Available online www.worldscientificnews.com

World Scientific News

An International Scientific Journal

WSN 212 (2026) 186-207 EISSN 2392-2192

A Lightweight Residual Separable Inception Network with Convolution-Only Attention for Explainable Solar Panel Soiling Classification

Olumba Confidence Chigozirim [1], Nneji Uchechi Joyce[2], Olumba Gladys Chinyere[3], Aririguzo Chibueze Favour [4]

[1]Department of Computer Science, Federal College of Education Obudu, Cross River State, Nigeria

[2]College of Humanities and Law, Chengdu University of Technology, China

[3]Department of Environmental Science and Engineering, Chengdu University of Technology, China

[4]Department of Civil Engineering, Lumkalt Energy Limited, Rukpokwu, Port Harcourt, Rivers State, Nigeria.

*Corresponding Author: olumbaconfidence3@gmail.com

https://doi.org/10.65770/CKUL9343

ABSTRACT

This paper deals with the challenge of classifying solar panel images as clean or dusty, at a scale that matters to reduce photovoltaic-related costs and inspection efforts. To solve this challenge in a transparent and computationally efficient way, we propose a novel residual depth-wise separable lightweight Inception-based model created from scratch. The backbone is based on bottleneck (1×1) convolutions and spatially separable (1×3) and (3×1) convolutions grouped in multi-branch Inception style residual blocks, and is also informed by a convolution only attention module that merges bottleneck and spatially separable convolutions to produce feature re-weighting maps without utilizing pre-existing attention frameworks such as SE or CBAM. The model is learned on realistic solar panel dust data with disjoint training, validation, and testing sets. The images are rescaled at a fixed resolution and augmented through geometric transformations so as to enhance generalization capabilities.

(Received 9 December 2025; Accepted 15 January 2026; Date of Publication 12 February 2026)

Quantitative experiments test the trained network on standard classification metrics, with the addition of confusion matrices and precision-recall curves to observe the performance of each class in detail, including accuracy, precision, recall, F1-score and ROC–AUC. Grad-CAM and LIME are embedded as XAI tools to show which areas of the solar panel images lead the model to its predictions. The results, in general, suggest that the lightweight and interpretable network design is capable of capturing discriminative dust related patterns, while the XAI analysis discloses the potential and existing limitations, as well as clear directions on the further enhancement of the methodology.

Keywords: Image classification, lightweight Inception network, depth-wise separable convolution, convolutional attention mechanism.

1. INTRODUCTION

Image classification plays a fundamental role in computer vision and serves as a foundation for many practical applications, such as quality inspection, medical diagnosis, and environmental monitoring. Convolutional neural networks (CNNs), a class of deep learning models, have demonstrated remarkable performance on large-scale image classification benchmarks in the past few years by learning hierarchical feature representations from raw data. Nevertheless, traditional CNN models are usually computationally heavy and act as "black boxes," hindering their application in resource-limited environments and bringing doubts on the reliability of their decisions.

Solar energy systems are vulnerable to environmental conditions, including dust, dirt, and stains that are deposited on the surface of photovoltaic (PV) modules. A moderate amount of soiling is enough to reduce energy production, leading to higher operation and maintenance costs. Manual inspection of large solar farms is labour-intensive and subjective, which motivates the development of automated computer vision systems that can reliably classify solar panel images as clean or dirty. For such systems to be adopted in practice, they must not only achieve reasonable predictive performance but also provide transparent and interpretable explanations of their decisions to engineers and operators.

The need for interpretability has driven increasing interest in explainable AI (XAI) methods and attention mechanisms. XAI methods like Gradient-weighted Class Activation Mapping (Grad-CAM) and Local Interpretable Model-agnostic Explanations (LIME) allow to identify which regions in a image or super-pixels lead the model to a specific prediction, potentially revealing spurious correlations and failure modes. In contrast, attention mechanisms are architectural primitives that physically re-weight feature maps to enable the network to attend more to useful features and less to noisy background patterns. When designed carefully, attention can improve both performance and interpretability, particularly in tasks where discriminative visual cues are localized, as in dust detection on solar panels.

In this paper, we design and implement from scratch a novel residual depth-wise separable lightweight Inception model for the specific application of solar panel dust image classification. The backbone is built solely by a sequence of bottleneck (1×1) and spatially separable (1×3) and (3×1) convolutions within multi-branch inception-style residual blocks. Based on these blocks, we design a novel, all Convolution-based attention module, which combines bottleneck and spatially separable convolutions to compute features re-weighting maps, discarding prior mechanisms like Squeeze-and-Excitation or CBAM.

The model is trained and tested on a real-world solar panel dust dataset with conventional train–validation–test splits and geometric data augmentation, and its property is further explored with Grad-CAM and LIME to offer visual explanations for right and wrong predictions.

A full end-to-end training and evaluation pipeline based on the solar panel dust dataset is built, including data preprocessing, splitting, model training with early stopping and learning rate scheduling, evaluating quantitatively with accuracy, precision, recall, F1-score, confusion matrices, and ROC–AUC. The integration of Grad-CAM and LIME into the analysis workflow to generate visual explanations that reveal which regions of solar panel images drive the model's decisions, thereby improving the interpretability and trustworthiness of the proposed system.

The remainder of this report is structured as follows. Section 2.0 Review of the Literature: this section presents the related work on image classification, ensemble models, mechanism of attention, and XAI. Section 3.0 outlines the proposed methodology, including detailed descriptions of the dataset, preprocessing operations, custom model architecture, attention mechanism, and the application of XAI methods. Section 4.0 reports the experiment outcomes, including quantitative results and visual interpretations based on XAI, and also discusses the strengths and limitations of the model. This report is concluded in Section 5.0 with a summary of its key findings and the the major limitation.

2. RELATED WORK

Dust deposition on the surface of solar panels is a common problem that can lead to significant energy loss – efficiency reduction of 30–40% and power loss of up to 86% in extreme drying conditions have been reported. This has led to the development of several investigations on the detection of dust and cleaning optimization for photovoltaic (PV) modules. Initial methods have been developed for panel cleaning decision-making from non-visual information using machine learning. For instance, a regression-tree-based dust estimation unit for predicting the dust level from environmental sensors (irradiance, temperature) and output power of the panel was introduced by Shaaban et al. (2020)[1] A dust accumulation estimation based on the panel output, irradiance, and temperature as input was also presented by Mokhtar et al. (2022)[2] which is an ANN system that estimates dust buildup and alarms for cleaning when a certain threshold is crossed. These data-driven approaches have shown that smart scheduling can decrease energy loss by ~60% and maintenance overhead by ~80% over routine cleaning. However, they require physical sensors and cannot determine at what depth in the panel dirt is attached, so they are not able to visually classify or locate dust on the panels.

Hence, most recent works rely on image-based dust detection on solar panels using CNNs, allowing for automatic visual analysis of panel cleanliness. Alçin et al. (2025)[3] proposed a novel lightweight CNN, SolPowNet tailored for the binary classification of solar panel images (clean vs dusty). Their model was trained on 842 panel images (502 clean, 340 dusty), and outperformed standard deep networks such as AlexNet, VGG16/19, ResNet50, InceptionV3 using the same dataset by achieving a classification accuracy of 98.8%. SolPowNet's architecture is optimized for complexity, with only ~11.17 million parameters – relatively tiny compared to typical off-the-shelf CNNs – thus facilitating real-time embedded application. This shows that a tasked CNN can be both accurate and computationally light for solar panel dust detection. One constraint is that SolPowNet was tested on a moderate-sized proprietary dataset; hence, its extrapolation to wider scenarios (different lightings, panel types) is not secured.

