Solar Panel Defect Detection using MobileViT-SE Attention Mechanism and YOLOv8

Sanjay kumar S
Department of Electronics and communication Engineering
Amrita school of Engineering
Coimbatore, TamilNadu
sanjay112311@gmail.com

Dinesh T
Department of Electronics and communication Engineering
Amrita school of Engineering
Coimbatore, TamilNadu
dineshtamil1812@gmail.com

Barathkumar S
Department of Electronics and communication Engineering
Amrita school of Engineering
Coimbatore, TamilNadu
s.barathkumar25@gmail.com

Dr. Karthika R
Department of Electronics and communication Engineering
Amrita school of Engineering
Coimbatore, TamilNadu
r_karthika@cb.amrita.edu

***Abstract*—We rely more on solar energy because it offers abundant electricity generation. Keeping the solar panels operating at peak efficiency is crucial. However, some environmental issues, such as hotspots and electrical damage, can reduce the effectiveness of solar panels. This paper presents a deep learning approach for detecting defects in solar panels. We tackle the problem from two perspectives: thermal and visual. For thermal images, we trained a YOLOv8 model to detect hotspots and diode failures with an accuracy of 90%. For visual images, we used a lightweight MobileViT with a Squeeze and Excitation attention block to identify defects like electrical damage and bird droppings, achieving an accuracy of 91.36%. Both models were developed using a dataset of healthy and faulty solar panel images. This research helps optimize solar power generation and ensures its reliability and longevity.**

I. INTRODUCTION

Due to the current worldwide situation of growing environmental pollution and increasing energy consumption, solar energy has been widely utilized for its low carbon, high power generation and sustainable development. Yet, operational challenges such as hotspots, micro-cracks, dust accumulation, bird droppings, and physical damage continuously affect panel efficiency and lifespan. These defects arise from non-uniform current densities, environmental stresses (shading, wind, humidity, snow), manufacturing imperfections, and installation faults. Hotspots, for instance, result from cells operating in reverse bias or experiencing abnormal thermal stress, causing local heating that sharply reduces power output and can permanently damage panel materials. Micro-cracks, often invisible to the naked eye, compromise internal circuitry, propagating with age and environmental wear, while surface-level faults like dust and debris block sunlight and introduce shading losses—sometimes resulting in up to 30% daily output decline.

Manual inspection of large-scale solar installations has proven impractical, highlighting the need for automated systems that can accurately detect defects. Imaging techniques, particularly thermal and visual methods, have shown to be an excellent options for spotting a range of PV faults. However, traditional single-domain approaches have limitations in scope and adaptability, especially when defects do not have uniform pattern across all modalities.

In this work, we present a defect detection framework for solar panels that seamlessly integrates thermal and visual modalities through deep learning. Our dual-system employs YOLOv8 for fast thermal anomaly localization and MobileViT-SE, a lightweight vision transformer with channel-wise attention for precise visual defect classification, covering categories such as Bird-drop, Clean, Dusty, Electrical-damage, Physical-damage, Snow-Covered, Hotspot, Cell Fault, and Bypass Diode. The selection of YOLOv8 is driven by its outstanding speed and accuracy in real-time object detection, ideally suited for the dynamic analysis of infrared thermal data and effective identification of heat-related faults in operational PV modules. MobileViT-SE was chosen for its mobile-friendly architecture and powerful hybridization of convolutional and transformer modules, alongside Squeeze-and-Excitation (SE) attention blocks that allow the model to focus on subtle yet critical surface defects in diverse visual conditions.

To unify the strengths of each model, our system leverages a late fusion strategy: predictive outputs from the thermal and visual domains are combined after independent analysis, ensuring that both heat-related and visually noticeable faults contribute to the final diagnosis. This approach not only addresses the limitations of single-modality frameworks but also delivers a robust, high-accuracy solution tailored for real-world solar farm management.

Custom datasets, collected using FLIR C5 thermal cameras across operational solar plants, contain thousands of annotated samples with multiple defect types and environmental scenarios. With preprocessing, the system achieves approximately 90–91% accuracy in both thermal and visual detection tasks, validated under challenging field conditions.

