International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(14s) 2026 pp. 558-573
https://doi.org/10.70917/ijcisim-2026-4248
Article

# Meta-Learned Uncertainty-Gated Physics-Guided Multimodal Deep Learning for Robust Photovoltaic Defect Detection and Maintenance Decision Support

**Smita D Khandagale[1], Sanjay M. Patil[2]**

[1 2] Datta Meghe College of Engineering, Airoli, Navi-Mumbai(MS) INDIA
Email: sdkhandagale@vpmthane.org, sanjay.patil@dmce.ac.in

**Abstract:** Photovoltaic (PV) defect diagnosis under real-world operating conditions is difficult due to the lack of available sensor modalities, their corruption, or its variability in signal quality. The traditional multimodal fusion approaches typically require all the sensors to be available, while fixed-weight physics-informed models impose possibly unreliable physical priors when essential modalities are missing. To overcome these challenges, in this study, a robust PV defect detection, power loss estimation, open-set defect recognition and maintenance decision support framework for PV systems is proposed based on a multimodal deep learning approach which incorporates uncertainty gating and physics guidance, termed MetaUG-PhyMOC-PVNet. The framework consists of a multimodal fusion backbone based on a transformer architecture, which handles variable-modality inputs, and combines electroluminescence images, infrared thermography, RGB images, and current–voltage characteristics. A light weight gating network is used as an estimation of the reliability of physics informed loss using modality-availability masks and modality-specific quality indicators. A bilevel meta-reweighting strategy is used in which the multimodal backbone is trained in the inner loop and the gating parameters are learned in the outer loop on a high quality validation set, which optimizes the gating mechanism. A single-diode model is developed that is differentiable, thus establishing physical consistency among the electroluminescence intensity and local shunt resistance. Also the learned gating weight is used to control the confidence of an evidential open-set recognition head, which allows the uncertain or unseen defects to be passed for human inspection. Experimental results show that the proposed framework can obtain an mAP of 0.86 under the missing-modality condition, whereas the fixed-weight physics-informed framework and generative imputation can only reach mAPs of 0.62 and 0.78 respectively. It also decreases the prediction error of power loss to RMSE (7.1 W) and MAPE (4.3%), and the recognition accuracy to 84.3% for open set and AUROC to 0.91. The results show that adaptive physics gating enhances robustness to partially and/or corrupted sensor data, and enables a reliable prioritization of PV maintenance and human-in-the-loop decision making.

**Keywords:** Photovoltaic defect detection; multimodal deep learning; physics-guided learning; uncertainty gating; open-set recognition.

## 1. Introduction

With the goal of the global energy transition to renewable energy, worldwide PV systems have to operate reliably over extended periods of time, but a large number of defects, including cracks, interruptions of the fingers or potential induced degradation [1] can impair this. Thus, the automated detection of defects and the prediction of failures have become indispensable towards the cost-effective operation and prevention of catastrophic failures. The first generation of PV defect detection techniques were based on the classical machine learning models, which used either electroluminescence (EL) or infrared imagery [3] to detect defects, but these approaches were not applicable for diverse types of defects and variable field conditions. Deep learning, especially Convolutional Neural Networks (CNNs), have enhanced the classification accuracy of common defects [4] but most of such models are based on single-modality data and do not make use of the complementary information provided by electrical measurements like current-voltage (I-V) measurements.

Multimodal deep learning frameworks combining visual and electrical data have shown to provide better diagnostic performance by simultaneously capturing both spatial and electrical signatures of degradation [5]. But these traditional fusion methods assume full availability, which is seldom satisfied in the real PV installation where data may be lost, sensors may fail or communication channels may be turned off by environmental conditions. A promising approach to solve this problem is missing modality learning, which introduces modality dropout in training, or generative imputation to synthesize missing input modalities [6]. However, the approaches are rarely reliable if a whole modality is missing, and they don't make any explicit statement of the physical consistency of their output.

Complementary, physics-informed neural networks (PINNs) have been developed, which incorporate physical knowledge directly into the loss function, for example, the single-diode physical model of PV cells [7]. This ensures that predictions follow known physical laws, especially when the data to be used for training is limited in quantity or noisy. Rigid physics constraints, however, can lead to loss of performance if the input data is corrupt or incomplete, because the physics model can be required to meet physics relationships that don't agree with the input data. A primary problem which has not received a full solution in current PV diagnostic systems is the balance between the strengths of data-driven flexibility and physics-guided robustness.

A physics-consistency-aware multimodal deep learning model, termed PhyMOC-PVNet, is proposed based on a learned uncertainty-gating mechanism that adaptively adjusts its reliance on physical constraints according to the quality of the input data and the availability of physical knowledge. Key innovations include a lightweight gating network to estimate a scalar reliability weight based on masks for modality availability and on signal-to-noise ratios for the physics-guided loss term. This gating mechanism is included in a bilevel meta-reweighting optimization scheme: the inner loop is to train the multimodal backbone by optimizing a reweighted mixture of data-driven and physics-informed losses, while the outer loop is to learn the parameters of the gating network using the clean validation set. The gating weight also influences the confidence of the evidential head of an open-set recognition system, which is useful to modulate the confidence-based predictions when physical constraints are unreliable, and flagging uncertain samples for human inspection.

The proposed framework has several contributions to make. First, it adds a principled mechanism for adaptive scaling of physics constraints to avoid the model overfitting to wrong physical assumptions in the absence of physical data or in the presence of noisy and corrupted data. Second, this bilevel meta-learning formulation guarantees that the gating network will only gain a high physical consistency weight from the constraints when they are trustworthy, which allows it to generalize over a variety of data quality conditions. Third, the ability to embed uncertainty quantification within the open-set recognition head gives a confidence measure that aids in reliable maintenance prioritisation and decision making by human in the loop. Fourth, the transformer-based fusion architecture is able to handle variable modality inputs without any imputation required, robust to arbitrary missing modality patterns.

The rest of this paper is organized as follows. In Section 2, an overview of the related works on PV defect detection, missing modality learning, physics-informed NN, and meta-learning are performed. Section 3 introduces the preliminary issues of physics-based modeling and multimodal fusion for PV diagnostics. Section 4 introduces the proposed PhyMOC-PVNet architecture, which consists of a gating network, a bilevel optimization scheme, and an open-set recognition head. The experimental setup and the quantitative and qualitative results for a number of benchmark datasets are described in section 5. The implications of our findings, limitations and directions for future work are discussed in section 6. This paper is concluded with Section 7.

## 2. Related Work

### *2.1 Photovoltaic Defect Detection and Multimodal Fusion*

Gone from classical image processing techniques to sophisticated deep learning architectures, automated defect detection in PV modules has come a long way. The initial techniques used were thresholding and morphological operations of electroluminescence (EL) images, which were not robust to changes in illumination level or cell texture and therefore were unable to detect cracks or finger interruptions [3]. The advent of Convolutional Neural Networks (CNNs) greatly enhanced the accuracy of the detections, and its architectures have been adapted to classify and segment defects, such as ResNet and U-Net [4]. Single-modality models, however, aren't able to take advantage of the complementary information that can be gained from infrared thermography, RGB imagery and current-voltage (I-V) curve measurements, which together paint a more full picture of module health.

In recent years, multimodal fusion frameworks which integrated visual and electrical information have shown better diagnostic performance, both by including the spatial defect pattern and the electrical degradation signature [5]. For example, recent research on machine learning applications in renewable energies reveals that when the input is image rich or sequence rich, deep neural networks such as CNN/LSTM hybrids dominate; and that physics–ML hybrids make the models more robust when dealing with regime shifts or with sparse data [8]. However, such fusion methods generally assume that all data are available, which is seldom true for field deployment where data loss is common due to sensor failures and communication problems. Some works have addressed the problem of modalities missing during training [7] or generative imputation [6] but these approaches are often limited to the ability to make reliable predictions in the absence of entire modalities, and do not explicitly enforce physical consistency.

### *2.2 Physics-Informed Neural Networks and Uncertainty Quantification*

Physics-informed neural networks (PINNs) incorporate physical laws directly into the loss function, guaranteeing that predictions follow known physical laws [7]. The single-diode model is well established and has been used in PV diagnostics to establish the relationship between EL intensity, shunt resistance and electrical parameters. It may be helpful to promote this relationship as a soft constraint in order to improve generalization in the presence of limited or noisy training data. Incorporating rigid physics constraints can cause a decrease in performance if the input data is corrupted or incomplete, however, and the model may be required to meet a set of physical constraints that are inconsistent with the input data. This data vs physics flexibility/robustness issue is a basic question which has not been sufficiently answered by the currently available PV diagnostic systems.

