Indian Journal of Pure & Applied Physics
Vol. 63, November 2025, pp. 979-993
DOI: 10.56042/ijpap.v63i11.20991

Hybrid Multiscale Contextual Framework for Enhanced Fault Detection in Photovoltaic Electroluminescence Imaging

Prabhakar Sharma[a, b*] & Ritesh Kumar Mishra[b]

[a]Department of Electronics and Communication Engineering, ITS Engineering College, Greater Noida 201 308, India

[b]Department of Electronics and Communication Engineering, National Institute of Technology, Patna 800 005, India

Received: 25th June 2025; accepted: 6th October 2025

Electroluminescence (EL) imaging has demonstrated efficacy in identifying cracks, inactive areas, and other concealed faults that are frequently undetectable in visible-spectrum examinations. Nonetheless, conventional deep learning models, such as standalone convolutional neural networks (CNNs), encounter limitations in generalization, sensitivity to complex features, and robustness at diverse fault sizes. This study introduces an innovative Hybrid Multiscale Contextual Framework (HMCF) architecture that combines two powerful networks, EfficientNetB0 and ResNet50 to extract diverse features. Proposed model also introduces a Multiscale Feature Fusion Block (MFFB) to handle diverse fault sizes. The findings validate the effectiveness of integrating hybrid CNN architectures with multiscale feature fusion for precise, scalable, and dependable fault classification in photovoltaic (PV) modules, facilitating the development of intelligent and automated solar farm inspection systems. The proposed architecture demonstrates a detection accuracy of 93.18%, significantly outperforming leading deep neural network approaches.

Keywords: Deep learning (DL), Multiscale feature fusion block (MFFB), Electroluminescence (EL), Hybrid multiscale contextual framework (HMCF), Photovoltaic (PV), Convolutional neural network (CNN)

1 Introduction

Solar energy serves as a clean and renewable substitute for conventional energy sources responsible for carbon emission in the environment. It is facilitating its extensive application in both daily life and industry. A notable development in 2024 was a significant increase in renewable energy capacity, which included the addition of 585 gigawatts (GW) of renewables, resulting to a 15.1% rise in the global renewable energy capacity. These renewable energy sources constituted a historic 92.5% of worldwide power increases, mostly driven by substantial advancements in solar and wind energy. PV technology predominates global solar power generation, accounting for 1,858,622 MW of the total 1,865,490 MW, which is around 99.63%[1]. This dominance is attributed to the reduction in costs, enhancement of efficiency, and improvement in the quality of PV solar panels over the past decades[2].

However, PV cells can have numerous defects and deficiencies that can impact the overall energy efficiency of the PV energy system. Therefore, it is necessary to analyse solar cells, beginning with their production phase while carrying out evaluations throughout their operational lifespan. Considering the recent developments in energy systems, it is essential to establish a reliable and comprehensive evaluation to guarantee the expected extraction of energy from solar PV modules. EL imaging is a standard inspection method for detecting defects in PV cells due to the high quality images it generates[3]. Defects in cells, such as cracks, fingers, hotspots, edge isolation, electrical-isolation, material-defects and back contact, can be identified by analysing the EL image of a PV cell, as illustrated in Fig. 1. According to the result of EL inspection, engineers replace defective cells with good ones.

Inspection conducted by experienced engineers to identify defects manually is prohibitively expensive, time-intensive and possibility of inspection errors. Consequently, there has been a demand for the automation of EL examination to find out defects. Efforts have been undertaken to utilize machine learning (ML) for the automation of defective cell inspection in PV modules. Conventional image classification techniques incorporate feature extraction, classification, and related processes. Feature extraction enables the collection of significant local descriptive information that is resistive to noise. The primary techniques for feature extraction encompass Scale-

*Corresponding author: E-mail: prabhakar.sh@gmail.com

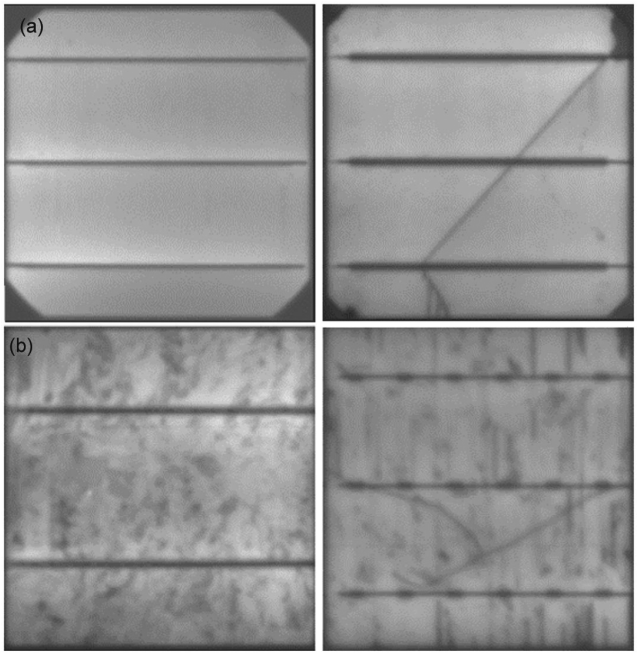

Fig. 1 — (a) Good and defective monocrystalline; and (b) polycrystalline Electroluminescence imaging of PV solar module

Invariant Feature Transform (SIFT)[4], Speeded Up Robust Features (SURF)[5], Histogram of Oriented Gradients (HOG)[6], and Local Binary Patterns (LBP)[7], among others. Additionally, certain studies advocate for the integration of various extraction techniques employed during the feature extraction process to mitigate the loss of valuable information. A fixed-dimension vector can be derived via feature encoding[8] following feature extraction and subsequently fed into the classifier for classification. The primary classifiers employed for image classification are K-Nearest Neighbor (KNN)[9], Support Vector Machine (SVM)[10], Random Forest (RF)[11] and several ensemble techniques.