By fusing modern deep learning with multimodal sensor data, our approach empowers early, automatic, and quantitative detection of complex solar panel faults, boosting system reliability, reducing maintenance costs, and supporting scalable deployment for the next generation of AI-driven PV monitoring solutions.

II. LITERATURE SURVEY

The available literatures only provide solution relying on single spectrum, either with Visual spectrum or InfraRed (IR) spectrum, So the model fails in the case where either Thermal or visual does not able to identify certain kind of defect. This hinders the quality of inspection.

Winston et al. presents a comprehensive fault detection technique using two distinct Machine Learning classifiers: Feed Forward Back Propagation Neural Network (FFBPNN) and the Support Vector Machine (SVM). The PV modules were classified into five categories: Healthy, One Hotspot, Two Hotspots, More than Two hotspots, and Microcracks. These categories were categorized based on six electrical parameters namely, Percentage of Power Loss (PPL), Open-circuit voltage (V_{OC}), Short circuit current (I_{SC}), Irradiance (I_{RR}), Panel temperature, and Internal impedance (Z). The SVM acquires accuracy of 99% while FFBPNN has accuracy of 87%. This model can be implemented at various environmental conditions[1].

Barnabé et al. proposes a novel approach that combines Image Processing and Vision Transformers (ViT) for flaw detection and classification in photovoltaic cells. The core contribution of this work is a lightweight flaw detection and Multi-class classifier version of Vision transformer named as ViT-μ model. The classifier addresses three classes: cracked cells, cells with cold solders, and undamaged cells. For cracks, the proportional damaged area is determined using an algorithm that checks if a pixel has a clear vertical path to a busbar, based on the principle that damage occurs when current paths are interrupted. Crack length is approximated by the diagonal of a boundary box around the detected crack after image binarization and thinning. For cold solders, a weakly supervised method utilizing the ViT- μ's attention maps is introduced to visualize and estimate the affected region. The proposed algorithm includes a Cell Cleansing process to produce both binary and RGB cell images to remove busbars. This process uses an adaptive Gaussian inverted threshold and Zhang-Suen thinning for image binarization. It also uses the Hough transform and an inpainting algorithm to highlight flaws while removing busbars. The model achieves highest accuracy of 98% for flaw detection(Binary Classification), For multi-class classification it achieves accuracy of 94%. This work implemented under controlled environment[2].

Gao et al. used YOLOv8 model composed of single stage object detection named Partial Spatial Attention (PSA) Mechanism to improve the feature extraction. PSA is the integration of Partial convolution and spatial attention. The model was validated using Panel-2 dataset that contains three types of defects: scratches, broken grids and deface, obtaining an mAP@50 of 87.2% and the Solar panel dataset,4 which contains three types of defects: bird droppings, cracks and dust, obtaining an mAP@50 of 72%[3].

Sujata et al. investigate how different pre-processing Methods affect the performance of defect identification in thermal PV images. This study uses filters and a Histogram Equalization (HE) to detect five types of faults: single-cell hotspot, multicell hotspot, diode fault, dust/shadow hotspot, and PID (potentially induced degradation) hotspot. Filters such as Mean, Median, Gaussian, and Bilateral were tested for noise removal and Histogram Equalization to improve the image contrast. There were two different segmentation processes involved: color thresholding using Histogram in the HSV color space and Thresholding based on separate color channels. Intersection Over Union (IOU) was the metric used to identify the efficiency. Bilateral filter with HE gives the best accuracy among all the filters tested [4].

Sameer Alam et al. focused on PV fault detection using Convolutional Neural Network model. The dataset comprises of 4000 thermal solar images equally split between two major classes: Faulty and Healthy. Under unhealthy, 11 anomalies that includes: Cell, Cell-multi, Hotspot, Hotspot-multi, Vegetation, Soiling, Offline-module, Diode, Diode-multi, Cracking, and Shadowing were identified. The model attains a validation accuracy of 89%[5].