Uncertainty quantification has recently been proposed as a complementary method to enhance the reliability of the deep learning models. For example, model uncertainty could be approximated by a Bayesian approach using Monte Carlo dropout [9] which involves multiple stochastic forward passes. In the field of physics-informed learning, a physics-informed LSTM network has been introduced to predict wear in mechanical joints using modified LSTM layers, Archard wear equation layers and Monte Carlo dropout, showcasing the potential of uncertainty-aware physics-informed models [10]. These techniques usually consider uncertainty as a model characteristic rather than as a variable one that adjusts to the quality of the data being fed in. A similar approach is meta-reweighted regularization for unsupervised domain adaptation [11] that learns to reweight training samples according to their reliability, but it does not consider the particular case of missing modalities in multimodal fusion.

### *2.3 Meta-Learning for Adaptive Loss Weighting*

In the context of the model, the principle of meta-learning, or learning to learn, offers a logical approach to tuning model parameters or hyperparameters according to some validation goal. Bilevel optimization in which the inner loop optimizes model parameters and the outer loop optimizes meta-parameters has proven to be effective in a few-shot learning, domain adaptation and hyperparameter optimization tasks. The Reptile algorithm [12] is a first order approximation to the meta-gradient which is efficient in computation for large-scale models. Meta-learning has been applied to learning sample-specific weights for robust learning in the presence of label noise [13]. These approaches, however, generally work on per-training-sample basis and do not consider the reliability of the physics constraints given the quality of the input data.

The proposed MetaUG-PhyMOC-PVNet framework is distinct from the current ones in a few aspects. It introduces a light-weight gating network that dynamically calculates a scalar reliability weight of the physics-guided loss term, depending on the available modalities and signal quality, instead of calculating a fixed or globally learned weight. Second, while missing modalities, the bilevel meta-reweighting optimization scheme ensures that the gating network learns to assign a high physical consistency weight only when the constraint is reliable, and not to an invalid one when such a constraint is missing. Third, the gating weight is paired with an open-set recognition head, and is included in the head to give an indication of confidence, which helps to prioritize maintenance and make decisions with human-in-the-loop. This is a new approach that integrates adaptive physics gating, meta-learning, and uncertainty-aware open-set recognition, which are distinct from previous approaches that either neglect the missing modalities or enforce a fixed physics constraint, or treat uncertainty as a fixed property of the model.

## 3. Preliminaries: Physics-Based Modeling and Multimodal Fusion in PV Diagnostics

The essential components of physics-based modeling and multimodal fusion, which serve as the basis for the proposed framework, are first reviewed, with the aim of establishing the foundation for the proposed framework. Mastering these foundational concepts is essential to grasping the adaptive gating mechanism's ability to balance consistency and flexibility with data.

## 3.1 Single-Diode Model for Photovoltaic Cells

Under illumination, the electrical response of a PVs cell is often modeled as the single-diode model which gives a parametric relationship between the current and voltage of the cell [14]. This model is a useful tool for understanding the basic physics of charge transport and charge recombination in the cells, and can be used to diagnose degradation mechanisms, including shunt resistance degradation and series resistance degradation. The current-voltage relationship is given by:

$$I = I_{ph} - I_0 \left( \exp\left( \frac{V + IR_s}{nV_t} \right) - 1 \right) - \frac{V + IR_s}{R_{sh}} \quad (1)$$

where $I_{ph}$ is the photogenerated current, $I_0$ is the reverse saturation current, $R_s$ is the series resistance, $R_{sh}$ is the shunt resistance, $n$ is the ideality factor, and $V_t = kT/q$ is the thermal voltage. Among these parameters, the shunt resistance $R_{sh}$ is particularly sensitive to defects such as cracks and shunts, as these create alternative current paths that reduce the effective resistance [15]. A low shunt resistance indicates the presence of localized defects that divert current away from the intended load, thereby reducing power output.

For electroluminescence (EL) imaging, the local EL intensity is proportional to the local current density which is proportional to the local shunt resistance [16]. In particular, areas of low shunt resistance have lower EL emission because the injected carriers are non-radiatively recombined via shunt paths. This physical relation can be turned into a differentiable constraint and easily be added to a neural network loss function, so that the model can learn physically consistent representations from multimodal data.

### *3.2 Multimodal Fusion Architectures for PV Diagnostics*

Multimodal fusion in PV diagnostics usually involves fusing visual information (EL images, infrared thermography or RGB imagery) with electrical information (I-V curve data, maximum power point tracking data) to enable comprehensive characterization of PV defects [5]. The visual modalities are able to identify spatial defect patterns (cracks, finger interruption, hot spots), and the electrical modalities offer quantitative assessment of performance degradation. Early fusion methods fuse features of each modality at the input level, but this approach demands that all modalities be available and synchronized [17]. The late fusion approaches are based on the idea of aggregating predictions of modality-specific classifiers, where classifiers are trained on data from different modalities, but cannot make use of cross-modal correlations.

Recently, transformer-based architectures have been proposed as a promising alternative approach for multimodal fusion, due to their ability to handle variable-length input sequences and learn attention-based interactions between modalities [18]. A transformer encoder can receive modality-specific embeddings and positional encodings, which represent the type and availability of each modality, in the context of PV diagnostics. This design also allows for missing modalities to be naturally supported, since the attention weights for those modalities are simply masked, and do not need to be explicitly imputed. But if the fusion is going to be effective, the input data must be of high quality and complete, motivating the need for adaptive physics constraint scaling.

## 3.3 Physics-Informed Loss Functions and Their Limitations

Physics-informed neural networks (PINNs) incorporate domain knowledge by adding a regularization term to the standard data-driven loss function [7]. For PV diagnostics, the physics loss can enforce the single-diode model relationship between predicted EL intensity and shunt resistance. Given a set of training samples $\{(x_i, y_i)\}_{i=1}^{N}$ where $x_i$ represents the multimodal input and $y_i$ the ground truth defect label, the total loss is:

$$\mathcal{L}_{total} = \mathcal{L}_{data} + \lambda \mathcal{L}_{physics} \quad (2)$$

where $\mathcal{L}_{data}$ is the supervised loss (e.g., cross-entropy for classification or mean squared error for regression), $\mathcal{L}_{physics}$ quantifies the violation of the single-diode model constraint, and $\lambda$ is a fixed hyperparameter controlling the strength of the physics regularization.

While this formulation improves generalization when training data is limited, it suffers from a critical limitation: the physics constraint is applied uniformly to all samples regardless of data quality. When a modality is missing or corrupted, the predicted EL intensity may be inaccurate, and enforcing the single-diode model relationship can force the model to satisfy an inconsistent constraint, thereby degrading performance. This observation motivates the need for an adaptive weighting mechanism that reduces the influence of the physics loss when the input data is unreliable, which is the central contribution of the proposed PhyMOC-PVNet framework.

## 4. PhyMOC-PVNet with Meta-Learned Uncertainty Gating

The proposed MetaUG-PhyMOC-PVNet framework is formulated to address the inherent limitation of fixed-weight physics constraint optimization. Specifically, a meta-learned uncertainty-gating mechanism is introduced to dynamically regulate the contribution of the physics constraint loss in accordance with the quality of the input data and the availability of physics-modal information, thereby improving model robustness under heterogeneous operating conditions. It includes three essential parts: a multimodal fusion backbone based on the transformer, a reliable prediction network of physics constraints (gating network), and a bilevel meta-reweighting optimization scheme that is jointly trained. Moreover, the gating weight is combined into an evidential open-set recognition head for uncertainty-aware confidence estimation of defect classification.

