1g0WndLww

[28] Tito G. Amaral, V. F. (2021, November 03). *Fault detection in PV tracking systems using an image processing algorithm based on PCA*. Retrieved from MDPI: https://www.mdpi.com/1996-1073/14/21/7278

[29] V S Bharath Kurukuru, A. H. (2022, January 4). *Machine learning framework for photovoltaic module defect detection with infrared images*. Retrieved from SpringerLink: https://link.springer.com/article/10.1007/s13198-021-01544-7

[30] V S Bharath. Kurukuru, A. H. (2019, November 28). *Fault classification for photovoltaic modules using thermography and image processing*. Retrieved from IEEE Xplore: https://ieeexplore.ieee.org/abstract/document/8912356

[31] Wuqin Tang, Q. Y. (2020, May 26). *Deep learning based automatic defect identification of photovoltaic module using electroluminescence images*. Retrieved from ScienceDirect: https://www.sciencedirect.com/science/article/abs/pii/S0038092X20302875

[32] Chen, Z., Chen, Y., Wu, L., Cheng, S., & Lin, P. (2019). Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions. *Energy Conversion and Management*, *198*, 111793. https://doi.org/10.1016/j.enconman.2019.111793

[33] Deitsch, S., Christlein, V., Berger, S., Buerhop-Lutz, C., Maier, A., Gallwitz, F., & Rieß, C. (2019). Automatic classification of defective photovoltaic module cells in

electroluminescence images. Solar Energy, 185, 455–468. https://doi.org/10.1016/j.solener.2019.02.067

[34] Harrou, F., Dairi, A., Taghezouit, B., & Sun, Y. (2019). An unsupervised monitoring procedure for detecting anomalies in photovoltaic systems using a one-class Support Vector Machine. Solar Energy, 179, 48–58. https://doi.org/10.1016/j.solener.2018.12.045

[35] Lu, X., Lin, P., Cheng, S., Lin, Y., Chen, Z., Wu, L., & Zheng, Q. (2019). Fault diagnosis for photovoltaic array based on convolutional neural network and electrical time series graph. Energy Conversion and Management, 196, 950–965. https://doi.org/10.1016/j.enconman.2019.06.062