Furthermore, although that work achieves high accuracy, it does not have an option for the user to ascertain why the model made the predictions that it does (we think of it as a black-box classifier with no visual explanations to the user).

Another study by Alatwi et al. (2024)[4] employed a pre-trained CNN models for an image classification perspective. Instead of introducing a novel architecture, they used 20 state-of-the-art CNN architectures (e.g. VGG, ResNet, MobileNet, DenseNet, etc.) as fixed feature extractors and trained an SVM classifier to tell clean vs dusty panels. Their dataset consisted of 1,068 images (405 clean, 663 dirty) collected from public sources in diverse conditions. The best accuracy is 86.79 % based on the features of DenseNet-169 with a linear SVM. This two-stage scheme (deep CNN features + SVM) substantiates that CNN-learned image representations are suitable for the application. However, the accuracy straightened out below 90%, suggesting that it could be improved – perhaps because they didn't fine-tune the CNNs on the solar panel dataset (some useful information may be lost by using a separate classifier). The approach is also relatively heavy since the best DenseNet is a huge model, which contradicts the lightweight deployment demand. On the plus side, Alatwi et al. – did consider model explainability: they used the LIME method to determine which part of the images affected the predictions for "dusty" vs "clean." This generated visual signs (highlighted panel patches) demonstrating where the model has "seen" dust in the panel, a useful progression toward interpretable AI in this field. Their contribution demonstrates the advantage of explainability, although it did not include an attention mechanism directly in the model - the explanation is post-hoc.

For monitoring panel soiling in the case of large-scale solar farms, Unmanned Aerial Vehicles (UAVs) with camera serve as a feasible solution. Gao and Li (2023)[5] introduce a deep learning approach for duster detection on PV panels based on UAV images using an enhanced YOLOv5 object detection model. Their adaptations tailored YOLOv5 to the application: a novel additional detection head to address the huge scale disparities of dust patches when drones fly over at different altitudes, among other custom "tricks" to enhance detection of dust spots on largescale images. The final model is pretty lightweight and real-time as it runs on a normal CPU for in situ analysis. In the experiments, their improved YOLOv5 shows a better performance on the standard model in terms of detection accuracy and F1-score and also increases the inference speed. It helps show how targeted network modifications can tune a general object detector to the needs of dust detection. A drawback is that the method yields bounding boxes of detected dust, but it does not explicitly infer the overall cleanliness condition of a panel; it's just focused on finding dust clusters. Also, like most object detectors, it does not natively provide an explanation for why a region is classified as dust other than by bounding the box. Still, the work of Gao and Li is significant in demonstrating that UAV-based inspection can be made automatic with efficient CNNs, and it underscores the importance of multi-scale attention to tiny objects (dust) in high-resolution images.

In a related vein, Naeem et al. (2025) [6] propose SDS-YOLO, an attention-based YOLO model to detect solar panel soiling in aerial images. Their system is designed for two prevalent soiling types – dust film and bird droppings – which have differing visual textures. To address the challenge of small defect detection, they introduced into the CNN a Convolutional Block Attention Module (CBAM), and proposed two dedicated detection heads, one for the dispersed dust region and one for small concentrated droppings. This multi-target, attention-enhanced strategy led to substantial improvements: the CBAM-based model increased mean average precision and F1-score by around 40% and 26%, respectively, for bird-dropping detection as compared with baseline YOLO, and also gave a slight improvement on dust detection.

Remarkably, these improvements in accuracy were achieved with 24% fewer model parameters after pruning unnecessary layers, making SDS-YOLO more applicable to edge devices. The attention module usage is crucial here, since narrowing down the CNN to focus on specific regions/features, helped reduce false positives/separates them better with small soiling features. This demonstrates how tailored attention-based methods can significantly improve the CNN trustworthiness for PV panel inspection. Nevertheless, similar to Gao's approach, this method is designed as an object detection problem with a rather complicated pipeline. It detects and locates soiling spots, but it does not give a simple yes/no cleanliness label for the entire panel, and it is not explicit in generating explanations that are friendly to humans (the attention is internal within the model). There remains a gap to converge such attention-guided accuracy with user-interpretable outputs for the end users.

In summary, existing literature shows a clear evolution from sensor-based, data-centric approaches that trigger cleaning decisions from electrical and environmental measurements to image-based deep learning models that directly detect dust on PV panels. Image-centric CNN methods, especially custom architectures like SolPowNet and pre-trained backbones combined with SVMs, achieve much higher accuracy in distinguishing clean and dusty panels than earlier non-visual methods. UAV-based detectors further extend this idea by localizing soiling patches in aerial images and optimizing YOLO-style architectures for efficiency on drones. However, these works either rely on heavy pre-trained networks, focus on detection rather than simple binary classification, or treat attention and features as internal black boxes with limited use of explainability tools. In contrast, this project targets a from-scratch, ultra-lightweight CNN built only from efficient convolutions (1×1 and separable layers) with an explicit convolution-only attention module and couples it with Grad-CAM and LIME to visualize how the model focuses on soiling. In doing so, it aims to combine the deployment advantages of lightweight architecture with stronger transparency and interpretability, providing a compact, explainable AI solution for solar panel dust detection that can support trustworthy maintenance decisions.

3. MATERIAL AND METHOD

3.1. Dataset

To build and evaluate the proposed model, we use a solar panel dust detection image dataset organised for a binary classification task. The goal is to discriminate between clean and dirty PV modules using RGB images acquired in real operating conditions. This configuration is selected to simulate realistic inspection scenarios and to verify the ability of proposed custom network to learn discriminative features between dust and stain accumulation.

The dataset is located in a local directory, organized in a way that is compatible with the Keras ImageDataGenerator API. The high-level data is split into two main folders: a train folder to build the training and validation sets, and a test folder, as an independent test set. Each of these two directories contains two subdirectories, clean and dirty, with the images from each class. This organisation allows class labels to be inferred directly from folder names, ensuring a simple and reproducible loading process.

During preprocessing, all images are resized to a fixed resolution of 224×224 pixels with three colour channels to match the input requirements of the proposed convolutional neural network. Pixel values are normalised to the range [0,1] by dividing by 255.0.

After the automatic splitting procedure described in Section 3.2, the resulting subsets contain 539 images for training, 134 images for validation, and 169 images for testing, with both clean and dirty samples present in each subset. Some randomly selected examples of clean and dirty solar panel images can be visualised to inspect the variety of lighting conditions, viewpoints, and dust patterns.

3.2. Proposed Model

The presented model is an original residual depth-wise separable lightweight Inception network crafted from ground up to the task of classifying dust on solar panels. The design follows the coursework constraint of only using bottleneck 1×1 convolutions and spatially separable convolutions in the form of 1×3 and 3×1 kernels, as described in the lectures. Unlike conventional approaches, no pre-trained backbone is exploited, and all weights are trained from scratch on the solar panel dataset.

The network can be conceptually divided into three main components: a stem block, a stack of Inception-style residual blocks with attention, and a classification head. The stem block consists of two spatially separable convolutions followed by max pooling. These layers quickly reduce the spatial resolution and extract low-level features from the input 224×224×3 RGB images.