III. PROPOSED METHODOLOGY

The proposed system is a multi-modal defect detection pipeline designed for solar panels. The core of the system relies on the processing of both thermal and visual data. The pipeline begins with data acquisition from FLIR C5 thermal imaging camera which captures the thermal image and its equivalent visual image. The thermal images are then processed by a YOLOv8 object detection model to identify defects like hotspots by placing bounding boxes around them. The visual images are processed by a MobileViT-SE model, which classifies the panel's condition into predefined categories. The outputs from these two models are then combined using a late fusion strategy, which leverages the strengths of both modalities to make a final, more reliable decision on the presence and nature of the defect.

A. Dataset

There were two different types of datasets used for training this model. These two different types of dataset are used for training the models of MobileVit-SE and YOLOv8. Faulty solar panel dataset which was taken from Kaggle contains 864 images which contains the defective images for dusty, electrical damage, physical damage, snow covered, bird drop and clean. It was used to train MobileVit-SE model which focuses only on visual images. For the thermal images, the dataset was obtained from a Roboflow source. It consists of 2,235 images, divided into 1,899 training images, 223 validation images, and 113 testing images. The dataset includes three distinct categories: Bypass Diode failure, Cell Fault and Hotspot.

In addition to these datasets to ensure robust training and validation of the proposed model, we took real world solar samples from The VC Green Energy Solar plant located at Vadavalli, Coimbatore and Amrita Vishwa Vidyapeetham, Ettimadai, Coimbatore. Around 3000 samples were collected from these two locations. We have created a dataset using these samples 1500 for thermal and 1500 for visual images which contains defect and clean images.

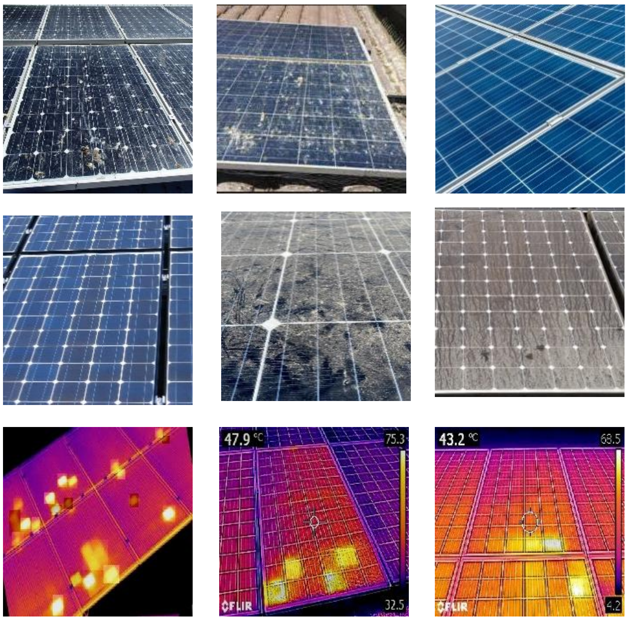
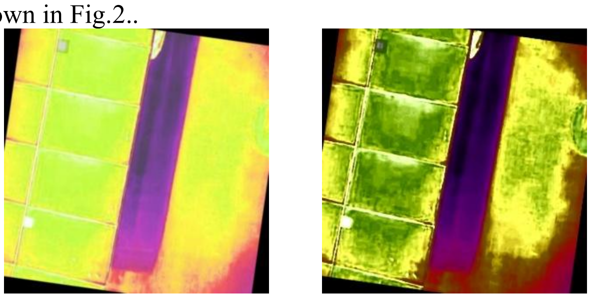


Fig 1: Samples of solar panels with defective and normal surfaces.

B. Dataset pre-processing

The Median Filter operation replaces a pixel's value with the median of all pixel values in its surrounding neighborhood. This process involves sorting all neighboring pixel values numerically and using the middle value of that sorted list to replace the center pixel. This method is highly effective for smoothing noise from an image while simultaneously preserving sharp edges and fine image details.

