E3S Web of Conferences 723, 01006 (2026) https://doi.org/10.1051/e3sconf/202672301006
AIEI 2026

Hierarchical Modeling for Low-Resolution and Ambiguous Thermal Fault Classification

Thanh Chi Nguyen[1,*] and Truc Thi Kim Nguyen[2,**]
[1]FPT School of Business & Technology, Vietnam
[2]University of Science and Technology - The University of Danang, Vietnam

Abstract

Thermal fault classification in photovoltaic modules remains challenging due to low-resolution infrared images, class imbalance, and strong visual ambiguity between fault types. Conventional approaches treat this task as a flat multi-class problem, which requires a single model to simultaneously separate heterogeneous classes and often leads to unstable decision boundaries. This paper presents a structured modeling approach that decomposes the classification process into a hierarchy of conditional decisions, including anomaly detection, family-level routing, and subtype classification. This formulation aligns the learning process with the intrinsic organization of fault categories and simplifies the underlying decision structure. Experiments conducted under a controlled setting show that the hierarchical formulation improves classification accuracy by up to 8.8% compared to a flat baseline using the same backbone and training conditions. Ablation results further demonstrate that each stage contributes to performance gains, while a collapsed-label analysis indicates that remaining errors are primarily concentrated in fine-grained subtype distinctions. These findings suggest that restructuring the classification problem is more effective than increasing model complexity for low-resolution and ambiguous thermal fault recognition.

1 Introduction

The rapid expansion of photovoltaic (PV) installations has made reliable and efficient monitoring increasingly critical. As global capacity continues to grow beyond the terawatt scale, even minor faults at the module level can accumulate into significant energy losses [1]. Thermal infrared (IR) imaging has emerged as a practical solution for large-scale inspection, as it enables the detection of temperature anomalies associated with defects such as hotspots, cracks, and diode failures [2, 3]. These advantages have driven the development of automated approaches for thermal fault classification in PV systems.

Despite recent advances in deep learning, thermal fault classification remains challenging due to the combined effects of low spatial resolution, severe class imbalance, and strong visual ambiguity between fault types. These factors hinder the ability of conventional models to learn stable and discriminative decision boundaries [4–6].

Most existing methods address this problem using a flat multi-class formulation, where a single model is trained to predict all fault categories simultaneously. Recent approaches based on convolutional neural networks and transformer architectures have demonstrated improved performance on thermal fault datasets [7, 8]. However, these models still operate within a unified decision space that must simultaneously separate both coarse and fine-grained classes. In practice, this often leads to confusion between visually similar fault types and limits performance under low-resolution and ambiguous conditions.

An alternative perspective is to explicitly consider the inherent structure of the problem. Fault categories are not independent; rather, they can be organized into meaningful groups based on their physical characteristics. Hierarchical classification has been explored in computer vision as a way to exploit such structure, allowing complex recognition tasks to be decomposed into simpler and more tractable subproblems [9]. More recent studies have further highlighted the importance of structured modeling for ambiguous and fine-grained classification tasks, particularly in challenging visual environments [10, 11].

Motivated by these observations, this paper investigates a hierarchical modeling framework for thermal fault classification. Instead of treating the task as a single-stage problem, the proposed approach decomposes it into multiple levels, including anomaly detection, fault family identification, and fine-grained subtype classification. This design reflects the natural organization of fault categories and enables each stage to focus on a more specific and coherent objective.

To evaluate the effectiveness of this formulation, we conduct a controlled experimental study on the InfraredSolarModules dataset [4], using consistent data splits and comparable backbone configurations. The results show that the hierarchical approach achieves a clear improvement over a flat baseline under identical conditions. Further analysis reveals that most remaining errors arise from confusion within fault subtypes rather than between broader categories, highlighting the importance of fine-grained discrimination.

Rather than benchmarking against heterogeneous experimental setups, this study focuses on a controlled evaluation to isolate and analyze the impact of hierarchical problem formulation under low-resolution and ambiguous conditions. Unlike prior work that primarily emphasizes architectural improvements, the proposed approach highlights the role of problem formulation by restructuring the classification process into a hierarchy of conditional decisions aligned with the intrinsic organization of fault categories. This perspective enables more effective modeling of heterogeneous decision boundaries and complements existing advances in model design.

