

Photovoltaic Panel Failure Detection Using Class-Conditioned Generative Adversarial Networks

Md Moshfigure Rahman

mr00157@mix.wvu.edu

West Virginia University

Paroma Chatterjee

West Virginia University

Varsha Sen

West Virginia University

Kamrul Hasan

West Virginia University

Prashnna Gyawali

West Virginia University

Anurag K. Srivastava

West Virginia University

Article

Keywords:

Posted Date: June 30th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9238947/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Scientific Reports on August 27th, 2026.

See the published version at <https://doi.org/10.1038/s41598-026-67806-1>.

Photovoltaic Panel Failure Detection Using Class-Conditioned Generative Adversarial Networks

Md Moshfiqur Rahman ¹, Paroma Chatterjee¹, Varsha Sen¹, Kamrul Hasan¹, Prashnna Gyawali¹, and Anurag K. Srivastava¹

¹Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, USA, . email: mr00157@mix.wvu.edu

ABSTRACT

Detection of defects and failure in photovoltaic (PV) modules increasingly relies on multi-modal data. Artificial intelligence models trained on imaging data can help identifying PV defects and failures. However, the effectiveness of these imaging-based diagnostic models is frequently limited by severe class imbalance in real-world datasets, where safety-critical fault categories are significantly underrepresented. This data scarcity leads to unreliable classification performance and poor probabilistic calibration. This study proposes an Auxiliary Classifier Generative Adversarial Network (AC-GAN) framework tailored for the class-conditioned synthesis of high-fidelity thermal images of PV defects. By integrating self-attention mechanisms and spectral normalization, the model achieves stable training and preserves the subtle structural signatures characteristic of diverse thermal anomalies. Evaluated on the Infrared Solar Modules dataset, the proposed model achieves a FID of 35.4 and outperforms class-weighted and oversampling baselines across multiple downstream metrics. While augmentation yields marginally higher classification accuracy, the AC-GAN framework achieves substantially better probabilistic calibration, with an Expected Calibration Error (ECE) of 1.5%, underscoring its advantage in uncertainty-aware fault diagnosis. These results confirm that fault-conditioned synthetic image generation provides high-quality, functionally informative training samples that effectively address minority class under-representation in PV inspection datasets. This methodology offers a scalable solution for enhancing the reliability of autonomous solar operations and maintenance (O&M), ensuring that diagnostic models remain robust even when field-acquired data for rare failure modes is unavailable.

Introduction

Solar photovoltaic (PV) energy is among the fastest-growing renewable technologies worldwide, with installed capacity expanding rapidly under global decarbonization efforts¹. Despite this expansion, PV system performance and safety depend on timely detection of module-level failures. Defects such as hot spots, cell cracks, bypass diode faults, interconnect degradation, and severe soiling can reduce power output and, in extreme cases, trigger irreversible damage or fire hazards². Early fault detection is therefore essential to limit energy losses, reduce maintenance costs, and ensure reliable long-term operation³.

Infrared (IR) thermography has become the dominant non-invasive method for PV fault/ failure/ defect inspection because it reveals localized thermal anomalies associated with defective cells and bypass diodes that are not visible in conventional imagery. Large-scale inspections increasingly employ Unmanned Aerial Vehicles (UAVs) equipped with calibrated thermal cameras to survey utility-scale PV plants efficiently, producing high-resolution thermographic mosaics and georeferenced fault annotations across thousands of modules. To improve scalability and reduce manual inspection effort, recent research has explored automated PV fault detection using machine learning, particularly deep convolutional neural networks applied to infrared imagery. Deep learning frameworks have been proposed for automated PV fault detection and classification, multimodal diagnostic systems integrating heterogeneous data sources, signal-analysis-driven AI approaches, Internet-of-Things monitoring platforms with machine learning classifiers, and visual inspection models for anomaly detection on PV modules⁴⁻⁹. These approaches have demonstrated promising performance for identifying common PV defects and offer the potential for scalable, automated diagnostics across large PV installations.

However, the effectiveness of supervised learning models depends heavily on the availability of balanced and representative training datasets. In practice, UAV-based IR inspection datasets exhibit severe class imbalance: most modules operate normally, whereas critical faults such as bypass diode failures or multi-cell hot spots occur rarely and are therefore sparsely represented in collected imagery. Supervised learning models trained on such skewed data tend to bias predictions toward the dominant nominal class, leading to poor recall for minority fault categories. In PV diagnostics, these missed detections are particularly

consequential, as rare faults can cause disproportionate energy losses, accelerated degradation, and safety risks.

Constructing large, high-quality datasets of rare photovoltaic (PV) faults remains challenging. Fault occurrence is stochastic and site dependent, controlled defect induction is costly and potentially destructive, and reliable thermographic imaging requires stable environmental conditions such as consistent irradiance and limited wind variation. These constraints make systematic acquisition of diverse rare-fault thermograms expensive and operationally complex. The imbalance problem is evident in commonly used datasets. The *Infrared Solar Modules* dataset¹⁰, for example, contains more than 10,000 healthy samples but only a few hundred instances for several anomaly classes. Such imbalance reduces model generalization, increases overfitting to majority classes, and limits reliable recognition of minority fault types while also degrading predictive uncertainty calibration¹¹.

The recent emergence of generative artificial intelligence has shown strong potential for addressing data scarcity and imbalance by synthesizing realistic training samples. Similar approaches have been widely applied for data augmentation in domains where labeled datasets are difficult to obtain, such as healthcare¹² and industrial inspection¹³. In the photovoltaic domain, several studies have begun exploring generative models to produce synthetic imagery for improving PV fault detection performance. Generative Adversarial Networks and diffusion-based models have been used for synthetic PV image generation, with studies reporting improved classification performance through GAN-based augmentation¹⁴. Related approaches have examined electroluminescence imagery and semi-supervised anomaly detection frameworks^{15,16}. However, many methods lack explicit class conditioning, operate outside the infrared thermographic domain, or generate global operational variations rather than fault-specific thermal signatures. Although diffusion models offer improved controllability and synthesis quality¹⁷, their computational requirements often limit use for low-resolution industrial infrared datasets.

Recent advances in conditional generative frameworks provide a mechanism for generating class-specific imagery by incorporating label information during training. Methods such as pix2pix and CycleGAN enable image-to-image translation, while Auxiliary Classifier GAN (AC-GAN) incorporates explicit class supervision within the adversarial training process^{18–20}. Such conditional generative models offer the potential to synthesize targeted fault representations and address class imbalance in photovoltaic datasets. Building on this capability, this work develops a fault-conditioned AC-GAN for infrared PV module image synthesis. The proposed framework learns class-specific thermal distributions to generate high-fidelity samples for minority fault categories, enabling dataset balancing without destructive fault induction. We evaluate the impact of synthetic data on downstream classification by comparing conventional augmentation methods, class-weighted loss baselines, and AC-GAN-based balancing. Beyond accuracy, we analyze predictive confidence, Expected Calibration Error (ECE), and probabilistic reliability to determine whether synthetic data improves both classification performance and uncertainty estimation under realistic inspection conditions.

