Impact of Combining RGB and Grayscale Images on Hotspot Detection in Solar Panels Using Inception Resnet V2 Architecture

Sandy Suryady[1], Busono Soerowirdjo[1], Sri Poernomo Sari[2*], Ernastuti[1]

[1] Department Information Technology, Gunadarma University, Depok 16424, Indonesia
[2] Department Mechanical Engineering, Gunadarma University, Depok 16424, Indonesia

Corresponding Author Email: sri_ps@staff.gunadarma.ac.id

ABSTRACT

Solar panels are a technology that converts solar energy into electricity through the photovoltaic effect. This photovoltaic technology is packaged into solar modules consisting of many solar cells arranged in series or parallel. Damage to these panels can be identified by detecting hotspots using a thermal camera. Hotspots can be classified into three categories of damage: No Damage, Minor Damage, and Severe Damage. This study applies the Inception ResNet V2 architecture from deep learning to automatically classify the level of damage based on thermal images. The novelty of this research is its implementation for real time monitoring of a structured array of 20 solar panels (5×4 panel), enabling early detection and reporting of damage conditions. The model includes several architectural enhancements such as Average Pooling, Flatten, and ReLU layers. Training was conducted using three different datasets: RGB, grayscale, and a combination of both. The RGB dataset achieved the highest accuracy at 98.62 percent, followed by the combined dataset at 98.44 percent, and the grayscale dataset at 96.93 percent. These high accuracy results demonstrate that the proposed system can effectively support preventive maintenance of solar panels. Specifically, the system is applicable for operational use at PT PLN Nusantara Power UP Cirata to improve reliability, reduce power loss, and enhance the overall efficiency of solar energy generation.

1. INTRODUCTION

The development of solar panel technology is a crucial component in the global effort to enhance renewable energy efficiency, an increasingly urgent priority in today's energy transition era. As an environmentally friendly energy source, solar panels must operate at maximum efficiency to achieve optimal energy production. However, a significant challenge in managing solar panels is the detection and handling of early damage, such as hotspots, which can directly impact efficiency and shorten the operational lifespan of the panels. These hotspots are often caused by various factors, such as micro-cracks in solar cells, dust or dirt accumulation, and shading that covers specific parts of the panel [1]. If hotspots are not addressed promptly, their performance can deteriorate significantly, affecting the overall energy output and potentially leading to further damage [2].

To tackle these challenges, this research adopts an image processing approach based on deep learning, focusing on the Inception ResNet V2 architecture. This model has proven effective in recognizing complex patterns and processing visual information with high accuracy [3]. By combining RGB images, rich in color information, with grayscale images, which emphasize texture and intensity, this model is expected to detect hotspots on solar panels more accurately [4]. RGB images play a crucial role in identifying objects based on

spectral variations, while grayscale images enhance detection under suboptimal lighting conditions [5, 6].

The Inception ResNet V2 architecture is designed to integrate information from various image spectrums, offering high flexibility in detection and classification processes [7]. The combination of RGB and grayscale images has proven to capture richer details, often missed when only using one type of image [8]. Consequently, this model provides more accurate results in detecting anomalies on solar panels, such as hotspots, which frequently occur due to environmental variations or physical disturbances on the panel surface [9, 10].

This study also highlights the importance of analyzing and comparing various image fusion approaches in classification. The goal is to find the most effective method for identifying different types of damage to solar panels, ensuring the model is not only more accurate but also more efficient in terms of time and computational resources [11, 12]. Although deep learning architectures like Inception ResNet V2 have been widely applied in general image classification, their use in hotspot detection on solar panels is still relatively limited [13]. Therefore, this research aims to fill this gap by developing and comparing the effectiveness of various neural network architectures in combining RGB and grayscale images for detecting damage on solar panels with higher accuracy and efficiency [14, 15].

The implementation of this technology is expected to make

a significant contribution to improving efficiency and reducing the operational and maintenance costs of solar panels [16]. Earlier and more accurate detection of hotspots will extend the lifespan of solar panels while maximizing energy output, which is crucial for the renewable energy industry [17, 18]. Furthermore, integrating RGB and grayscale image analysis opens up opportunities for developing more advanced diagnostic systems that can be applied in various contexts for real-time solar panel condition monitoring [19]. This research not only focuses on enhancing detection accuracy but also explores the broader application of this technology to address operational challenges in the field [20, 21].

Hotspots may arise from micro-cracks in solar cells, accumulated dust or dirt, soldering defects, or partial shading, and are often difficult to detect through manual inspection alone. The real-world consequences of undetected hotspots are severe: they not only reduce energy yield but can also escalate into thermal runaway, posing fire hazards and accelerating material degradation [22, 23]. Studies have shown that hotspots can decrease the energy output of photovoltaic modules by up to 30% and, in extreme cases, cause permanent cell damage or electrical arcing [24].

To address these challenges, this study adopts a deep learning approach using the Inception ResNet V2 architecture, known for its robustness in complex visual recognition tasks. The research focuses on combining RGB and grayscale images to improve the accuracy of hotspot detection on solar panels. RGB images are rich in color and spectral information, aiding in identifying temperature variations, while grayscale images emphasize contrast and structural texture, which are particularly useful under low-light or uneven illumination conditions [25].

The novelty of this study lies in the fusion of RGB and grayscale modalities to enhance the model's ability to capture subtle anomalies that are often missed by single-spectrum inputs. Previous works have typically used only one image type or applied generic CNNs without optimizing for spectral variation or real-time field applications. Furthermore, although several deep learning models have been explored in photovoltaic diagnostics, the use of Inception ResNet V2 for hotspot classification with dual-spectral input remains underexplored in current literature [26].

This research not only seeks to bridge that gap but also evaluates the efficiency of the proposed method in terms of processing speed and computational cost important factors for real-time deployment in solar farms. The proposed system is applied to monitor a structured layout of 20 solar panels (5×4 panel), generating early diagnostic reports that support preventive maintenance, particularly for implementation at PT PLN Nusantara Power UP Cirata.

By integrating multiple image sources and a powerful deep learning model, this study contributes toward improving the accuracy, speed, and practicality of hotspot detection. This can lead to reduced operational costs, extended panel lifespan, and most importantly, minimized risks of energy loss or fire outbreaks in solar installations [27].

2. LITERATURE REVIEW

This research also explores various new approaches in image processing, including the use of hybrid methods and multispectral imaging to enhance the detection and classification of damage in solar panels [28]. In this context,

architectures like Inception ResNet V2 not only offer efficient solutions in terms of computational cost but also provide high accuracy, making them an ideal choice for applications that require high-resolution image processing [29, 30]. Through this approach, this study aims to achieve significant improvements in the early detection of hotspots, ultimately supporting global efforts to maximize the utilization of solar energy [31].

In Inception ResNet, each block consists of several parallel convolutional paths that are merged into a single output and then combined with a residual connection.

$$A^{(l+n)} = ReLU\left(\sum_i (W^{(l,i)} * A^{(l-1)})\right) + b^{(l)} + A^{(l)} \qquad (1)$$

To reduce spatial dimensions while increasing or maintaining the number of filters [32]. Inception ResNet architecture uses a dimension reduction strategy to decrease spatial dimensions while increasing or maintaining the number of filters. This strategy involves using pooling layers (such as max-pooling or average-pooling) or convolutions with a stride greater than 1. This process can be formulated as:

$$A^{(l)} = Pooling\,(A^{(l-1)}) \qquad (2)$$

or

$$A^{(l)} = W^{(l)} * A^{(l-1)} + b^{(l)} \qquad (3)$$

where, Pooling $(A^{(l-1)})$ is a pooling operation that reduces the spatial dimensions of $(A^{(l-1)})$ while maintaining or even increasing the number of filters as needed for the subsequent layers. This is crucial for maintaining computational efficiency and avoiding overfitting, especially in very deep networks like Inception ResNet [33].

Furthermore, to control the output size generated by each block, various regularization techniques are employed, such as dropout and batch normalization. Batch normalization can be formulated as:

$$\hat{Z}^{(l)} = \frac{Z^{(l)} - \mu^{(l)}}{\sqrt{(\sigma^{(1)})^2 + \epsilon}} \qquad (4)$$

$$A^{(l)} = ReLU(\hat{Z}^{(l)}) \qquad (5)$$

where, $\hat{Z}^{(l)}$ is the normalized output, $\mu^{(l)}$ is the mini-batch mean of the output $\hat{Z}^{(l)}$, σ is the mini-batch standard deviation, and ϵ is a small value to prevent division by zero [34]. To optimize the entire network, a commonly used loss function for classification tasks is categorical cross-entropy, which is formulated as:

$$L = -\sum_{i=1}^{N} y_i \log(\hat{y}_i) \qquad (6)$$

where, y_i is the labelling, and $\hat{y}_i$ is the probability predicted by the model for class i. This loss function measures the difference between the predicted probability distribution of the model and the actual distribution, and it is used to update the model parameters through backpropagation [35]. Overall, the Inception ResNet architecture combines the flexibility of

Inception blocks with the stability of residual connections, resulting in a robust model capable of handling complex image classification tasks [34].

3. METHODOLOGY

In this research, the Inception ResNet V2 architecture showed outstanding performance in image classification, achieving high accuracy and efficient computational processing. The model successfully utilizes the integration of Inception design and residual connections, which enables the capture of a wide range of deep features across multiple scales. The following are the stages conducted in this research:

First, the model was trained using carefully selected hyperparameters to optimize performance. The detailed hyperparameter settings used during the training process are presented in Table 1. These parameters, such as learning rate, batch size, number of epochs, and dropout rate, played a crucial role in achieving stable and high-performing results.