The main contributions of this work are summarized as follows:

- We propose a structured probabilistic formulation for thermal fault classification that explicitly decomposes

* e-mail: thanhnc25msa33053@fpt.edu.vn
** e-mail: ntktruc@dut.udn.vn

https://doi.org/10.1051/e3sconf/202672301006

Table 1. Class Taxonomy and Family Assignment

Family	**Class**	**Characteristic**
Electrical	Cell	Single-cell fault
	Cell-Multi	Multi-cell fault
	Cracking	Cell cracking
	Diode	Single diode fault
	Diode-Multi	Multi-diode fault
Thermal-Local	Hot-Spot	Single hotspot
	Hot-Spot-Multi	Multiple hotspots
Env.-System	Shadowing	External shadow
	Soiling	Surface contamination
	Vegetation	Plant obstruction
	Offline-Module	Module offline
Normal	No-Anomaly	No fault

the prediction process into anomaly detection, family-level routing, and subtype classification.

- We demonstrate through controlled experiments that restructuring the classification problem yields significant performance gains over flat multi-class formulations, without increasing model complexity.
- We provide a comprehensive analysis of error patterns, showing that the proposed hierarchy effectively separates coarse-level categories, while remaining challenges are concentrated in fine-grained subtype distinctions.
- We highlight the importance of problem formulation in low-resolution and ambiguous visual recognition tasks, offering insights that extend beyond photovoltaic fault classification.

The remainder of this paper is organized as follows. Section II describes the dataset and challenges. Section III presents the proposed method. Section IV details the experiments and results. Section V concludes the paper.

2 Dataset and Challenges

This study uses the InfraredSolarModules dataset [2, 4], which consists of thermal images of PV modules captured from aerial inspections. Each sample corresponds to a cropped module with a very low spatial resolution of 24×40 pixels, providing only limited structural detail for visual analysis.

The dataset contains 12 classes (11 fault types and one normal class). As summarized in Table 1, these classes exhibit a clear semantic organization, where faults can be grouped into electrical, thermal-local, and environmental/system-related categories. However, this structure is not explicitly encoded in the learning objective and must therefore be implicitly inferred by the model.

Thermal fault classification in this setting is fundamentally challenging due to the interaction of three factors. First, the low resolution severely limits the availability of discriminative features, making it difficult to capture subtle spatial variations between fault patterns [5]. Second, the dataset is highly imbalanced, with several critical fault types having very few samples, which biases the learning process toward dominant classes [12,13]. Third, many fault categories exhibit strong intra-family similarity, where subclasses differ primarily in the extent or multiplicity of thermal signatures (e.g., Cell vs. Cell-Multi, Hot-Spot vs. Hot-Spot-Multi) [6,8].

These challenges are inherently interconnected. Low resolution amplifies ambiguity between visually similar subclasses, while class imbalance further reduces the model's ability to learn reliable decision boundaries for minority classes.

This observation suggests that the difficulty of the problem lies not only in feature extraction, but also in how the classification task is formulated. Although the label space naturally follows a hierarchical organization, treating it as a flat classification problem fails to exploit this structure.

Motivated by this limitation, we consider a structured reformulation that explicitly aligns the classification process with the taxonomy in Table 1. By decomposing the task into stages corresponding to different levels of abstraction, the proposed approach reduces the complexity of each sub-task and provides a more effective framework for handling low resolution, imbalance, and ambiguity.

As shown in Fig. 1, the dataset exhibits a pronounced long-tail distribution at both class and family levels. In particular, the No-Anomaly class dominates the dataset, while several fault subtypes contain only a few hundred samples.

Importantly, the imbalance is not uniform. Most minority classes correspond to fine-grained subtypes that are also visually similar to their corresponding major classes. At the family level, thermal-local faults are significantly underrepresented compared to electrical and environmental categories.

This multi-level imbalance further motivates the use of hierarchical modeling, where different stages can address class imbalance and ambiguity at appropriate levels of abstraction.