The core of the network is formed by three lightweight Inception residual blocks. Each block uses only the basic building elements defined in the notebook: batch normalisation followed by ReLU activation, bottleneck 1×1 convolutions, and spatially separable 1×3 and 3×1 convolutions. Within each block, the input feature map is processed by multiple parallel branches: one branch applies a single 1×1 convolution, a second branch applies a 1×1 bottleneck followed by one spatially separable convolution, and a third branch applies a 1×1 bottleneck followed by two stacked spatially separable convolutions.

A second branch applies a 1×1 bottleneck followed by one spatially separable convolution, and a third branch applies a 1×1 bottleneck followed by two stacked spatially separable convolutions. The outputs of these branches are concatenated along the channel dimension and passed through another 1×1 bottleneck convolution to control the number of channels. A residual shortcut connection is then added, using either identity mapping or a 1×1 projection if the number of channels or the spatial resolution does not match. This architecture is Inception-based in that it performs multi-branch feature extraction but also uses residual connections to enable better flow of gradients.

To further enhance the model's focus on dust-related regions, a custom convolution-only attention block is attached after each Inception residual block. This attention module is also implemented exclusively with bottleneck 1×1 convolutions and spatially separable convolutions. Given an input feature map, the module first reduces the number of channels with a 1×1 convolution, applies batch normalisation and ReLU activation, and then processes the reduced features with a spatially separable convolution to capture local spatial dependencies. Another 1×1 convolution restores the original channel dimension, and a sigmoid activation produces an attention map with values between 0 and 1. The attention map is multiplied element-wise with the original input feature map to generate re-weighted features in which informative locations and channels are emphasised and less relevant ones are suppressed. Importantly, this mechanism does not rely on existing attention designs such as Squeeze-and-Excitation or CBAM and thus fully satisfies the requirement of being a custom, lecture-compliant attention design.

After the last attention block, the feature maps are pooled by a Global Average Pooling layer, which reduces the spatial dimensions to a single feature vector for each channel. A dropout layer with dropout rate of 0.4 is applied to prevent overfitting by randomly setting to zero a fraction of features in training. Then, the probability that the input image is a dirty image is outputted by a fully connected layer with a single neuron and sigmoid activation. The proposed model has about 504, 161 trainable parameters, which makes it lightweight for possible usage in mobile/edge devices.

In addition to the core architecture, the notebook integrates Grad-CAM and LIME as post hoc explainable AI modules. Grad-CAM generates class-specific activation heatmaps by backpropagating gradients from the output to the last convolutional layer, while LIME approximates the model's behaviour around a particular input with a simple interpretable model using perturbed super-pixels. These tools are applied to selected test images to visualise which regions of the solar panels most strongly influence the model's predictions.

3.3. Evaluation Strategy

The performance of the proposed model is assessed using a combination of quantitative classification metrics and qualitative visual explanations. All quantitative metrics are computed on the independent test set using the predictions produced by the trained network.

An example has been provided below:

Accuracy (ACC) measures the overall proportion of correctly classified images and is defined as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision (P) for the positive (dirty) class quantifies how many images predicted as dirty are truly dirty:

$$P = \frac{TP}{TP + FP}$$

Recall (R) or sensitivity measures the proportion of truly dirty panels that are correctly identified:

$$R = \frac{TP}{TP + FN}$$

F1 − score (F1) is the harmonic mean of precision and recall and is given by:

$$F1 = \frac{2 \cdot TP}{2 \cdot TP + FP + FN}$$

A confusion matrix is also computed to provide a detailed view of how many clean and dirty images are correctly classified or misclassified. This matrix directly shows the distribution of TP, TN, FP, and FN and allows a more fine-grained analysis of the model's behaviour on both classes.

Furthermore, the Receiver Operating Characteristic (ROC) curve and its corresponding Area Under the Curve (ROC–AUC) are used to evaluate the model's discriminative ability over different decision thresholds. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR), where

$$TPR = \frac{TP}{TP + FN}, \quad FPR = \frac{FP}{FP + TN}$$

The ROC–AUC is then computed numerically from the ROC curve and summarises the trade-off between TPR and FPR.

A Precision–Recall (PR) curve is also plotted by computing precision and recall at various thresholds on the predicted probabilities for the dirty class. This curve is especially useful for imbalanced class distributions or when the positive class is more important to predict accurately than the overall accuracy.

These metrics are calculated in the notebook by evaluating the model on the test generator to get the test loss and accuracy, then getting predicted probabilities and labels to create a confusion matrix, a detailed classification report (precision, recall, F1-score), and finally ROC and PR curve plots with ROC–AUC. Qualitative results: Grad-CAM and LIME are used to produce explanations on a set of representative test images, which allows us to visually interpret which areas the model focuses on to make its predictions.

3.4. Environment Execution

All the experiments in this work have been carried out on a personal computer with Windows 11 operating system, and a Python environment managed by Anaconda was used. The implementation was in Python with TensorFlow 2.x as the Keras high-level API. Also, NumPy was used for numerical calculations, scikit-learn to calculate classification metrics and curves, matplotlib to generate plots of training histories and evaluation curves, LIME, and scikimage for producing and visualizing local explanations.

The hardware environment was a multi-core processor, and the system memory was large enough for the mini-batch training using the batch size of 16. We did not use a dedicated GPU; training and inference were done on the CPU. With this setup, training the custom lightweight Inception model for up to 50 epochs using early stopping took about 20 minutes (wall-clock time), as the notebook shows. Although there is no GPU acceleration, the relatively small number of parameters (around 0.5 million) allows this model to be trained and tested within reasonable time, which validates that the proposed architecture is suitable for implementation in execution environments with limited resources.

4. EXPERIMENTAL RESULTS

In this section, the results of the experiments conducted on the solar panel dust image dataset using the proposed residual depth-wise separable lightweight Inception module are discussed. The model is trained from scratch on the training and validation splits described in Section 3 and evaluated on the unseen test split.

Training is done for a maximum of 50 epochs with early stopping after 19 epochs when the validation loss no longer improves using the Adam optimizer. Figure 1 illustrates the progression of training and validation loss and accuracy. The training accuracy quickly rises from 0.90 to nearly 0.98 and stays high, while the training loss also rapidly drops from about 0.29 to under 0.05.

On the other hand, the validation curves are significantly more volatile: the validation accuracy oscillates between about 0.37 and 0.72, the validation loss decreases at first but then it suddenly shoots up after epoch 13. This suggests the network may be fitting the training data well, but the small size of the dataset may lead to overfitting, which supports the use of early stopping and learning-rate reduction.