Table 1. Hyperparameter values in the training process of Inception ResNet V2

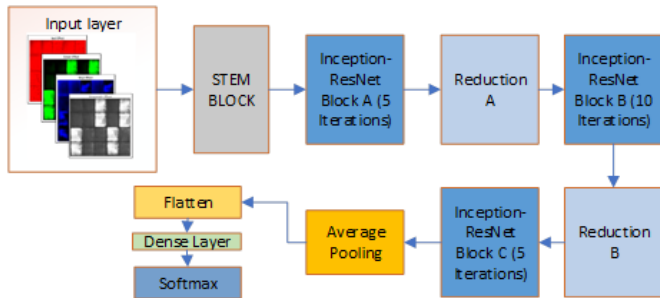| No.	Parameters	Capability Values
1	Input Shape	299×299
2	Learning Rate	0.0001
3	Dropout	0.5
4	Batch size	32
5	Epochs	200
6	FC Layer	512
7	Activation Function	ReLu
8	Optimizer	Adam

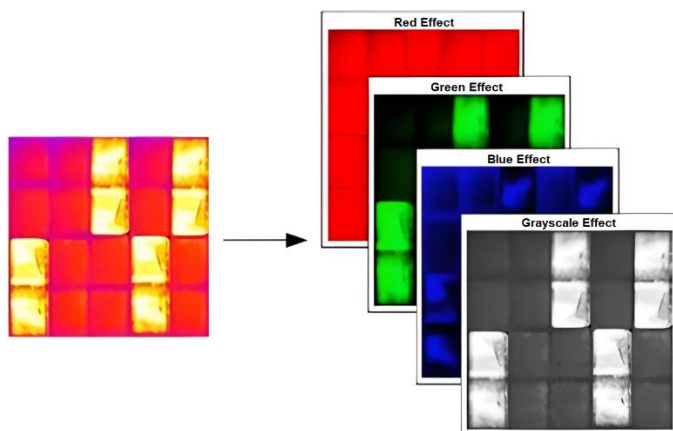**Figure 1.** The process flow of the Inception ResNet V2 architecture

Figure 2. Splitting RGB and grayscale images

Figure 1 illustrates the process flow of the Inception ResNet V2 architecture, starting with the "Input Layer" that receives the image data, followed by the Stem Block, which performs initial feature extraction using convolution and pooling. The data then passes through Inception ResNet Block A, Reduction A, Inception ResNet Block B, and Reduction B, with appropriate iterations and dimensionality reduction. The process continues until the data is pooled, flattened, and classified using a Dense Layer and Softmax.

This research focuses on the integration and analysis of an image dataset consisting of 570 RGB images, 570 grayscale images, and 1140 combined images, categorized into three types of damage. Before training, all images are normalized and undergo data augmentation techniques such as shifting, flipping, and scaling. The process of organizing and splitting the RGB and grayscale images prior to model training is illustrated in Figure 2. This figure provides a clear visualization of how the dataset was separated and prepared for input into the model, ensuring that each category was properly balanced and systematically managed.

The process concludes with an output layer that uses the softmax activation function to produce the final classification into three different classes.

The selection of learning rate=0.0001, dropout=0.5, and 200 epochs was based on initial experiments and supported by recent research in the field. A learning rate of 0.0001 was chosen to allow stable and gradual updates during fine-tuning of the pre-trained Inception ResNet V2 model. This is important in transfer learning, where larger learning rates can disrupt the pre-trained weights. The dropout rate of 0.5 was applied to reduce the risk of overfitting, especially considering the moderate dataset size and the complexity of the model. We set the number of training epochs to 200 to give the model enough time to learn complex features from the RGB, grayscale, and combined image datasets. However, we also implemented early stopping to automatically stop training when the validation loss no longer improved. This helped ensure that the model was both accurate and efficient. Overall, this set of parameters provided the most stable and reliable performance across all experiments.

This research evaluates the effectiveness of image segmentation using RGB, grayscale, and combined images to enhance object recognition and classification accuracy. RGB images are rich in color information, while grayscale images emphasize texture and intensity. The combined approach is expected to optimize the advantages of both methods, and the results indicate that this approach consistently outperforms single methods in terms of segmentation accuracy and robustness against noise, significantly improving image processing performance.

4. RESULTS AND DISCUSSION

This section presents the results and analysis related to the implementation of the model, as well as the evaluation of the Inception ResNet V2 architecture after the data augmentation process has been performed. The results show a total of 55,125,219 parameters, with 4,919,267 being trainable parameters and 50,205,952 non-trainable parameters, spread across 2048 layers. The detailed results of the Inception ResNet V2 model analysis are summarized in Table 2, which presents the key statistics obtained during the model's running process.

The performance of the Inception ResNet V2 architecture can be evaluated using outputs such as Training Process Output, Model Accuracy and Loss Graphs, Confusion Matrix, Test Accuracy, Precision, Recall, and F1-score. Visualization results of the modeling from all three databases are displayed in the images below. The image augmentation process is implemented to enhance and expand the diversity of the dataset, thereby improving the robustness and generalization capabilities of the desired image model. By applying various transformations to the original images, such as rotation, flipping, scaling, and brightness adjustments, this process creates new, synthetic variations of the dataset. These augmented images mimic real-world variations, enabling the model to better recognize and understand patterns even when the input data differs in orientation, scale, or lighting conditions. After augmentation, the dataset is divided into training and validation sets using an 80:20 ratio, where 80 percent of the data is used for training and the remaining 20 percent is used for validation to evaluate the model's performance during the learning process. The following augmentation techniques were applied to achieve this enhancement.

Table 2. Results of Inception ResNet V2 analysis running

Steps	Layer and Output Shape	Description
Input	299×299×3	Citra
Stem Block		Initial process for basic feature extraction
Conv2D	149×149×32	Kernel: 3×3, Stride: 2×2, Padding: valid
BatchNormalization	149×149×32	Batch normalization
ReLU	149×149×32	Activation function
Conv2D	147×147×32	Kernel: 3×3, Stride: 1×1, Padding: valid
BatchNormalization	147×147×32	Batch normalization
ReLU	147×147×32	Activation function
Conv2D	147×147×64	Kernel: 3×3, Stride: 1×1, Padding: same
BatchNormalization	147×147×64	Batch normalization
ReLU	147×147×64	Activation function
MaxPooling2D	73×73×64	Pool size: 3×3, Stride: 2×2, Padding: valid
Inception ResNet Block A		Repeated 5 times, each block integrates residual connections for enhanced stability.
Mixed Layer	35×35×256	Combination of multiple parallel convolutions with residual connections
Reduction A	17×17×896	Reducing spatial dimensions while maintaining or increasing depth.
Inception ResNet Block B		Repeated 10 times, deepening feature extraction using convolutions and residual connections.
Mixed Layer	17×17×1152	More complex combination of parallel convolutions.
Reduction B	8×8×1152	Dimensionality reduction to prepare for further feature processing in subsequent blocks
Inception ResNet Block C		Repeated 5 times, integrating the final features before the final classification.
Mixed Layer	8×8×2048	Aggregation of features from various scales using residual connections.
	Average Pooling2D 1×1×2084	Dropout 0.5
Classification	Flatten Nonex2048 Dense: Nonex2048 Dropout: Nonex2048 Dense 1/Output/Softmax Nonex3	**Fully Connected Layer**
	Total params: 55,125,219	
	Trainable params: 4,919,267	
	Non-trainable params: 50,205,952	

Table 3. Image augmentation

Parameter	Value
Augmentation Rotation	30
Width Shift	0.2
Height Shift	0.2
Horizontal Flip	True
Vertical Flip	True
Brightness	[0.4, 1.5]
Zoom range	0.3
Rescale	1./255
Validation split	0.2

As a result of the image augmentation process, the dataset size increases significantly (excluding the validation split). Augmentation techniques applied to the dataset introduce new variations of the existing images, enhancing the model's ability to generalize across different conditions and scenarios. These transformations include changes in orientation, scale, brightness, and positioning, which simulate real-world variability in data. The details of the augmentation techniques applied are summarized in Table 3, providing an overview of the strategies implemented to diversify the dataset and strengthen the model's robustness.

Below is a summary of the outcomes produced by the augmentation process. The results demonstrate the significant expansion of the dataset, along with the number of images generated for each class and dataset type after applying various augmentation techniques. These details are provided in Table 4, which outlines the augmentation values obtained after the completion of the augmentation process.

Table 4. Augmentation values after augmentation

Dataset	Value	Data Augmentation
RGB	570	4,560
Grayscale	570	4,560
Combine	1140	9,120

4.1 Results for RGB

The augmentation process is carried out based on the nature of the damage, and the images are transformed accordingly to match the specific augmentations applied. This procedure helps to enrich the existing dataset by generating a wider range of variations from the original data, making the model more robust and capable of handling different scenarios. Through these augmentations, the dataset becomes more comprehensive, allowing the model to better understand and detect the specific types of damage it is being trained on. The results of the augmentation process for the RGB dataset are illustrated in Figure 3, highlighting the enhanced diversity and variation achieved through augmentation techniques.
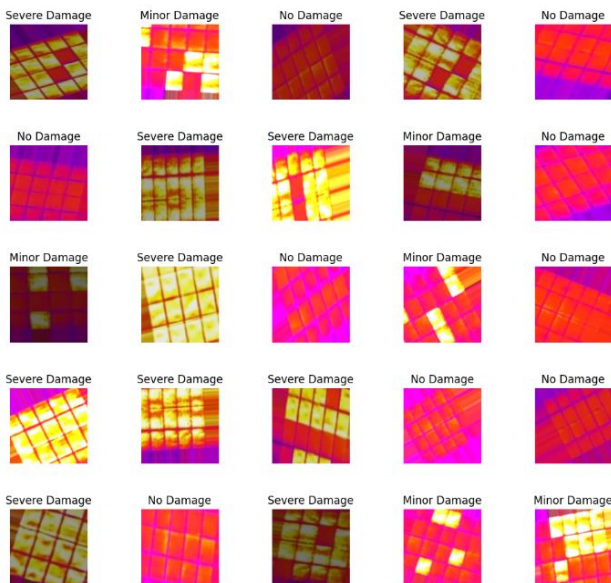

Figure 3. Results augmentation for dataset RGB

The training process for the model was conducted using Google Colaboratory. The results obtained can be seen in the images below.