The main contributions of this work are summarized as follows:

- We propose a fault-conditioned Auxiliary Classifier Generative Adversarial Network (AC-GAN) specifically designed for infrared thermographic imagery of photovoltaic modules, enabling class-aware synthesis of rare fault signatures and addressing severe dataset imbalance without destructive fault induction.
- We conduct a rigorous comparative evaluation of generative augmentation against conventional data augmentation, oversampling strategies, and class-weighted loss formulations, systematically quantifying the impact of each balancing approach on downstream fault classification performance across both majority and minority fault categories.
- We demonstrate that GAN-based synthetic augmentation substantially improves predictive uncertainty quantification, achieving a 1.5% ECE, a critical property for deployment in safety-sensitive diagnostic workflows where overconfident misclassifications carry significant operational and economic consequences.

Background and Related Work

Solar module anomalies manifest as distinct thermal signatures in infrared imaging, making thermography one of the preferred inspection technique in operational PV plants²¹. Various failure mechanisms produce characteristic temperature patterns that can be observed in IR imagery. Common anomaly types include:

- **Cell-level anomalies:** Individual or multiple cells exhibiting abnormal temperature patterns due to microcracks, manufacturing defects, or material degradation.
- **Hot spots:** Localized regions with significantly elevated temperatures, often caused by partial shading, internal defects, or localized current mismatch.

- **Diode failures:** Malfunctioning bypass diodes that produce distinctive thermal signatures due to abnormal current flow.
- **Physical defects:** Structural damage such as cracks, delamination, or encapsulation failure.
- **External factors:** Environmental effects including shading, soiling, vegetation coverage, or module disconnection.

Traditional anomaly detection methods relied primarily on manual inspection or threshold-based algorithms²². Although these approaches can identify obvious defects, they lack scalability and robustness for large-scale PV installations. Recent advances in deep learning have enabled automated classification using convolutional neural networks (CNNs)²³, achieving promising results under controlled conditions. However, their performance degrades substantially when confronted with the severe class imbalance present in real-world solar farm datasets.

Class imbalance is a fundamental challenge in machine learning, where minority classes contribute disproportionately little to the overall loss function, resulting in biased models²⁴. Several strategies have been proposed to address this issue:

- **Resampling techniques:** Oversampling minority classes (e.g., SMOTE²⁵) or undersampling majority classes.
- **Cost-sensitive learning:** Applying class-weighted loss functions to heavily penalize misclassification of minority classes²⁶.
- **Data augmentation:** Increasing training diversity by applying transformations such as rotation, scaling, and flipping²⁷.

While these methods provide partial mitigation, they often introduce additional challenges. Oversampling may lead to overfitting, undersampling discards useful information, and cost-sensitive approaches require careful tuning of class weights. Consequently, synthetic data generation using deep generative models has emerged as an alternative strategy capable of increasing both the quantity and diversity of training samples.

Proposed Architecture

Generative Adversarial Networks

Generative Adversarial Networks (GANs)²⁸ have revolutionized synthetic data generation by learning to produce realistic samples through an adversarial training process. A GAN consists of two neural networks: a generator G that maps random noise to synthetic samples, and a discriminator D that attempts to distinguish between real and generated samples. The networks are trained through the minimax objective:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

Conditional GANs (cGAN) extend this framework by conditioning both the generator and discriminator on auxiliary information y (e.g., class labels)²⁹, enabling controlled generation of samples from specific classes. The conditional GAN objective becomes:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{data}(x)} [\log D(x|y)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z|y)))] \quad (2)$$

Auxiliary Classifier GANs (AC-GANs)³⁰ further enhance conditional GANs by requiring the discriminator to not only distinguish real from fake samples but also classify them into the correct category. This dual-objective design encourages the generation of high-quality, class-specific synthetic images while maintaining diversity within each class. For solar module thermal imaging, this architecture addresses the severe class imbalance problem by enabling controlled generation of minority class samples.

Proposed AC-GAN Generator Architecture

As outlined in fig. 1a below, the generator network $G : \mathbb{R}^{100} \times \{0, \dots, 11\} \rightarrow \mathbb{R}^{1 \times 24 \times 40}$ transforms a latent noise vector $\mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$ and a class label c into a synthetic thermal image $\tilde{\mathbf{x}}$. The discrete class label is embedded into a continuous representation:

$$\mathbf{e}_c = \text{Embed}(c) \in \mathbb{R}^{12} \quad (3)$$

This embedding is concatenated with the noise vector:

$$\mathbf{z}' = [\mathbf{z}, \mathbf{e}_c] \in \mathbb{R}^{112} \quad (4)$$

A fully connected layer projects this to the initial feature map:

$$\mathbf{h}_0 = \text{reshape}(\text{Linear}(\mathbf{z}')) \in \mathbb{R}^{128 \times 6 \times 10} \quad (5)$$

The generator employs two upsampling blocks, each consisting of batch normalization, bilinear upsampling, convolution, and LeakyReLU activation:

$$\mathbf{h}_1 = \text{LeakyReLU}(\text{BN}(\text{Conv}(\text{Upsample}_{\times 2}(\mathbf{h}_0)))) \in \mathbb{R}^{128 \times 12 \times 20} \quad (6)$$

$$\mathbf{h}_2 = \text{LeakyReLU}(\text{BN}(\text{Conv}(\text{Upsample}_{\times 2}(\mathbf{h}_1)))) \in \mathbb{R}^{64 \times 24 \times 40} \quad (7)$$

Between upsampling blocks, a self-attention layer captures long-range spatial dependencies. Given input features $\mathbf{h}_1$, we compute:

$$\mathbf{Q} = \text{Conv}_{1 \times 1}(\mathbf{h}_1) \in \mathbb{R}^{d \times N}, \quad d = \frac{C}{8} \quad (8)$$

$$\mathbf{K} = \text{Conv}_{1 \times 1}(\mathbf{h}_1) \in \mathbb{R}^{d \times N} \quad (9)$$

$$\mathbf{V} = \text{Conv}_{1 \times 1}(\mathbf{h}_1) \in \mathbb{R}^{C \times N} \quad (10)$$

where $C = 128$ is the number of channels and $N = H \times W$ is the number of spatial locations.

The attention map and output are:

$$\mathbf{A} = \text{softmax}(\mathbf{Q}^T \mathbf{K}) \in \mathbb{R}^{N \times N} \quad (11)$$

$$\mathbf{o} = \text{reshape}(\mathbf{V} \mathbf{A}) \in \mathbb{R}^{C \times H \times W} \quad (12)$$

The output is combined with the input through a learnable parameter γ :

$$\mathbf{h}'_1 = \gamma \cdot \mathbf{o} + \mathbf{h}_1 \quad (13)$$

where γ is initialized to 0, allowing the model to gradually learn the importance of attention.

The final layer produces the grayscale thermal image:

$$\tilde{\mathbf{x}} = \tanh(\text{Conv}_{3 \times 3}(\mathbf{h}_2)) \in \mathbb{R}^{1 \times 24 \times 40} \quad (14)$$

The tanh activation constrains outputs to $[-1, 1]$, matching the normalized range of real thermal images.

Proposed AC-GAN Discriminator Architecture

The discriminator $D : \mathbb{R}^{1 \times 24 \times 40} \rightarrow [0, 1] \times \Delta^{11}$ performs two simultaneous tasks: determining image authenticity and predicting the anomaly class, where Δ^{11} denotes the 11-dimensional probability simplex. The detailed diagram for the discriminator is provided in fig. 1b below.