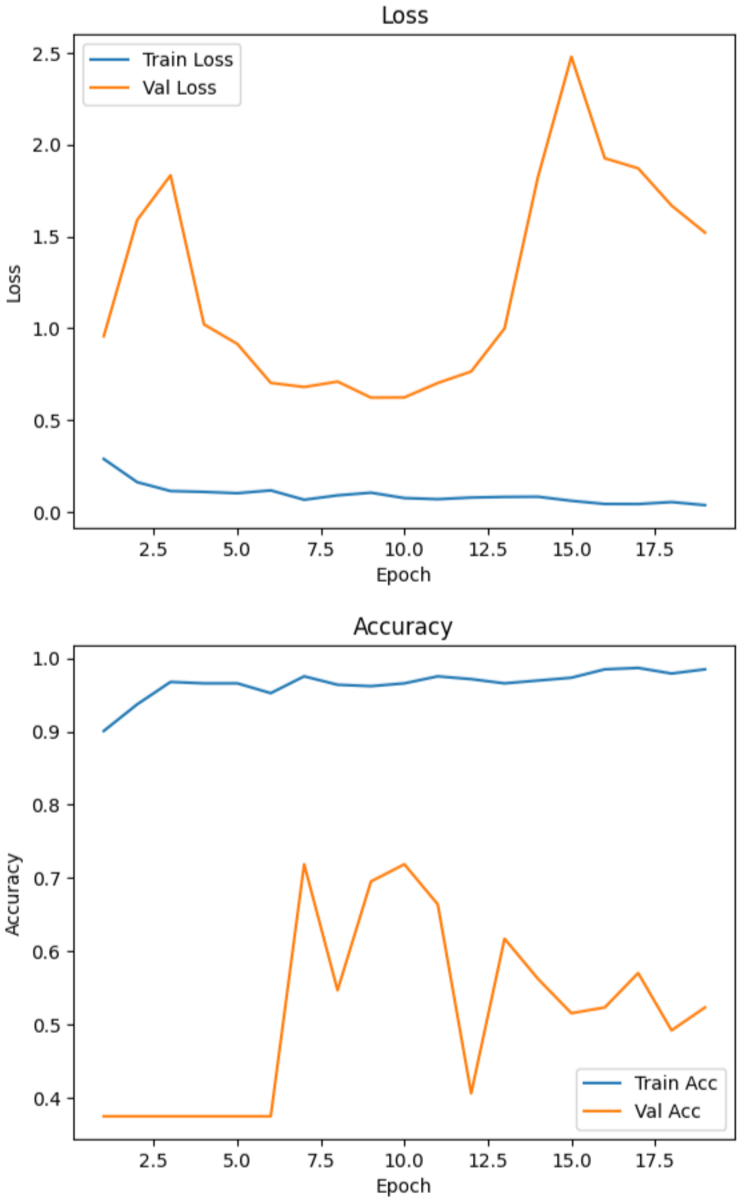


Figure 1. Train/validate the loss & accuracy curve

After training, the best model checkpoint according to the validation performance is selected and evaluated on the test set. The Keras evaluate function reports a test loss of 0.9459 and a test accuracy of 0.4911, meaning that at the default decision threshold of 0.5 the classifier correctly labels slightly less than half of the test images. To obtain a more detailed view of the behaviour on each class, we compute additional metrics and construct the confusion matrix, ROC curve, and precision–recall (PR) curve, as described in Section 3.3.

4.1. Performance Results Using the Dataset

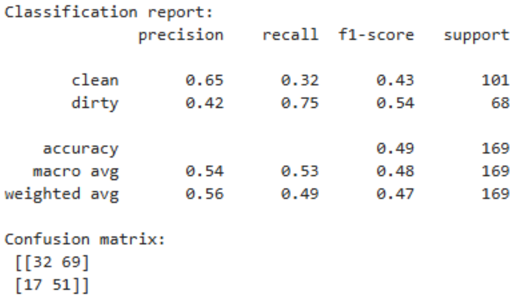
Provide the performance result of your work. An example has been provided below:

The quantitative performance of the proposed model on the test subset (169 images) is summarised in Table 1. Here, the positive class is dirty, and the negative class is clean. From the confusion matrix, the numbers of true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are TP = 51, TN = 32, FP = 69, FN = 17. Using these values, the overall accuracy, precision, recall, specificity, and F1-score are computed according to the formulas in Section 3.3. The ROC–AUC is obtained from the ROC curve.

Table 1. Performance of the proposed model on the solar panel dust test set.

Metric	**Value(%)**
Accuracy	**49.11**
Precision (dirty class)	**42.50**
Recall / Sensitivity	**75.00**
Specificity (clean class)	**31.68**
F1-score (dirty class)	**54.26**
ROC - AUC	**71.12**

The corresponding confusion matrix is visualised in Figure 2. Among the 101 clean panels, only 32 are correctly classified as clean, while 69 are misclassified as dirty. Among the 68 dirty panels, 51 are correctly classified, and 17 are incorrectly labelled as clean. This distribution matches the class-wise recall reported by the classification report: recall for the clean class is 0.32, whereas recall for the dirty class reaches 0.75.



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

       clean       0.65      0.32      0.43       101
       dirty       0.42      0.75      0.54        68

    accuracy                           0.49       169
   macro avg       0.54      0.53      0.48       169
weighted avg       0.56      0.49      0.47       169