Figure 4 shows the training log of a deep learning model over multiple epochs, with a focus on accuracy and loss for both training and validation datasets. The model was initially set to train for 200 epochs but was stopped early at epoch 37 due to the implementation of early stopping, which helps prevent overfitting by halting training when no further significant improvements are observed. By epoch 37, the model achieved a training accuracy of 98.62% and a validation accuracy of 96.88%, with a training loss of 0.0289 and a validation loss of 0.2019.

The two graphs illustrate the training and validation (test) accuracy and loss of a deep learning model over 37 epochs. In the accuracy graph, both training and test accuracies show an overall upward trend, with training accuracy consistently hovering around 95–100%, indicating effective learning from the training data. Test accuracy also remains high, despite some fluctuations, suggesting that the model generalizes well

to unseen data. The training and validation accuracy over epochs are presented in Figure 5, demonstrating the model's strong and consistent performance.

In the loss graph, training loss remains low and stable throughout the epochs, indicating that the model is learning effectively without overfitting. The training and validation loss over epochs are shown in Figure 6, highlighting the model's stability and its ability to maintain low error rates across training iterations.
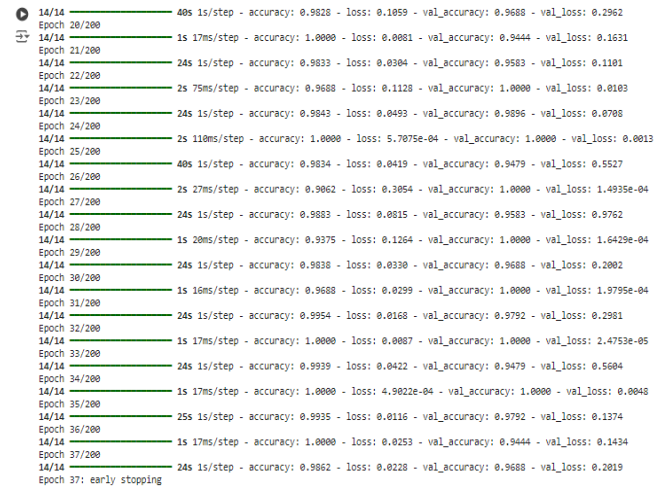

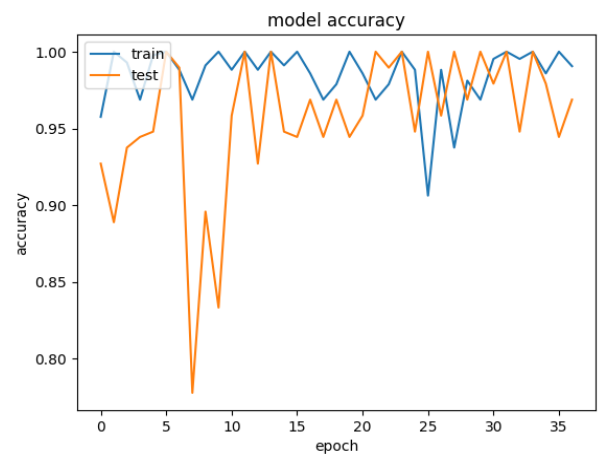**Figure 4.** Training and validation process for RGB

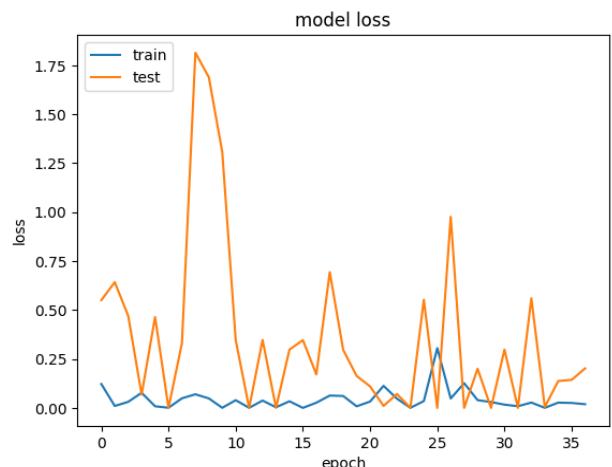**Figure 5.** Accuracy vs. epoch graph for RGB

Figure 6. Loss vs. epoch graph for RGB

The classification model was designed to distinguish between three categories of damage severity: Class 0 (No Damage), Class 1 (Minor Damage), and Class 2 (Severe Damage). These labels were used to interpret model predictions and assess performance metrics. The confusion matrix for the RGB dataset, which visualizes the distribution of correct and incorrect predictions across these classes, is presented in Figure 7.

Furthermore, the model's balance between precision and recall for each class is depicted in Figure 8, through the recall versus precision curve for the RGB dataset. This curve provides additional insight into the model's classification performance, particularly its ability to maintain high precision and recall simultaneously.
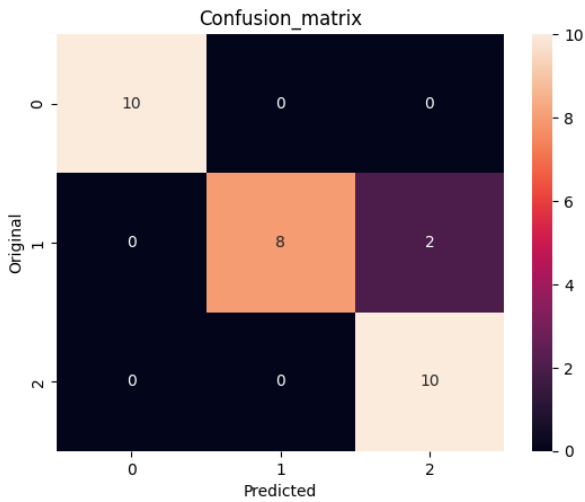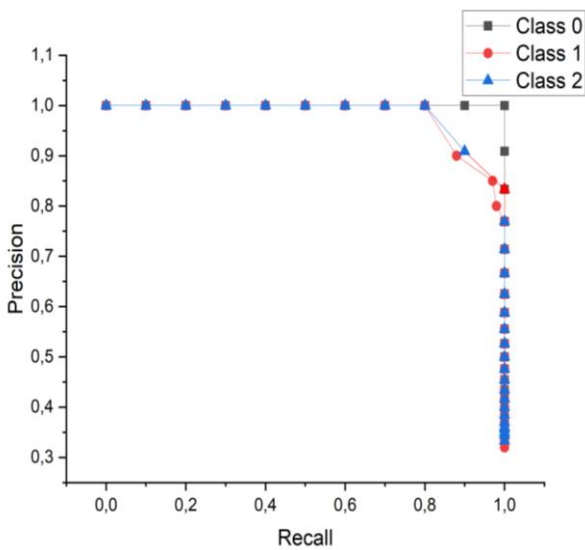

Figure 7. Confusion matrix for RGB

Figure 8. Recall vs. precision curve for RGB

The confusion matrix and classification report show that the model has strong performance, with an overall test accuracy of 93.33%. The model perfectly classified instances of class 0 and class 2, but misclassified 2 instances of class 1 as class 2. Precision, recall, and F1-score for class 0 are all 1.00, while class 1 has a precision of 1.00 but a lower recall of 0.80, resulting in an F1-score of 0.89. Class 2 has a precision of 0.83 and recall of 1.00, with an F1-score of 0.91. Overall, macro and weighted averages for precision, recall, and F1-score are

approximately 0.93, indicating consistent and reliable classification performance. The overall test accuracy achieved by the model for the RGB dataset is illustrated in Figure 9, providing a visual confirmation of the high accuracy obtained.

```
⤓  Test Accuracy : 0.9333333333333333
              precision    recall  f1-score   support

          0       1.00      1.00      1.00        10
          1       1.00      0.80      0.89        10
          2       0.83      1.00      0.91        10

   accuracy                           0.93        30
  macro avg       0.94      0.93      0.93        30
weighted avg       0.94      0.93      0.93        30
```

Figure 9. Test accuracy for RGB

4.2 Results for grayscale

The image augmentation process was applied to a dataset focused on detecting damage levels on surfaces. This augmentation process aims to generate new variations of the dataset by applying transformations. These variations enable the model to better recognize and detect different types of damage under various conditions. The results of the augmentation process for the grayscale dataset are illustrated in Figure 10, showing examples of the generated variations that help improve the model's robustness.

Figure 11 shows the training log of a deep learning model over multiple epochs, focusing on grayscale image processing. The training process was initially set to run for 200 epochs but was halted early at epoch 68 due to early stopping criteria, indicating that further training would not yield significant improvement. By the end of the training, the model achieved a training accuracy of 99.03% with a corresponding loss of 0.027, while the validation accuracy reached 95.83% with a validation loss of 0.1311.