3 Structured Probabilistic Fault Modeling

A closer examination of the dataset suggests that the difficulty of thermal fault classification arises not only from limited visual information, but also from how the problem is formulated. In particular, normal samples, fault families, and fine-grained subtypes exhibit distinct statistical behaviors. Treating all classes within a single flat classification space forces a model to approximate a highly heterogeneous decision boundary, which becomes especially challenging under low-resolution and ambiguous conditions.

To address this issue, we adopt a structured probabilistic formulation that decomposes the prediction into a sequence of conditional decisions aligned with the taxonomy in Table 1. This formulation separates decisions across different semantic levels and allows each component to be learned under a more coherent and homogeneous distribution.

3.1 Hierarchical Posterior Decomposition

Let x denote an input image and c the corresponding class label. We introduce a binary anomaly variable $a \in \{0,1\}$ and a family mapping $f(c)$.

For the normal class, we define:

$$P(\text{No-Anomaly} \mid x) = P(a = 0 \mid x). \tag{1}$$

For any anomalous class c, the posterior can be decom-

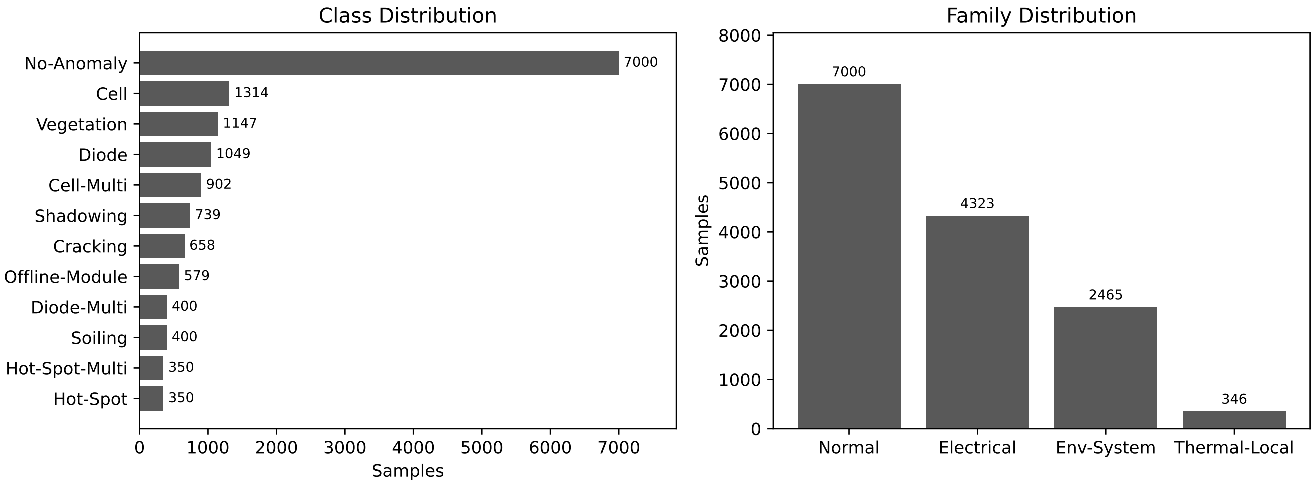

Figure 1. Distribution of training samples across classes (left) and fault families (right). Exact sample counts are shown on each bar. The dataset exhibits a pronounced long-tail distribution, with severe imbalance at both class and family levels.

posed as:

$$P(c \mid x) = P(a=1 \mid x)\,P(f(c) \mid x, a=1) \cdot P(c \mid f(c), x, a=1), \quad c \neq \text{No-Anomaly}. \tag{2}$$

This decomposition transforms the original multi-class prediction into a hierarchy of conditional sub-problems. By explicitly conditioning on intermediate variables, the model reduces the effective complexity of the decision boundary. In contrast, a flat classifier must implicitly learn all dependencies within a single mapping, which is more difficult under heterogeneous data distributions.

3.2 Hierarchical Prediction Model

The above decomposition is implemented using three sequential predictors, as illustrated in Fig. 2.