Confusion matrix:
 [[32 69]
 [17 51]]
```

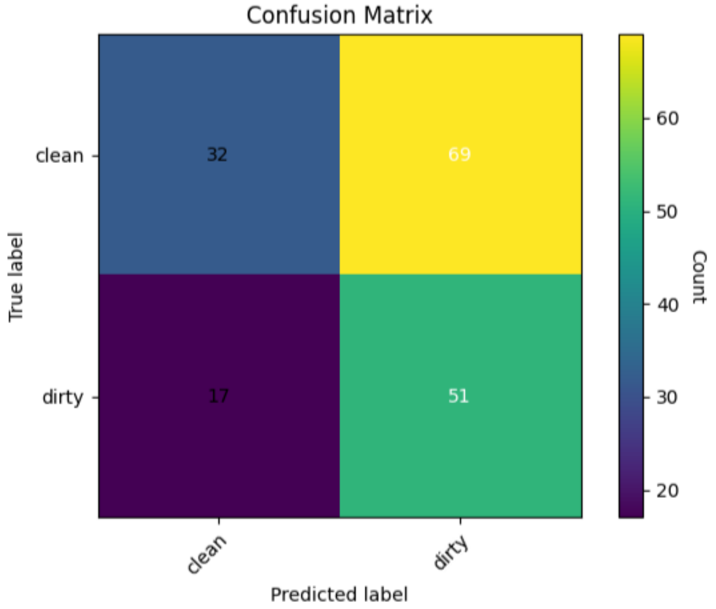


Figure 2. Confusion matrix of the proposed model on the solar panel dust test set.

To assess threshold-independent discriminative ability, the ROC curve and precision–recall curve are plotted in Figure 3 and 4, respectively. The ROC curve in Figure 3 yields an area under the curve (ROC–AUC) of 0.7212, showing that the model is able to rank dirty panels ahead of clean ones substantially better than random guessing, even though the fixed-threshold accuracy is modest.

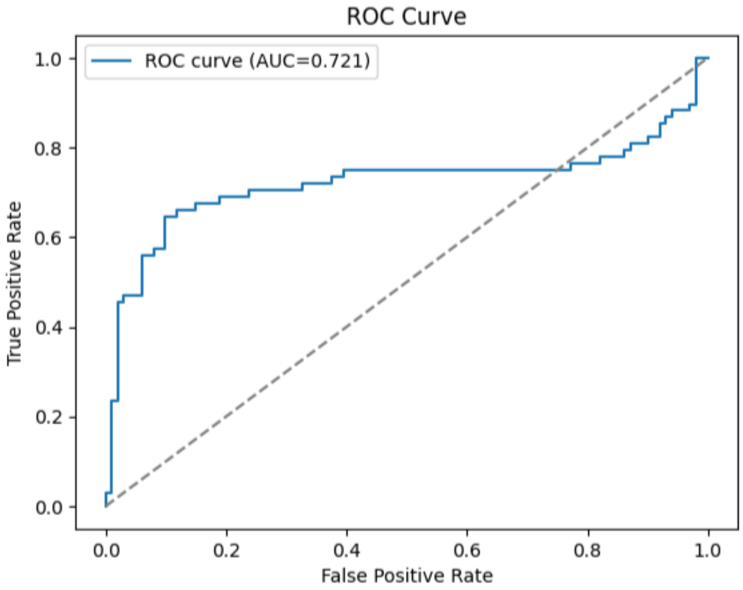


Figure 3. ROC curve.

The precision–recall curve for the dirty class in Figure 4 illustrates the trade-off between precision and recall as the decision threshold varies. At low thresholds, recall is high, but precision drops below 0.5, reflecting many false positives (clean panels predicted as dirty). As the threshold increases, precision improves at the cost of reduced recall. This behaviour is consistent with the confusion matrix analysis, where the model favours detecting dirty panels but struggles to avoid mislabeling clean ones.

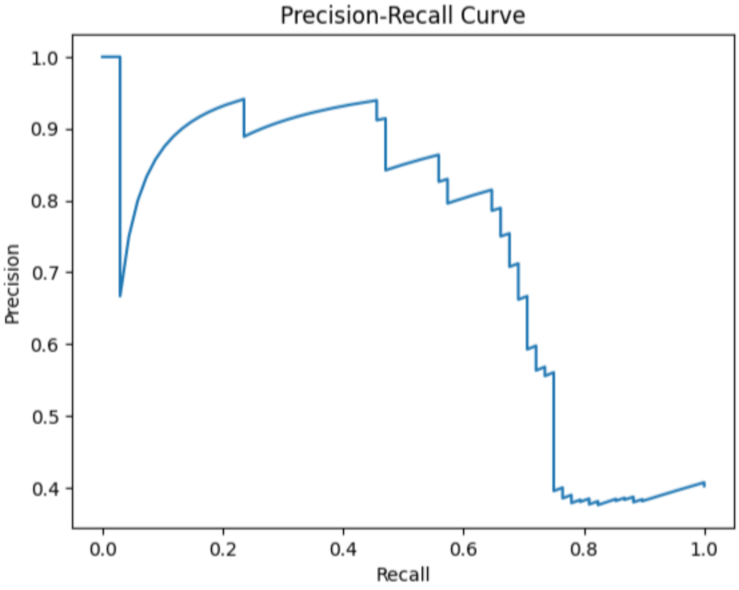


Figure 4. Precision–recall curve.

Taken together, Table 1 and Figures 4–6 show that the proposed lightweight model has a meaningful ability to distinguish dirty panels from clean ones in terms of ranking (ROC–AUC), but its operating point at a 0.5 threshold leads to many false alarms, which depresses accuracy and specificity.

4.2. Discussion

This subsection provides an in-depth discussion of the obtained results and analyses the behavior of the proposed model using both quantitative metrics and explainable AI visualizations.

Before everything else, the training curves in Figure 1 clearly exhibit a strong unfairness towards very high training accuracy (≈0.98) on the one hand, and significantly lower and unstable validation accuracy (0.37–0.72) on the other. The training error continually decreases to below 0.05, but the validation error starts increasing, indicating that the network is overfitting on the training data even though data augmentation is applied. This is expected since the custom-designed architecture is relatively large compared to the dataset. Early stopping/lr scheduling policies can alleviate vel but not completely remove such overfitting.

The test results in Table 1 confirm this limitation. While the ROC–AUC of 72.12% shows that the model has learned some useful discriminative structure, the overall accuracy of 49.11% and the low specificity of 31.68% demonstrate that the classifier frequently mislabels clean panels as dirty. From a maintenance perspective, such behaviour leads to a conservative system that rarely misses truly dirty panels (recall 75.00%) but generates many false positives and, consequently, unnecessary cleaning operations. Depending on the application scenario, this trade-off might or might not be acceptable; it suggests that threshold tuning or cost-sensitive training could be explored in future work.

The asymmetry is further highlighted by the confusion matrix in Figure 2 and the PR curve in Figure 4. Since the model over-classifies the dirty class, precision is moderate even for high recall values. This behavior is more noticeable in the tail of the PR curve at recall values close to 1.0 with precision dropping to less than 0.4. To achieve a more balanced performance, extra regularization and/or further architectural simplifications may be needed in conjunction with gathering more clean panels with challenging lighting and background conditions environment samples.

To get a sense of the model's decisions, Grad-CAM and LIME are run on a few test images in each class. Figure 5 displays a Grad-CAM for the true positive dirty panel. The heatmap of the last convolutional layer focuses on the central module cells, which contain visible dust and darker lines, and the overlay shows that these areas have the highest contribution to the dirty prediction. This coincidence of the highlighted regions with the dust locations as perceived by humans indicates that in correct predictions, the model is frequently basing its decision on useful visual evidence.

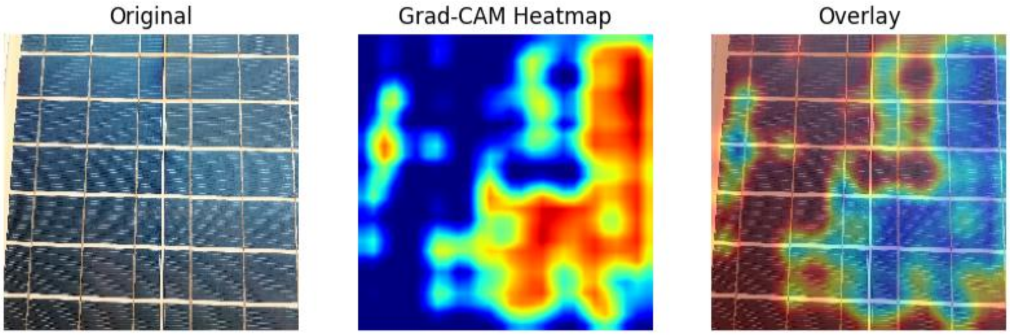


Figure 5. Grad-CAM triptych.

Figure 6 presents a LIME explanation for another example. The super-pixels outlined in yellow correspond to regions that LIME identifies as having a strong positive contribution to the dirty prediction. These regions largely coincide with darker cells and structural patterns on the panel surface. For correctly classified dirty panels, this again suggests that the model focuses on genuinely soiled areas. However, inspection of misclassified clean images (not shown here) reveals that LIME sometimes highlights super-pixels along high-contrast edges, reflections, or shadows that superficially resemble dust, indicating that the model can be misled by such artefacts.

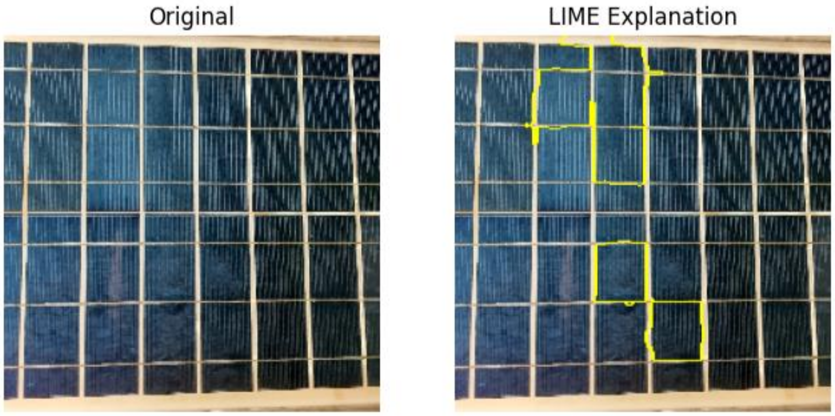


Figure 6. LIME two-part graph.

In summary, the quantitative results and XAI visualizations demonstrate that the lightweight proposed Inception model successfully captures relevant dust-related features on the solar panels, especially for obviously dirty scenarios, but it surprisingly also reveal several challenges in consistently classifying clean solar panels. These results suggest multiple avenues for future work, including development of an attention mechanism that effectively suppresses background edges and reflections, stronger regularization, and larger dataset with more variations in lighting conditions and panel types. These problems are further discussed in the conclusion and future work section.

4.3. Fair Comparison with Other Deep Learning Models

To fairly position the proposed residual depth-wise separable lightweight Inception model against other deep learning approaches, it is important to recognise both the differences in task formulation and data and the constraints imposed in this coursework. In this work, a custom binary classifier is trained from scratch on a relatively small solar panel dust dataset (539 training, 134 validation, 169 testing images) using only bottleneck (1×1) and spatially separable (1×3 / 3×1) convolutions. Under these constraints, the best checkpoint achieves a test accuracy of 49.11%, with a precision of 0.42, recall of 0.75 and F1-score of 0.54 for the dirty class, and a ROC–AUC of 0.72. As discussed in Section 4.2, the model clearly overfits the training data but still learns a non-trivial ranking between clean and dirty panels. This performance is not competitive in absolute terms with state-of-the-art image-based dust detectors, but the comparison below highlights why such a gap is expected and how the present work remains useful as a lightweight, interpretable baseline.

