

# UNDERGRADUATE PROJECT PROGRESS REPORT

|                  |   |
|------------------|---|
| Project Title:   | Deep Learning for Predicting Solar Cell Degradation using Thermal Imaging |
| Surname:         | Gao   |
| First Name:      | Rui (Olivia)  |
| Student Number:  | 202118010407  |
| Supervisor Name: | Dr. Happy Nkanta Monday   |
| Module Code:     | CHC 6096  |
| Module Name:     | Project   |
| Date Submitted:  | 30 December 2024  |

## Table of Contents

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Introduction .....</b>                        | <b>4</b> |
| 1.1      | Background .....                                 | 4        |
| 1.2      | Aim.....   | 4        |
| 1.3      | Objectives .....                                 | 4        |
| 1.3.1    | Implemented part.....                            | 5        |
| 1.3.2    | Upcoming part.....                               | 5        |
| 1.4      | Project Overview .....                           | 5        |
| 1.4.1    | Scope.....                                       | 5        |
| 1.4.2    | Audience.....                                    | 5        |
| <b>2</b> | <b>Background Review.....</b>                    | <b>7</b> |
| <b>3</b> | <b>Technical Progress .....</b>                  | <b>9</b> |
| 3.1      | Approach.....                                    | 9        |
| 3.1.1    | Inception module.....                            | 9        |
| 3.1.2    | Attention module.....                            | 10       |
| 3.1.3    | BiLSTM Module .....                              | 10       |
| 3.1.4    | The Architecture Diagram of Proposed model ..... | 12       |
| 3.2      | Dataset.....                                     | 12       |
| 3.2.1    | Dataset Information .....                        | 12       |
| 3.2.2    | Data Separation .....                            | 13       |
| 3.2.3    | Data Resize .....                                | 13       |
| 3.2.4    | Data Balance.....                                | 13       |
| 3.2.5    | Label Processing.....                            | 13       |
| 3.2.6    | Data Augmentation .....                          | 13       |
| 3.3      | Technology .....                                 | 14       |
| 3.4      | Testing and Evaluation Plan .....                | 14       |
| 3.5      | Design and Implementation .....                  | 17       |
| 3.5.1    | Accuracy Curve .....                             | 17       |
| 3.5.2    | Loss Curve .....                                 | 18       |
| 3.5.3    | Precision.....                                   | 19       |
| 3.5.4    | Recall and Sensitivity.....                      | 19       |
| 3.5.5    | F1-Score.....                                    | 19       |
| 3.5.6    | ROC Curve & AUC.....                             | 19       |

|       |                                     |    |
|-------|-------------------------------------|----|
| 3.5.7 | Specificity .....                   | 20 |
| 3.5.8 | Confusion Matrix .....              | 20 |
| 3.5.9 | Classification Report .....         | 21 |
| 4     | Project Management .....            | 22 |
| 4.1   | Activities .....                    | 22 |
| 4.2   | Schedule .....                      | 23 |
| 4.3   | Project Version Management .....    | 23 |
| 4.4   | Project Data Management .....       | 24 |
| 4.5   | Project Deliverables .....          | 24 |
| 5     | Professional Issues and Risk: ..... | 25 |
| 5.1   | Risk Analysis .....                 | 25 |
| 5.2   | Professional Issues .....           | 25 |
| 6     | References .....                    | 27 |

## **1 Introduction**

### **1.1 Background**

People now focus more on renewable and clean resource due to the impact of global warming and fossil fuels. Solar cell can generate solar energy. In addition, photovoltaic systems have numerous advantages, including extended lifespan, sustainability, low-noise performance, and cleanliness [1].

However, photovoltaic systems are vulnerable to damage during operation, especially during errors during installation or under long-term operating stress. Unavoidable defects can greatly diminish the photoelectric conversion efficiency and lifespan of the modules, resulting in significant economic losses [2]. For example, hot-spots represent areas within specific regions of a solar cell where the temperature is elevated, which can cause a significant increase in the of the solar cells, consequently affecting the overall performance of the temperature module [3]. Additionally, when bypass diodes fail, the current cannot bypass the faulty cells or modules properly, leading to localized overheating within the affected cells or modules. This can result in hotspots, burn marks, and, in the worst-case scenario, fires [4].

### **1.2 Aim**

This project aims to develop an advanced model based on deep learning to accurately detect the aging and degradation of solar cells by deeply analyzing thermal imaging data of solar panels. This method enables the timely identification of potential issues, assisting maintenance personnel in taking targeted measures to prevent further damage to the equipment. By leveraging this intelligent diagnostic approach, it is possible to significantly improve the efficiency and accuracy of maintenance work, effectively extend the lifespan of solar panels, and reduce replacement costs.

### **1.3 Objectives**

This project will collect thermal image of solar cell from online source, includes datasets like PHOTOVOLTAIC THERMAL IMAGES DATASET for automated fault detection and analysis in large photovoltaic systems using photovoltaic module fault detection and some datasets that will soon be used for the model.

### **1.3.1 Implemented part**

The first model is classification model include Inception, Attention and BiLSTM to identify the type of anomaly (one anomalous cell, more than one anomalous cell, or a contiguous series of anomalous cells).

Additionally, the model's evaluation will encompass metrics like "Accuracy" and "Loss." Performance will also be analyzed using "Precision", "Recall", "F1-Score", "ROC Curve", "AUC", "Specificity," "Sensitivity," and the "Confusion Matrix."

### **1.3.2 Upcoming part**

Next, the model will add the Time Series Prediction to the model to predict the degradation of solar cells. In addition, the model's evaluation will encompass "Mean Squared Error (MSE)", "Root Mean Squared Error (RMSE)", "Mean Absolute Error (MAE)", "Mean Absolute Percentage Error (MAPE)", "Coefficient of Determination ( $R^2$ )".

## **1.4 Project Overview**

### **1.4.1 Scope**

The goal of this project is to develop a deep learning-based model designed to detect degradation in solar cells. This model will be trained to identify failure points which can negatively impact the performance of solar panels. By leveraging this model, failures can be detected early, enabling timely maintenance and repairs. As a result, the lifespan of solar panels can be significantly extended, and their energy production efficiency can be optimized.