The first stage estimates the anomaly probability:

$$P(a=1 \mid x) = \sigma(z^{(1)}(x)). \tag{3}$$

Conditioned on $a = 1$, the second stage predicts the family-level distribution:

$$\mathbf{p}^{(2)}(x) = \text{softmax}(\mathbf{z}^{(2)}(x)). \tag{4}$$

The third stage assigns the final class within each family:

$$\mathbf{p}^{(3,f)}(x) = \text{softmax}(\mathbf{z}^{(3,f)}(x)). \tag{5}$$

Here, $z^{(1)}(x)$ denotes the logit of the anomaly classifier, $\mathbf{z}^{(2)}(x)$ represents logits over fault families, and $\mathbf{z}^{(3,f)}(x)$ corresponds to logits over subtypes within family f.

3.3 Model Instantiation

All stages share a common backbone $\phi(x)$ for feature extraction. In this work, we adopt standard vision backbones (e.g., Swin-Tiny and EfficientNet-B3) to evaluate the proposed formulation.

Task-specific heads are attached on top of the shared features:

- Stage 1: binary classifier with sigmoid output
- Stage 2: multi-class classifier over fault families
- Stage 3: separate classifiers for each family

This design ensures that any observed improvement arises from the structured formulation rather than architectural differences.

3.4 Flat vs. Hierarchical Formulation

For comparison, the flat baseline uses the same backbone $\phi(x)$ followed by a single classifier:

$$\mathbf{p}(x) = \text{softmax}(\mathbf{z}(x)). \tag{6}$$

In this setting, the model must simultaneously separate normal samples, fault families, and subtypes within a single decision space. In contrast, the hierarchical formulation decomposes this problem into simpler conditional tasks, each operating on a more structured and coherent subset of the data. This decomposition is expected to improve learning under both ambiguity and imbalance.

Handling Class Imbalance. The proposed formulation explicitly addresses class imbalance at multiple levels of the hierarchy. At the anomaly stage, weighted binary cross-entropy compensates for the dominance of normal samples. At the family level, logit adjustment shifts decision boundaries according to class priors, thereby reducing bias toward majority families. At the subtype level, balanced softmax further mitigates long-tail effects by incorporating class frequency into the normalization term.

This multi-level strategy differs from conventional approaches that apply a single imbalance correction globally. Instead, each stage adapts to the specific distribution of its corresponding sub-task, resulting in more stable and balanced learning.

3.5 Stage-Aware Optimization

Each stage corresponds to a distinct learning problem and is optimized using a task-specific objective.

At the anomaly stage, weighted binary cross-entropy is employed to handle imbalance between normal and faulty samples. At the family level, logit-adjusted cross-entropy [14] is applied:

$$\tilde{z}_c = z_c + \tau \log \pi_c, \tag{7}$$

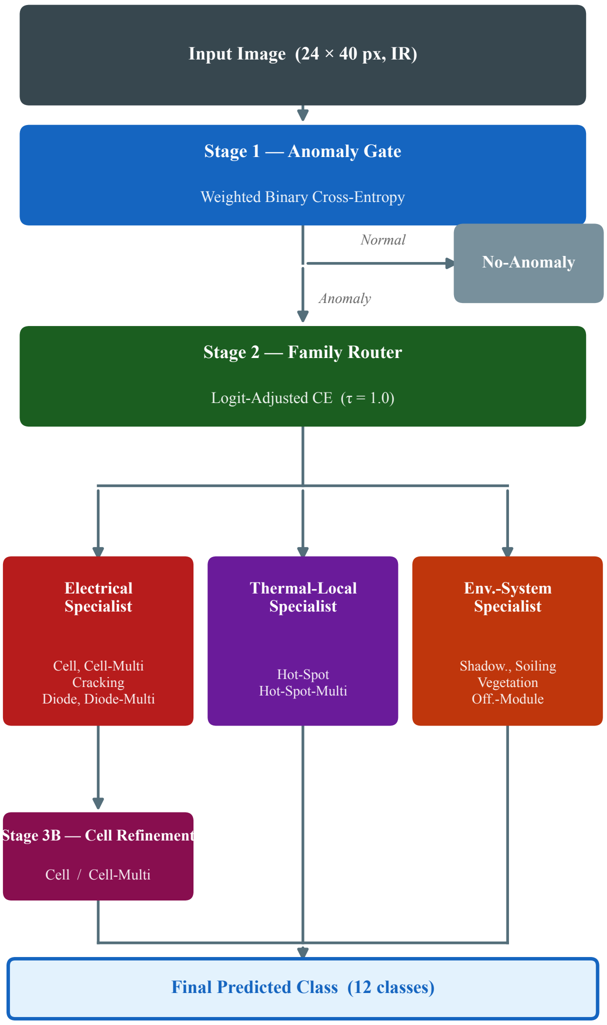

Figure 2. Overview of the proposed hierarchical formulation. The model decomposes prediction into anomaly detection, family-level routing, and family-specific classification, with stage-aware optimization.