Following the median filtering, Histogram equalization is applied to the entire dataset which is a technique to increase the contrast of an image. It increases the intensity difference among objects and back ground This image preprocessing stage includes cropping and resizing of each image to 640x640 regions. An example of this preprocessing stage is shown in Fig.2..



(a) (b)

Fig 2: a) without pre-processing b) with pre-processing

C. MobileVit-SE model

Objective of employing this model is to detect the defects present in solar panel by using the visual images. A standard Vision Transformer (ViT) model, employs Squeeze-and-excitation block. A standard Vision Transformer (ViT) model first reshapes the input image into a sequence of flattened patches and learns inter-patch representations using a stack of L transformer blocks. For scaling quadratically with number of patches in traditional ViTs,The computational cost is high. Also the Light-weight Convolutional Neural Networks (CNNs) are spatially local, which is a major drawback for learning long-range dependencies. For further enchancement the discriminative power of the features extracted by the MobileViT-XXS backbone, we incorporate a Squeeze-and-Excitation (SE) attention block immediately before the final classification layers. The SE block introduces a lightweight, channel-wise attention mechanism to adaptively recalibrate feature responses[6].

- **Squeeze Operation (Global Average Pooling):** The output feature map from the MobileViT backbone is first passed through an Adaptive Global Average Pooling 7 layer. This generates a compact channel descriptor by aggregating global spatial information for each channel.
- **Excitation Operation (Channel Gating):** The channel descriptor is processed by a small bottleneck fully-connected network (Linear→ReLU→Linear→Sigmoid). This mechanism learns non-linear, channel-wise dependencies, producing a vector of attention weights (0 to 1) for each channel.
- **Scale Operation**: The attention weights are then broadcast and multiplied (scaled) channel-wise onto the input feature map. This process adaptively emphasizes informative feature channels specific to solar panel defects while suppressing less relevant ones

D. YOLOv8 model

Objective of employing YOLOv8 model is to detect the defects that are present in the solar panel using thermal images. There are several defects in solar panel that cannot be seen using visual images but seen in thermal images. Those defects reduce the efficiency and working of the solar panels. So, identifying those defects are very crucial. Few such defects are Bypass diode failure, cell fault and hotspot. The YOLO (You Only Look Once) series represents a core technology in object detection, known for its outstanding balance of speed and accuracy. YOLOv8, introduced by Ultralytics in 2023, offers state-of-the-art (SOTA) performance and a unified framework for various computer vision tasks. The YOLOv8 model comprises four principal components: the Input Layer, the Backbone Network Layer, the Neck Structure Layer, and the Output Layer (Head)[7].

- **Backbone Network:** Primarily responsible for feature extraction, utilizing modules such as Conv, C2f, and SPPF.
- **Neck Structure:** Connects the backbone and the output layer, integrating and fusing multi-scale features via the Feature Pyramid Network (FPN) and Path Aggregation Network (PAN) structures to enhance context information.
- **Output Layer (Head):** Generates prediction results, including bounding boxes, category labels, and confidence scores.

The proposed model adopts the YOLOv8s network structure due to its strong performance-to-lightness ratio, which is crucial for real-time monitoring on resource constrained platforms like UAVs[7].

IV. Results and Discussion

A. *Visual spectrum (MobileVit-SE model)*

The training was conducted using the MobileVit with Squeeze and Excitation attention block on 885 images categorized as training and validation. Each category has 5 distinct defects of solar panels, namely, Bird-drop, Dusty, Electrical-damage, Physical-damage, and Snow-covered. The dataset also contains healthy solar panel images named as clean. The images in the dataset were split into 80% and 20% for training and validation, respectively, for each category of defect. The 20% validation images is for validating the trained MobileVit-SE model. Figure-7 illustrate the training process of the MobileVit-SE, considering accuracy and loss metrics, respectively. Both training and validation accuracy show an increasing trajectory such that the model effectively learns the training data and generalizes to validation data. The training and validation loss decrease over iterations, indicating that the model adjusts the parameters during training. Table 1 shows the classification report of the trained model that includes the evaluation metrics such as Precision, Recall, F1 score, and accuracy. The model is trained over 100 epochs and achieves a training accuracy of 99% and a validation accuracy of 91.56%. Figure-3 displays the confusion matrix of the validation dataset. This matrix shows the visual representation of misclassifications of the data from the trained model.