In this process, the results are better with RGB, even though some outcomes are similar. However, during training and validation, RGB showed superior performance compared to other methods.
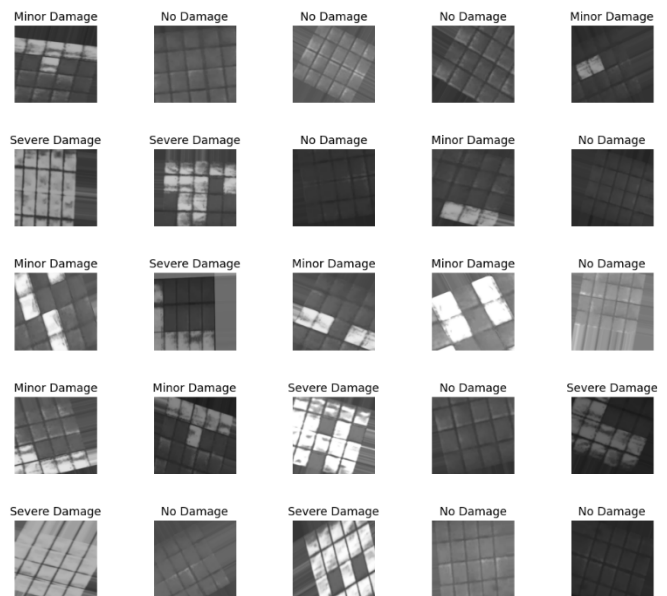

Figure 10. Results augmentation for dataset grayscale

```
  ▶  14/14 [==============================] - 2s 1s/step - loss: 0.0323 - accuracy: 0.9906 - val_loss: 0.1622 - val_accuracy: 0.9792
 ⇄  Epoch 51/200
     14/14 [==============================] - 2s 2s/step - loss: 0.0609 - accuracy: 0.9835 - val_loss: 0.3142 - val_accuracy: 0.9792
     Epoch 52/200
     14/14 [==============================] - 2s 2s/step - loss: 0.1009 - accuracy: 0.9788 - val_loss: 0.4734 - val_accuracy: 0.9479
     Epoch 53/200
     14/14 [==============================] - 2s 2s/step - loss: 0.0576 - accuracy: 0.9906 - val_loss: 0.1051 - val_accuracy: 0.9583
     Epoch 54/200
     14/14 [==============================] - 2s 2s/step - loss: 0.0315 - accuracy: 0.9933 - val_loss: 0.0534 - val_accuracy: 0.9792
     Epoch 55/200
     14/14 [==============================] - 2s 2s/step - loss: 0.0353 - accuracy: 0.9933 - val_loss: 0.0258 - val_accuracy: 0.9896
     Epoch 56/200
     14/14 [==============================] - 2s 2s/step - loss: 0.0501 - accuracy: 0.9882 - val_loss: 0.2425 - val_accuracy: 0.9792
     Epoch 57/200
     14/14 [==============================] - 2s 2s/step - loss: 0.0364 - accuracy: 0.9906 - val_loss: 0.1183 - val_accuracy: 0.9792
     Epoch 58/200
     14/14 [==============================] - 2s 1s/step - loss: 0.0369 - accuracy: 0.9811 - val_loss: 0.1477 - val_accuracy: 0.9688
     Epoch 59/200
     14/14 [==============================] - 2s 2s/step - loss: 0.0337 - accuracy: 0.9953 - val_loss: 0.0735 - val_accuracy: 0.9688
     Epoch 60/200
     14/14 [==============================] - 2s 1s/step - loss: 0.0383 - accuracy: 0.9929 - val_loss: 0.2517 - val_accuracy: 0.9583
     Epoch 61/200
     14/14 [==============================] - 2s 2s/step - loss: 0.0233 - accuracy: 0.9906 - val_loss: 0.2556 - val_accuracy: 0.9583
     Epoch 62/200
     14/14 [==============================] - 2s 2s/step - loss: 0.0268 - accuracy: 0.9955 - val_loss: 0.2793 - val_accuracy: 0.9688
     Epoch 63/200
     14/14 [==============================] - 2s 2s/step - loss: 0.0162 - accuracy: 0.9976 - val_loss: 0.2523 - val_accuracy: 0.9479
     Epoch 64/200
     14/14 [==============================] - 2s 1s/step - loss: 0.0386 - accuracy: 0.9953 - val_loss: 0.1180 - val_accuracy: 0.9792
     Epoch 65/200
     14/14 [==============================] - 2s 2s/step - loss: 0.0220 - accuracy: 0.9929 - val_loss: 0.1787 - val_accuracy: 0.9688
     Epoch 66/200
     14/14 [==============================] - 2s 2s/step - loss: 0.0762 - accuracy: 0.9835 - val_loss: 0.8412 - val_accuracy: 0.9375
     Epoch 67/200
     14/14 [==============================] - 2s 2s/step - loss: 0.0613 - accuracy: 0.9811 - val_loss: 0.0011 - val_accuracy: 0.9896
     Epoch 68/200
     14/14 [==============================] - 2s 2s/step - loss: 0.0827 - accuracy: 0.9693 - val_loss: 0.2931 - val_accuracy: 0.9583
     Epoch 68: early stopping
```

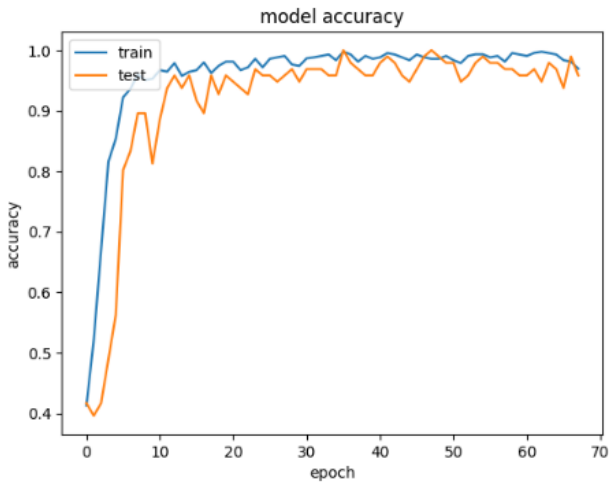**Figure 11.** Training and validation process for grayscale

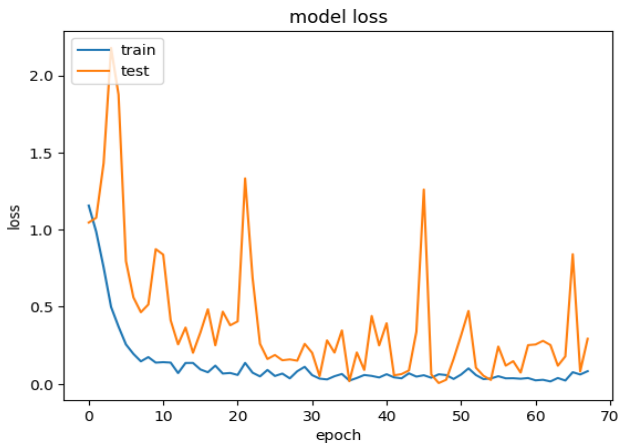**Figure 12.** Accuracy vs. epoch graph for grayscale

Figure 13. Loss vs. epoch graph for grayscale

The graphs illustrate the training and validation (test) accuracy and loss of a deep learning model using grayscale images over 68 epochs. In the accuracy graph, both training and test accuracies show a significant improvement, with training accuracy quickly reaching nearly 100% and test accuracy stabilizing above 95% after the initial epochs. This indicates effective learning and good generalization. The training and validation accuracy trends over the epochs are shown in Figure 12, providing a clear depiction of the model's

progressive learning behavior.

The loss graph shows that training loss decreases steadily, remaining low and stable, suggesting that the model is efficiently minimizing error on the training data. The trend of training and validation loss over epochs is illustrated in Figure 13, highlighting the model's ability to maintain low error rates throughout the learning process.

The confusion matrix and classification report indicate the performance of the model on the grayscale image test set. The confusion matrix, shown in Figure 14, reveals that the model correctly classified all instances of class 0 (10/10) and class 2 (10/10), but misclassified 2 instances of class 1 as class 2, achieving 8 correct classifications for class 1.

The classification report complements these findings, with perfect precision, recall, and F1-score (1.00) for class 0, while class 1 shows high precision (1.00) but a lower recall (0.80), leading to an F1-score of 0.89. Class 2 has a precision of 0.83, perfect recall (1.00), and an F1-score of 0.91.

Further insight into the model's balance between precision and recall for each class can be seen in Figure 15, which displays the recall versus precision curve for the grayscale dataset.