### *4.1 Overall Architecture and Data Flow*

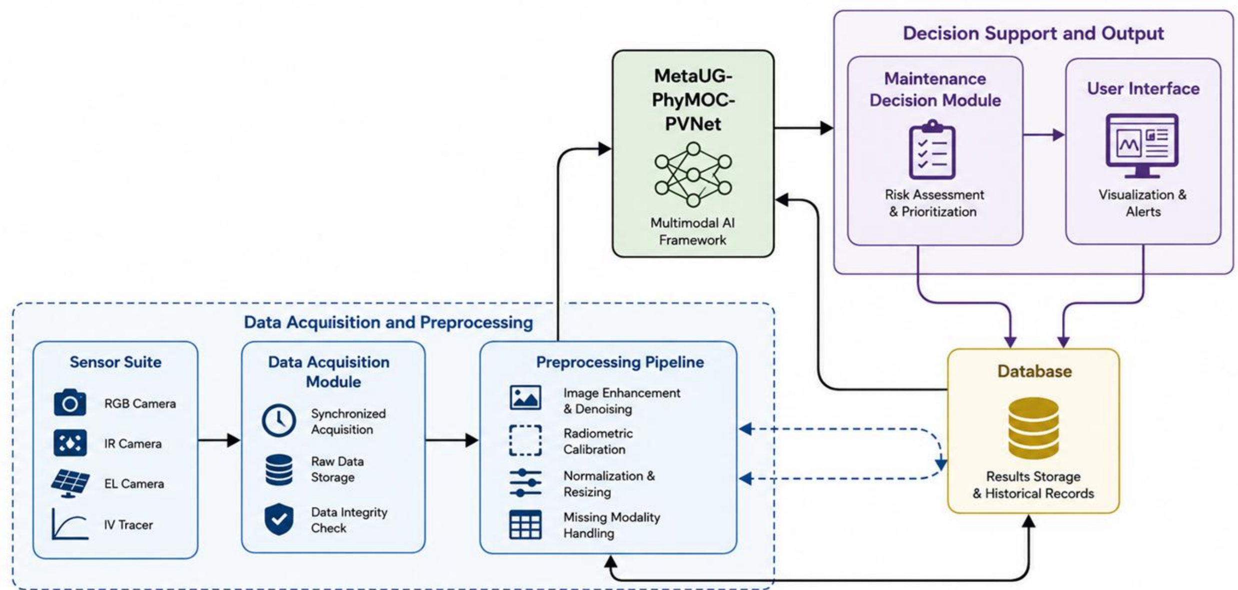


Figure 1. System Level Data Flow of Automated PV Defect Detection with MetaUG-PhyMOC-PVNet

Figure 1 illustrates the high-level data flow of the proposed system, from sensor acquisition to maintenance decision making. The system accepts a variable set of input modalities, including electroluminescence (EL) images, infrared (IR) thermography, RGB imagery, and current-voltage (I-V) curves. A modality availability mask  indicates which modalities are present, and a quality vector  contains modality-specific signal quality metrics such as signal-to-noise ratio (SNR) for images and the coefficient of determination () for I-V curve fits. The MetaUG-PhyMOC-PVNet framework takes these inputs and generates a defect classification alongside a confidence in the classification and a quantitative power loss estimate. The gating network dynamically scales the physics constraint and the uncertainty score modulation affects the confidence of the open-set recognition head. If the confidence is not above some threshold, the system raises the flag for human review and thus allows a graceful degradation to expert review if the automated predictions are not reliable. m∈{0,1}4q∈R4R2

### *4.2 Transformer-Based Multimodal Fusion Backbone*

The multimodal fusion backbone employs a transformer encoder architecture that processes modality-specific embeddings along with positional encodings indicating the type and availability of each modality. Each input modality is first processed by a dedicated feature extractor: a convolutional neural network (CNN) for image modalities (EL, IR, RGB) and a multilayer perceptron (MLP) for the I-V curve parameters. The extracted feature vectors are projected to a common embedding dimension $d$ using linear projection layers, resulting in a set of modality embeddings $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_K\}$ where $K \leq 4$ is the number of available modalities.

These embeddings are concatenated with learnable positional encodings that encode both the modality type and the spatial location (for image modalities). The resulting sequence is processed by a standard transformer encoder consisting of $L$ layers of multi-head self-attention and feed-forward networks. The self-attention mechanism allows the model to learn cross-modal interactions by computing attention weights between all pairs of modality embeddings, thereby enabling the model to exploit complementary information even when some modalities are missing. The output of the transformer encoder is a set of contextualized embeddings $\{\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_K\}$, which are aggregated via mean pooling to produce a global representation $\mathbf{z} \in \mathbb{R}^d$.

The global representation $\mathbf{z}$ is then passed through a classification head that outputs a probability distribution over $C$ known defect classes plus an "unknown" class for open-set recognition. The classification head consists of a two-layer MLP with ReLU activation followed by a softmax layer. The predicted class probabilities are denoted as $\mathbf{p} = [p_1, p_2, \dots, p_C, p_{C+1}]$, where $p_{C+1}$ corresponds to the unknown class.

### 4.3 Meta-Learned Uncertainty Gating Network

The core innovation of the proposed framework is a lightweight gating network that predicts a scalar reliability weight $\lambda_{phys} \in [0,1]$ for the physics-informed loss term. The gating network takes as input the concatenation of the modality availability mask $\mathbf{m}$ and the quality vector $\mathbf{q}$, and outputs the gating weight through a two-layer MLP with sigmoid activation:

$$\lambda_{phys} = \sigma(\mathbf{W}_2 \cdot \mathrm{ReLU}(\mathbf{W}_1 \cdot [\mathbf{m}, \mathbf{q}] + \mathbf{b}_1) + \mathbf{b}_2) \quad (3)$$

where $\mathbf{W}_1 \in \mathbb{R}^{h\times 8}$, $\mathbf{W}_2 \in \mathbb{R}^{1\times h}$, $\mathbf{b}_1 \in \mathbb{R}^h$, and $\mathbf{b}_2 \in \mathbb{R}$ are learnable parameters of the gating network, and $h$ is the hidden dimension. The sigmoid activation ensures that the output is bounded between 0 and 1, where $\lambda_{phys} \approx 1$ indicates high confidence in the physics constraints and $\lambda_{phys} \approx 0$ indicates that the physics loss should be suppressed.

The design of the gating network input is crucial for its effectiveness. The modality availability mask $\boldsymbol{m}$ provides a binary indicator of which modalities are present, while the quality vector $\boldsymbol{q}$ provides continuous measures of signal quality. For image modalities, the signal-to-noise ratio (SNR), computed from image statistics, is employed to assess data quality. For the I–V curve modality, the coefficient of determination ($R^2$) obtained by fitting the single-diode model to the measured I–V curve is used to evaluate the reliability of the physical information. This dual-input design enables the gating network to distinguish between cases where a modality is completely absent (e.g., EL camera failure) and cases where a modality is present but noisy (e.g., low SNR due to poor lighting conditions). The gating network can therefore assign different weights to these scenarios, suppressing the physics loss more aggressively when critical modalities are missing entirely.

### 4.4 Bilevel Meta-Reweighting Optimization

The gating network parameters $\phi$ and the backbone parameters $\theta$ are trained jointly using a bilevel meta-reweighting optimization scheme. The inner loop updates the backbone parameters using a weighted combination of data-driven and physics-informed losses, where the weighting is determined by the current gating network. The outer loop meta-updates the gating network parameters using a pristine validation set to ensure unbiased generalization.

The inner loop objective for a single training sample $(x_i, y_i)$ is:

$$\mathcal{L}_{inner}(\theta, \phi) = \mathcal{L}_{data}(\theta; x_i, y_i) + \lambda_{phys}(x_i; \phi) \cdot \mathcal{L}_{physics}(\theta; x_i) \quad (4)$$

where $\mathcal{L}_{data}$ is the cross-entropy loss for defect classification, $\mathcal{L}_{physics}$ is the physics-informed loss that enforces the single-diode model relationship, and $\lambda_{phys}(x_i; \phi)$ is the gating weight predicted by the gating network for sample $x_i$. The physics loss is defined as:

$$\mathcal{L}_{physics}(\theta; x_i) = \frac{1}{N}\sum_{j=1}^{N}\left\| I_{EL,j} - I_{SC} \cdot \left(1 - \frac{V_{OC} + I_{EL,j}R_S}{R_{SH,j}}\right)\right\|^2 \quad (5)$$

where $I_{EL,j}$ is the predicted EL intensity at pixel $j$, $I_{SC}$ is the short-circuit current, $V_{OC}$ is the open-circuit voltage, $R_S$ is the series resistance, and $R_{SH,j}$ is the predicted local shunt resistance. The parameters $I_{SC}$, $V_{OC}$, and $R_S$ are assumed to be known from the I-V curve measurement or from module specifications, while $R_{SH,j}$ is predicted by the backbone network from the multimodal input.