Class	Precision	Recall	F1-Score	Support
Bird-drop	0.92	0.91	0.91	65
Clean	0.88	0.92	0.9	39
Dusty	0.93	0.93	0.93	106
Electrical-damage	0.84	0.96	0.9	28
Physical-Damage	0.9	0.9	0.9	39
Snow-Covered	1	0.75	0.86	24
accuracy			0.91	301
macro avg	0.91	0.9	0.9	301
weighted avg	0.91	0.91	0.91	301

Table 1: Classification report of MobileVit model

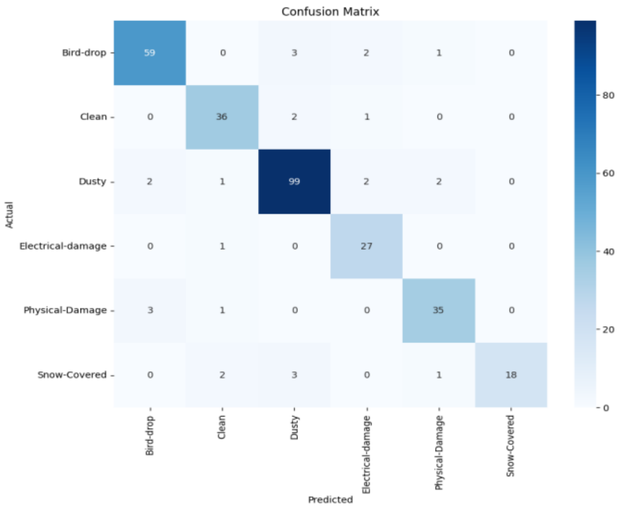


Fig 3: Confusion matrix for MobileVit-SE model trained for visual images

B. *IR Spectrum (YOLOv8 model)*

The thermal images of solar panel was trained using the YOLOv8s model on 2,235 images categorized as training, validation, and testing. The dataset includes 3 distinct thermal defect, namely: Bypass Diode, Cell Fault, and Hotspot. The images in the dataset were split into 85% for training (1,899 images), 10% for validation (223 images), and 5% for testing (113 images). The model's performance was monitored using the validation set, while the separate test set was used for the final performance evaluation. Figure-4 illustrates the training and validation process over 100 epochs. Both the training and validation loss curves (e.g., val/box_loss, val/cls_loss) show a sharp decrease in the initial epochs and then stabilize, indicating that the model effectively adjusted its parameters. The validation metric curves, mAP50(B) and mAP50-95(B), show a steady increasing trajectory, confirming that the model learned to generalize to unseen data without significant overfitting. Table-2 show the detailed classification report for the trained model. This includes the key evaluation metrics for object detection: Precision, Recall, and mean Average Precision (mAP). The model achieves a final Precision of 63.6%, a Recall of 61.2%, and a mean Average Precision (mAP@0.5) of 61.4% on the test set.

epoch	time	train/ box_loss	train/ cls_loss	train/ dfl_loss	metrics/ precision (B)	metrics/ recall(B)	metrics/ mAP50(B)	metrics/ mAP50-95(B)	val/ box_loss	val/ cls_loss	val/ dfl_loss
1	48.5715	2.38874	3.003	1.89182	0.03562	0.07807	0.01834	0.00631	3.30838	12.9318	4.54435
2	94.0336	2.36927	2.62834	1.94257	0.08056	0.21114	0.06305	0.02111	2.55627	6.58515	2.20382
3	142.9	2.31619	2.51269	1.87767	0.49432	0.17243	0.13778	0.04401	2.40896	5.31995	2.02142
4	191.112	2.24975	2.49549	1.86945	0.1361	0.25586	0.12597	0.04544	2.50866	3.37499	2.05138
...	...	...	...	...	...	...	...	...	...	...	...
96	4577.47	1.68724	1.11745	1.45634	0.63696	0.60396	0.60692	0.29425	1.9812	1.42215	1.67471
97	4622.24	1.67585	1.09768	1.43659	0.62916	0.61925	0.61831	0.29397	1.97113	1.39275	1.67526
98	4667.67	1.67316	1.1026	1.4428	0.61805	0.60053	0.60827	0.29721	1.98584	1.39368	1.68236
99	4717.05	1.66473	1.09191	1.43539	0.63385	0.60945	0.61545	0.29831	1.98746	1.39683	1.68359
100	4764.59	1.66879	1.09323	1.4487	0.63503	0.61212	0.61328	0.29922	1.97716	1.39496	1.67665