The discriminator employs four convolutional blocks with spectral normalization. For a weight matrix $\mathbf{W}$, spectral normalization constrains the maximum singular value:

$$\bar{\mathbf{W}} = \frac{\mathbf{W}}{\sigma(\mathbf{W})} \quad (15)$$

where $\sigma(\mathbf{W})$ is the spectral norm (largest singular value). Each block applies the transformation:

$$\mathbf{f}_i = \text{Dropout}_{0.25}(\text{LeakyReLU}_{0.2}(\text{BN}(\text{Conv}_{\text{SN}}(\mathbf{f}_{i-1})))) \quad (16)$$

resulting in progressive downsampling: $\mathbf{f}_1 \in \mathbb{R}^{16 \times 12 \times 20}$, $\mathbf{f}_2 \in \mathbb{R}^{32 \times 6 \times 10}$, $\mathbf{f}_3 \in \mathbb{R}^{64 \times 3 \times 5}$, and $\mathbf{f}_4 \in \mathbb{R}^{128 \times 2 \times 3}$.

To ensure consistent feature dimensions, adaptive average pooling is applied:

$$\mathbf{f} = \text{flatten}(\text{AdaptiveAvgPool2d}(\mathbf{f}_4, (1, 1))) \in \mathbb{R}^{128} \quad (17)$$

The shared feature representation feeds into two separate output heads:

1. **Adversarial:** $p_{\text{real}} = \sigma(\mathbf{W}_{\text{adv}}\mathbf{f} + b_{\text{adv}})$
2. **Auxiliary classification:** $\mathbf{p}_{\text{class}} = \text{softmax}(\mathbf{W}_{\text{aux}}\mathbf{f} + \mathbf{b}_{\text{aux}})$

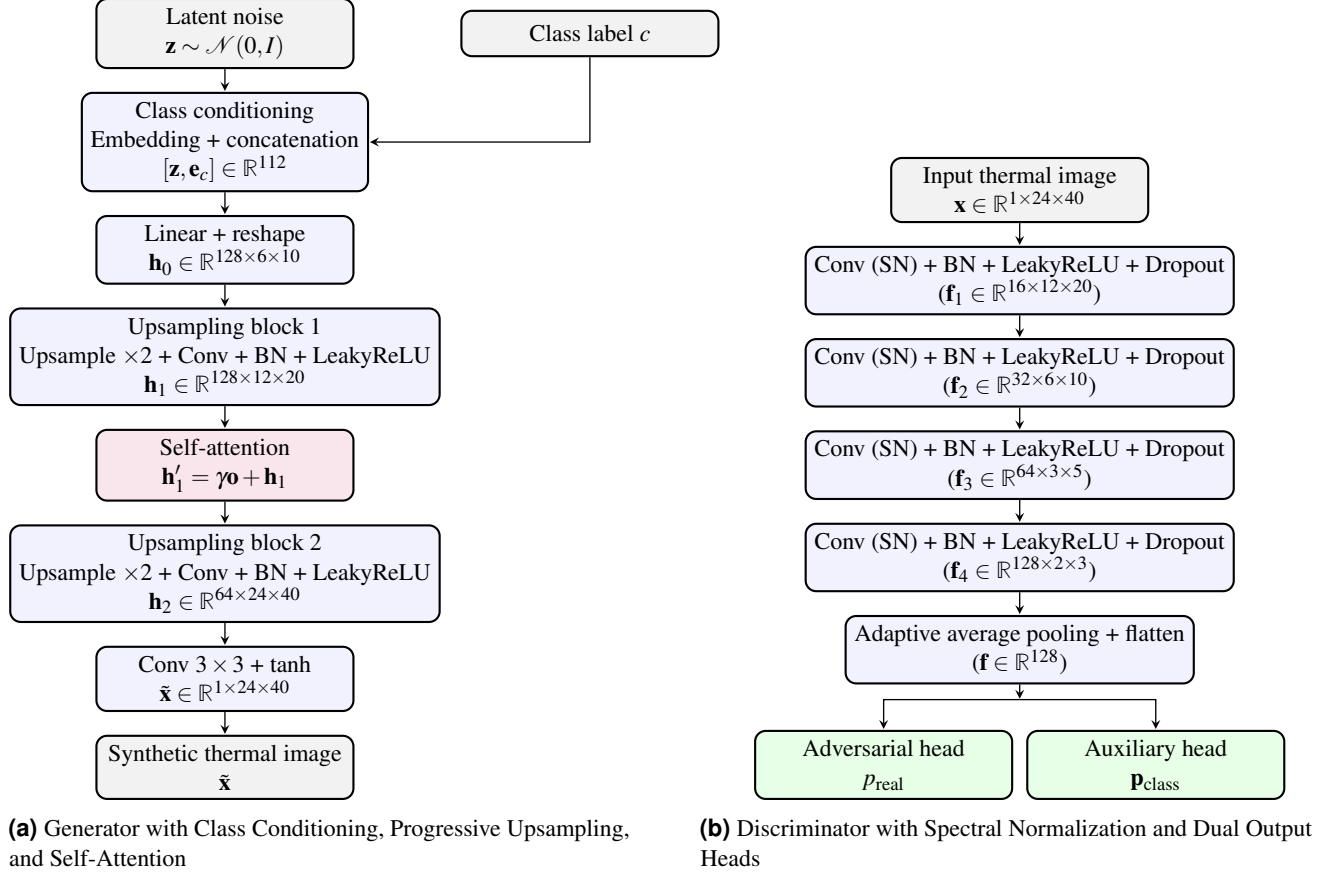


Figure 1. Illustration of Generator and Discriminator Architectures for the proposed AC-GAN Model

Loss Function Design

The AC-GAN training objective combines adversarial learning with an auxiliary classification task to encourage both realistic image synthesis and class-consistent thermal patterns. The generator and discriminator are optimized using composite loss functions consisting of adversarial and auxiliary components:

$$\mathcal{L}_G = \frac{1}{2} (\mathcal{L}_G^{\text{adv}} + \mathcal{L}_G^{\text{aux}}), \quad \mathcal{L}_D = \frac{1}{2} (\mathcal{L}_D^{\text{adv}} + \mathcal{L}_D^{\text{aux}}). \quad (18)$$

The adversarial losses follow the standard binary cross-entropy formulation. For the generator,

$$\mathcal{L}_G^{\text{adv}} = -\mathbb{E}_{\mathbf{z}, c} [\log D_{\text{real}}(G(\mathbf{z}, c))], \quad (19)$$

which encourages the generator to produce samples that the discriminator classifies as real. The auxiliary generator loss

$$\mathcal{L}_G^{\text{aux}} = -\mathbb{E}_{\mathbf{z}, c} [\log P_D(c|G(\mathbf{z}, c))] \quad (20)$$

forces the generated samples to exhibit class-consistent features.

For the discriminator, the adversarial objective is defined as

$$\mathcal{L}_D^{\text{adv}} = -\mathbb{E}_{\mathbf{x} \sim p_{\text{data}}} [\log D_{\text{real}}(\mathbf{x})] - \mathbb{E}_{\mathbf{z}, c} [\log(1 - D_{\text{real}}(G(\mathbf{z}, c)))], \quad (21)$$

while the auxiliary classification objective is

$$\mathcal{L}_D^{\text{aux}} = -\mathbb{E}_{\mathbf{x}, c \sim p_{\text{data}}} [\log P_D(c|\mathbf{x})] - \mathbb{E}_{\mathbf{z}, c} [\log P_D(c|G(\mathbf{z}, c))]. \quad (22)$$

The auxiliary loss weight $\lambda_{\text{aux}} = 0.5$ balances the objectives of generating realistic thermographic images and preserving class-specific thermal characteristics during training.