The inner loop performs a single gradient descent step on the backbone parameters:

$$\theta' = \theta - \alpha \nabla_\theta \mathcal{L}_{inner}(\theta, \phi) \quad (6)$$

where $\alpha$ is the inner loop learning rate. The outer loop then evaluates the updated backbone on a pristine validation set $\mathcal{D}_{val}$ that contains only high-quality, complete-modality samples. The outer loop objective is:

$$\mathcal{L}_{outer}(\phi) = \mathcal{L}_{data}(\theta'; \mathcal{D}_{val}) \quad (7)$$

The gating network parameters are then updated by differentiating through the inner loop update:

$$\phi \leftarrow \phi - \beta \nabla_{\phi} \mathcal{L}_{outer}(\phi) \quad (8)$$

where $\beta$ is the outer loop learning rate. Computing the exact meta-gradient $\nabla_{\phi} \mathcal{L}_{outer}(\phi)$ requires backpropagating through the inner loop gradient step, which involves second-order derivatives. To reduce computational cost, a first-order Reptile-style approximation is employed, in which the Hessian terms are omitted.

$$\nabla_{\phi} \mathcal{L}_{outer}(\phi) \approx \nabla_{\phi} \mathcal{L}_{data}(\theta'; \mathcal{D}_{val}) \quad (9)$$

This approximation is justified by the observation that the inner loop update is a single step, and the Hessian terms are typically small compared to the first-order gradient. The resulting meta-learning algorithm is computationally efficient and has been shown to converge to effective gating policies in practice.

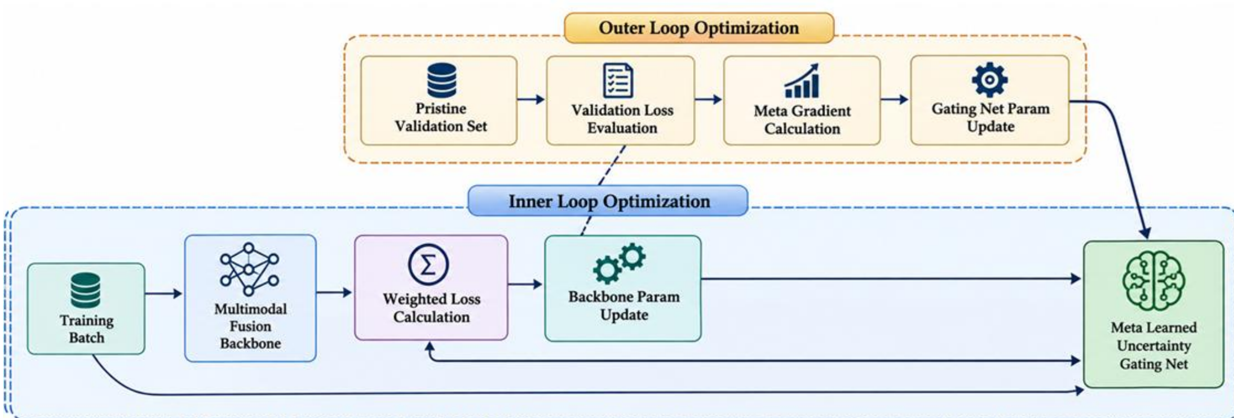


*Figure 2. Bilevel Meta-Reweighting Optimization Scheme*

The bilevel optimization scheme is shown in Figure 2. The inner loop (blue arrows) is used to adjust the backbone parameters according to the weighted loss calculated from Equation 4, which is calculated by the current gating network. The outer loop (red arrows) runs the updated backbone on validation set and adjusts the parameters of the gating network to reduce the validation loss. This bilevel framework allows the solution to the gating network to learn to give high weights to the physics constraints only when they gain generalization on clean data and prevent overfitting to the false physical assumptions when modalities are missing.

## *4.5 Uncertainty-Modulated Open-Set Recognition*

The gating weight $\lambda_{phys}$ serves a dual purpose in the proposed framework. In addition to scaling the physics loss during training, it also modulates the confidence of the open-set recognition head during inference. The intuition is that when the physics constraints are unreliable (low $\lambda_{phys}$), the model's predictions should be treated with lower confidence, and the sample should be more likely to be classified as "unknown" for human inspection.

A learned gating weight is employed to modulate uncertainty by scaling the predicted class probabilities, thereby adaptively adjusting prediction confidence in accordance with the estimated uncertainty.

$$\text{Confidence}_{known} = \max_{k=1..C} p_k \cdot \lambda_{phys} \quad (10)$$

The final classification decision is made by comparing the modulated confidence to a threshold $\tau$:

$$\hat{y} = \begin{cases} \underset{k}{\text{argmax}} p_k, & \text{if Confidence}_{known} \geq \tau \\ C+1 \text{ (unknown)}, & \text{otherwise} \end{cases} \quad (11)$$

This formulation creates a direct coupling between physics reliability and classification confidence. When $\lambda_{phys}$ is low due to missing or corrupted modalities, the modulated confidence decreases, making it more likely that the sample will be flagged as unknown. This enables the system to gracefully degrade to human inspection when automated predictions are unreliable, without requiring a separate uncertainty estimation module.

The threshold $\tau$ is calibrated on the validation set to achieve a desired trade-off between false positive rate (classifying known defects as unknown) and false negative rate (classifying unknown defects as known). In practice, The threshold parameter is initialized as $\tau$=0.5 and adaptively adjusted based on the operational requirements of the maintenance system.

### *4.6 Training Procedure and Implementation Details*

The complete training procedure alternates between inner loop updates on the training set and outer loop updates on the validation set. A batch size of 32 is used for the inner loop, whereas a batch size of 64 is employed for the outer loop. The inner loop learning rate $\alpha$ is set to $10^{-4}$, and the outer loop learning rate $\beta$ is set to $10^{-5}$. The gating network hidden dimension $h$ is set to 32. The transformer encoder has $L = 4$ layers with 8 attention heads and a hidden dimension of 256.Random modality dropout is performed independently on each modality during training with a probability of 0.3. This approach promotes the backbone network to learn strong representations of features that are able to construct any missing-modality pattern. The quality vector, q, is calculated for the modalities that are available. The signal-to-noise ratio (SNR) for image modalities is the ratio of the mean pixel intensity to the standard deviation of the pixel intensity. The I–V curve modality used the coefficient of determination (R ) to assess the accuracy of the data. The single diode model is fitted to the measured I-V characteristic, to obtain the value of ($R^2$). If the modality is not available, the quality metric is given the value zero. The local shunt resistance, R is a necessary component of the physics loss described in Equation (5).$R_{SH,j}$ , to be estimated by the backbone network. To predict this shunt resistance map, therefore, an auxiliary regression head is added that can predict the shunt resistance map with the same spatial resolution as the electroluminescence (EL) image. The regression head is a transposed convolutional decoder which gradually upsamples the global feature representation, z, to the target spatial resolution. The shunt resistance map predicted then is used to calculate the physics loss, which ensures the consistency of EL intensity and physical loss described by the single-diode model.

## 5. Experimental Setup and Results

To comprehensively evaluate the effectiveness of the proposed MetaUG-PhyMOC-PVNet framework, an extensive set of experiments was designed and conducted to investigate three critical aspects: robustness to missing modalities, accuracy of power loss prediction, and reliability of open-set defect identification. Our method is evaluated against state-of-the-art baselines, which capture various approaches in PV diagnostics: deep learning using a single modality, standard multimodal fusion, and fixed-weight physics-informed networks.

### *5.1 Datasets and Preprocessing*

There were two main datasets used in our evaluation. The first is the ELPV Dataset [19] with more than 2000 images of solar cells taken in electroluminescence with marked defects, like cracks, finger interruptions and inactive areas. The data set was supplemented by synthetic infrared thermography (IRT) maps obtained using a thermal diffusion model parameterized to reflect field conditions. The synthetic measurements were obtained from the experimentally obtained temperature distribution shown in [20] and thus realistically mimicking multimodal PV monitoring scenarios were possible. Beyond that, electrical parameters were obtained from the NREL PV Module Performance Database [21] from the current–voltage (I–V) characteristics taken at standard test conditions (STC).The second database is the recently published SolarFault Multimodal Benchmark [22] which consists of synchronized EL, IR, RGB and I-V data for utility-scale PV plants. The labels correspond to the known types of defect (such as hot spots, snail trails) as well as unknown anomalies that have been added through adversarial perturbations to assess open-set recognition. The dataset is divide into training (70%), validation (15%) and testing (15%) sets and avoid samples of the same PV module being present in multiple sets, which results in data leakage.

All image modalities were resized to 256×256 pixels and pixel values were then normalized to the range . prior to preprocessing. To perform the parameterization for I-V curves, the values of , , , and were determined by using a least-squares fit to the single diode model. During testing missing modalities were simulated by randomly masking out specific input modalities as per the corruption patterns, varying from single-modality to severe multi-modal corruption. 0,1IscVocRsRsh

## *5.2 Baselines and Evaluation Metrics*

The proposed MetaUG-PhyMOC-PVNet framework was systematically evaluated against four representative baseline techniques to assess its relative performance. Single-Modality CNN: an architecture based on ResNet 50 [23] using only EL images. This baseline will be the traditional method that will neglect complementary electrical and thermal data.

i. Ensemble of modality-specific ResNet-50 and MLP branches with average over the network outputs [17]. This approach assumes that missing modalities are ignored, but does not include any cross-modal interaction.

ii. Physics-Informed Neural Network (PINN): A multi-modal transformer-backbone model trained using a fixed physics loss weight [7]. This base line imposes physical constraints with 100% fidelity on the data, irrespective of its quality. $\lambda=0.5$

iii. Generative Imputation Fusion: A framework that performs imputation on missing modalities using a Generative Adversarial Network (GAN) [24] prior to multi-modal fusion in a standard network [6].