Table 2: Performance metrices for YOLOv8 model trained for Thermal images

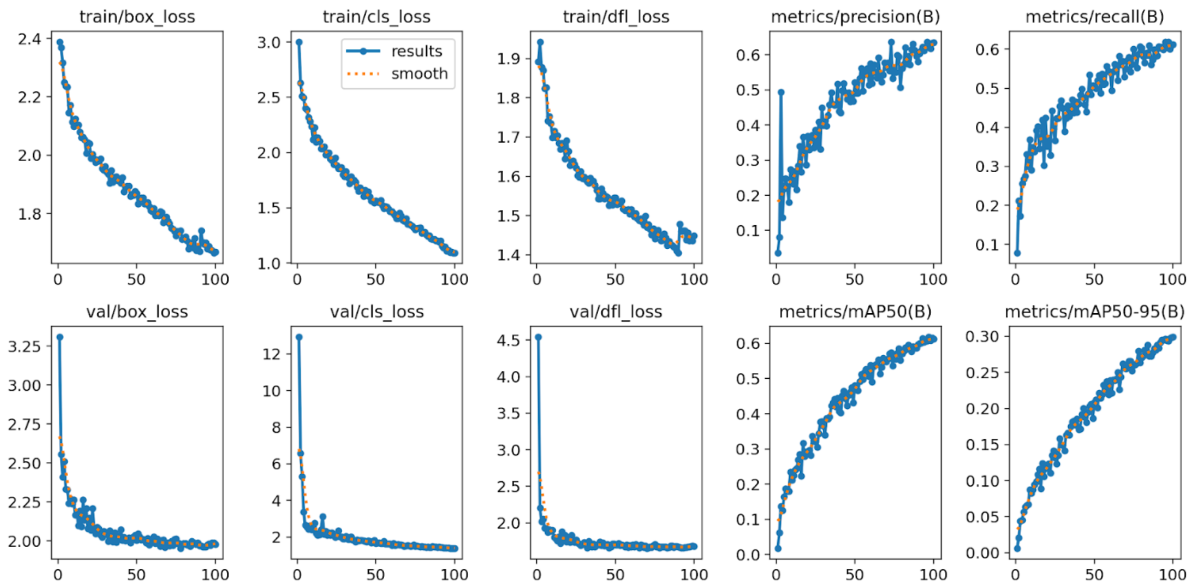


Fig 4: Performance curves during the training and validation process for YOLOv8

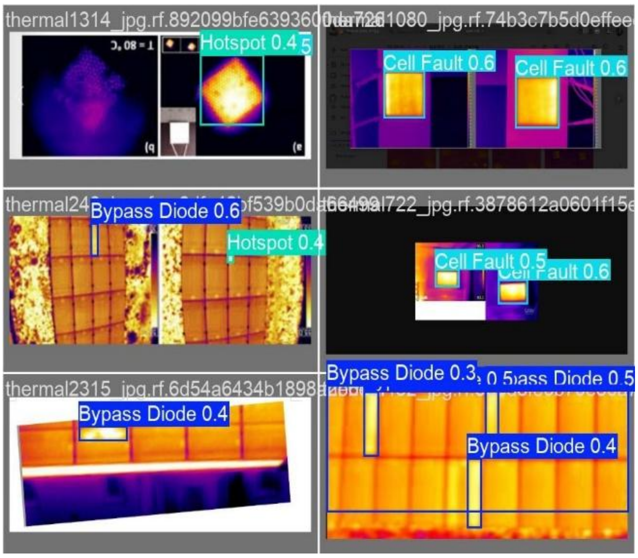


Fig 5: Result of YOLOv8 model

Fig 6: Results of MobileVit-SE model

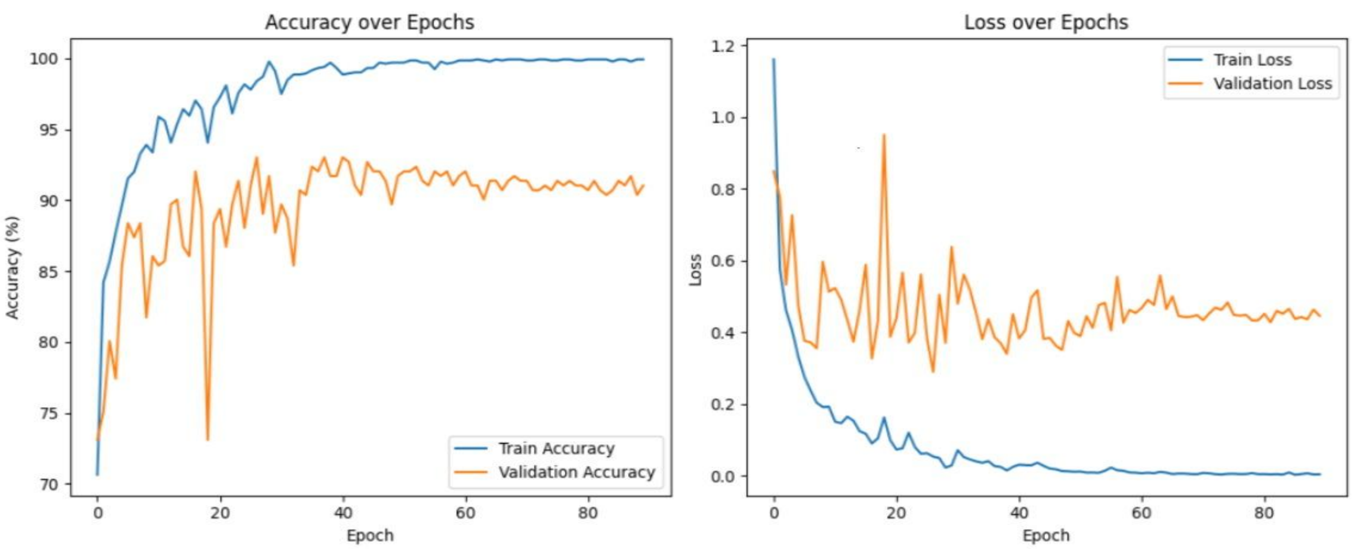


Fig 7: Accuracy and loss plot of MobileVit-SE model

V. CONCLUSION

While our system for finding solar panel defects works well, it has its limits.As We intentionally chose faster, lightweight models over the most powerful ones for visual images, testing new panel samples could be a challenge. we sacrificed some potential accuracy that larger models could offer.The system also isn't ready for messy, real-world situations like moving shadows, or for handling panels that have multiple overlapping defects at the same time.The future direction of this research is on improving implementing a Late Fusion of MobileVit and YOLOv8 model and develop a real-time hybrid detection system for defect classification captured in FLIR C5 thermal imaging camera.

This work has been successful in identifying defects in solar panel inspections by an two part system. For thermal images, our YOLOv8 model proved highly effective, demonstrating high precision and recall in locating defects like hotspots. For Visual images, the MobileVit-SE model used to identify issues such as dust, cracks, and electrical damages. These outcomes shows that the dual-model approach is powerful and reliable solution for modern solar farm maintenance.

VI. REFERENCES

[1] D. P. Winston *et al.*, 'Solar PV's Micro Crack and Hotspots Detection Technique Using NN and SVM', *IEEE Access*, vol. 9, pp. 127259–127269, 2021, doi: 10.1109/ACCESS.2021.3111904.