This study holds great significance because the early detection of solar cell degradation is crucial to maintaining both the efficiency and longevity of solar energy systems. Specifically, by identifying and addressing issues before they escalate, it can lead to reduced maintenance costs, improved performance of solar panels, and ultimately contribute to the sustainability and cost-effectiveness of solar energy solutions. Moreover, this approach can help mitigate downtime and enhance the overall reliability of solar power systems, making them more viable and efficient for long-term use.

### **1.4.2 Audience**

The outcomes of this project will benefit several key stakeholders, including:

1. Solar energy companies and operators: This model can help reduce operational costs, improve solar energy efficiency, and extend the lifespan of solar panels, ultimately leading to better long-term performance.
2. Countries and governments: Governments and nations can leverage the project's results to advance the development of renewable and clean energy solutions. By reducing dependence on fossil fuels, countries can enhance energy security and sustainability. Additionally, this technology can help governments lower the overall costs associated with energy production.
3. Consumers of solar energy: End users will benefit from improved solar panel efficiency and extended lifespan, which can lead to lower energy costs over time. With more reliable and efficient systems, consumers can enjoy long-term savings and increased energy independence.
4. Environmental organizations: By improving the efficiency and longevity of solar panels, the project supports the broader goal of reducing environmental impact and promoting renewable energy sources that help protect the planet.

## 2 Background Review

There have already been several studies focused on detecting the degradation of solar panels. Through advanced data analysis and early detection, these studies aim to improve the maintenance of solar panels, extend their lifespan, and enhance overall energy efficiency, contributing to the sustainability and cost-effectiveness of solar energy systems.

Koshy et al. [5] proposed a deep model integrating infrared (IR) imaging, including ResNet and custom convolutional neural networks (CNN). The test results show that our method is effective, with an average prediction accuracy of 94% and a classification accuracy of 86% for 12 parameters. Le et al. [6] proposed a hybrid approach combining deep learning computer vision and thermal analysis for detecting defects in photovoltaic (PV) panels using aerial thermal images captured by drones. Their method integrates panel segmentation based on the Mask Region-Based Convolutional Neural Network (Mask-RCNN) with temperature distribution analysis to achieve automatic fault detection. This approach has been applied to an operational 50 MW solar power plant, and the results demonstrate the potential of integrating deep learning with thermal imaging for proactive maintenance of solar energy systems. Dunderdale et al. [7] proposed a method based on deep learning and features to detect and classify defective photovoltaic modules using thermal infrared images. They used VGG-16, MobileNet and SVM in the research. Vega Díaz et al. [8] proposed two approaches in automated solar power station inspection: one based on classical techniques (Edge detection and classification) and the other based on deep learning (CNN). And both panel detection methods are highly effective in complex backgrounds.

| Authors          | Methods & Models                           | Result   |
|------------------|--|--|
| Koshy et al. [5] | IR imaging, ResNet, and custom CNNs.       | 94% prediction accuracy and 86% classification accuracy for 12 parameters. |
| Le et al. [6]    | Deep learning, Mask-RCNN, thermal analysis | Successful in proactive maintenance for                                    |

|                       |   |  |
|-----------------------|---|--|
|                       |   | a 50 MW solar plant.   |
| Dunderdale et al. [7] | VGG-16 + MobileNet +SVM                                   | <p>Overall Accuracy: 91.2%</p> <p>VGG-16 using SGD optimizer and data augmentation: accuracy was 85.8%</p> <p>MobileNet using Adam optimizer: accuracy was 89.5%</p> |
| Vega Díaz et al. [8]  | <p>Edge detection and classification</p> <p>CNN + SVM</p> | <p>1. Classical techniques:<br/>Precision: 0.997<br/>Recall: 0.970</p> <p>2. Deep learning:<br/>Precision: 0.996<br/>Recall: 0.981</p>                               |

Table 1. Summary of Related Works



### 3 Technical Progress

#### 3.1 Approach

##### 3.1.1 Inception module

Inception is a convolutional neural network (CNN) architecture aimed at improving network performance with a more efficient structure, achieving higher accuracy even with limited computational resources. The key concept behind Inception is to apply multiple convolutional kernels of varying sizes in parallel and combine their outputs.

In my model, Inception module learns from spatial features at different scales by using convolution kernels of different sizes (1x1, 3x3, 5x5) and max pooling operations. By applying multiple convolution operations in parallel, the Inception module is able to simultaneously capture different features in the image, thereby enhancing the model's representational capacity. Below is the architecture diagram of the model.

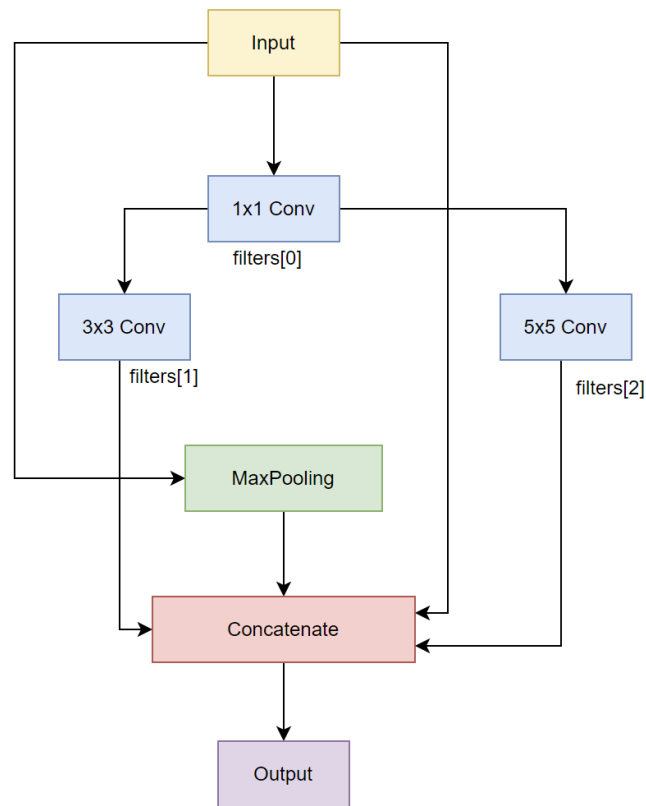


Figure 1. The Architecture Diagram of Inception Module

### 3.1.2 Attention module

The Attention Mechanism is a technique that enables neural networks to "selectively" concentrate on the most relevant parts of the input data. This ability to focus on key information allows attention mechanisms to greatly enhance the model's performance on complex tasks.