The overall test accuracy for grayscale images, confirming the model's strong classification performance, is illustrated in Figure 16.
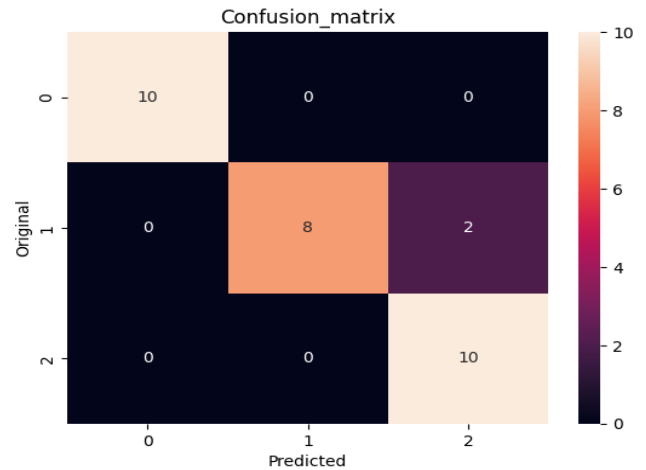

Figure 14. Confusion matrix for grayscale

Figure 15. Recall vs. precision curve for grayscale
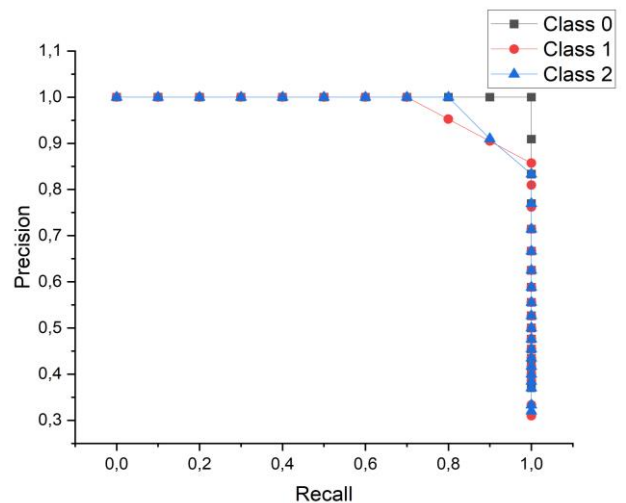
```
    Test Accuracy : 0.9333333333333333
              precision    recall  f1-score   support

           0       1.00      1.00      1.00        10
           1       1.00      0.80      0.89        10
           2       0.83      1.00      0.91        10

    accuracy                           0.93        30
   macro avg       0.94      0.93      0.93        30
weighted avg       0.94      0.93      0.93        30
```

Figure 16. Test accuracy for grayscale

4.3 Results for combine

The results from the augmentation process applied to this combined dataset resulted in a twofold increase in data size, bringing the total to 9,120 images. This significant expansion of the dataset enhances the model's ability to generalize and improve performance across various conditions and scenarios. By creating diverse variations of the original images, the augmentation process strengthens the model's capacity to recognize patterns, detect damage, and handle different levels of severity more effectively. The analyzed outcomes from this augmentation process are illustrated in Figure 17, highlighting the enriched dataset and its potential impact on model performance.
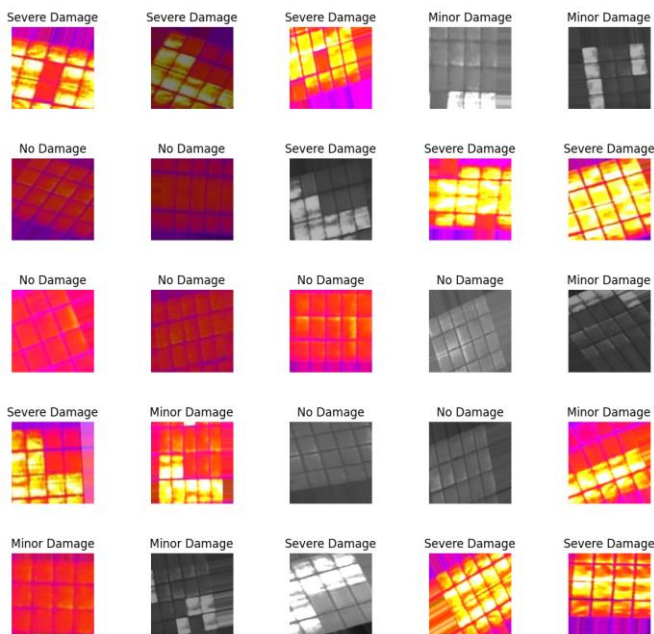

Figure 17. Results augmentation for dataset combine

This process shows the results of combining RGB and grayscale data, highlighting the differences that arise from this integration. Figure 18 shows the training log of a deep learning model over multiple epochs, showing the performance of the model using combined RGB and grayscale data. The training process was initially set for 200 epochs, but early stopping was triggered at epoch 58, indicating optimal performance without further improvement. By the end of training, the model achieved a training accuracy of 98.44% with a training loss of 0.0567, while the validation accuracy reached 97.22% with a validation loss of 0.0555. These results demonstrate the model's strong learning capability and effective generalization to validation data.

The graphs depict the training and validation (test) accuracy

and loss of a deep learning model over 58 epochs, utilizing combined RGB and grayscale image data. In the accuracy graph, both training and validation accuracy steadily improve, with training accuracy approaching 100% and validation accuracy stabilizing around 95-97%. This demonstrates the model's ability to learn effectively and generalize well. The training and validation accuracy trends over epochs for the combined dataset are illustrated in Figure 19.

The loss graph shows a steady decline in training loss, indicating effective optimization and reduced errors in predictions. However, validation loss exhibits some fluctuation, with a general downward trend, suggesting occasional variability in the test data but overall good performance. The training and validation loss patterns are presented in Figure 20, offering additional insight into the model's convergence behavior.

The confusion matrix for the combined dataset (Figure 21) illustrates the classification performance across three classes: Class 0 (No Damage), Class 1 (Minor Damage), and Class 2 (Severe Damage).

The model perfectly classified all instances of Class 0 and Class 2, with 20 correct predictions each. However, for Class 1, out of 20 total instances, 15 were correctly classified, 1 instance was misclassified as Class 0, and 4 instances were misclassified as Class 2. This indicates that the model had some difficulty distinguishing Class 1 from the others, though overall classification performance remains strong.
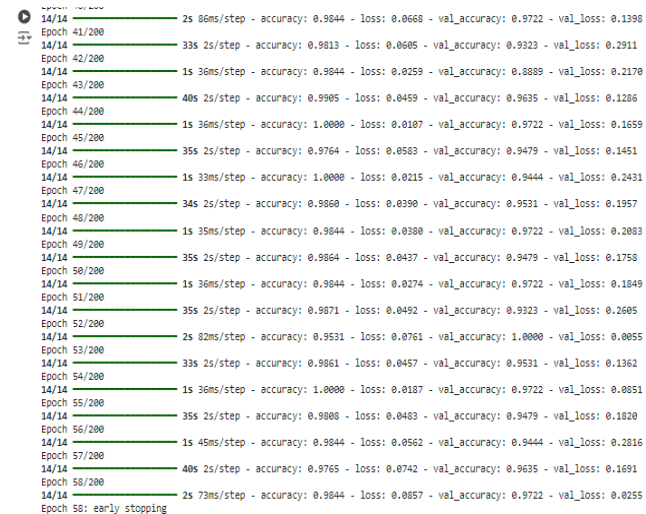

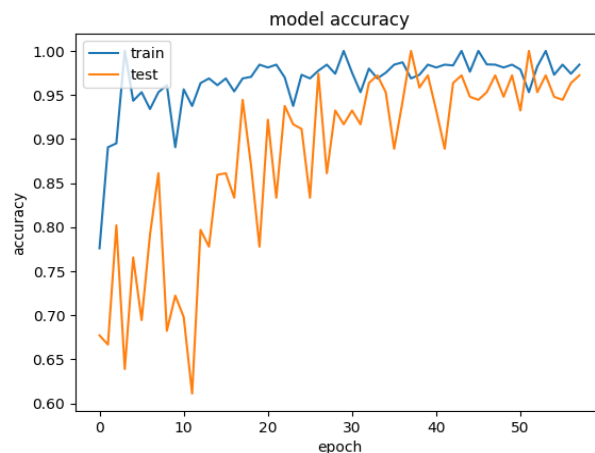**Figure 18.** Training and validation process for combine

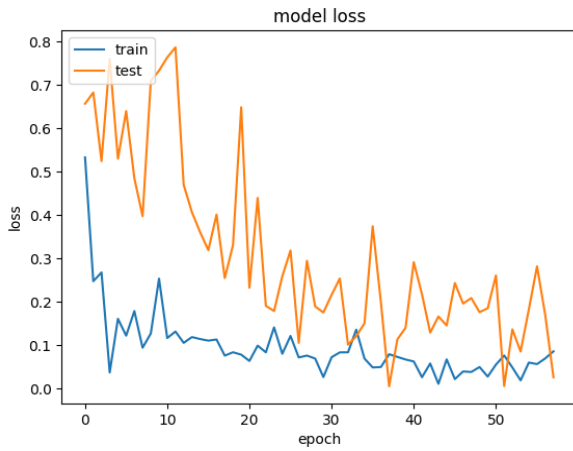**Figure 19.** Accuracy vs. epoch graph for RGB combine
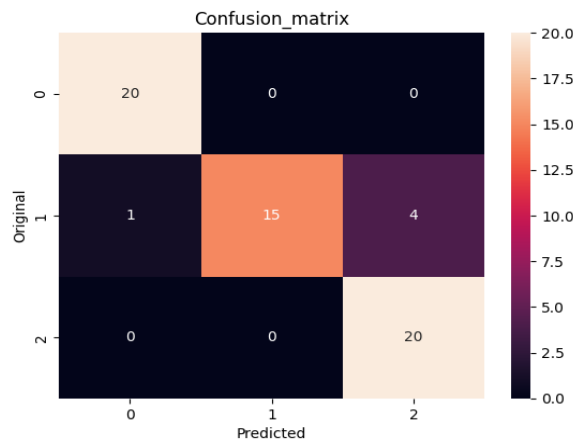
Figure 20. Loss vs. epoch graph for combine
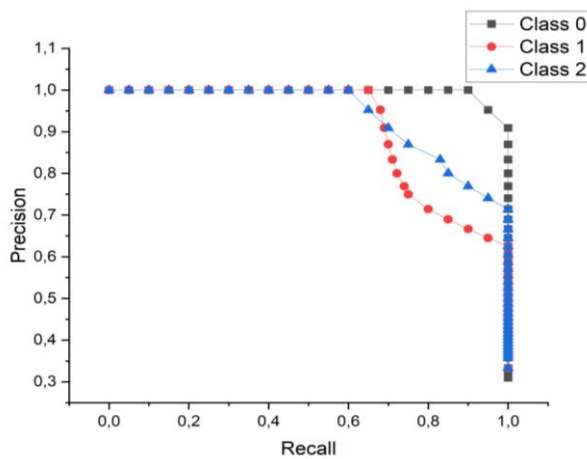