Design Principles

To ensure stable adversarial training and prevent common GAN failure modes such as discriminator collapse and gradient instability, several regularization strategies are employed. First, label smoothing is applied by setting real and fake targets to 0.9 and 0.1, respectively, which prevents the discriminator from becoming overconfident. Second, Gaussian noise with standard deviation $\sigma = 0.05$ is injected into real images, $\mathbf{x}_{\text{noisy}} = \mathbf{x} + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, 0.05^2 \mathbf{I})$, improving discriminator robustness to small perturbations. Third, gradients are clipped to a maximum ℓ_2 norm of 1.0, $\nabla \theta \leftarrow \frac{\nabla \theta}{\max(1, \|\nabla \theta\|_2)}$, to prevent unstable parameter updates during training. Finally, optimization is performed using the Adam optimizer³¹ with learning rates $\alpha_G = 10^{-4}$ for the generator and $\alpha_D = 4 \times 10^{-4}$ for the discriminator, and momentum parameters $\beta_1 = 0.5$ and $\beta_2 = 0.999$.

Table 1 summarizes the complete architecture specifications. The generator and discriminator contain approximately 1.2M and 0.8M parameters, respectively, yielding a total of 2.0M trainable parameters. This compact architecture is well-suited to the relatively small thermal image resolution while retaining sufficient capacity to capture complex thermal patterns.

Table 1. AC-GAN Architecture Specifications

Component	Layer	Output Shape
Generator	Embedding	$\mathbb{R}^{12}$
	Linear	$\mathbb{R}^{128 \times 6 \times 10}$
	Upsample + Conv	$\mathbb{R}^{128 \times 12 \times 20}$
	Self-Attention	$\mathbb{R}^{128 \times 12 \times 20}$
	Upsample + Conv	$\mathbb{R}^{64 \times 24 \times 40}$
	Conv + Tanh	$\mathbb{R}^{1 \times 24 \times 40}$
Discriminator	Conv (SN)	$\mathbb{R}^{16 \times 12 \times 20}$
	Conv (SN) + BN	$\mathbb{R}^{32 \times 6 \times 10}$
	Conv (SN) + BN	$\mathbb{R}^{64 \times 3 \times 5}$
	Conv (SN) + BN	$\mathbb{R}^{128 \times 2 \times 3}$
	AdaptiveAvgPool	$\mathbb{R}^{128}$
	FC (Adversarial)	$\mathbb{R}^1$
	FC (Auxiliary)	$\mathbb{R}^{12}$

The architecture is designed to reflect both the physical characteristics of thermographic PV imagery and the stability requirements of adversarial training. First, a self-attention module is included in the generator to capture long-range spatial dependencies in thermal patterns, which often extend across large portions of a PV module. Second, spectral normalization is applied in the discriminator to constrain the Lipschitz constant, preventing rapid discriminator dominance and improving training stability. Third, the auxiliary classification objective encourages the generator to synthesize class-specific thermal signatures rather than generic temperature distributions.

The network architecture is further adapted to the characteristics of the inspection data. The relatively low spatial resolution (24×40) reflects the thermal camera specifications commonly used in aerial solar farm inspections and motivates a compact network design with limited parameter complexity. Finally, because infrared thermography represents temperature distributions rather than color information, the model operates on single-channel inputs and outputs, enabling more efficient representation learning for thermal patterns.

Optimization Strategy

Optimization is performed using the Adam optimizer³¹ with distinct learning rates for the generator ($\alpha_G = 2 \times 10^{-4}$) and discriminator ($\alpha_D = 5 \times 10^{-4}$). The discriminator is trained with a slightly higher learning rate to ensure sufficiently informative gradients for the generator, following common GAN training practice³². The momentum parameters are set to $\beta_1 = 0.5$ and

$\beta_2 = 0.999$. Training is conducted for N epochs with a minibatch size of 64, and the procedure outlined in Algorithm 1 is executed at each iteration. To further improve training stability, gradients are clipped to an ℓ_2 norm of 1.0, and spectral normalization³³ is applied to the discriminator convolutional layers to constrain the Lipschitz constant and mitigate mode collapse.

Algorithm 1: AC-GAN Training Procedure

```

1: for  $epoch = 1$  to  $N$  do
2:   for each minibatch  $(x_{real}, y_{real})$  do
3:     Sample noise  $z \sim \mathcal{N}(0, I)$ 
4:     Sample labels  $y_{gen} \sim \text{Uniform}\{0, \dots, 11\}$ 
5:     Generate fake images:  $x_{fake} = G(z, y_{gen})$ 
6:     Update Discriminator
7:      $\mathcal{L}_D \leftarrow$  Compute discriminator loss on real and fake
8:     Backpropagate  $\mathcal{L}_D$  and update  $D$ 
9:     Update Generator
10:    Sample new noise  $z' \sim \mathcal{N}(0, I)$ 
11:     $\mathcal{L}_G \leftarrow$  Compute generator loss
12:    Backpropagate  $\mathcal{L}_G$  and update  $G$ 
13:   end for
14: end for

```

Table 2 presents the ablation study results. The quality of the generated images is evaluated using the Fréchet Inception Distance (FID)³⁴, which measures the similarity between the feature distributions of real and generated samples. Let $\mathbf{f}_r$ and $\mathbf{f}_g$ denote feature embeddings extracted from an InceptionV3 network for real and generated images, respectively. Across all experimental configurations, FID values range between 30 and 80, indicating strong alignment between the distributions of synthetic and real thermographic images.

Table 2. Ablation Study Results

Configuration	FID
Full Model (Ours)	35.4
w/o Self-Attention	52.3
w/o Spectral Normalization	68.7
w/o Label Smoothing	47.2
w/o Gradient Clipping	61.5
Standard Conditional GAN	124.8

Downstream Classification Evaluation and Calibration Analysis

In this section, we evaluate the utility of the generated images by measuring their impact on a downstream diagnostic classifier. Beyond standard top-1 classification accuracy, we conduct a granular analysis of how synthetic data influences the model’s decision-making process. Specifically, we investigate the relationship between classifier confidence and empirical correctness through reliability diagrams and confidence-versus-correctness curves. By calculating the Expected Calibration Error (ECE), we quantify the extent to which the generative framework improves the model’s “uncertainty awareness”—a critical requirement for deploying autonomous inspection systems in safety-critical PV environments.

Classifier and Training Procedure

The classification backbone is based on MobileNetV3-Large, a state-of-the-art lightweight convolutional neural network optimized for efficient visual recognition³⁵. It was selected due to its strong feature extraction capability on small- and medium-scale image datasets, as well as its proven robustness in transfer learning scenarios. These characteristics are particularly important for infrared photovoltaic imagery, where anomaly patterns can be subtle, and class distributions are often imbalanced.

The backbone was initialized with ImageNet-pretrained weights to leverage generic low-level and mid-level visual features. The final classification layer was replaced with a task-specific classifier head, consisting of a fully connected layer matching the number of anomaly classes. This modification enables the network to adapt to the photovoltaic anomaly classification task while retaining the representational strengths of the pretrained backbone.

To ensure stable optimization and effective utilization of both real and synthetic data, training was conducted in three distinct phases using a staged fine-tuning strategy. This approach progressively increases model flexibility while mitigating the risk of overfitting, particularly when synthetic data is introduced.