which adjusts decision boundaries according to class priors.

At the subtype level, balanced softmax [15] is used:

$$P(c \mid x, f) = \frac{\pi_c e^{z_c}}{\sum_j \pi_j e^{z_j}}. \tag{8}$$

Notation. Here, π_c denotes the empirical class prior, τ controls the strength of prior adjustment, and τ_a is the anomaly threshold used during inference.

3.6 Preserving Spatial Structure

Let the original image size be $H \times W$. We define:

$$s = \frac{S}{\max(H,W)}, \tag{9}$$

and resize the image to:

$$(H', W') = (\lfloor sH \rfloor, \lfloor sW \rfloor). \tag{10}$$

The resized image is then padded to form a square input, preserving spatial geometry while avoiding distortion.

Algorithm 1 Hierarchical Training Procedure

Require: Training dataset $\mathscr{D}$
Train anomaly gate g_1
Extract anomalous subset $\mathscr{D}_a$
Train family router g_2 on $\mathscr{D}_a$
for each family f **do**
    Train specialist $g_3^{(f)}$ on subset $\mathscr{D}_f$
end for
Train refinement head if needed

Algorithm 2 Hierarchical Inference

Require: Input image x
Preprocess x
$p_a \leftarrow g_1(x)$
if $p_a < \tau_a$ **then**
    return No-Anomaly
end if
$\hat{f} \leftarrow \arg\max g_2(x)$
$\hat{c} \leftarrow \arg\max g_3^{(\hat{f})}(x)$
return $\hat{c}$

Training Procedure

The hierarchical formulation naturally induces a stage-wise training strategy. Instead of optimizing all decisions jointly, each component is trained on a subset of the data corresponding to its role in the hierarchy. This reduces interference between tasks and allows each stage to focus on a more coherent data distribution.

The overall process is summarized in Algorithm 1. The anomaly gate is first trained to separate normal and faulty samples. The resulting anomalous subset is then used to train the family router, followed by family-specific classifiers trained on their respective subsets. This sequential refinement ensures that each stage operates on increasingly structured data, improving stability under imbalance and ambiguity.

Inference

At inference time, prediction follows the same hierarchical structure. Rather than evaluating all classes simultaneously, decisions are made progressively, with each stage conditioning on the output of the previous one.

This process is summarized in Algorithm 2. The anomaly gate first determines whether a sample is normal. For anomalous samples, the family router selects the most likely category, and the corresponding specialist classifier produces the final prediction. By restricting fine-grained classification to relevant subsets, the model reduces confusion between unrelated fault types and yields more stable predictions.

Discussion

The proposed approach demonstrates that performance improvements can be achieved by restructuring the learning problem rather than increasing model complexity. By decomposing prediction into conditional stages, the model operates on simpler and more structured sub-tasks, leading to improved robustness under low-resolution and ambiguous conditions.

Table 2. Comparison of flat and hierarchical models

Model	Acc.	wF1	mF1
Flat Swin-Tiny	0.7480	0.7625	0.6560
Hierarchical Swin-Tiny	**0.8360**	**0.8351**	**0.6976**

4 Experiments and Results

4.1 Experimental Setup

All experiments were conducted on Kaggle GPU kernels (NVIDIA T4) using mixed-precision training. The implementation is based on PyTorch with the `timm` library for model initialization.

All models were trained and evaluated using a fixed 70/15/15 data split with a consistent random seed. Hierarchical models were trained sequentially, where each stage was optimized independently and its best checkpoint was used for subsequent stages. This controlled setup enables a consistent evaluation of the proposed hierarchical formulation under identical conditions.