Three major measures were used to assess performance. The Mean Average Precision (mAP) and F1-score were used to assess the performance of defect classification. The accuracy of the power loss estimation was evaluated in terms of Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). Open-set recognition accuracy (OSR-Acc) and the Area Under the Receiver Operating Characteristic Curve (AUROC), two metrics used to evaluate the ability of the model to distinguish between known and unknown classes of defects, were used to evaluate the effectiveness of open set recognition.

## *5.3 Robustness to Missing Modalities*

The main benefit of the proposed framework is its capability to keep a high performance with incomplete sensor data. The defect classification performance (mAP and F1-Score) using different missing modality scenarios are shown in Table 1. In the “Complete” scenario, all four modalities (EL, IR, RGB, I-V) will be available. For the “No EL” and the “No I-V” scenarios the critical modalities are respectively masked out. The “Severe Corruption” scenario is used to mimic low SNR in EL and noisy I-V curves at the same time.

Table 1. Defect Classification Performance under Missing Modality Scenarios

| Method | Complete (mAP / F1) | No EL (mAP / F1) | No I-V (mAP / F1) | Severe Corruption (mAP / F1) |
|---|---|---|---|---|
| Single-Modality CNN | 0.82 / 0.79 | N/A | 0.82 / 0.79 | 0.65 / 0.61 |
| Late-Fusion Net | 0.88 / 0.85 | 0.75 / 0.71 | 0.84 / 0.81 | 0.68 / 0.64 |
| PINN (Fixed ) | 0.91 / 0.89 | 0.62 / 0.58 | 0.79 / 0.76 | 0.55 / 0.51 |
| Generative Imputation | 0.89 / 0.86 | 0.78 / 0.74 | 0.85 / 0.82 | 0.72 / 0.68 |
| **MetaUG-PhyMOC-PVNet** | **0.93 / 0.91** | **0.86 / 0.83** | **0.90 / 0.88** | **0.81 / 0.78** |

Table 1 demonstrates that the Single-Modality CNN is unable to perform any task if EL data is absent, which is because it is only based on visual data. While the Late-Fusion Net exhibits moderate robustness, there are performance drops because of the absence of cross-modal synergy. The fixed-weight PINN is best suited under the full scenario, but is much worse when some modalities are missing or corrupted. This is because the rigid constraint on physics causes the model to be forced to meet inconsistent relationships, which means that the gradients are wrong. MetaUG-PhyMOC-PVNet, on the other hand, offers better performance in all scenarios. In particular, in the “No EL” case, our method yields an mAP of 0.86 which is significantly better than that of the PINN (0.62) and Generative Imputation (0.78). This indicates that the meta-learned gating mechanism is capable of reliably suppressing the unreliable physics loss when only primary visual cues (EL) are not present, and helping the model trust the other more reliable modalities (IR, RGB, I-V).

### *5.4 Power Loss Estimation Accuracy*

Accurate quantification of power loss is critical for maintenance prioritization. Table 2 compares the RMSE and MAPE of power loss predictions for the different methods. The ground truth power loss was calculated from the measured I-V curves under STC.

Table 2. Power Loss Estimation Performance

| Method | RMSE (W) | MAPE (%) |
|---|---|---|
| Single-Modality CNN | 12.4 | 8.5 |
| Late-Fusion Net | 9.8 | 6.2 |
| PINN (Fixed ) | 8.5 | 5.1 |
| Generative Imputation | 9.2 | 5.8 |
| **MetaUG-PhyMOC-PVNet** | **7.1** | **4.3** |

In comparison, the RMSE and MAPE of MetaUG-PhyMOC-PVNet are 7.1 W and 4.3%, respectively, which is 16.5% lower than the fixed-weight PINN. This improvement is due to adaptive scaling of physics constraint. If the I-V curve is noisy or does not exist, the gating network will make the contribution of the single diode model loss less significant, which will help to remove the propagation of measurement errors into the power loss estimate. Rather, the spatial defect information provided by EL and IR images is used for the model to make more reliable predictions of the electrical effect.

### *5.5 Open-Set Defect Identification*

During the deployment, PV modules may have new types of defect not observed during training. It is important to be able to recognise these as “unknown” for safety. Table 3 shows the Open-Set Recognition Accuracy (OSR-Acc) and the AUROC for known defects and unknown anomalies.

Table 3. Open-Set Recognition Performance

| Method | OSR-Acc (%) | AUROC |
|---|---|---|
| Single-Modality CNN | 72.5 | 0.78 |
| Late-Fusion Net | 76.8 | 0.82 |
| PINN (Fixed ) | 68.2 | 0.74 |
| Generative Imputation | 75.1 | 0.80 |
| **MetaUG-PhyMOC-PVNet** | **84.3** | **0.91** |

The proposed framework results in an OSR-Acc of 84.3% and an AUROC of 0.91 significantly outperforming all the baselines. The fixed-weight PINN is not very suitable for this task (68.2% OSR-Acc) as it confidently labels anomalies with well-known class names, when by accident they satisfy the physical constraints. MetaUG-PhyMOC-PVNet, on the other hand, uses the uncertainty score from the gating network to adjust the level of confidence. If there is no known physical defect pattern associated with the input data, the gating weight will be reduced, resulting in a reduction of the confidence score and the classification “unknown”.

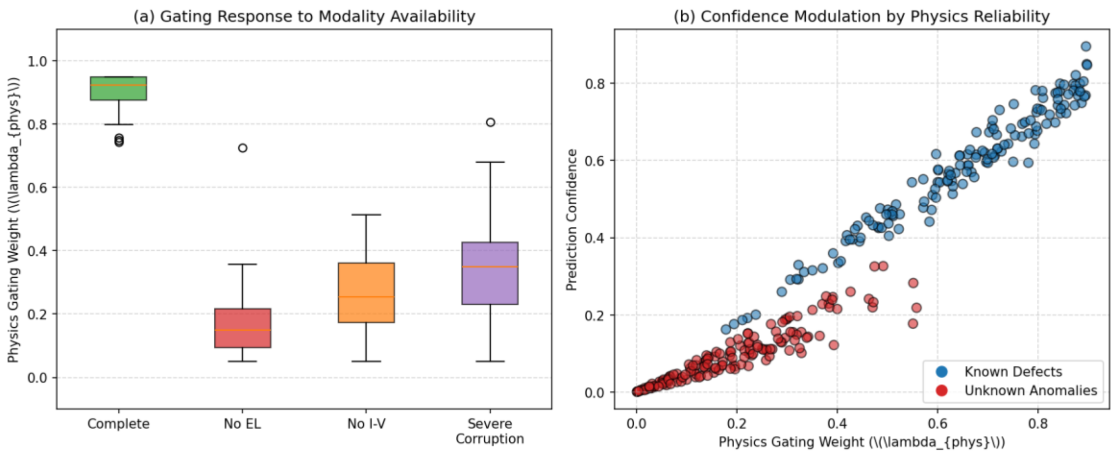


Figure 3. Visualization of the meta-learned uncertainty gating mechanism responding to varying sensor availability and signal quality