Figure 21. Confusion matrix for combine

Figure 22. Recall vs. precision curve for combine

Figure 23. Test accuracy for combine

This resulted in class 1 having a recall of 0.75, indicating

some difficulty in distinguishing class 1 from others. The overall test accuracy is 91.67%. The classification report shows precision, recall, and F1-score values across classes, with class 0 achieving near-perfect metrics and class 2 also showing high performance. Class 1, despite high precision, has a lower recall, affecting its F1-score. The macro and weighted averages for precision, recall, and F1-score are all around 0.92, reflecting robust overall model performance with consistent reliability across different classes. The balance between precision and recall across the classes for the combined dataset is visualized in Figure 22, through the recall versus precision curve. Meanwhile, the overall test accuracy for the combined dataset is depicted in Figure 23, confirming the high reliability of the model's performance.

4.4 Comparison with baseline models

This section presents a comparative analysis between the proposed Inception ResNet V2 architecture and several baseline models to evaluate its effectiveness in hotspot detection on solar panels. A comparative analysis was conducted to evaluate the performance of Inception ResNet V2 against two baseline models, Inception V3 and Xception, using three types of datasets: RGB, grayscale, and combined. The comparative performance results across these models are summarized in Table 5, highlighting the testing accuracy and performance distinctions observed in the study.
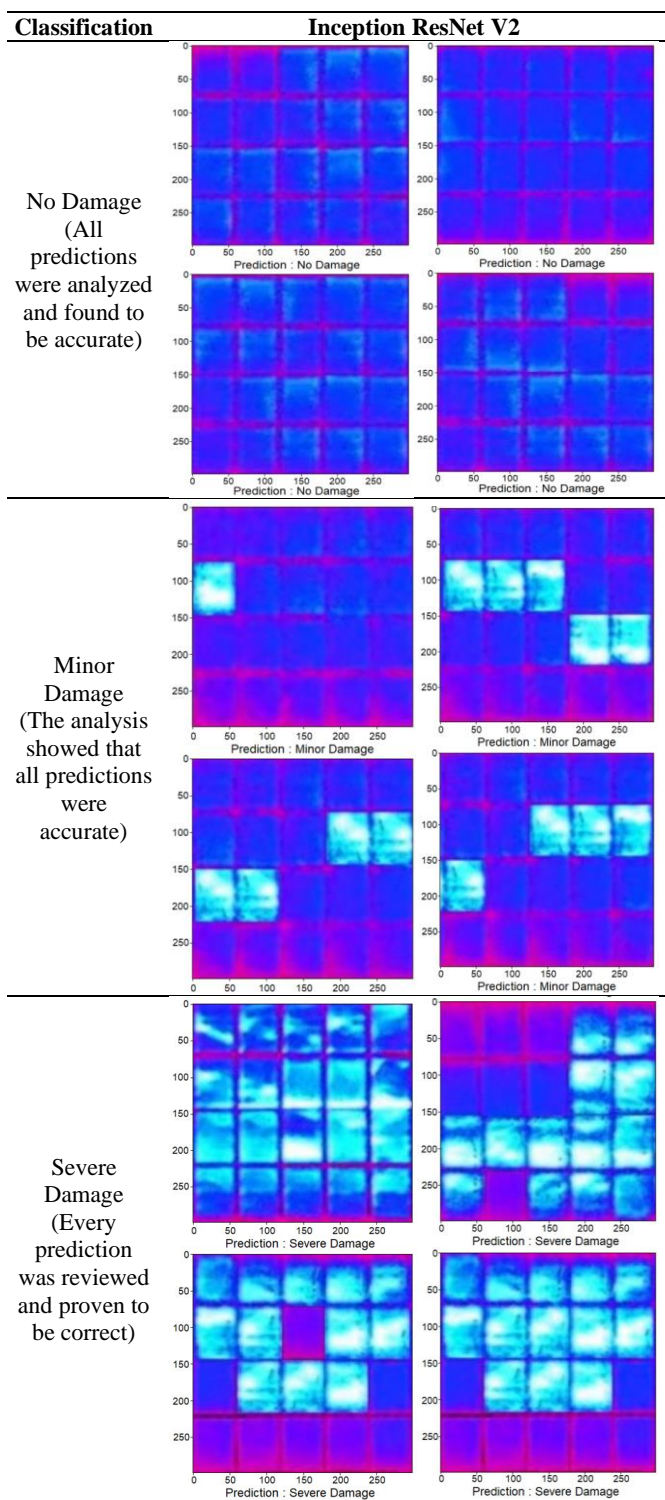
Table 5. Comparison value in baseline models

Architecture	Dataset	Accuracy Value			Remarks
		Training	Validasi	Testing	
Inception V3	RGB	99.29%	96.88%	93.33%	Bestfit
	Grayscale	100%	77.78%	90.00%	Overfit
	Combine	98.52%	78.12%	66.66%	Overfit
Xception	RGB	99.11%	97.92%	80.00%	Overfit
	Grayscale	98.58%	95.83%	93.33%	Bestfit
	Combine	99.09%	97.32%	85.00%	Overfit
Inception ResNet V2	RGB	98.62%	96.88%	93.33%	Bestfit
	Grayscale	96.93%	95.83%	93.33%	Bestfit
	Combine	98.44%	97.22%	91.66%	Bestfit

The results revealed that Inception ResNet V2 achieved the highest performance, particularly on the combined dataset, with a testing accuracy of 91.66%, outperforming Inception V3 (66.66%) and Xception (85.00%). This model also demonstrated consistent results across training, validation, and testing phases, effectively avoiding overfitting, which was observed in other models, especially those trained with grayscale images. Although the model achieved high classification performance, it is important to acknowledge that the dataset consisted of only 570 images per class. This relatively limited data size may affect the generalizability of the model, especially when applied to more diverse or real-world datasets.

Future studies should consider expanding the dataset and incorporating additional data augmentation techniques to enhance the model's robustness and reduce potential overfitting. The use of the combined dataset proved to enhance feature richness, and only the Inception ResNet V2 architecture was able to process it effectively without compromising generalization capabilities. This is further supported by the "Bestfit" remark assigned exclusively to this model, whereas the others showed signs of overfitting. Therefore, Inception ResNet V2 is considered the most reliable and stable architecture for hotspot detection on solar panels based on the observed performance outcomes.

Table 6. Prediction results for RGB dataset

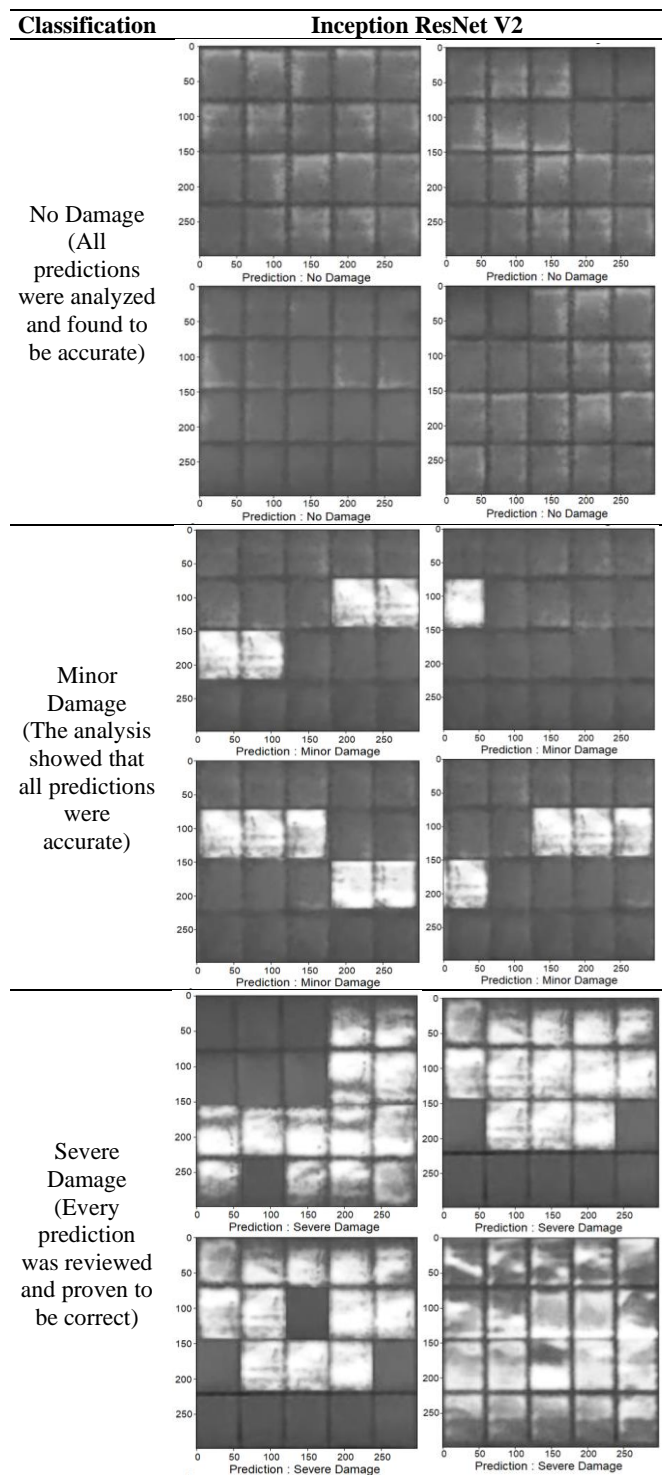| Classification	Inception ResNet V2
No Damage (All predictions were analyzed and found to be accurate)	
Minor Damage (The analysis showed that all predictions were accurate)	
Severe Damage (Every prediction was reviewed and proven to be correct)	

Although the dataset was identical across experiments, performance differences were mainly due to architectural variations, as each model processes features differently. This highlights the crucial role of model design in classification performance. The prediction results for the RGB dataset across models are summarized in Table 6.