Alatwi et al. [4]tackled a conceptually similar problem—binary classification of panel images into clean and dusty—but relied exclusively on pre-trained CNNs as feature extractors. They evaluated 20 popular architectures (e.g. VGG, ResNet, EfficientNet, DenseNet) on a dataset of over one thousand images and trained a linear SVM on the deep features of each network. Their best configuration, DenseNet-169 + linear SVM, achieved an accuracy of about 86.79%, with MobileNet and other models also reaching mid-80% accuracy. Compared to these results, the accuracy of the proposed coursework model is much lower. However, this difference is largely explained by (i) the use of ImageNet pre-training in[4], which provides rich generic feature representations before fine-tuning; (ii) a larger and more diverse training set; and (iii) the absence of strict architectural constraints. In contrast, the coursework model is intentionally small (~0.5M parameters) and trained from scratch on a smaller dataset. More importantly, [4]treat the CNNs as black-box feature generators and do not integrate any internal attention mechanism, whereas the present work explicitly designs a convolution-only attention block and later analyses its behaviour using Grad-CAM and LIME. Thus, while the raw accuracy is lower, the coursework contributes an interpretable, resource-constrained alternative that is more aligned with the assignment specification.

Alçin et al. [3]proposed SolPowNet, a custom CNN built specifically for solar panel dust classification. Like the network in this coursework, SolPowNet is trained end-to-end on panel images (502 clean, 340 dusty), but it uses a deeper and more flexible architecture without being restricted to only bottleneck and separable convolutions. On their dataset, SolPowNet achieves a test accuracy of 98.82% and an F1-score above 0.98, clearly outperforming large pre-trained baselines such as AlexNet, VGG16/19, ResNet50 and even InceptionV3. In other words, SolPowNet shows that a carefully designed custom CNN can surpass heavy pre-trained models both in accuracy and computational efficiency. Compared with SolPowNet, the model in this coursework is significantly lighter (roughly an order of magnitude fewer parameters) but also significantly less accurate. A fair interpretation is that this project is closer to an exploratory prototype under strong architectural and data constraints, whereas SolPowNet represents a production-ready design optimized via extensive experimentation. Another key distinction is interpretability: SolPowNet mainly reports aggregate metrics and qualitative examples, while the coursework explicitly integrates XAI tools to inspect the learned attention and failure cases. From this perspective, the proposed model trades accuracy for strict architectural simplicity and richer explainability analysis.

Fair comparison with object-detection-based approaches must be even more cautious because they address a different task: localising dust or soiling patches rather than simply classifying an image. Gao and Li [5] improved YOLOv5s for detecting dust spots in UAV images of large solar farms. Their modified detector adds an extra prediction head and architectural tweaks, achieving very high recall, precision, and F1-scores (all close to or above 0.95) and a mean Average Precision around 0.95 on their test set. Naeem et al. [6]developed SDS-YOLO, which integrates a CBAM attention module and dual detection heads for dust and bird droppings. On an aerial soiling dataset, SDS-YOLO delivers F1-scores above 0.75 for the dust class and around 0.70 for bird droppings, with mAP50 values in the mid-0.7 to high-0.7 range, while simultaneously reducing parameters by about 24% compared with the baseline YOLOv5. Both detectors dramatically outperform the coursework classifier in terms of detection quality, but they also: (i) operate on high-resolution UAV imagery rather than close-range panel photos; (ii) rely on strong pre-training on large datasets such as COCO; (iii) use more complex backbones and attention mechanisms (e.g. CBAM) that are not permitted in the coursework; and (iv) require more sophisticated training pipelines and computing resources. Moreover, neither [5] nor [6] integrates post-hoc XAI methods like Grad-CAM or LIME to explain individual detections. The attention used in SDS-YOLO is internal and not directly interpretable by end users.

The non-image-based deep learning works in the set of seven papers also provide context for fairness. Mokhtar and Shaaban's ANN-based cleaning scheduler [2]and Shaaban et al.'s regression-tree dust estimator [1]reach high prediction accuracy on their respective tabular datasets and are highly practical for cleaning decision support. However, they do not perform visual dust recognition at all, and thus are complementary rather than competitive to the image-based classifier developed here. In particular, they cannot show where the dust lies on a panel or whether apparent performance loss is due to soiling or other factors (e.g. shading or module degradation), whereas an image-based approach with Grad-CAM and LIME can, in principle, provide that level of spatial insight.

In summary, a rough comparison to the models of deep learning suggests that our proposed coursework model is not competitive in terms of accuracy or F1-score, especially compared to pre-trained CNNs and state-of-the-art YOLO versions trained on big datasets. But those models typically assume they can pre-train, use more data, and have fewer architectural constraints, and they almost never produce fine-grained, instance-level justifications. Contrarily, the current study toys with a very restricted, minimalist design consisting exclusively of bottleneck and spatially separable convolutions paired with a novel attention unit and explicit Grad-CAM/LIME evaluation. The key takeaway, then, is not that we beat existing pre-trained deep models for PL images, but rather that we show how even a tiny, interpretable network performs when confronted with realistic data constraints, surfacing its strengths (good ranker, high recall for dirty panels, meaningful attention maps) and weaknesses (overfitting, low specificity). These results provide a foundation for future work wherein aspects of more highly performing pre-trained or attention-augmented models from the literature may be judiciously inserted into more powerful, yet still interpretable dust detection systems.

4.4. Comparison with Existing Literature

Provide direct comparison of your work with other prior works. An example has been provided below:

Work	Input/Task	Model type	Pre-training	Main metrics	Interpretability	Relation to my work
[1]	Sensor&electrical data; dust level estimation	Regression-based dust estimation unit (fine tree and others)	No	High dust-level estimation accuracy; effective cleaning decision support	No explicit XAI	Data-based dust estimation; complementary, non-visual (no image-based dust detection)
[2]	Sensor&electrical data; cleaning decision support	ANN-based cleaning decision model (regression / scheduling)	No	High correlation with expert decisions; reduced energy loss and cleaning cost	No explicit XAI	ANN cleaning scheduler; non-visual, complementary
[3]	Near-field panel images; binary	Custom CNN classifier (SolPowNet)	No	Test accuracy ≈ 98.8%, F1-score>0.98	Limited qualitative examples; no formal XAI analysis	Same task (clean vs dusty); strong custom CNN baseline
[4]	Near-field panel images; binary clean vs dusty	Pre-trained CNN feature extractors (VGG, ResNet, MobileNet, DenseNet, etc.) + SVM	Yes	Best: DenseNet-169 + SVM, accuracy ≈ 86.8%; several models in mid-80% range	Can be combined with post-hoc XAI; not core focus	Same task; heavy ImageNet backbones; focuses on transfer learning