In this model, I have implemented an attention module with a residual connection, which combines the attention mechanism and residual connections. This structure helps capture important information in the input and alleviates the vanishing gradient problem in deep networks through residual connections, thereby improving the model's training effectiveness.

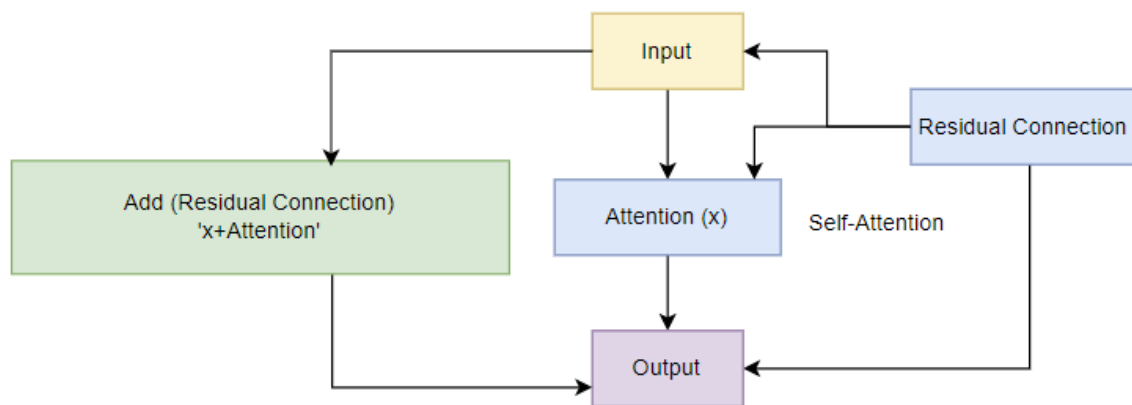


Figure 2. The Architecture Diagram of Attention Module

### 3.1.3 BiLSTM Module

BiLSTM (Bidirectional Long Short-Term Memory) is an extension of the traditional LSTM (Long Short-Term Memory) network model, designed to process sequential data. By simultaneously learning features in both the forward and backward directions of the sequence, it is able to capture more information. In addition, BiLSTM uses two LSTM networks simultaneously, one processing the sequence in the forward direction (from left to right) and the other in the backward direction (from right to left).

Below is the architecture diagram of BiLSTM in my model.

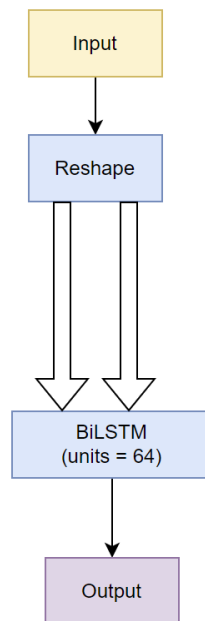


Figure 3. The Architecture Diagram of BiLSTM Module

### 3.1.4 The Architecture Diagram of Proposed model

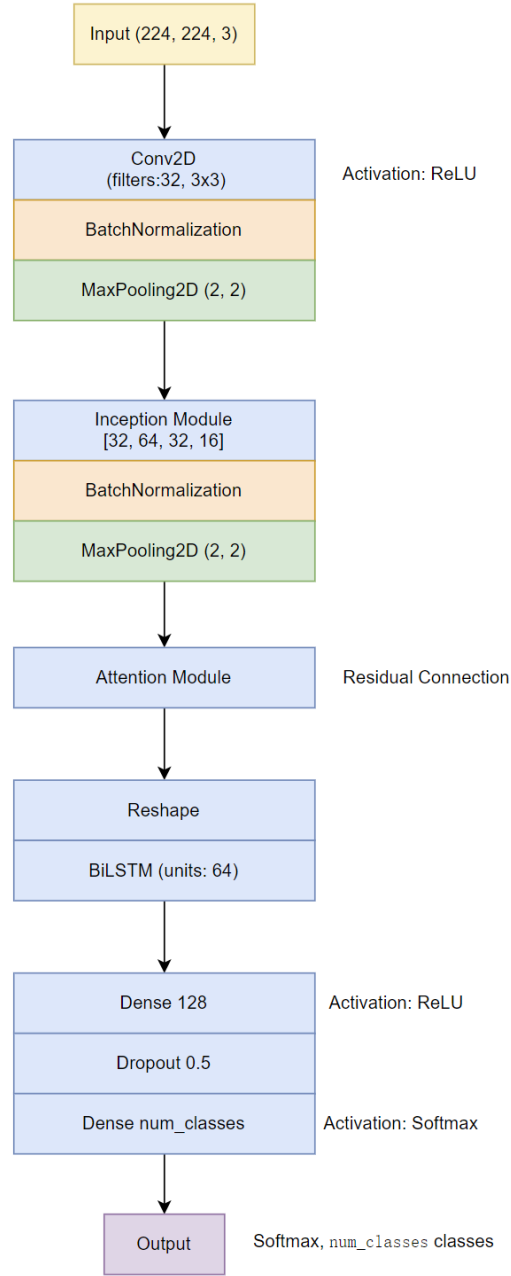


Figure 4. The Architecture Diagram of proposed model

## 3.2 Dataset

### 3.2.1 Dataset Information

This dataset consists of 1009 thermal images related to photovoltaic cells, with each image having a size of 512×640 pixels. The data is stored across three numpy array files: `imgs_temp.npy`, `imgs_mask.npy`, and `imgs_check.npy`. The `imgs_temp.npy` file contains

the thermal images, where the pixel values represent the temperature in Celsius at each (x, y) coordinate. The `imgs_mask.npy` file contains the corresponding anomaly mask images, with pixel values of 0 indicating no anomaly and 1 indicating the presence of an anomaly at that pixel. Lastly, the `imgs_check.npy` file contains a 3-value label for each image, denoting the type of anomaly present: 0 indicates one anomalous cell, 1 indicates more than one anomalous cell, and 2 indicates a contiguous series of anomalous cells. This dataset is specifically designed for detecting and analyzing anomalies in photovoltaic cells, which can aid in identifying faulty modules, hot spots, and other issues that impact system performance.