Building upon these findings, the next step involves conducting further tests to predict and validate the real-world performance of the proposed architecture. These evaluations are essential for determining the model's ability to generalize and accurately detect the targeted patterns. The results from such tests will provide valuable insights into the overall reliability, robustness, and precision of the model, guiding potential improvements for future implementations. The prediction results specifically for the grayscale dataset are presented in Table 7, offering a detailed assessment of the model's performance on grayscale images.

Table 7. Prediction results for grayscale dataset

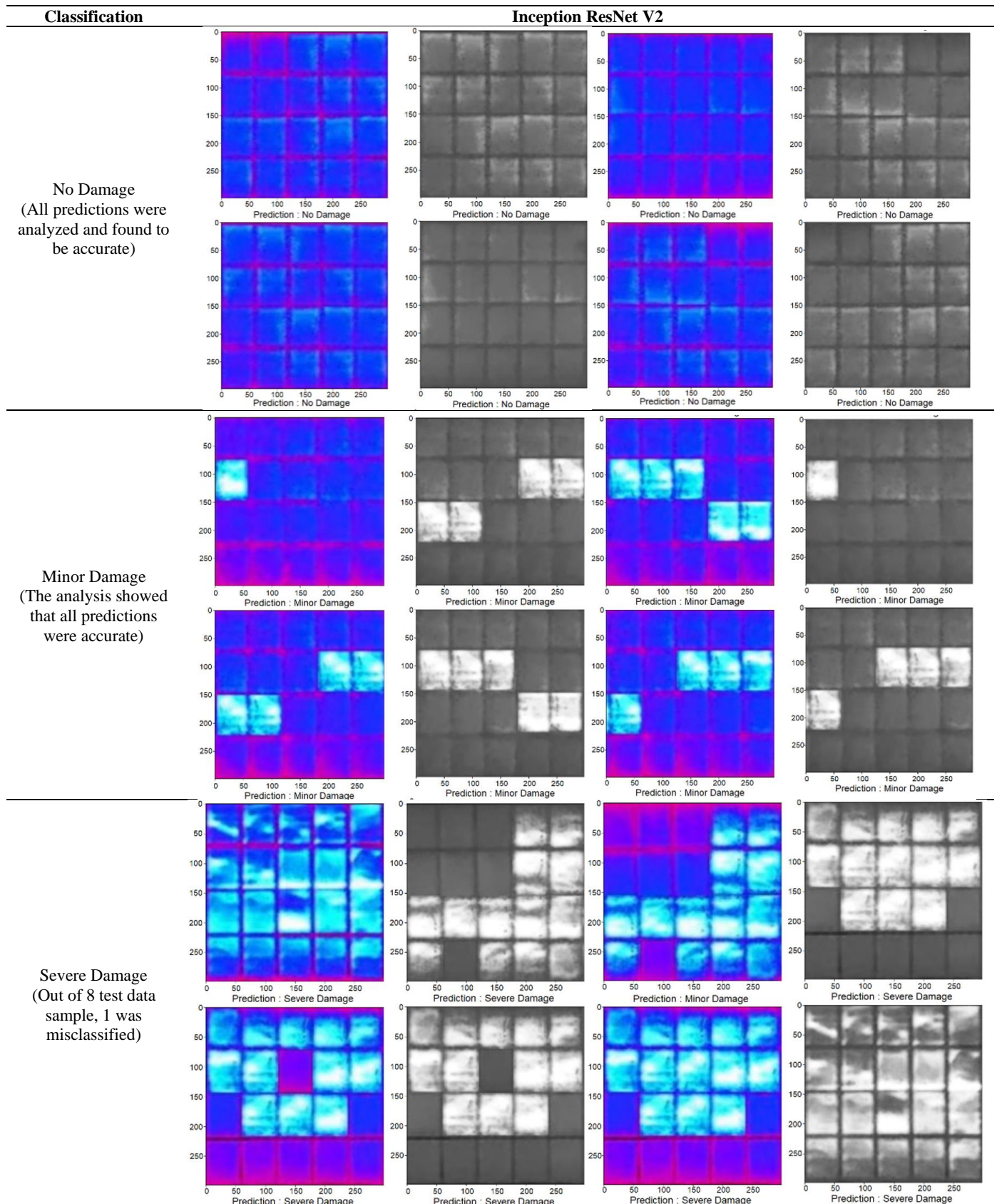| Classification	Inception ResNet V2
No Damage (All predictions were analyzed and found to be accurate)	
Minor Damage (The analysis showed that all predictions were accurate)	
Severe Damage (Every prediction was reviewed and proven to be correct)	

Despite expectations that the combined dataset would enhance model performance by incorporating both RGB and grayscale representations, the results show that this dataset did not consistently outperform the individual formats across all architectures. This may be attributed to potential feature conflicts or redundancies, where overlapping or contradictory information from RGB (color-based features) and grayscale (intensity-based features) can lead to ineffective learning,

particularly in models not optimized for multi-channel input integration. For instance, Inception V3 and Xception exhibited lower testing accuracy with the combined dataset compared to their performance with RGB-only or grayscale-only inputs. This finding suggests that the success of combining multiple image formats depends heavily on the model's ability to process heterogeneous feature types effectively. The prediction results for the combined dataset across the evaluated architectures are presented in Table 8, summarizing the classification outcomes.

Table 8. Prediction results for combine dataset

Classification	Inception ResNet V2			
No Damage (All predictions were analyzed and found to be accurate)				
Minor Damage (The analysis showed that all predictions were accurate)				
Severe Damage (Out of 8 test data sample, 1 was misclassified)				

One of the main reasons is the high visual similarity between panels with minor damage and those with no damage. Minor anomalies such as faint hotspots, micro-cracks, or subtle discoloration often appear nearly identical to normal panels, particularly when viewed under consistent lighting or low contrast conditions. These subtle features are difficult for the model to distinguish, especially when relying on RGB or grayscale images alone, leading to frequent misclassification between Class 0 and Class 1.

Furthermore, Minor Damage represents a transitional state that visually overlaps with both the No Damage and Severe Damage classes. This overlap reduces the separability of features learned by the model, making it harder to assign accurate class boundaries. The high intra-class variability within Minor Damage in terms of the shape, size, and intensity of the defects combined with the inter-class similarity with No Damage, contributes to reduced classification accuracy. Addressing this challenge may require enhancing the dataset with more diverse samples of minor damage, using higher-resolution imagery, or integrating attention-based mechanisms to help the model focus on subtle but relevant local features.

4.5 Statistical analysis of RGB and grayscale dataset performance

To statistically validate the performance difference between RGB and grayscale image datasets, an independent samples t-test was conducted using 37 accuracy values for each group. The mean classification accuracy for the RGB dataset was 98.5224% (SD=1.8996), notably higher than the grayscale dataset, which achieved a mean accuracy of only 93.0176% (SD=12.8013), as shown in Table 9.

Table 9. Independent samples T-Test results

Assumption	t	df	Sig. (2-Tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference
Equal variances assumed	2.587	72	0.012	5.5049	2.1276	[1.2636, 9.7461]
Equal variances not assumed	2.587	37.585	0.014	5.5049	2.1276	[1.1963, 9.8135]

Levene's Test for Equality of Variances indicated a significant difference in variance ($F=13.303$, $p=0.000$), suggesting that the assumption of equal variances is not met. Therefore, the t-test results under "equal variances not assumed" were used. The test showed a statistically significant difference between the two groups with $t=2.587$, $df \approx 37.585$, and $p=0.014$ ($p<0.05$). The mean difference in accuracy was 5.50486%, with a 95% confidence interval ranging from 1.19626% to 9.81347%.

These findings clearly demonstrate that the RGB dataset significantly outperforms the grayscale dataset in hotspot detection accuracy. The lower variance and higher mean accuracy of the RGB group suggest that using RGB imagery results in more stable, reliable, and consistent model performance for solar panel anomaly classification compared to grayscale input.

5. CONCLUSION

This research has demonstrated the effectiveness of the Inception ResNet V2 architecture in detecting and classifying hotspot damage on solar panels by using RGB, grayscale, and combined image datasets. Among the evaluated models, Inception ResNet V2 consistently achieved high performance across all datasets, with the RGB input yielding the most stable and accurate results. The fusion of RGB and grayscale images further enhanced the model's ability to extract both spectral and textural features, enabling more reliable detection of subtle anomalies. However, classification challenges remained for the "Minor Damage" category due to its high visual similarity to the "No Damage" class. Statistical analysis using an independent t-test confirmed a significant performance difference between RGB and grayscale datasets, reinforcing the advantage of using color information in hotspot detection.