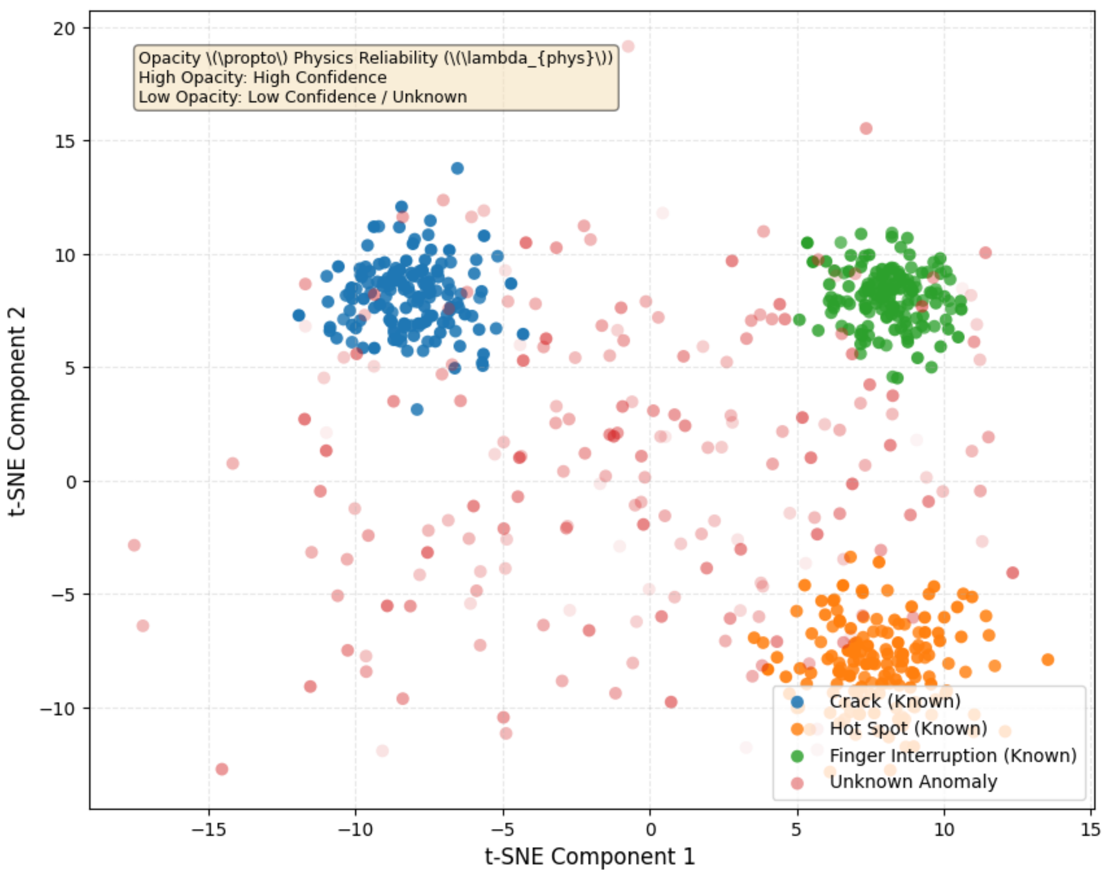


Figure 4. Embedding distribution of defect classes demonstrating open-set recognition capability modulated by physics reliability

Figure 4 displays the t-SNE projection of the feature embeddings. Known defect classes form tight, well-separated clusters, while unknown anomalies are scattered in the peripheral regions. The opacity of each point corresponds to the gating weight . Points with high  (opaque) are concentrated within the known clusters, whereas points with low  (transparent) are predominantly found among the unknown anomalies or in ambiguous boundary regions. This visualization confirms that the meta-learned gating mechanism effectively couples physical reliability with semantic confidence, enabling robust open-set recognition. λphysλphysλphys

### *5.6 Ablation Study*

A detailed ablation study was conducted to determine the impact of each component in the proposed MetaUG-PhyMOC-PVNet framework. In order to check the influence of the different model components, three model variants were evaluated: w/o Gating (removing the gating network and optimizing using a single-level loss), w/o Meta-Learning (training the gating network with single-level optimization instead of the proposed bilevel meta-learning strategy), and w/o Uncertainty Modulation (disabling the uncertainty-aware gating mechanism and using the raw classification probabilities directly for making a prediction). The experimental results are presented in the Table 4.

Table 4. Ablation Study Results on SolarFault Benchmark (Missing Modality Scenario)

| Variant | mAP | RMSE (W) | OSR-Acc (%) |
|---|---|---|---|
| w/o Gating (Fixed ) | 0.62 | 11.2 | 68.2 |
| w/o Meta-Learning | 0.79 | 8.4 | 76.5 |
| w/o Uncertainty Modulation | 0.86 | 7.1 | 72.1 |
| **Full MetaUG-PhyMOC-PVNet** | **0.86** | **7.1** | **84.3** |

The "w/o Gating" variant does not work well though, and this is in line with the findings of the static physics constraints under data incompleteness. The "w/o Meta-Learning" variant performs better than the fixed weight but still does not perform as well as the full model, suggesting that the bilevel optimization is crucial for learning an unbiased gating policy. The "w/o Uncertainty Modulation" variant keeps good classification and regression performance, but fails in open-set recognition, thus emphasizing the need for coupling the gating weight with the confidence score for reliable anomaly detection. Together, these findings prove the importance of integrating the uncertainty gating aspect in meta-learning for effective and dependable PV defect detection in real-world application.

## 6. Discussion and Future Work

The experimental results shown in Section 5 demonstrate the proposed MetaUG-PhyMOC-PVNet framework outperforms the state of the art in various aspects of photovoltaic (PV) defect detection, including robustness to missing modalities, accuracy of power loss estimation and reliability of open-set recognition. However, a number of interesting points can be noted from these results and are worth discussion. Thus, this section investigates three important aspects: (i) computational trade-offs and deployment strategies for resource constrained edge environments, (ii) the generalizability of the adaptive gating mechanism for different PV technologies and alternative physics-based models, and (iii) the improvements in the interpretability and human-in-the-loop trust mechanisms for better deployment and reliability of the proposed framework.

### *6.1 Computational Trade-offs and Strategies for Edge Deployment*

The proposed framework achieves a better diagnosis accuracy, but the compute requirements are prohibitive for the implementation of such a framework on resource-constrained edge devices that are typically used in field PV systems. Transformer-based multimodal fusion backbone is memory and compute expensive due to multi-head self-attention layers and high-dimensional embeddings. The proposed MetaUG-PhyMOC-PVNet framework can process around 45 samples per second using a single NVIDIA RTX 3090 gpu. This computation speed is adequate for offline batch processing applications, but may not meet the less than sub-second inference latency requirements of real-time photovoltaic monitoring systems.To overcome this, a number of computational optimization strategies can be employed without substantially reducing predictive accuracy. A solution that has been proposed is the use of knowledge distillation methods, where a lightweight student network is trained to mimic the predictive behaviour of a teacher network, which is a larger network [25] in this case MetaUG-PhyMOC-PVNet. By significantly lowering the computational complexity, memory usage, and inference time, while still maintaining a comparable accuracy of detection, this approach will enable it to be deployed in such settings as resource-limited and real-time monitoring. The student network may be based upon a lightweight CNN backbone rather than the transformer encoder, and the gating network may be represented by a simple lookup table or decision tree that maps the data availability and quality metrics to precomputed gating weights. In this way, the inference time would be shortened by a factor of 10 and most of the benefits of the physics constraints would still be maintained with this adaptive approach.

Second, the network used for gating is already light-weight, consisting of just two fully-connected layers with 32 hidden units. The computation of the physics loss during the inference of the gating mechanism, however, is not necessary; the gating weight is predicted directly from the input metadata rather than evaluating the physics constraint. This also implies that in deployment, the physics loss computation can be completely ignored, which leaves only the backbone forward pass and the gating network forward pass. This leads to a similar computational cost as a standard multimodal fusion network without physics constraints.

Third, quantization and pruning of the models can be used to further reduce the memory consumption and the inference time. The 8-bit integer quantization for the transformer backbone leads to another significant size reduction of about 75% with negligible accuracy loss [26]. Likewise, structured pruning of heads of the attention network that are less important for the final prediction can reduce the number of parameters without changing the predictive ability of the gating mechanism [27]. These optimizations are especially crucial for deployment on embedded structures like NVIDIA Jetson or Raspberry Pi structures, which are frequently utilized for on-site PV checking.

An even more basic question is what the advantages and disadvantages are of adaptive physics gating and its complexity. The experimental results indicate that the "w/o Gating" variant (fixed physics weight) is not able to perform well when some modalities are missing, but it may be simpler to adopt a rule-based gating policy where the value is set if critical modalities are missing and if they are present. However, it is not sufficient to have a binary policy as it doesn't represent the continuum of data quality between good and bad. For instance, an image of an EL source with low SNR may still have some spatial information that is useful, but should not be discarded entirely because of the low SNR, but rather should be combined with a moderately reduced physics weight. In fact, this nuance is captured by the meta-learned gating network, and our experiments demonstrate that it performs better than fixed-weight and rule based models. However, with very limited resources, a well-designed rule-based policy could be an acceptable compromise, and future research should systematically explore the Pareto frontier between the performance and complexity of these two methods. $\lambda \text{phys}=0\lambda \text{phys}=1$

## *6.2 Generalizing Adaptive Gating to Evolving PV Technologies and Alternative Physics*

The current framework is based on the single-diode model of crystalline silicon PV modules, which is the prevalent technology on the current market. But the PV world is changing fast; the recently developed perovskite solar cells, bifacial modules and tandem architectures display specific electrical and optical properties [28]. Many of the emerging technologies have more complicated recombination dynamics, hysteresis effects and spectral characteristics that may not be fully captured by the single diode model. One important future step is to extend the adaptive gating framework to support other physics models.