We report three evaluation metrics: accuracy (Acc.), weighted F1 (wF1), and macro F1 (mF1). Accuracy is used as the primary metric, while weighted and macro F1 provide additional insight into class imbalance and minority-class performance.

Backbone Models and Training Details. We consider two backbone architectures: Swin-Tiny and EfficientNet-B3, both initialized from ImageNet-pretrained weights. Swin-Tiny [16] is a transformer-based model that captures long-range spatial dependencies through shifted-window attention, while EfficientNet-B3 [17] serves as a strong convolutional baseline with high parameter efficiency.

This comparison is particularly relevant for low-resolution thermal images, where fine-grained spatial cues are limited and global context may play a more important role. By evaluating both architectures under identical conditions, we assess whether attention-based representations provide additional benefits in capturing subtle thermal patterns.

For fairness, the same backbone is used in both flat and hierarchical settings. The flat baseline employs a single classifier over all classes, whereas the hierarchical models use stage-specific heads as described in Section III, ensuring that performance differences arise from the prediction structure rather than model capacity.

Data Augmentation. Given the extremely low resolution of the input images, data augmentation is applied conservatively to avoid distorting thermal patterns. The pipeline includes random horizontal flipping, small-angle rotations, and mild brightness and contrast adjustments. All images are resized using aspect-ratio-preserving scaling followed by padding.

4.2 Main Results and Analysis

Table 2 summarizes the main results. The most critical comparison is between the flat and hierarchical Swin-Tiny models, which share the same backbone and differ only in prediction structure.

Table 3. Backbone comparison under hierarchical modeling

Backbone	Acc.	wF1	mF1
EfficientNet-B3	0.8200	0.8170	0.6765
Swin-Tiny	**0.8360**	**0.8351**	**0.6976**

Table 4. Ablation study of hierarchical components

Configuration	Acc.	wF1	mF1
Flat baseline	0.7480	0.7625	0.6560
+ Anomaly gate	0.7920	0.8015	0.6683
+ Family routing	0.8210	0.8237	0.6842
Full hierarchy	**0.8360**	**0.8351**	**0.6976**

The hierarchical formulation improves accuracy by +0.0880 under identical conditions, indicating that the gain is primarily due to structured prediction rather than increased model capacity.

We further evaluate the effect of backbone choice under the hierarchical formulation.

Swin-Tiny consistently outperforms EfficientNet-B3 across all metrics, suggesting that attention-based representations better capture spatial relationships in upsampled low-resolution thermal images.

4.3 Ablation and Error Analysis

To validate the contribution of each component, we conduct an ablation study by progressively adding stages of the hierarchical formulation.

Performance improves consistently as hierarchical components are introduced. The anomaly gate simplifies the decision space by separating normal samples, while family-level routing reduces inter-class interference by restricting predictions to semantically related categories.

To further analyze residual errors, we perform a collapsed-label experiment by merging visually similar subtype pairs.

The improvement across all metrics indicates that a significant portion of the remaining error arises from fine-grained subtype distinctions, particularly between visually similar patterns such as single-instance and multi-instance faults.

Fig. 3 provides further insight through confusion matrix analysis. Most misclassifications occur within the same fault family rather than across different families. For example, confusion is mainly observed between Cell and Cell-Multi, and between Hot-Spot and Hot-Spot-Multi, while cross-family errors are relatively rare.

These observations confirm that the model effectively captures coarse-level structure, while the remaining challenge lies in fine-grained discrimination. Overall, the results demonstrate that performance gains are achieved by restructuring the learning problem rather than increasing model complexity.