Despite these promising outcomes, the study was conducted on a relatively limited dataset consisting of 570 images per format, which may restrict the model's generalizability in more diverse and complex environments. Therefore, future work should involve testing the proposed architecture on larger, more varied datasets to ensure robustness under real-world conditions. Additionally, integrating other modalities such as LiDAR or thermal imaging is recommended to enrich the feature space and improve sensitivity to early-stage physical or thermal anomalies. These enhancements could pave the way for real-time, intelligent monitoring systems that improve operational efficiency and lifespan of solar panel installations.

REFERENCES

[1] Han, S.H., Rahim, T., Shin, S.Y. (2021). Detection of faults in solar panels using deep learning. In 2021 international conference on electronics, information, and communication (ICEIC), Jeju, Korea, pp. 1-4. https://doi.org/10.1109/ICEIC51217.2021.9369744

[2] Karima, N.N., Rimon, K., Molla, M.S., Hasan, M., Bhuyan, M.H. (2023). Advanced image processing based solar panel dust detection system. In 2023 26th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, pp. 1-6. https://doi.org/10.1109/ICCIT60459.2023.10441647

[3] Korkmaz, D., Acikgoz, H. (2022). An efficient fault classification method in solar photovoltaic modules using transfer learning and multi-Scale convolutional

neural network. Engineering Applications of Artificial Intelligence, 113: 104959. https://doi.org/10.1016/j.engappai.2022.104959

[4] Xiao, Y., Huang, X., Liu, K. (2021). Model transferability from ImageNet to lithography hotspot detection. Journal of Electronic Testing, 37: 141-149. https://doi.org/10.1007/s10836-021-05925-5

[5] Mansouri, M., Trabelsi, M., Nounou, H., Nounou, M. (2021). Deep learning-based fault diagnosis of photovoltaic systems: A comprehensive review and enhancement prospects. IEEE Access, 9: 126286-126306. https://doi.org/10.1109/ACCESS.2021.3110947

[6] Latoui, A., Daachi, M.E.H. (2023). Real-time monitoring of partial shading in large PV plants using convolutional neural network. Solar Energy, 253: 428-438. https://doi.org/10.1016/j.solener.2023.02.041

[7] Li, Q., Lin, T., Yu, Q., Du, H., Li, J., Fu, X. (2023). Review of deep reinforcement learning and its application in modern renewable power system control. Energies, 16(10): 4143. https://doi.org/10.3390/en16104143

[8] Li, P., Zhou, K., Lu, X., Yang, S. (2020). A hybrid deep learning model for short-term PV power forecasting. Applied Energy, 259: 114216. https://doi.org/10.1016/j.apenergy.2019.114216

[9] Yousif, H., Al-Milaji, Z. (2024). Fault detection from PV images using hybrid deep learning model. Solar Energy, 267: 112207. https://doi.org/10.1016/j.solener.2023.112207

[10] Gaviria, J.F., Narváez, G., Guillen, C., Giraldo, L.F., Bressan, M. (2022). Machine learning in photovoltaic systems: A review. Renewable Energy, 196: 298-318. https://doi.org/10.1016/j.renene.2022.06.105

[11] Huuhtanen, T., Jung, A. (2018). Predictive maintenance of photovoltaic panels via deep learning. In 2018 IEEE Data Science Workshop (DSW), Lausanne, Switzerland, pp. 66-70. https://doi.org/10.1109/DSW.2018.8439898

[12] Duranay, Z.B. (2023). Fault detection in solar energy systems: A deep learning approach. Electronics, 12(21): 4397. https://doi.org/10.3390/electronics12214397

[13] Herraiz, A.H., Marugán, A.P., Márquez, F.P.G. (2020). Photovoltaic plant condition monitoring using thermal images analysis by convolutional neural network-based structure. Renewable Energy, 153: 334-348. https://doi.org/10.1016/j.renene.2020.01.148

[14] Klaiber, J., Van Dinther, C. (2023). Deep learning for variable renewable energy: A systematic review. ACM Computing Surveys, 56(1): 1. https://doi.org/10.1145/3586006

[15] Chen, H., Pang, Y., Hu, Q., Liu, K. (2020). Solar cell surface defect inspection based on multispectral convolutional neural network. Journal of Intelligent Manufacturing, 31(2): 453-468. https://doi.org/10.1007/s10845-018-1458-z

[16] Segovia Ramírez, I., Das, B., Garcia Marquez, F.P. (2022). Fault detection and diagnosis in photovoltaic panels by radiometric sensors embedded in unmanned aerial vehicles. Progress in Photovoltaics: Research and Applications, 30(3): 240-256. https://doi.org/10.1002/pip.3479

[17] Haidari, P., Hajiahmad, A., Jafari, A., Nasiri, A. (2022). Deep learning-based model for fault classification in solar modules using infrared images. Sustainable Energy Technologies and Assessments, 52: 102110. https://doi.org/10.1016/j.seta.2022.102110

[18] Roumpakias, E., Stamatelos, T. (2022). Health monitoring and fault detection in photovoltaic systems in central greece using artificial neural networks. Applied Sciences, 12(23): 12016. https://doi.org/10.3390/app122312016

[19] Kumar, U., Mishra, S., Dash, K. (2023). An IoT and semi-supervised learning-based sensorless technique for panel level solar photovoltaic array fault diagnosis. IEEE Transactions on Instrumentation and Measurement, 72: 1-12. https://doi.org/10.1109/TIM.2023.3287247

[20] Thakfan, A., Bin Salamah, Y. (2024). Artificial-intelligence-based detection of defects and faults in photovoltaic systems: A survey. Energies, 17(19): 4807. https://doi.org/10.3390/en17194807

[21] Mehta, S., Saini, R. (2024). Enhancing solar panel maintenance: A CNN-GAN approach for dust detection and classification. In 2024 IEEE International Conference on Intelligent Signal Processing and Effective Communication Technologies (INSPECT), pp. 1-5. https://doi.org/10.1109/INSPECT63485.2024.10896001

[22] Juarez-Lopez, J.M., Franco, J.A., Hernandez-Escobedo, Q., Muñoz-Rodríguez, D., Perea-Moreno, A.J. (2023). Analysis of a novel proposal using temperature and efficiency to prevent fires in photovoltaic energy systems. Fire, 6(5): 196. https://doi.org/10.3390/fire6050196

[23] Boubaker, S., Kamel, S., Ghazouani, N., Mellit, A. (2023). Assessment of machine and deep learning approaches for fault diagnosis in photovoltaic systems using infrared thermography. Remote Sensing, 15(6): 1686. https://doi.org/10.3390/rs15061686

[24] Al Mahdi, H., Leahy, P.G., Alghoul, M., Morrison, A.P. (2024). A review of photovoltaic module failure and degradation mechanisms: Causes and detection techniques. Solar, 4(1): 43-82. https://doi.org/10.3390/solar4010003

[25] Kuo, C.F.J., Chen, S.H., Huang, C.Y. (2023). Automatic detection, classification and localization of defects in large photovoltaic plants using unmanned aerial vehicles (UAV) based infrared (IR) and RGB imaging. Energy Conversion and Management, 276: 116495. https://doi.org/10.1016/j.enconman.2022.116495

[26] Al-Otum, H.M. (2024). Classification of anomalies in electroluminescence images of solar PV modules using CNN-based deep learning. Solar Energy, 278: 112803. https://doi.org/10.1016/j.solener.2024.112803

[27] Hanson, E., Elete, T.Y., Nwakile, C., Esiri, A.E., Erhueh, O.V. (2024). Risk-Based maintenance and inspection in energy infrastructure: Future lessons for safety and efficiency. International Journal of Engineering Research and Development, 20(11): 823-844.

[28] Itako, K., Alhabib, A. (2020). A new method of detecting hot spots in PV generation system utilizing AI. IOP Conference Series: Earth and Environmental Science, 581(1): 012006. https://doi.org/10.1088/1755-1315/581/1/012006

[29] Shaik, A., Balasundaram, A., Kakarla, L.S., Murugan, N. (2024). Deep learning-based detection and segmentation of damage in solar panels. Automation, 5(2): 128-150. https://doi.org/10.3390/automation5020009

[30] Chen, H., Pang, Y., Hu, Q., Liu, K. (2020). Solar cell surface defect inspection based on multispectral convolutional neural network. Journal of Intelligent

Manufacturing, 31(2): 453-468. https://doi.org/10.1007/s10845-018-1458-z

[31] Ahmed, W., Ali, M.U., Mahmud, M.P., Niazi, K.A.K., Zafar, A., Kerekes, T. (2023). A comparison and introduction of novel solar panel's fault diagnosis technique using deep-Features shallow-blassifier through infrared thermography. Energies, 16(3): 1043. https://doi.org/10.3390/en16031043

[32] Shafiq, M., Gu, Z. (2022). Deep residual learning for image recognition: A survey. Applied Sciences, 12(18): 8972. https://doi.org/10.3390/app12188972

[33] Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. (2017). Inception-V4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI Conference on Artificial Intelligence, 31(1): 4278-4884. https://doi.org/10.1609/aaai.v31i1.11231

[34] Ioffe, S., Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International Conference on Machine Learning, pp. 448-456.

[35] Goodfellow, I., Bengio, Y., Courville, A. (2016). Deep Learning. The MIT Press.