### **3.2.2 Data Separation**

In the separation of the dataset, the number of the training set is 726, the number of the validation set is 81 and the number of the test set is 202. The percentage of the data separation is: training set is 72%, validation set is 8% and the test set is 20%. So, the proportion is reasonable. At last, save the preprocessed data for the model training.

### **3.2.3 Data Resize**

The original size of the dataset is 512x640 pixels which is not suitable for the model, so resize them to 224x224.

### **3.2.4 Data Balance**

To process the data balance, the Synthetic Minority Over-sampling Technique (SMOTE) is used to oversampling the data to deal with the label imbalance problem. First, each image is flattened into a one-dimensional array in order to use SMOTE. Then, after SMOTE processing, reshape is used to restore the oversampled data to the original image shape.

### **3.2.5 Label Processing**

The dataset labels are stored in the 'npy' format. Upon loading the file, obtain the number of images, the pixel data, and the temperature range. After checking the distribution of labels, find that label 0 has 841 instances, label 1 has 116 instances, and label 3 has 52 instances. The next step is to convert these labels into one-hot encoding, where the labels will be represented as 0, 1, and 2 for each class.

### **3.2.6 Data Augmentation**

The specified parameters introduce various transformations to the images to artificially expand the dataset and improve model generalization. These transformations include

random rotations (up to 30 degrees), width and height shifts (up to 20% of the image size), shearing, zooming, and horizontal flipping. The fill\_mode='nearest' argument ensures that any empty areas after transformation are filled with the nearest pixel values. This augmentation helps the model become more robust to variations in the data.

### 3.3 Technology

|          |                               |   |
|----------|-------------------------------|---|
| Software | Framework                     | Tensorflow  |
|          | Language                      | Python  |
|          | Libraries and Application     | Numpy, Keras, Matplotlib  |
| Hardware | Central processing unit (CPU) | Intel® Core(TM) i9-14900HX processor, with a base clock speed of 2.20 GHz and a maximum turbo frequency of 5.40 GHz |
|          | Graphic Processing Unit (GPU) | Nvidia GeForce RTX 4070 GPU   |

Table 2. The technology used in the project

### 3.4 Testing and Evaluation Plan

In this part, the model is tested by 10 metrics, including:

TN: the number of samples correctly predicted as negative by the model.

TP: The number of samples correctly predicted as positive by the model.

FN: The number of samples incorrectly predicted as negative by the model when they are actually positive.

FP: The number of negative samples incorrectly predicted as positive by the model.

#### 1. Accuracy

Accuracy represents the proportion of samples that the model correctly predicts. The higher the accuracy, the stronger the model's predictive ability.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

#### 2. Loss

$$L = -\sum_{i=1}^C y_i \log(p_i) \quad (2)$$

$C$  is the number of classes.

$y_i$  is the actual label (in one-hot encoding form), where if the true label of the sample is class  $i$ , then  $y_i=1$ , otherwise  $y_i=0$ .

$p_i$  is the predicted probability of class  $i$  by the model (i.e., the model's output).

### 3. Precision

Precision is used to measure the proportion of predicted positive samples that are actually positive in a classification model. The higher the precision, the fewer false positives the model produces.

$$Precision = \frac{TP}{TP+FP} \quad (3)$$

### 4. Recall and Sensitivity

These are used to measure the model's ability to correctly identify positive class samples. The higher the recall, the more accurately the model identifies positive class samples.

$$Sensitivity = Recall = \frac{TP}{TP+FN} \quad (4)$$

### 5. F1-Score

The F1-score ranges between 0 and 1, with values closer to 1 indicating higher precision and recall, and thus better model performance.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} = \frac{2TP}{2TP+FP+FN} \quad (5)$$

### 6. ROC Curve

The ROC curve plots the relationship between the True Positive Rate (TPR) and the False Positive Rate (FPR) of a model.

$$TPR = \frac{TP}{TP+FN} = Recall \quad (6)$$

$$FPR = \frac{FP}{FP+TN} = 1 - Specificity \quad (7)$$

In an ideal scenario, the ROC curve should be as close as possible to the top-left corner, indicating a high True Positive Rate (TPR) and a low False Positive Rate (FPR).

### 7. AUC (Area Under the ROC Curve)

AUC = 1: This indicates that the model perfectly classifies all samples without any misclassification.

AUC = 0.5: This means the model has no discriminative power, essentially performing as random guessing.

AUC < 0.5: This suggests that the model's performance is worse than random guessing, meaning the model is making incorrect predictions (inverse predictions).

$$AUC = \int_0^1 TPR(FPR) d(FPR) \quad (8)$$

## 8. Specificity

Specificity measures the ability of a classification model to correctly identify negative class samples. It represents the proportion of actual negative samples that are correctly predicted as negative by the model. The higher the specificity, the better the model can effectively exclude negative class samples and reduce the occurrence of false positives.

$$Specificity (SP) = \frac{TN}{TN+FP} \quad (9)$$

## 9. Confusion Matrix

A confusion matrix is a table that contains four basic elements. It lists the comparison between the model's predicted results and the actual labels.