[2] J. P. C. Barnabé, L. P. Jiménez, G. Fraidenraich, E. R. De Lima, and H. F. D. Santos, 'Quantification of Damages and Classification of Flaws in Mono-Crystalline Photovoltaic Cells Through the Application of Vision Transformers', *IEEE Access*, vol. 11, pp. 112334–112347, 2023, doi: 10.1109/ACCESS.2023.3322653.

[3] 'A Single-Stage Photovoltaic Module Defect Detection Method Based on Optimized YOLOv8 | IEEE Journals & Magazine | IEEE Xplore'. Accessed: Nov. 03, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/10877827

[4] S. P. Pathak and S. A. Patil, 'Evaluation of Effect of Pre-Processing Techniques in Solar Panel Fault Detection', *IEEE Access*, vol. 11, pp. 72848–72860, 2023, doi: 10.1109/ACCESS.2023.3293756.

[5] S. Alam, S. Kaushik, S. M. Shaique, and N. Rafiuddin, 'PV Fault Detection Using CNN For Enhancing Reliability Of Solar Power Plants', in *2024 IEEE Third International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES)*, Apr. 2024, pp. 913–917. doi: 10.1109/ICPEICES62430.2024.10719330.

[6] S. Mehta and M. Rastegari, 'MobileViT: Light-weight, General-purpose, and Mobile-friendly Vision Transformer', Mar. 04, 2022, *arXiv*: arXiv:2110.02178. doi: 10.48550/arXiv.2110.02178.

[7] L. Zhang, X. Wu, Z. Liu, P. Yu, and M. Yang, 'ESD-YOLOv8: An Efficient Solar Cell Fault Detection Model Based on YOLOv8', *IEEE Access*, vol. 12, pp. 138801–138815, 2024, doi: 10.1109/ACCESS.2024.3466209.

[8] 'Classification of High Resolution Remote Sensing Images using Deep Learning Techniques | IEEE Conference Publication | IEEE Xplore'. Accessed: Nov. 03, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/8554605

[9] N. Drir, A. Mellit, and M. Bettayeb, 'A Novel Ensemble CNN Framework With Weighted Feature Fusion for Fault Diagnosis of Photovoltaic Modules Using Thermography Images', *IEEE Journal of Photovoltaics*, vol. 15, no. 1, pp. 146–154, Jan. 2025, doi: 10.1109/JPHOTOV.2024.3492283.

[10] R. G, D. S. Manish, R. Narasimhan A, A. U. Dhavale, and R. R. John, 'SparkNet–A Solar Panel Fault Detection Deep Learning Model', *IEEE Access*, vol. 13, pp. 75599–75617, 2025, doi: 10.1109/ACCESS.2025.3564714.

[11] B. Tejesh, G. U. Baskar, B. N. Krishna, K. Deepa, P. V. Manitha, and V. Sailaja, 'A Comparative Analysis of Power Generation Using Solar and Fuel Cell for Charging EVs', in *2022 International Virtual Conference on Power Engineering Computing and Control: Developments in Electric Vehicles and Energy Sector for Sustainable Future (PECCON)*, May 2022, pp. 1–5. doi: 10.1109/PECCON55017.2022.9851087.

[12] P. V. Varma, B. Ramachandra Rao, M. Sarath, M. B. Kumar, G. Rahul, and K. L. Nisha, 'Dual-Axis Solar Tracking System for Enhanced Energy Utilization in Grid-Tied Systems With Net-Metering', in *2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)*, Jul. 2024, pp. 1205–1210. doi: 10.1109/AIC61668.2024.10731115.

[13] Y. Palaniappan, V. A. Vishanth, N. Santhosh, R. Karthika, and M. Ganesan, 'R - Peak Detection using Altered Pan-Tompkins Algorithm', in *2020 International Conference on Communication and Signal Processing (ICCSP)*, Jul. 2020, pp. 1501–1505. doi: 10.1109/ICCSP48568.2020.9182298.