[5]	UAV aerial images of PV plants; dust spot detection	Improved YOLOv5 object detector with extra prediction head	Yes	Precision/recall/F1 ≳ 0.95; mAP ≈ 0.95	No Grad-CAM/LIME; internal feature maps only	Different task: detection and localisation of dust regions from UAV images
[6]	UAV aerial images; dust&bird droppings detection	SDS-YOLO (YOLOv5 + CBAM attention + dual detection heads and pruning)	Yes	Dust F1>0.75; bird droppings F1 ≈ 0.70; mAP50 in mid–high 0.7 range	Internal CBAM attention, not turned into XAI plots	Different task: multi-class soiling detection on aerial images
My work	Near-field panel images; binary clean vs dirty	Residual depth-wise separable lightweight Inception classifier with custom conv-only attention	No	Test accuracy 49.11%; precision 0.42; recall 0.75; F1-score 0.54; ROC–AUC 0.72	Explicit Grad-CAM and LIME explanations; custom conv-only attention	Binary classifier under strong architectural constraints; focuses on lightweight design and XAI rather than SOTA accuracy

To provide a direct comparison between the proposed model and prior deep learning approaches, it is necessary to consider differences in input modality, task formulation, model size, and training regime. In this coursework, the proposed residual depth-wise separable lightweight Inception network is a binary image classifier that distinguishes clean from dirty solar panels. It is trained from scratch on a relatively small near-field image dataset (539 training, 134 validation, and 169 testing samples), and is strictly constructed from bottleneck (1 × 1) and spatially separable (1 × 3 / 3 × 1) convolutions with a custom convolution-only attention block.

Under these constrained conditions, the best checkpoint achieves a test accuracy of 49.11%, a precision of 0.42, a recall of 0.75 and an F1-score of 0.54 for the dirty class, with a ROC–AUC of 0.72. As discussed earlier, this reflects clear overfitting and modest overall performance, but also a non-trivial ability to rank dirty panels ahead of clean ones.

In contrast, some of the prior works do not use images at all but instead focus on data-driven dust estimation and cleaning decisions. Shaaban et al. [1] proposed a data-based dust estimation unit for PV panels using regression models trained on electrical and environmental measurements. Their best "fine tree" regressor achieves high accuracy in estimating dust levels and can reliably trigger cleaning events when a threshold is exceeded [1]. Mokhtar and Shaaban [2], [7] further designed an ANN-based cleaning approach that takes panel output, irradiance, and temperature as inputs to estimate dust accumulation and schedule cleaning only when necessary. Their case study shows that such an ANN-driven strategy can significantly reduce energy loss and cleaning costs compared to routine cleaning [2], [7]. These models are highly effective for cleaning scheduling, but they do not perform visual dust recognition and cannot show where dust is located on the panel surface. From the perspective of this coursework, they address a complementary problem: they optimise maintenance policies based on sensor data, whereas the current work explores an image-based CNN under strict architectural constraints.

Among image-based classifiers, Alçin et al. [3] and Alatwi et al. [4] are the closest to the present work. Alçin et al. [3] proposed SolPowNet, a custom CNN architecture designed specifically for classifying panel images into clean or dusty categories. SolPowNet is trained end-to-end on a dataset of 502 clean and 340 dusty panel images and achieves a test accuracy of 98.82% and an F1-score above 0.98, clearly outperforming standard pre-trained networks such as AlexNet, VGG16/19, ResNet50 and InceptionV3 evaluated on the same data [3]. The network is also computationally efficient, with significantly fewer parameters and lower inference cost than these large backbones. Compared with SolPowNet, the model in this coursework is even more constrained in its building blocks (only 1 × 1 and separable convolutions, with a shallower depth) and is trained on fewer samples. As a result, its accuracy is much lower. A fair interpretation is that SolPowNet represents a mature, performance-optimised custom architecture suitable for deployment, whereas the coursework model is an exploratory prototype designed primarily to satisfy coursework constraints and to support explainability analysis. Another important difference is that SolPowNet mainly reports aggregate metrics and qualitative predictions, while the current work explicitly integrates Grad-CAM and LIME to obtain instance-level visual explanations for correct and incorrect classifications.

Alatwi et al. [4] addressed almost the same task as this coursework—binary dust detection from panel photographs—but took a different modelling strategy. They evaluated twenty state-of-the-art pre-trained CNNs (including VGG, ResNet, MobileNet, EfficientNet, and DenseNet) as fixed feature extractors and trained a linear SVM classifier on their deep features. On a dataset of over one thousand images, the best configuration (DenseNet-169 features plus a linear SVM) achieved an accuracy of 86.79%, with several other backbones also reaching accuracies in the mid-80% range [4]. Compared with these numbers, the 49.11% test accuracy of the coursework model is clearly much lower. This gap is largely explained by the use of ImageNet pre-training and high-capacity backbones in [4], as well as a larger and more diverse dataset.

The models in [4] also do not face the strict architectural restrictions imposed in this coursework. On the other hand, the pre-trained CNNs in [4] are primarily used as black-box feature generators; internal attention mechanisms are not explicitly designed, and interpretability is limited to post-hoc tools if used. In contrast, the coursework model is intentionally tiny (around 0.5M parameters), built only from the allowed convolutional operations, and explicitly coupled with Grad-CAM and LIME to study how its custom attention module focuses on dusty regions and where it fails on clean panels.

Object-detection-based approaches, such as those from Gao and Li [5] and Naeem et al. [6], solve a related but different task: they localise soiling or dust patches in images, often captured from unmanned aerial vehicles (UAVs), instead of simply classifying an entire panel as clean or dirty. Gao and Li [5] proposed a deep learning method based on an improved YOLOv5 detector, adding an extra prediction head and architectural refinements tailored to PV plants seen from drones. Their model achieves very high detection performance on UAV imagery, with precision, recall, and F1-scores close to or above 0.95 and a mean Average Precision around 0.95 on their test set [5]. Naeem et al. [6] developed SDS-YOLO, which integrates a Convolutional Block Attention Module (CBAM) and dual detection heads to separately detect dust and bird droppings on panels. On an aerial soiling dataset, SDS-YOLO attains F1-scores above 0.75 for the dust class and around 0.70 for bird droppings, with mAP(50) values in the mid-0.7 to high-0.7 range, while simultaneously reducing the number of parameters by about 24% compared with the baseline YOLOv5 [6]. In absolute terms, these detectors far surpass the coursework classifier in their respective detection tasks. However, they rely on high-resolution UAV imagery, pre-trained YOLO backbones, more complex attention modules (e.g. CBAM, which is not permitted in this coursework), and more sophisticated training pipelines and computational resources. Moreover, although SDS-YOLO uses internal attention, neither [5] nor [6] incorporates post-hoc explainability techniques such as Grad-CAM or LIME to provide human-readable explanations for individual detections.