5 Conclusion

This paper presented a hierarchical formulation for thermal fault classification in low-resolution infrared images. By decomposing prediction into anomaly detection, family-level routing, and subtype classification, the proposed approach aligns the learning process with the structure of the label

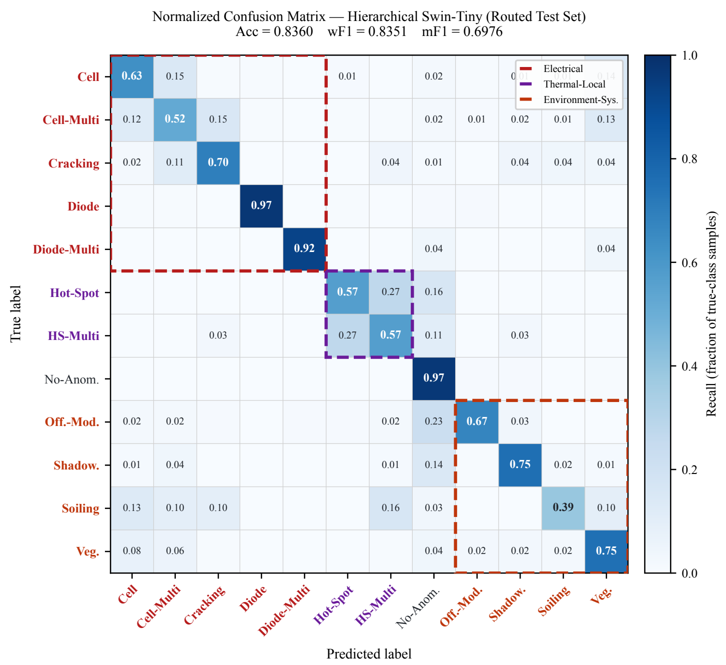

Figure 3. Normalized confusion matrix of the hierarchical model. Most errors are concentrated within the same fault family, indicating effective coarse-level separation.

Table 5. Collapsed-label analysis

Setting	Acc.	wF1	mF1
12-class	0.8360	0.8351	0.6976
9-class	**0.8643**	**0.8635**	**0.7453**

space. Experimental results demonstrate that this formulation consistently improves performance over flat classification under identical conditions. The findings suggest that, in challenging scenarios with ambiguity and imbalance, improving problem formulation can be more effective than increasing model complexity. Future work will focus on improving fine-grained subtype discrimination and exploring more integrated training strategies.

References

[1] J. Polo, N. Martín-Chivelet, and M. Alonso-García, "Automatic fault classification in photovoltaic modules using convolutional neural networks," *Energies*, 2022.

[2] C. e. a. Buerhop-Lutz, "A benchmark for infrared-based fault classification in photovoltaic systems," *Progress in Photovoltaics*, 2022.

[3] S. e. a. Deitsch, "Automatic classification of defective photovoltaic module cells," *Solar Energy*, 2019.

[4] R. Maps, "Infrared solar modules dataset," 2020, kaggle.

[5] H. e. a. Wu, "Feature-space balancing for solar fault classification," *Engineering Applications of Artificial Intelligence*, 2024.

[6] M. Aksoy, "Solar panel fault detection using swin transformer," *arXiv preprint*, 2025.

[7] A. e. a. Pamungkas, "Infrared solar module fault detection using deep learning," *Sensors*, 2023.

[8] A. e. a. Ramadan, "Photovoltaic fault classification using transformer neural networks," *Energy Conversion and Management*, 2024.

[9] J. e. a. Deng, "Imagenet: A large-scale hierarchical image database," in *CVPR*, 2009.

[10] Y. Zhang and X. Liu, "Hierarchical representation learning for fine-grained recognition," *IEEE TPAMI*, 2026.

[11] T. Nguyen and H. Pham, "Structured modeling for ambiguous visual classification," *Pattern Recognition*, 2026.

[12] Y. Cui, M. Jia, T.-Y. Lin, Y. Song, and S. Belongie, "Class-balanced loss based on effective number of samples," *CVPR*, 2019.

[13] K. e. a. Cao, "Learning imbalanced datasets with label-distribution-aware margin loss," *NeurIPS*, 2019.

[14] A. K. Menon, S. Jayasumana, A. S. Rawat, H. Jain, A. Veit, and S. Kumar, "Long-tail learning via logit adjustment," in *ICLR*, 2021.

[15] J. Ren, C. Yu, X. Ma, H. Zhao, S. Yi, and H. Li, "Balanced meta-softmax for long-tailed visual recognition," in *NeurIPS*, 2020.

[16] Z. Liu *et al.*, "Swin transformer: Hierarchical vision transformer using shifted windows," in *ICCV*, 2021.

[17] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *ICML*, 2019.