Phase 1 Classifier Training: All backbone layers were frozen, and only the classifier head was trained. Optimization was performed using the Adam optimizer with an initial learning rate of 3×10^{-4} . A ReduceLROnPlateau scheduler was employed with a reduction factor of 0.3 and a patience of 3 epochs, allowing adaptive learning rate control based on validation performance.

Phase 2 Partial Fine-Tuning: Feature blocks 6 through 12 of the MobileNetV3 backbone were unfrozen to enable partial adaptation of higher-level representations. Training continued with a reduced learning rate of 1×10^{-4} , using the same scheduler configuration. This phase allows the model to refine discriminative features relevant to photovoltaic anomalies while preserving stable lower-level representations.

Phase 3 Full Fine-Tuning: All network layers were unfrozen for global fine-tuning. The model was optimized using the Adam optimizer with a further reduced learning rate of 5×10^{-5} . A CosineAnnealingLR scheduler was employed during this phase to facilitate smooth convergence and prevent abrupt parameter updates.

To improve probabilistic calibration and reduce overconfidence, label smoothing with a factor of 0.1 was applied during training. This regularization technique encourages softer target distributions, which is particularly beneficial when training with synthetic data and evaluating confidence-based metrics such as ECE. This staged training strategy was applied consistently across both augmentation-based and AC-GAN-based datasets, ensuring that observed performance differences arise from data source characteristics rather than optimization discrepancies.

Dataset and Class Balancing

The experiments use the *InfraredSolarModules* dataset released by Raptor Maps Inc.¹⁰, which was developed for large-scale photovoltaic inspection using infrared thermography. The dataset contains approximately 20,000 UAV-acquired thermal images of individual solar modules captured during aerial inspection of utility-scale solar farms. Each image has a spatial resolution of 24×40 pixels, reflecting the limited spatial detail typical of drone-based thermal imaging systems. The dataset defines 12 classes, including 11 anomaly categories and a *No-Anomaly* class representing nominal module operation. The anomaly classes span a range of defect types, including cell-level hot spots, multi-cell hot spots, cracking, shading, bypass diode activation, vegetation occlusion, soiling, and offline modules. The dataset exhibits pronounced class imbalance consistent with real-world inspection distributions. The *No-Anomaly* class contains 10,000 samples (48.54%), whereas several critical fault categories contain only a few hundred images. The rarest class, *Diode-Multi*, contains 175 samples, corresponding to an imbalance ratio of approximately 57:1 between the majority and minority classes. This extreme imbalance motivates the synthetic data generation framework developed in this work.

For each anomaly class, the dataset was divided into a training set (80%) and a strictly held-out test set (20%). We evaluated three training regimes. First, an imbalanced baseline was considered using two approaches: training on the original dataset with class-weighted cross-entropy to penalize minority-class errors, and oversampling minority-class images to partially mitigate imbalance. Second, augmentation-based class balancing was performed using label-preserving transformations, including horizontal flipping, random rotation ($\pm 10^\circ$), translation (up to 5%), mild Gaussian noise injection, and intensity scaling with clamping. Third, GAN-based data generation was applied by augmenting the training set with synthetic samples produced by a pretrained class-conditional generative adversarial network. Generator checkpoints from different training stages were used to analyze how generative quality influences downstream classification performance.

Experimental Results

The comparative performance of the four training strategies is summarized in Table 3. While top-1 accuracy provides a general overview, a granular examination of the confusion matrices in Figure 2 reveals how each strategy handles the high-dimensional complexity of PV anomalies.

Table 3. Overall performance comparison for the 12-class classifier

Training Strategy	Accuracy (%)	Final Validation Loss
Baseline Class-Weighted	67.97%	2.1747
Baseline Oversampling	70.50%	1.0260
Augmentation-Based	82.12%	0.9064
AC-GAN-Based	80.45%	0.6987

Performance without Synthetic Balancing Training on the original imbalanced dataset resulted in the lowest overall performance, with the class-weighted cross-entropy approach yielding only 67.97% accuracy. As shown in the confusion matrices in Figures 2a and 2b, these baseline models struggle significantly with minority classes, often misclassifying rare faults as majority classes. While class weighting penalizes errors, it fails to provide the intra-class diversity required for robust decision boundaries. Oversampling (70.50% accuracy) similarly falls short, as it leads to overfitting on repeated samples rather than introducing new informative variations.

The lack of representative data is further evidenced in Figure 3(a). The confidence vs. correctness analysis reveals a severely miscalibrated model where predicted scores are heavily concentrated in lower probability bins, even for correct predictions. This indicates a model that is consistently uncertain, resulting in near-random behavior for underrepresented fault categories.

Augmentation-Based Balancing Conventional geometric augmentation significantly improved accuracy to 82.12%. By perturbing existing samples, the model achieves better class separability, as seen in the cleaner diagonal of the confusion matrix in Figure 2c. However, this strategy achieved a higher validation loss (0.9064) compared to the GAN approach. This suggests that while simple perturbations like rotation and flipping help the model memorize boundaries, they do not necessarily refine the model’s internal probability estimates or capture the true underlying distribution of thermal anomalies.

GAN-Based Balancing The GAN-based strategy yielded a competitive accuracy of 80.45% and, crucially, achieved the lowest validation loss of all tested methods (0.6987). As illustrated in Figure 2d, the AC-GAN balancing allows the model to maintain high precision across both majority and minority classes without the overfitting associated with oversampling.

Most importantly, the GAN-supplemented model exhibited superior probabilistic reliability. As illustrated in Figure 3(b), the GAN-based model demonstrates a much stronger alignment between confidence and correctness. It is significantly more confident when correct and shows a steep decline in confidence when incorrect. This confirms that the AC-GAN introduces new, physically plausible thermal patterns that help the classifier define more meaningful probabilistic boundaries—a critical requirement for autonomous PV health monitoring.

11-Class Classifier

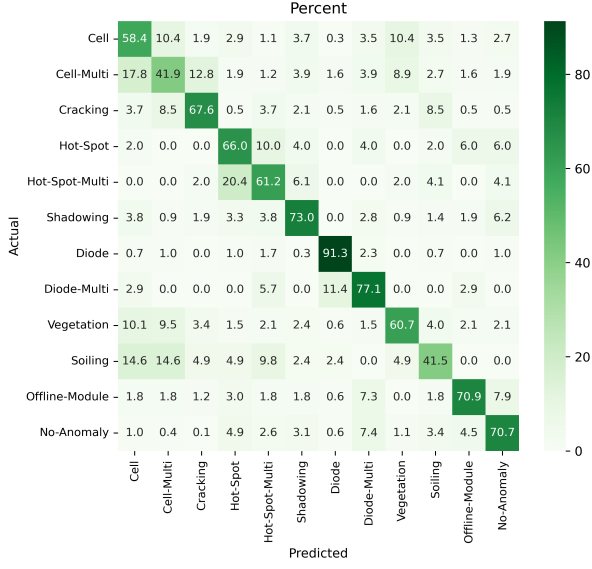
The second evaluates an **11-class classifier**, where the normal class is excluded and only anomaly categories are considered. This separation allows a more focused analysis of anomaly discriminability without the influence of dominant normal samples. Detailed per-class performance results are reported in Tables 4 and 5. The augmentation-based model achieves a slightly higher overall accuracy (69% vs. 68%), while the AC-GAN-based model exhibits comparable class-wise performance across most anomaly categories.