$$\begin{bmatrix} TP & FP \\ FN & TN \end{bmatrix} \quad (10)$$



### 3.5 Design and Implementation

#### 3.5.1 Accuracy Curve

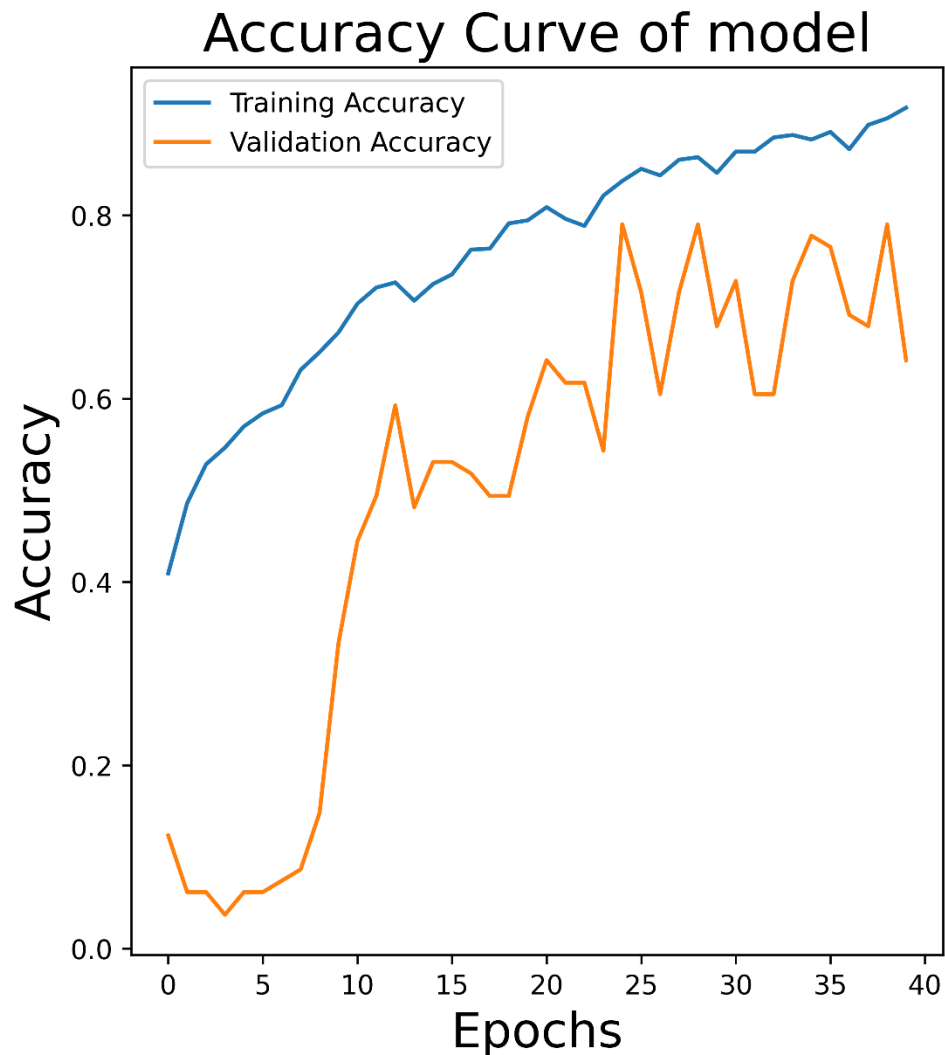


Figure 5. The accuracy curve of the model

In the early stages of training, both training accuracy and validation accuracy increase gradually with little difference between them, indicating that the model is balancing learning and generalization. As training progresses, training accuracy increases significantly while validation accuracy grows more slowly, causing the gap between them to widen, which suggests overfitting. This means the model has overlearned the training data, reducing its ability to generalize to new data. After a certain number of training cycles, training accuracy plateaus, while validation accuracy continues to rise, indicating that the model has learned generalized features, improving its adaptability to unseen data.

### 3.5.2 Loss Curve

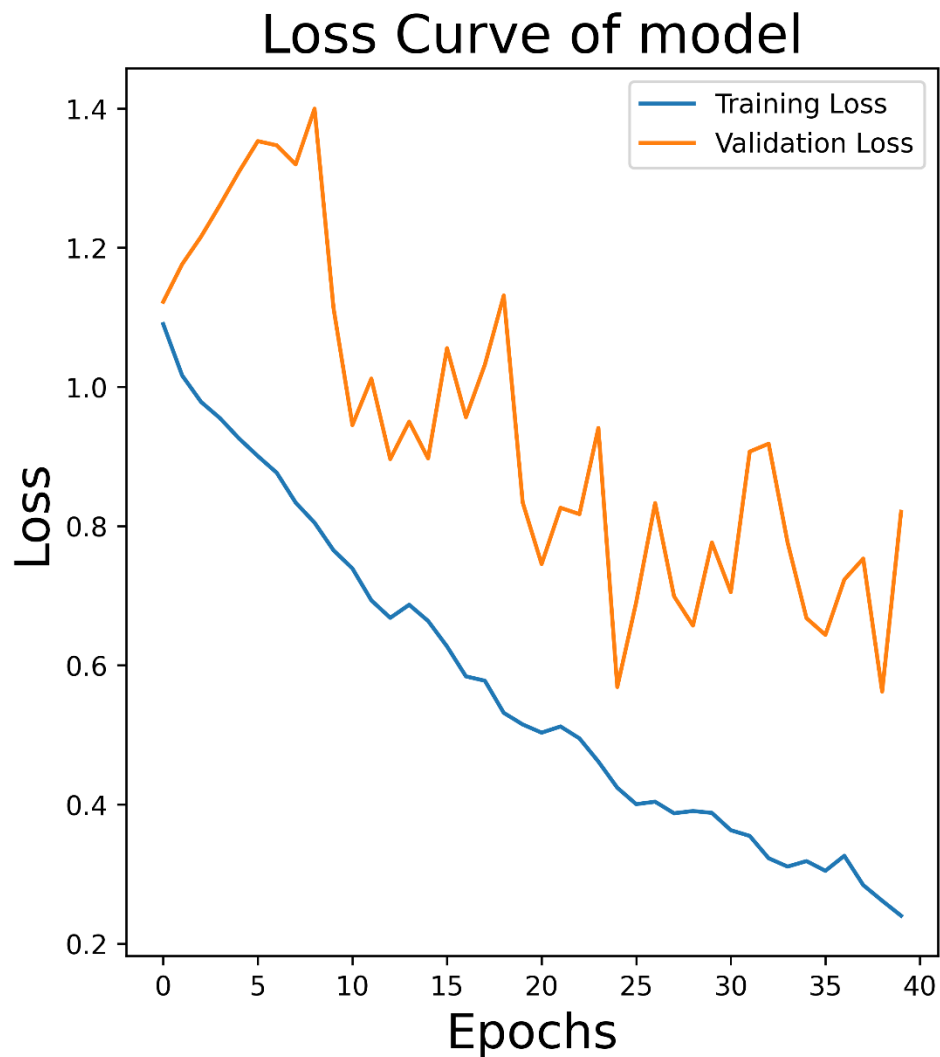


Figure 6. The loss curve of the model

in the early stages, both lines show an upward trend, indicating that the model might have experienced overfitting. As training progresses, the blue line (training loss) gradually decreases and stabilizes, while the orange line (validation loss) first decreases, then fluctuates and slightly increases. This suggests that the model has improved its generalization ability to some extent and is better able to adapt to new, unseen data samples. However, since the validation loss does not continue to stabilize or decrease further, it indicates potential issues such as underfitting or an overly large learning rate, which may require adjustments in the optimization strategy to improve overall performance.