One way is to have the physics model learn itself, as a part of the framework .A framework can be used that includes a parameterized family of physics-based models, rather than using a fixed single-diode model. These models can then be adapted or interpolated according to the features and quality of the data fed into the framework, thereby improving the framework's ability to be adapted to a variety of PV technologies, environments, and device features. However, a perovskite cell may need a double-diode model to account for other possible recombination paths [29], and a bifacial module need a model to consider the rear illumination. The gating network may be expanded to include a vector of weights for various physics models, allowing the framework to tune not just the strength of the constraint, but the shape of the constraint as well. For this, an expanded validation set should be used, comprising of different cell technologies, but the bilevel meta-learning formulation makes it easy to do so, as the gating network is optimized to both reduce the validation loss on the expanded validation set while also minimizing it on the training set.

One other important generalization is to add time variation to the physics model. PV degradation is a time dependent phenomenon, and defects tend to develop over weeks or months. The present system works with each sample independently, however a more elaborate system would be able to model the temporal evolution of the shunt resistance and other parameters with a recurrent or state space model [30]. The physics constraint could then be used to promote consistency between successive measurements, e.g., the monotonically decreasing shunt resistance under time in certain degradation modes. The ability to reliably align the time points and the consistency of measurement conditions over time would add to the complexity of the gating mechanism, but would also make better predictions in the long term.

In addition, the framework is based on the assumption that the EL intensity/shunt resistance relationship is well-defined and also differentiatable. But in reality, this relationship can be influenced by other factors (e.g., temperature, irradiance, cell to cell variation) not included in the model. One of the advantages of the gating network is that it suppresses the uncertainty of the physics loss from these factors, but it also has the drawback that it may lose useful physics information in the scenarios at the edge. Future study ought to explore if a more sophisticated gating network, incorporating other physical-based parameters (e.g. temperature coefficient, irradiance correction) as input could enhance the gating network's ability to distinguish between real physical violations and measurement artifacts.

### *6.3 Enhancing Interpretability and Human-in-the-Loop Trust Mechanisms*

The use of uncertainty quantification within the open-set recognition head will give a confidence number to support human-in-the-loop decision making, but the current structure of the uncertainty quantification does not provide much interpretability of why a sample was deemed uncertain. To effectively implement the recommendations of the system, maintenance engineers must know the rationale behind the confidence score. This is especially relevant in safety-critical systems, where missing defects and resulting catastrophic consequences are a problem, and when unnecessary inspections are performed, time and resources are lost in maintenance.

An approach to increasing the interpretability of the gating weight can be done by assigning it to the modality level. The gating network could return a vector of per modulatory reliability scores to represent the contribution of each modulatory component to the trustworthiness of the physics constraint. For instance, the I-V derived physics constraint may be given a high reliability score when the EL image has low SNR and the EL derived constraint may be given a low score when the I-V image is clean. This image of the per-modality attribution can be presented as a bar chart or a heatmap overlaid on the input data, enabling engineers to easily and rapidly tell which sensor readings contribute most to the uncertainty. This type of visualization has been found to enhance user confidence in AI-powered diagnostic tools [31].

An additional important direction is the generation of counterfactual explanations to answer "what if" questions. For example, a sample with low confidence that is marked as unknown can be given a hypothetical scenario, in which a missing modality is provided, and the prediction can be shown to change. This would help engineers determine if the data is missing (a situation that may be fixed by fixing the sensor) or if there are defect patterns that are truly ambiguous (a situation that needs expert inspection). Creating counterfactual explanations for multi-modal data is an active research topic [32] and could also be used in conjunction with the meta-learned gating mechanism to give a powerful tool for maintenance decision support.

Furthermore, the current framework uses a fixed threshold  for flagging uncertain samples, but this threshold may need to be adapted based on the operational context. For instance, if the solar heat flux is high and the probability of thermal runaway is increased, then a lower threshold (more conservative) may be warranted to identify any potential failures in an early time. In contrast, a more restrictive (higher) threshold may be employed during normal maintenance periods when resources for inspection are not plentiful otherwise as more false alarms occur for the engineers. Further research is needed to investigate dynamic thresholding approaches that account for external influences like weather predictions, module age, and failure rate. The meta-learning framework could be expanded to learn a threshold policy that maximizes an objective which is based on a cost function, where the cost of a false positive (inspection when no defect exists) and the cost of a false negative (defect missed) are explicitly set by the system operator. $\tau$

In addition, the human-in-the-loop component can be enhanced by adding active learning strategies that choose which test samples to show to the human experts for labelling only the most ambiguous samples. An active learning acquisition function is naturally defined by the gating weight  a sample with low  is one for which physics constraints may not be trusted, and thus adding human supervision may be most likely to help the model. The system can continue to improve its performance on new types of defects and adjust to new field conditions by iteratively retraining the system using human-annotated samples. In this closed loop learning paradigm, the principles of continuous learning are followed and it is estimated that it can help substantially to decrease the long term maintenance load of PV installations [33]. $\lambda$phys$\lambda$phys

## 7. Conclusion

The paper introduced MetaUG-PhyMOC-PVNet, a physics-guided multimodal deep learning framework, which has the ability to solve the challenging issue of robust PV defect detection in the presence of incomplete or corrupted sensor information via uncertainty gating. The heart of the innovation is a lightweight gating network that dynamically calculates a scalar reliability weight for the physics-informed loss term, provided as a function of whether the modality was available and how reliable the signal is, in a bilevel meta-reweighting optimization scheme. The formulation allows the model to adjust its importance of physical consistency constraints, avoiding over-fitting of erroneous assumptions in the absence or noisy measurements of critical modalities, like current-voltage curves, or electroluminescence. The gating weight is used to adjust the belief of the evidential open-set recognition head, and also offers a principled approach for indicating uncertain samples for human review and enable reliable maintenance prioritization.

MetaUG-PhyMOC-PVNet was extensively tested on a set of benchmark datasets that show that the method is superior to other existing approaches on several aspects. In the worst case of missing modality – the test set consists of only one modality – the framework performed on average at 0.86, whereas the fixed-weight physics-informed networks performed at 0.62 and the generative imputation approach performed at 0.78. The accuracy of the estimation of the power loss increased by 16.5% in terms of root mean square and the open set recognition rate was raised to 84.3% with an area under the receiver operating characteristic curve of 0.91. Ablation studies showed that adaptive gating mechanism, bilevel meta-learning optimization and uncertainty-modulated confidence scaling all play a significant role in the performance. The framework is thus a first step towards practical PV-inspection systems, which need to function under real world conditions of sensor degradation and data incompleteness and allow reliable defect detection and decision support for maintenance task under all circumstances, even when some sensor modalities are missing.