Table 4. Per-class performance (Augmented)

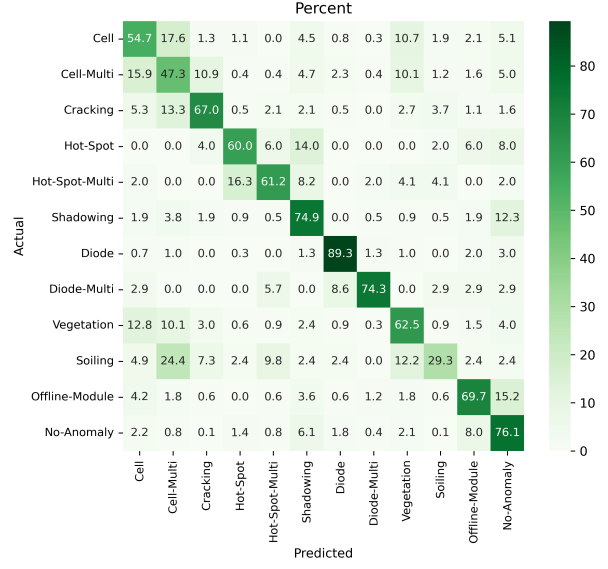
Class	P	R	F1	Sup
Cell	0.64	0.60	0.62	375
Cell-Multi	0.50	0.48	0.49	258
Cracking	0.69	0.71	0.70	188
Hot-Spot	0.62	0.66	0.64	50
Hot-Spot-M	0.60	0.51	0.55	49
Shadowing	0.74	0.85	0.79	211
Diode	0.91	0.92	0.92	300
Diode-Multi	0.81	0.71	0.76	35
Vegetation	0.67	0.66	0.67	328
Soiling	0.39	0.29	0.33	41
Offline-Mod	0.73	0.82	0.77	165
Acc.	0.69			

Table 5. Per-class performance (AC-GAN)

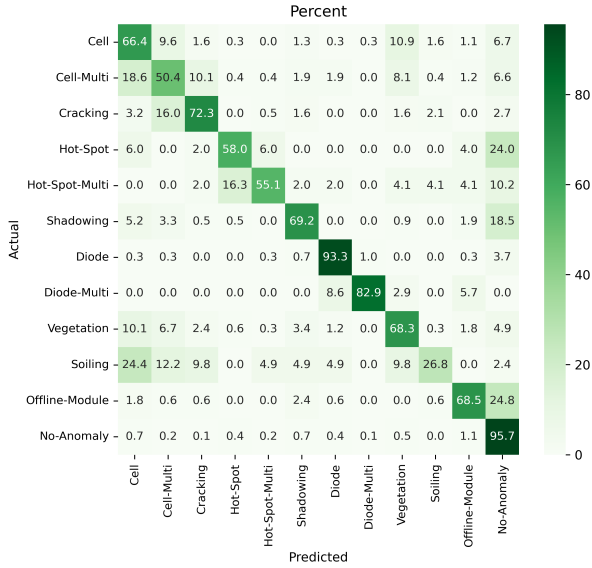
Class	P	R	F1	Sup
Cell	0.64	0.60	0.62	375
Cell-Multi	0.52	0.50	0.51	258
Cracking	0.68	0.71	0.69	188
Hot-Spot	0.60	0.62	0.61	50
Hot-Spot-M	0.63	0.53	0.58	49
Shadowing	0.66	0.80	0.72	211
Diode	0.91	0.91	0.91	300
Diode-Multi	0.79	0.63	0.70	35
Vegetation	0.69	0.64	0.66	328
Soiling	0.36	0.22	0.27	41
Offline-Mod	0.70	0.81	0.75	165
Acc.	0.68			



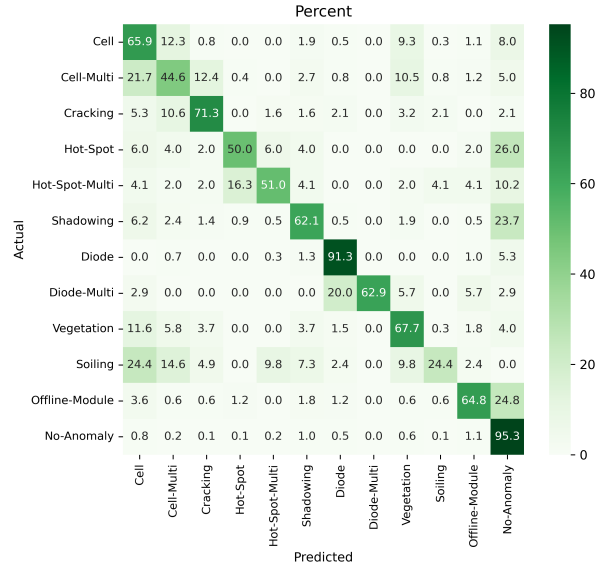
(a) Baseline: Class-Weighted



(b) Baseline: Oversampling



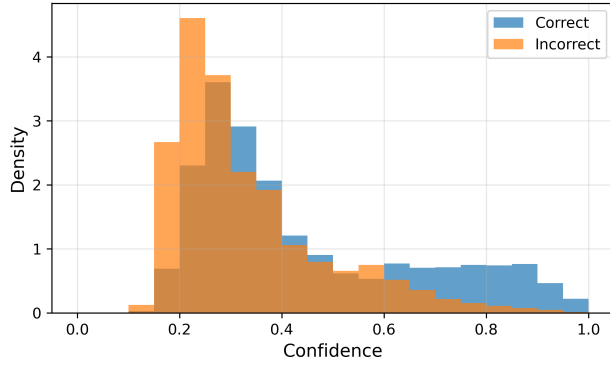
(c) Augmentation-Based



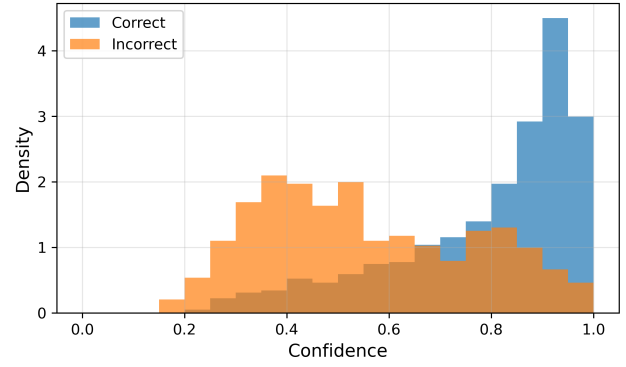
(d) AC-GAN-Based

Figure 2. Confusion matrices for the four training strategies. Each matrix shows the classification performance across the 12 anomaly classes.

Across both, augmentation-based training demonstrates a consistent advantage in top-1 accuracy. In contrast, AC-GAN-based training yields lower validation loss and more stable confidence behavior, indicating improved alignment between predicted probabilities and true labels. These results suggest that while augmentation remains effective for maximizing discriminative performance, AC-GAN-generated samples provide complementary benefits not fully captured by accuracy alone, particularly in terms of optimization stability and confidence reliability.