### 3.5.3 Precision

The Precision is 0.3951, this means that only 39.51% of the samples predicted as positive by the model are actually correct, suggesting that the model has a significant number of false positive predictions. Low precision may mean that the model is overestimating the number of positive samples, which could lead to unnecessary maintenance or false alarms in photovoltaic system fault detection.

### 3.5.4 Recall and Sensitivity

Recall (Sensitivity) is 0.3940 which indicates that model correctly identified 39.40% of the actual positive samples.

### 3.5.5 F1-Score

F1-Score is 0.3944 which means that the model has a low balance between precision and recall.

### 3.5.6 ROC Curve & AUC

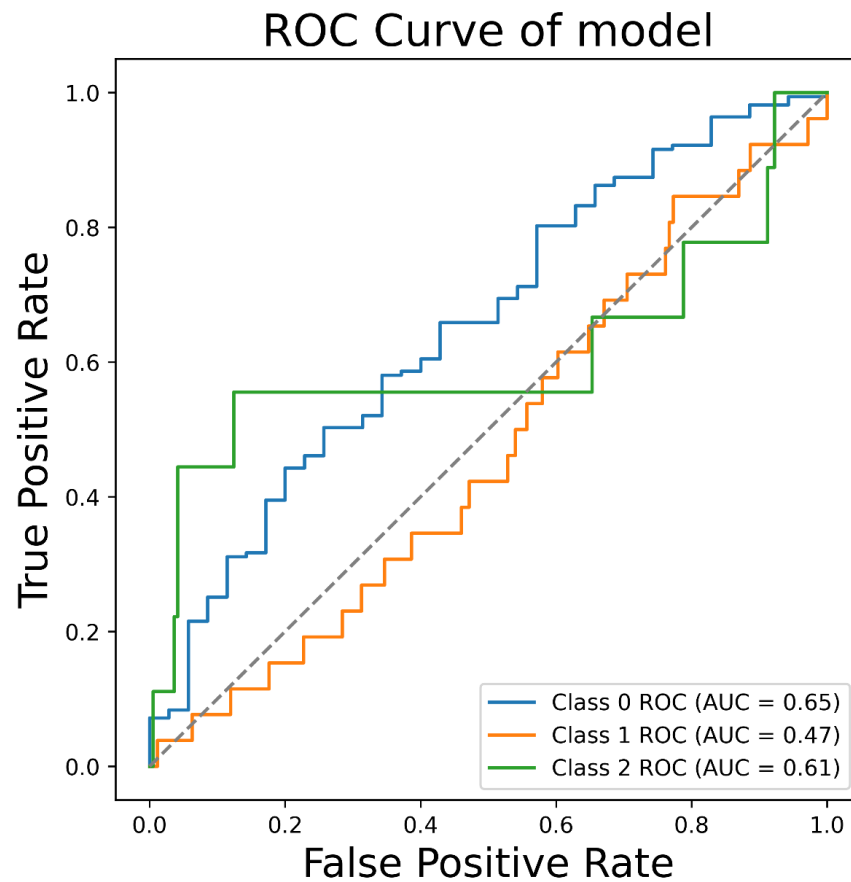


Figure 7. The ROC curve of the model

Class 0: 0.65, Class 1: 0.47, Class 2: 0.61.

In summary, the model performs better for Class 0 and Class 2 but needs improvement for Class 1.

### 3.5.7 Specificity

The overall specificity is 0.8614 which indicates that the model correctly identified 86.14% of the actual negative samples. The model performs well in avoiding false positives and effectively classifies negative samples as negative.