Overall, these comparisons show that the proposed coursework model is not competitive in raw accuracy or F1-score when placed alongside pre-trained CNN feature extractors [4], high-capacity custom architectures like SolPowNet [3], or attention-enhanced YOLO detectors [5], [6]. This performance gap is expected given the smaller dataset, the absence of pre-training, and the strict architectural limitations. The main contribution of the present work lies instead in its lightweight and interpretable design: it explores how a very small, constraint-driven CNN with a custom convolution-only attention block behaves under realistic data limitations, and how its decisions can be analysed via Grad-CAM and LIME. In particular, the model still provides useful ranking ability (ROC–AUC 0.72) and high recall for dirty panels, while its confusion matrix and XAI visualisations clearly expose weaknesses such as low specificity and susceptibility to reflections and background structures. These insights form a basis for future work in which elements from stronger pre-trained or attention-based models in the literature can be selectively incorporated into new architectures that are both more accurate and remain interpretable for solar panel dust detection.

5. CONCLUSION, LIMITATION, FUTURE WORK

This coursework explored image-based dust detection on photovoltaic panels using a custom residual depth-wise separable lightweight Inception model with a convolution-only attention mechanism and XAI tools. The model was built entirely from bottleneck 1×1 and spatially separable 1×3/3×1 convolutions and trained from scratch on a small near-field image dataset for binary classification (clean vs. dirty). On the test set, it achieved an accuracy of 49.11%, with precision 0.42, recall 0.75 and F1-score 0.54 for the dirty class, and ROC–AUC 0.72. Although the overall accuracy is modest, the model shows a tendency to prioritize detecting dirty panels (higher recall), which is desirable for maintenance, and Grad-CAM/LIME visualizations indicate that, in many correct predictions, the network focuses on dust-relevant regions of the panel surface.

However, several limitations remain. The dataset is relatively small and not highly diverse in terms of locations, lighting, panel types, and soiling patterns, which leads to clear overfitting and low generalization to unseen data. The architecture, while lightweight, is still expressive relative to the data size, and only limited hyperparameter tuning and regularization were performed within the coursework time frame. Evaluation is based on a single train/validation/test split, without external validation, and the XAI analysis is mainly qualitative, relying on a limited number of Grad-CAM and LIME examples. In addition, the current work focuses only on binary classification and does not address other practically important tasks such as multi-class soiling recognition or localization of dust regions.

Future work should first expand and diversify the dataset and consider multi-class labels for different soiling types, which would help both accuracy and robustness. On the modelling side, systematic ablation studies and stronger regularization (e.g. Heavier augmentation, dropout, weight decay) could be used to reduce overfitting, and lightweight pre-training or transfer learning strategies might be explored while still keeping the parameter count low. The XAI component could be extended with additional methods (e.g. Integrated Gradients or SHAP for images) and more quantitative evaluation of explanation quality, for example, by comparing highlighted regions with manually annotated dust areas. Finally, integrating the proposed classifier into a broader decision-support framework (combining image-based outputs with sensor-based dust estimators) and extending it towards detection or segmentation of soiling in UAV imagery would move the approach closer to real-world deployment while keeping interpretability as a central design goal.

References

[1] Mostafa. F. Shaaban, A. Alarif, M. Mokhtar, U. Tariq, A. H. Osman, and A. R. Al-Ali, ‘A New Data-Based Dust Estimation Unit for PV Panels,” *Energies*, vol. 13, no. 14, p. 3601, July 2020, doi: 10.3390/en13143601.

[2] M. Mokhtar and M. F. Shaaban, ‘A New ANN-Based Cleaning Approach for Photovoltaic Solar Panels’, in *2022 9th International Conference on Electrical and Electronics Engineering (ICEEE)*, Alanya, Turkey: IEEE, Mar. 2022, pp. 260–263. doi: 10.1109/ICEEE55327.2022.9772579.

[3] Ö. F. Alçin, M. Aslan, and A. Ari, ‘SolPowNet: Dust Detection on Photovoltaic Panels Using Convolutional Neural Networks’, *Electronics*, vol. 14, no. 21, p. 4230, Oct. 2025, doi: 10.3390/electronics14214230.

[4] A. M. Alatwi, H. Albalawi, A. Wadood, H. Anwar, and H. M. El-Hageen, ‘Deep Learning-Based Dust Detection on Solar Panels: A Low-Cost Sustainable Solution for Increased Solar Power Generation’, *Sustainability*, vol. 16, no. 19, p. 8664, Oct. 2024, doi: 10.3390/su16198664.

[5] Y. Gao and S. Li, ‘A deep learning-based method detects dust from solar PV panels through Unmanned Aerial Vehicles’, *J. Phys.: Conf. Ser.*, vol. 2584, no. 1, p. 012019, Sept. 2023, doi: 10.1088/1742-6596/2584/1/012019.

[6] U. Naeem, K. Chadda, S. Vahaji, J. Ahmad, X. Li, and E. Asadi, ‘Aerial Imaging-Based Soiling Detection System for Solar Photovoltaic Panel Cleanliness Inspection’, *Sensors*, vol. 25, no. 3, p. 738, Jan. 2025, doi: 10.3390/s25030738.

[7] M. Fang, P. M. Rodrigo, L. R. Narvarte, G. Makrides, and G. E. Georghiou, “A deep learning model for photovoltaic soiling loss estimation,” *Applied Energy*, vol. 376, pt. B, art. no. 124335, Dec. 2024, doi: 10.1016/j.apenergy.2024.124335.

[8] G. Cavieres, A. Barraza, A. Estay, M. Bilbao, and A. Valdivia-Lefort, “Automatic soiling and partial shading assessment tool for photovoltaic power plants based on convolutional neural networks,” *Applied Energy*, vol. 306, art. no. 117964, Jan. 2022, doi: 10.1016/j.apenergy.2021.117964.

[9] H. Zhang, Y. Zhang, S. Chen, D. Zhou, M. T. M. Emam, and L. Yu, “SoilingEdge: PV soiling power loss estimation at the edge using surveillance cameras,” *IEEE Transactions on Sustainable Energy*, vol. 15, no. 1, pp. 556–566, Jan. 2024, doi: 10.1109/TSTE.2023.3320690.

[10] P. Naskar, A. Chowdhury, S. Adhikari, and C. Chakraborty, “A deep learning regression framework for robust PV soiling quantification using multi-source drone-ground imaging,” *Expert Systems with Applications*, vol. 306, art. no. 130944, Apr. 2026, doi: 10.1016/j.eswa.2025.130944.

[11] L. Micheli, E. F. Fernández, A. Smets, and J. A. del Cañizo, “Variability and associated uncertainty analysis of image-based soiling characterization,” *Solar Energy Materials and Solar Cells*, vol. 259, art. no. 112437, Jun. 2023, doi: 10.1016/j.solmat.2023.112437.

[12] M. Yang, W. Javed, B. Guo, and J. Ji, “Estimating PV soiling loss using panel images and a feature-based regression model,” *IEEE Journal of Photovoltaics*, vol. 14, pp. 661–668, 2024, doi: 10.1109/JPHOTOV.2024.3388168.