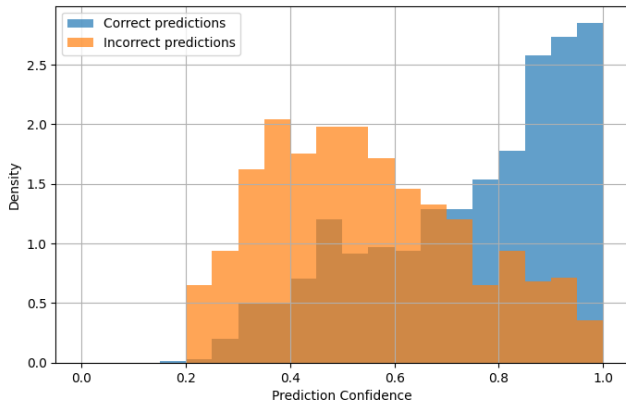


(a) Baseline: Class-Weighted

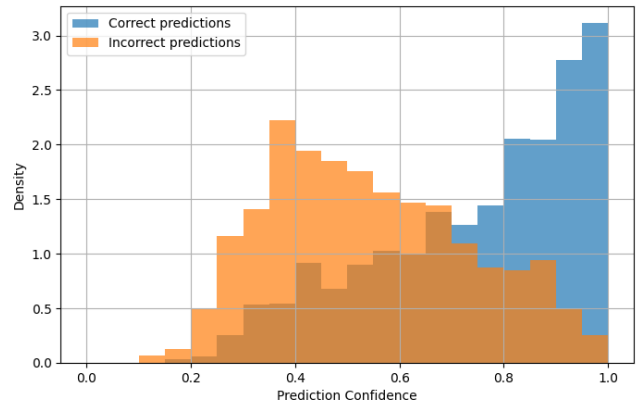


(b) GAN-Based Balancing

Figure 3. Confidence vs. Correctness (Reliability) diagrams. (a) Without synthesis, the model shows low confidence even when correct. (b) With GAN-based synthesis, the model shows high confidence for correct predictions and appropriately low confidence for errors.



(a) Augmentation-based training.



(b) AC-GAN-based training.

Figure 4. Confidence vs. Correctness (Reliability) diagrams under augmentation-based and AC-GAN-based training for 11 class classifier.

Confidence Calibration Analysis

Beyond accuracy, confidence calibration was analyzed to assess the reliability of predicted probabilities, which is critical in anomaly diagnosis where overconfident errors may lead to unsafe decisions. All calibration analyses were conducted using the 11-class anomaly-only classifier to isolate fault discrimination behavior.

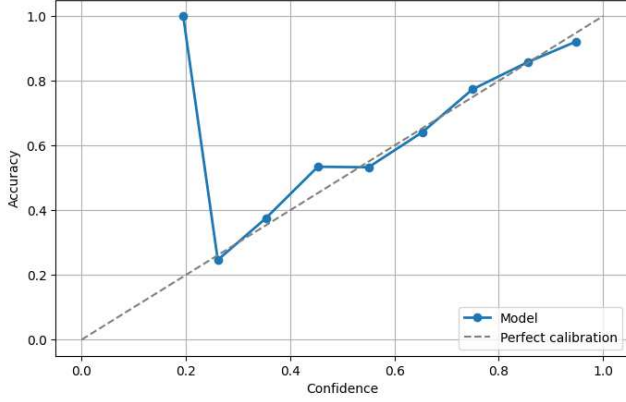
Confidence Distribution: Correct vs. Incorrect Predictions

The distribution of maximum softmax confidence was examined separately for correctly and incorrectly classified test samples to evaluate whether model confidence aligns with prediction correctness.

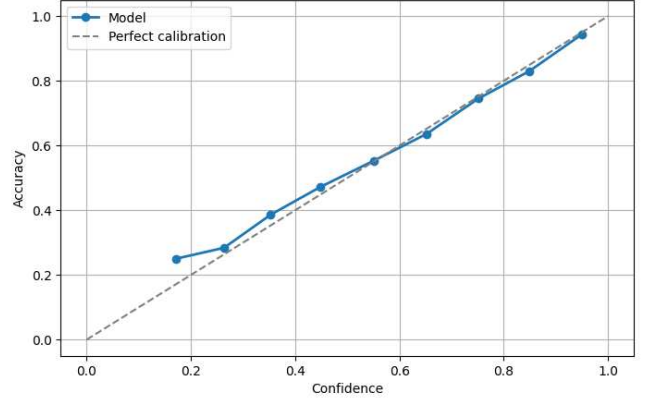
As shown in Fig. 4, the AC-GAN-based classifier exhibits clearer separation between correct and incorrect confidence distributions, with misclassified samples receiving lower confidence scores. In contrast, the augmentation-based model shows greater overlap, indicating a higher degree of overconfidence.

Reliability Diagrams

Reliability diagrams were used to assess the agreement between predicted confidence levels and empirical classification accuracy. Model predictions were grouped into M disjoint confidence bins spanning the interval $[0, 1]$. For each bin, the empirical accuracy was computed as the fraction of correctly classified samples, while the average predicted confidence was



(a) Augmentation-based training.



(b) AC-GAN-based training.

Figure 5. Reliability diagrams for augmentation-based and AC-GAN-based classifiers. The diagonal indicates perfect calibration.

obtained by averaging the maximum softmax probabilities of samples within the bin. A perfectly calibrated classifier exhibits a diagonal reliability curve, indicating alignment between predicted confidence and true correctness likelihood.

As shown in Fig. 5, the AC-GAN based classifier follows the ideal calibration trend more closely, particularly in the mid-to-high confidence regime, indicating improved probability estimation despite slightly lower overall accuracy.

Table 6. Expected Calibration Error (ECE) comparison.

Training Strategy	ECE
Augmentation-Based	0.0256
AC-GAN-Based	0.0150

Expected Calibration Error (ECE)

Calibration quality was further quantified using ECE³⁶, which measures the average discrepancy between predicted confidence and empirical accuracy across all confidence bins. Lower ECE values indicate better alignment between predicted probabilities and true outcomes.

Discussion

This research addresses a fundamental bottleneck in the automated inspection of solar infrastructure: the scarcity of high-fidelity, labeled data for rare but critical PV failure modes. By developing and validating a class-conditioned AC-GAN, this work provides a robust framework for synthesizing infrared (IR) thermography data that effectively mitigates the challenges of class imbalance. The achieved FID of 35.4 indicates that the integration of self-attention mechanisms and spectral normalization allows the model to capture the nuanced structural signatures of rare faults such as PID or cell-level cracks, that are often blurred or lost in standard generative models. Furthermore, the integration of these synthetic samples into a downstream classifier resulted in superior performance compared to weighted cross-entropy and oversampling baselines across all evaluated metrics. Specifically, the GAN-based approach achieved a classification accuracy of 80.45% with a validation loss of 0.6987, significantly outperforming the class-weighted baseline (67.97% accuracy; 2.1747 loss) and the oversampling baseline (70.50% accuracy; 1.0260 loss). These quantitative improvements are further substantiated by the confidence-versus-correctness analysis, which demonstrates that the generative framework produces more reliable probability estimates than the baselines.

Crucially, the results highlight a vital trade-off between raw classification accuracy and probabilistic calibration. While traditional augmentation techniques may yield marginally higher accuracy scores, they often produce models that exhibit high ECE. In safety-critical applications like autonomous aerial PV inspections, a model with slightly lower raw accuracy but superior calibration (as evidenced by the 1.5% ECE) is fundamentally more reliable than a higher-accuracy model that is poorly calibrated. This framework ensures that predicted confidence scores closely align with the true likelihood of a fault, providing a necessary layer of trust for field-scale deployment.

However, it is important to acknowledge that there are certain limitations of this study. First, the current study is limited to

the Infrared Solar Modules dataset; the generalizability of these synthetic signatures across varying camera resolutions and different climatic conditions remains a subject for further validation. Furthermore, this work primarily focuses on GAN-based approaches and does not explore other generative models such as diffusion models, largely due to their higher computational requirements. However, with the growing body of research aimed at improving the computational efficiency of diffusion-based models^{37,38}, future work should consider leveraging these powerful generative frameworks as well.