### 3.5.8 Confusion Matrix

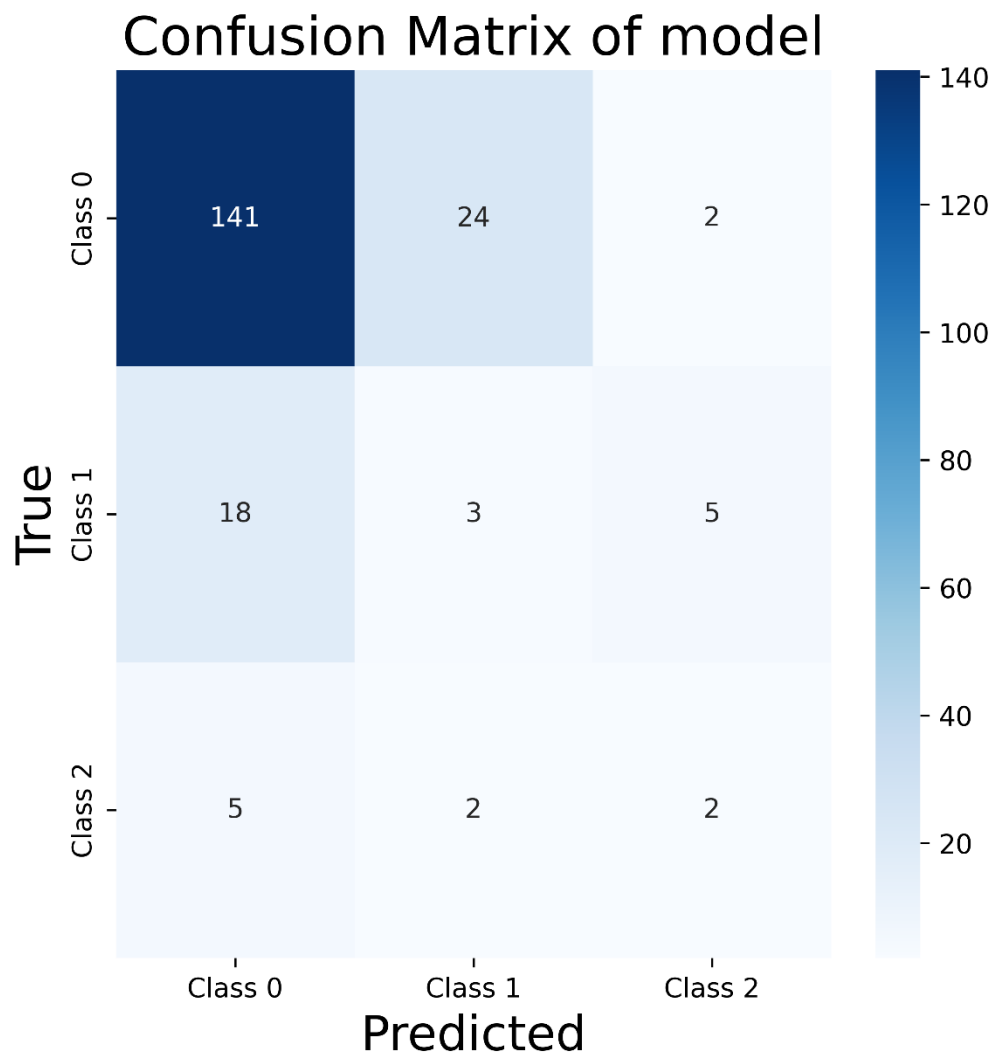


Figure 8. The confusion matrix of model

### 3.5.9 Classification Report

| Classification Report of model |           |        |          |         |
|--------------------------------|-----------|--------|----------|---------|
|                                | precision | recall | f1-score | support |
| Class 0                        | 0.86      | 0.84   | 0.85     | 167     |
| Class 1                        | 0.10      | 0.12   | 0.11     | 26      |
| Class 2                        | 0.22      | 0.22   | 0.22     | 9       |
| accuracy                       |           |        | 0.72     | 202     |
| macro avg                      | 0.40      | 0.39   | 0.39     | 202     |
| weighted avg                   | 0.73      | 0.72   | 0.73     | 202     |

Figure 9. The classification report of model

Category 0 has the highest precision, recall, and F1 score, and also has the largest support, indicating the best performance. Category 2 has the lowest values for all three metrics, suggesting that the model performs the worst for this category. Although Category 1's overall performance is not as good as the other two categories, its F1 score is not significantly different from the others, meaning it may still have some value in certain situations.

## 4 Project Management

### 4.1 Activities

| Phase   | Activities   |
|---|--|
| <b>1. Understand the aim of the project and have a brief idea about the project. (Completed)</b>  | <ol style="list-style-type: none"><li>1. Read the title of the project and find the keywords of the project.</li><li>2. Search the keywords in the internet to learn about the background about the topic.</li><li>3. Search the deep learning models and learn the basic knowledge about the models</li></ol> |
| <b>2. Research the existing researches and the models they used. (Completed)</b>  | <ol style="list-style-type: none"><li>1. Search the keywords in the internet to find 10 existing works.</li><li>2. Research the existing works and find the models they used.</li><li>3. Analyze the result and accuracy of their models.</li></ol>  |
| <b>3. Collect the dataset of thermal imaging data of solar panels and preprocess the data. (Uncompleted: need more suitable dataset for prediction)</b> | <ol style="list-style-type: none"><li>1. Find the dataset which is suitable for the topic on the website.</li><li>2. View the dataset and check its usability.</li><li>3. Preprocess the data of the thermal images of solar panels.</li></ol>   |
| <b>4. Select and design the model for analyzing thermal images. (Uncompleted: need to add prediction model)</b>   | <ol style="list-style-type: none"><li>1. Find and select the suitable models for the topic.</li><li>2. Learn more about the deep knowledge of CNN models include the basic concepts, components, working principles, advantages, applications, etc.</li><li>3. Design the deep learning model.</li></ol>       |

|   |  |
|---|--|
| 5. Train and test the model with the dataset and optimize it. (Uncompleted) | <ol style="list-style-type: none"> <li>1. Train the model with the dataset and find the result of the accuracy and the loss.</li> <li>2. Analyze the result and summarize the advantages and limitations</li> <li>3. Optimize the model result.</li> </ol> |
| 6. Present the result of the project. (Uncompleted)                         | <ol style="list-style-type: none"> <li>1. Conclude all the contribution and result.</li> <li>2. Prepare the final report and presentation.</li> </ol>  |

Table 3. Completed or Uncompleted Activities of the Project

## 4.2 Schedule

NB: 1.1 represent Objective 1, Activity 1.

Green is completed part, yellow is upcoming part and red is uncompleted.

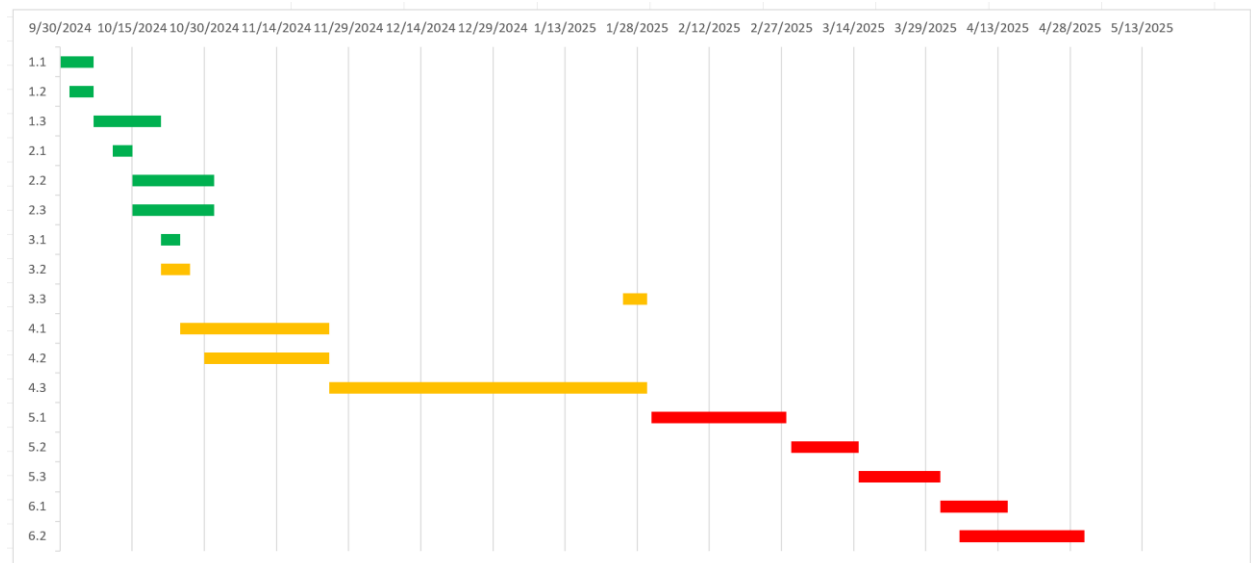


Figure 10. The schedule of the project

## 4.3 Project Version Management

To manage the different versions of codes modification, I plan to use Github as the version management tools.