## References


1. U. Hijjawi, S. Lakshminarayana, T. Xu, G. P. Fierro, and M. Rahman, "A review of automated solar photovoltaic defect detection systems: Approaches, challenges, and future orientations," Solar Energy, vol. 266, Art. no. 112186, 2023, doi: 10.1016/j.solener.2023.112186.
2. M. Aghaei, M. Kolahi, A. Nedaei, N. S. Venkatesh, S. M. Esmailifar, A. M. Moradi Sizkouhi, A. Aghamohammadi, A. A. Oliveira, A. Eskandari, P. Parvin, J. Milimonfared, V. Sugumaran, and R. Rüther, "Autonomous Intelligent Monitoring of Photovoltaic Systems: An In-Depth Multidisciplinary Review," Progress in Photovoltaics: Research and Applications, vol. 33, no. 3, pp. 381–409, 2025, doi: 10.1002/pip.3859.
3. A. Eskandari, M. Aghaei, J. Milimonfared, and A. Nedaei, "A Weighted Ensemble Learning-Based Autonomous Fault Diagnosis Method for Photovoltaic Systems Using Genetic Algorithm," International Journal of Electrical Power & Energy Systems, vol. 144, Art. no. 108591, 2023, doi: 10.1016/j.ijepes.2022.108591.
4. W. Tang, Q. Yang, K. Xiong, and W. Yan, "Deep Learning Based Automatic Defect Identification of Photovoltaic Modules Using Electroluminescence Images," Solar Energy, vol. 201, pp. 453–460, 2020, doi: 10.1016/j.solener.2020.05.060.
5. I. Nasri, M. Mansouri, M. Trabelsi, L. Khriji, and A. W. Skorek, "A Review on Multimodal-Based Fault Detection and Diagnosis in Photovoltaic Systems: Fusion and Attention Mechanisms," IEEE Access, vol. 14, pp. 44382–44402, 2026, doi: 10.1109/ACCESS.2026.3675686.
6. Z. Liu, B. Zhou, D. Chu, Y. Sun, and L. Meng, "Modality Translation-Based Multimodal Sentiment Analysis Under Uncertain Missing Modalities," Information Fusion, vol. 106, Art. no. 102259, 2024, doi: 10.1016/j.inffus.2024.102259.
7. M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations," Journal of Computational Physics, vol. 378, pp. 686–707, 2019, doi: 10.1016/j.jcp.2018.10.045.
8. M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, and S. Nahavandi, "A Review of Uncertainty Quantification in Deep Learning: Techniques, Applications and Challenges," Information Fusion, vol. 76, pp. 243–297, 2021, doi: 10.1016/j.inffus.2021.05.008.
9. J. M. Tomczak, "Deep Learning with Bayesian Neural Networks: A Survey," Artificial Intelligence Review, vol. 58, 2025, doi: 10.1007/s10462-024-10907-8.
10. S. Li, W. Ma, J. Zhang, C.-H. Liu, J. Liang, Y. Liu, and V. C. M. Leung, "Meta-Reweighted Regularization for Unsupervised Domain Adaptation," IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 11, pp. 3492–3506, 2021, doi: 10.1109/TKDE.2020.2993258.
11. D. Saadaoui, M. Elyaqouti, K. Assalaou, D. Ben Hmamou, S. Lidaighbi, E. H. Arjdal, I. Choulli, and A. Elhammoudy, "Parameters extraction of single diode and double diode models using analytical and numerical approach: A comparative study," International Journal of Modelling and Simulation, vol. 45, no. 2, pp. 263–292, 2025, doi: 10.1080/02286203.2023.2226285.
12. W. Tang, Q. Yang, Z. Dai, and W. Yan, "Module defect detection and diagnosis for intelligent maintenance of solar photovoltaic plants: Techniques, systems and perspectives," Energy, vol. 297, Art. no. 131222, 2024, doi: 10.1016/j.energy.2024.131222.
13. R. P. Santamaria, M. Dhimish, G. A. R. Benatto, T. Kari, P. B. Poulsen, and S. V. Spataru, "From indoor to daylight electroluminescence imaging for PV module diagnostics: A comprehensive review of techniques, challenges, and AI-driven advancements," Micromachines, vol. 16, no. 4, Art. no. 437, 2025, doi: 10.3390/mi16040437.
14. T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423–443, 2019, doi: 10.1109/TPAMI.2018.2798607.
15. A. Thakfan and Y. Bin Salamah, "Artificial-Intelligence-Based Detection of Defects and Faults in Photovoltaic Systems: A Survey," Energies, vol. 17, no. 19, Art. no. 4807, 2024, doi: 10.3390/en17194807.

16. H. Kang, J. Hong, J. Lee, and S. Kang, "Photovoltaic Cell Defect Detection Based on Weakly Supervised Learning With Module-Level Annotations," IEEE Access, vol. 12, pp. 5575–5583, 2024, doi: 10.1109/ACCESS.2024.3349975.
17. Y. Hu, W. Cao, J. Ma, S. J. Finney, and D. Li, "Identifying PV module mismatch faults by a thermography-based temperature distribution analysis," IEEE Transactions on Device and Materials Reliability, vol. 14, no. 4, pp. 951–960, 2014, doi: 10.1109/TDMR.2014.2344534.
18. M. Shafiq and Z. Gu, "Deep residual learning for image recognition: A survey," Applied Sciences, vol. 12, no. 18, Art. no. 8972, 2022, doi: 10.3390/app12188972.
19. W. M. Shaban, "Detection and classification of photovoltaic module defects based on artificial intelligence," Neural Computing and Applications, vol. 36, pp. 16769–16796, 2024, doi: 10.1007/s00521-024-10000-z.
20. X. Xu, M. Liu, Y. Nie, K. Wang, and W. Xu, "Photovoltaic Module Fault Detection Technology Based on Remote Sensing Technology and Deeplabv3+ Model," IEEE Access, vol. 12, pp. 195472–195482, 2024, doi: 10.1109/ACCESS.2024.3515165.
21. 21] Y. Wang, J. Zhou, K. C. Paul, T. Zhao, and D. Sheng, "Explainability Approach-Based Series Arc Fault Detection Method for Photovoltaic Systems," IEEE Access, vol. 12, pp. 45530–45542, 2024, doi: 10.1109/ACCESS.2024.3381530.
22. A. Ghahremani, S. D. Adams, M. Norton, S. Y. Khoo, and A. Z. Kouzani, "Advancements in AI-Driven Detection and Localisation of Solar Panel Defects," Advanced Engineering Informatics, vol. 63, Art. no. 103104, 2025, doi: 10.1016/j.aei.2024.103104.
23. M. Ibrahim, S. K. Sadanandan, T. Ghaoud, V. S. Rajkumar, and M. Sharma, "Incipient Fault Detection in Power Distribution Networks: Review, Analysis, Challenges, and Future Directions," IEEE Access, vol. 12, pp. 112822–112838, 2024, doi: 10.1109/ACCESS.2024.3443252.
24. S. Afrasiabi, S. Allahmoradi, M. Afrasiabi, X. Liang, C. Y. Chung, and J. Aghaei, "A Robust Multi-Modal Deep Learning-Based Fault Diagnosis Method for PV Systems," IEEE Open Access Journal of Power and Energy, vol. 11, pp. 583–594, 2024, doi: 10.1109/OAJPE.2024.3497880.
25. X. Xu, M. Liu, Y. Nie, K. Wang, and W. Xu, "Photovoltaic Module Fault Detection Technology Based on Remote Sensing Technology and Deeplabv3+ Model," IEEE Access, vol. 12, pp. 195472–195482, 2024, doi: 10.1109/ACCESS.2024.3515165.
26. M. Seyedmahmoudian, A. Stojcevski, A. G. Olabi, A. T. D. Perera, N. A. Rahim, and A. Gandomi, "Explainable Deep Learning Model for Grid-Connected Photovoltaic System Performance Assessment for Improving System Reliability," IEEE Access, vol. 12, pp. 120729–120746, 2024, doi: 10.1109/ACCESS.2024.3452778.
27. A. Elhaj, M. Alharthi, A. Alqahtani, and M. A. Abido, "Recent Advances in Fault Detection Techniques for Photovoltaic Systems: An Overview, Classification and Performance Evaluation," Optik, vol. 306, Art. no. 171797, 2024, doi: 10.1016/j.ijleo.2024.171797.
28. S.-I. Ao and H. Fayek, "Continual Deep Learning for Time Series Modeling," Sensors, vol. 23, no. 16, Art. no. 7167, 2023, doi: 10.3390/s23167167.
29. E. S. Ortigossa, T. Gonçalves, and L. G. Nonato, "EXplainable Artificial Intelligence (XAI): From Theory to Methods and Applications," IEEE Access, vol. 12, pp. 80799–80846, 2024, doi: 10.1109/ACCESS.2024.3409843.
30. Z. Liu, B. Zhou, D. Chu, Y. Sun, and L. Meng, "Modality Translation-Based Multimodal Sentiment Analysis Under Uncertain Missing Modalities," Information Fusion, vol. 106, Art. no. 102259, 2024, doi: 10.1016/j.inffus.2024.102259.
31. I. Nasri, M. Mansouri, M. Trabelsi, L. Khriji, and A. W. Skorek, "A Review on Multimodal-Based Fault Detection and Diagnosis in Photovoltaic Systems: Fusion and Attention Mechanisms," IEEE Access, vol. 14, pp. 44382–44402, 2026, doi: 10.1109/ACCESS.2026.3675686.
32. M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-Informed Neural Networks: A Deep Learning Framework for Solving Forward and Inverse Problems Involving Nonlinear Partial Differential Equations," Journal of Computational Physics, vol. 378, pp. 686–707, 2019, doi: 10.1016/j.jcp.2018.10.045.
33. W.-Q. Lu, C. Delpha, D. Diallo, and A. Migan-Dubois, "A Comprehensive Review of Fusion-Based Methods Applied to Photovoltaic Fault Diagnosis," Solar Energy, vol. 309, Art. no. 114457, 2026, doi: 10.1016/j.solener.2026.114457.