Conclusion

This study introduced a class-conditioned AC-GAN framework tailored for the synthesis of high-fidelity thermal images to address severe class imbalance in PV defect datasets. By incorporating self-attention and spectral normalization, the model successfully generates rare fault signatures that improve the diagnostic capabilities of downstream classifiers. The experimental results demonstrate that the generative-based approach provides a critical advantage over traditional data augmentation by significantly improving probabilistic calibration.

References

1. Kakoulaki, G. *et al.* Mapping europe's rooftop photovoltaic potential with a building-level database. *Nat. Energy* **11**, 324–333 (2026).
2. Rahman, M. M. & Srivastava, A. K. Detection, classification, and localization of faults and failures in photovoltaic arrays: A review, comparative analysis, and path forward. *Discov. Energy* (2026).
3. Basnet, B. & Sen, V. Robust fault detection and classification in power systems via physics-informed and data-driven learning. *J. Electr. Eng.* **7**, 241–259 (2025).
4. Ahmed, W. *et al.* Photovoltaic panels classification using isolated and transfer learned deep neural models using infrared thermographic images. *Sensors* **21**, 5668 (2021).
5. Talaat, F. M., Salem, M. & Shaban, W. M. Ai-driven fault detection and classification in photovoltaic systems using deep learning techniques. *Sci. Reports* **16**, 8727 (2026).
6. Yu, S. *et al.* Multimodal fault detection model for photovoltaic modules. *Sci. Reports* **16**, 3278 (2026).
7. Seghiour, A. *et al.* Fault detection and diagnosis in photovoltaic systems using artificial intelligence and time–frequency analysis. *Sci. Reports* **16**, 39386 (2026).
8. Srisiri, W. *et al.* Artificial intelligence-based fault classification on photovoltaic plants using a low-cost open-source iot system. *Sci. Reports* **16**, 1110 (2026).
9. Jin, H., Li, M., Lv, X. & Peng, J. Foreign object detection on photovoltaic panels based on dhl-net. *Sci. Reports* **16**, 8145 (2026).
10. Millendorf, M., Obropta, E. & Vadhavkar, N. Infrared solar module dataset for anomaly detection. In *International Conference on Learning Representations* (2020).
11. Tang, C., Ren, H., Xia, J., Wang, F. & Lu, J. Automatic defect identification of pv panels with ir images through unmanned aircraft. *IET Renew. Power Gener.* **17**, 3108–3119 (2023).
12. Kim, M. *et al.* Synthesizing realistic high-resolution retina image by style-based generative adversarial network and its utilization. *Sci. Reports* **12**, 17307 (2022).
13. Deng, F. *et al.* Dg2gan: improving defect recognition performance with generated defect image sample. *Sci. Reports* **14**, 14787 (2024).
14. da Silveira Junior, C. R., Sousa, C. E. R. & Alves, R. H. F. Automatic fault classification in photovoltaic modules using denoising diffusion probabilistic model, generative adversarial networks, and convolutional neural networks. *Energies* **18**, 776 (2025).
15. Ebied, M. A. *et al.* Advanced deep learning modeling to enhance detection of defective photovoltaic cells in electroluminescence images. *Sci. Reports* **15**, 31640 (2025).
16. Lu, F., Niu, R., Zhang, Z., Guo, L. & Chen, J. A generative adversarial network-based fault detection approach for photovoltaic panel. *Appl. Sci.* **12**, 1789 (2022).
17. Pan, H., Pei, W., Li, X. & He, Z. Unified conditional image generation for visible–infrared person re-identification. *IEEE Transactions on Inf. Forensics Secur.* **19** (2024).

18. Isola, P., Zhu, J.-Y., Zhou, T. & Efros, A. A. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1125–1134 (2017).
19. Zhu, J.-Y., Park, T., Isola, P. & Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision*, 2242–2251 (2017).
20. Odena, A., Olah, C. & Shlens, J. Conditional image synthesis with auxiliary classifier gans. In *Proceedings of the International Conference on Machine Learning*, 2642–2651 (2017).
21. Aghaei, M. *et al.* Review of degradation and failure phenomena in photovoltaic modules. *Renew. Sustain. Energy Rev.* **159**, 112160 (2022).
22. Kovashka, A., Russakovsky, O., Fei-Fei, L. & Grauman, K. Crowdsourcing in computer vision. *Foundations Trends Comput. Graph. Vis.* **10**, 177–243 (2015).
23. Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **25** (2012).
24. He, H. & Garcia, E. A. Learning from imbalanced data. *IEEE Transactions on Knowl. Data Eng.* **21**, 1263–1284 (2009).
25. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. Smote: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).
26. Thai-Nghe, N., Gantner, Z. & Schmidt-Thieme, L. Cost sensitive learning for imbalanced classification problems. *J. Mach. Learn. Res.* **11**, 383–392 (2010).
27. Galar, M., Fernandez, A., Barrenechea, E., Bustince, H. & Herrera, F. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Syst. Man, Cybern. Part C (Applications Rev.)* **42**, 463–484 (2012).
28. Goodfellow, I. *et al.* Generative adversarial nets. In *Advances in Neural Information Processing Systems*, vol. 27 (2014).
29. Mirza, M. & Osindero, S. Conditional generative adversarial nets. In *arXiv preprint arXiv:1411.1784* (2014).
30. Odena, A., Olah, C. & Shlens, J. Conditional image synthesis with auxiliary classifier gans. In *International Conference on Machine Learning*, 2642–2651 (PMLR, 2017).
31. Kingma, D. P. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
32. Salimans, T. *et al.* Improved techniques for training gans. In *Advances in Neural Information Processing Systems*, vol. 29 (2016).
33. Miyato, T., Kataoka, T., Koyama, M. & Yoshida, Y. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations* (2018).
34. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B. & Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, vol. 30 (2017).
35. Howard, A. *et al.* Searching for mobilenetv3. In *Proceedings of the IEEE/CVF international conference on computer vision*, 1314–1324 (2019).
36. Naeini, M. P., Cooper, G. & Hauskrecht, M. Obtaining well calibrated probabilities using bayesian binning. In *Proceedings of the AAAI conference on artificial intelligence*, vol. 29 (2015).
37. Yin, T. *et al.* One-step diffusion with distribution matching distillation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 6613–6623 (2024).
38. Zhou, L., Ermon, S. & Song, J. Inductive moment matching. *arXiv preprint arXiv:2503.07565* (2025).

Data availability

The dataset is publicly available at: Millendorf, M., Obropta, E. & Vadhavkar, N. Infrared solar module dataset for anomaly detection. In *International Conference on Learning Representations* (2020). <https://github.com/RaptorMaps/InfraredSolarModules?tab=readme-ov-file>

Acknowledgments (not compulsory)

Author contributions statement

M.M.R., P.C., V.S. and K.H. contributed to the conceptualization, methodology, software, visualization, and investigation. A.K.S. and P.G. contributed to the conceptualization, provided resources, were involved in writing—review & editing, and provided supervision and project administration. All authors have read and agreed to the published version of the manuscript.

Competing interests

The authors declare no competing interests.

Funding

This work was supported in part by the US Department of Energy under Grant DE-EE0010166.

Additional information

To include, in this order: **Accession codes** (where applicable);