URL: <https://github.com/Olivia-Gao/Deep-Learning-Project>

| Version Number  | Code Name            | Contents  | Results  |
|-----------------|----------------------|---|--|
| 1 (Completed)   | Classification model | Classification deep learning model, architecture diagrams, Datasets | The model result for classification with all measured metrics. |
| 2 (Uncompleted) | Prediction model     | Prediction deep learning model, architecture diagrams, Datasets     | The model result for prediction with all measured metrics.     |

Table 4. Project Version Management

#### 4.4 Project Data Management

All files will be replicated into three copies for fail safe, one on local computer, one on hard drive, one on GitHub.

Following are documents of the Project for uploading and synchronization:

- Reports (Weekly, Proposal, Progress, Final) & Presentation PPT
- Model architecture diagrams
- References
- Datasets Link
- Model evaluation documents
- Model codes (Different versions)

#### 4.5 Project Deliverables

1. Code
2. Datasets
3. Ethics for Project
4. Final Report
5. Presentation
6. Progress Report
7. Proposal
8. Reference
9. Testing
10. Weekly Report



## 5 Professional Issues and Risk:

### 5.1 Risk Analysis

| Current Progress  | Resolved risks  | success of the mitigation strategy   | Changes to project plan as a result of risks   |
|-------------------|---|--|--|
| Dataset selection | <ol style="list-style-type: none"><li>1. The dataset is not clear, the pixel is low</li><li>2. The dataset is hard to preprocess.</li></ol> | <ol style="list-style-type: none"><li>1. Find more datasets which is suitable for the model.</li><li>2. Try many methods to process, if the dataset is incorrect, change it.</li></ol> | <ol style="list-style-type: none"><li>1. More time spend to find the dataset</li><li>2. Use more time to process the dataset</li></ol> |
| Model Training    | <ol style="list-style-type: none"><li>1. Out of Memory</li><li>2. The model accuracy is low.</li></ol>                                      | <ol style="list-style-type: none"><li>1. Reduce the batch_size, change the model etc.</li><li>2. Change the model, add more layers etc.</li></ol>                                      | More time to change the model and train the model.   |

Table 5. The risk analysis of the project

#### Future Risks:

1. The datasets for prediction model are not suitable.
2. The model doesn't work well.
3. The Out of Memory problem.
4. Poor time management.

### 5.2 Professional Issues

1. Legal Issues: Ensure compliance with data privacy laws (e.g., GDPR) when using thermal images, obtain consent for data usage, and adhere to intellectual property rights. Consider liability concerns if the model's predictions lead to incorrect maintenance decisions that cause damage to infrastructure.

2. Social Issues: Automation of maintenance tasks may lead to job displacement for workers traditionally involved in inspections. The accessibility of advanced technology should be considered to prevent widening the digital divide, ensuring smaller companies and developing regions can benefit from it.
3. Ethical Issues: Address biases in deep learning models to ensure fairness and accuracy. Ensure transparency and accountability in predictions, and provide explainability of AI decisions. Obtain informed consent from solar plant owners, ensuring they are aware of data usage and analysis.
4. Environmental Issues: The project promotes sustainability by improving solar panel efficiency. However, the energy consumption of deep learning model training and potential electronic waste from AI devices should be considered, ensuring responsible disposal and reducing the environmental footprint of the technology.
5. Professional Code of Conduct: Follow the BCS and ACM codes, focusing on integrity, transparency, privacy, and public welfare. The project must ensure ethical use of AI, safeguard against misuse, and prioritize the welfare of individuals and communities affected by the technology's deployment.

## 6 References

- [1] A. Shaik, A. Balasundaram, L. S. Kakarla, and N. Murugan, "Deep Learning-Based Detection and Segmentation of Damage in Solar Panels," *Automation*, vol. 5, no. 2, pp. 128–150, May 2024, doi: 10.3390/automation5020009.
- [2] D. Lang and Z. Lv, "A PV cell defect detector combined with transformer and attention mechanism," *Sci. Rep.*, vol. 14, no. 1, p. 20671, Sep. 2024, doi: 10.1038/s41598-024-72019-5.
- [3] M. Dhimish, M. Theristis, and V. d'Alessandro, "Photovoltaic hotspots: A mitigation technique and its thermal cycle," *Optik*, vol. 300, p. 171627, Apr. 2024, doi: 10.1016/j.ijleo.2024.171627.
- [4] M. Dhimish and A. M. Tyrrell, "Photovoltaic Bypass Diode Fault Detection Using Artificial Neural Networks," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–10, 2023, doi: 10.1109/TIM.2023.3244230.
- [5] L. Koshy, M. V. Vaishnav, S. Sunil, S. V. Abraham, and S. Vidhyadharan, "A Thermal Image-based Fault Detection System for Solar Panels," in *2024 International Conference on Cybernation and Computation (CYBERCOM)*, Dehradun, India: IEEE, Nov. 2024, pp. 246–250. doi: 10.1109/CYBERCOM63683.2024.10803145.
- [6] "Integration\_of\_Aerial\_Thermal\_Imaging\_and\_Deep\_Learning\_for\_Fault\_Detection\_in\_Photovoltaic\_Panels\_A\_Study\_at\_Thinh\_Long\_Solar\_Power\_Plant."
- [7] C. Dunderdale, W. Brettenny, C. Clohessy, and E. E. Van Dyk, "Photovoltaic defect classification through thermal infrared imaging using a machine learning approach," *Prog. Photovolt. Res. Appl.*, vol. 28, no. 3, pp. 177–188, Mar. 2020, doi: 10.1002/pip.3191.
- [8] J. J. Vega Díaz, M. Vlaminc, D. Lefkaditis, S. A. Orjuela Vargas, and H. Luong, "Solar Panel Detection within Complex Backgrounds Using Thermal Images Acquired by UAVs," *Sensors*, vol. 20, no. 21, p. 6219, Oct. 2020, doi: 10.3390/s20216219.