With the advancement of deep learning, CNNs are increasingly employed in image classification, utilizing cell images as input to determine their defectiveness[12-16]. Defect detection technologies are primarily classified into Photoluminescence (PL)[17,18] and EL. As certain faults in solar cells can only be seen through EL imaging of PV modules, the majority of present-day methods employ EL for-defect detection in solar cells. Tan *et al*.[19] employ the Reconfigurable Adaptive Focus Background Suppression Detector (RAFBSD), which possesses exceptional proficiency in accurately detecting faults across 12 distinct classifications of PVEL images. Qian *et al*.[20] successfully accomplished crack identification by combining features from various short-term and long-term deep features. Furthermore, Abdullah-Vetter *et al*.[21] utilized ML methodologies for identifying defects in EL images. Conversely, Ge *et al*.[22] presented the Hybrid Fuzzy Convolutional Neural Network (HFCNN), effectively integrating conventional fuzzy theory with CNN technology to achieve significant improvements in EL image processing. Nonetheless, these researches are confined to basic defect identification in EL images.

Su *et al*.[23-25] included attention modules into the Faster R-CNN framework and employed an efficient BAFPN to identify three distinct types of defects in EL images. Acikgoz *et al*.[26] presented an innovative approach employing a deep evolutionary neural network model to precisely identify seven distinct categories of defects in EL images. Wang *et al*.[27] utilized the coordinate attention (CA) mechanism on feature maps and implemented ResNet152-Xception to proficiently classify eight distinct categories of defects in EL images. Fu and Cheng[28] effectively included Efficient Long-Range Convolutional Network (ELCN) into the YOLOv7 algorithm, facilitating the efficient detection of four distinct categories of defects in EL images. Lu *et al*.[29] integrated CA with the YOLOv5 algorithm, resulting in notable performance in identifying nine categories of defects in EL images. Li *et al*.[30] attained significant outcomes by integrating the Global Channel and Spatial Context (GCSC)module with the YOLOv7 algorithm for the effective identification of four distinct classes of defects in EL images. Fast and precise detection of these defects is essential for preserving the durability and power generating capacity of PV systems. The demand for automated, accurate, and scalable solutions in defect identification is ncreasingly apparent. As the number of solar plant increases, deep learning (DL) has become an efficient tool for managing the associated challenges. DL has the capacity to analysis extensive datasets or recognize complex patterns to identify faults in solar cells with exceptional accuracy. Defects should be found as soon as feasible to reduce losses and enhance solar plant performance[31]. With the utilization of CNNs and other advanced deep learning architectures, automated systems may accurately identify issues and categorize them, hence providing valuable insights for maintenance and quality control[32]. Recent advancements in AI-assisted sensing

devices[33–37] exhibit enhanced sensitivity and data dependability across many fields, underscoring the significance of artificial intelligence methodologies, such as CNNs, for effective fault detection in photovoltaic modules utilizing Electroluminescence images.

Automated technologies substantially decrease the time and expense associated with problem discovery; hence, their utilization is very useful in large solar plants. Moreover, deep learning models improve defect detection by continuously enhancing their performance through the combination of supplementary data and intensive training, significantly increasing both the accuracy and efficiency of solar PV plant operations[38-39]. The popular CNN models employed in PV defect classification include ResNet, SqueezeNet, AlexNet[40], DenseNet, MobileNet[41], visual geometry group (VGGNet)[42], EfficientNet[43], Vision Transformer (ViT)[44], U-Net[45] andYOLO[46]. Manual visual evaluation is unfeasible in large solar PVenergy systems, which require the inspection of thousands of PV modules. The proposed HMCF framework enhances reliability, efficiency, and scalability of photovoltaic modules under real-world constraints through fault detection, similar to how tunnel field-effect transistors (TFET) have shown that material and structural optimization can markedly improve device reliability, efficiency, and scalability across various operating conditions[47-51].

This paper introduces an autonomous evaluation method employing EL images to detect defects in monocrystalline and polycrystalline solar PV modules byusing a novel deep learning architecture. In summary, the key contributions of this work are as follows:

i The proposed structure adopted a unique hybridmodel architecture, exhibits a sophisticated integration of advanced neural network components, precisely developed to enhance image classification efficiency. An input image is concurrently analysed by two robust feature extractors: ResNet50, distinguished for its deep residual learning, and EfficientNetB0, noted for its efficient scaling and computing efficacy. This dual-pathway approach ensures the acquisition of a diverse and comprehensive set of characteristics.

ii Subsequently, the Convolutional Block Attention Module (CBAM) is applied to the outputs of both networks, augmenting the feature maps by sequentially emphasizing the most salient channels and spatial locations hence improving the quality of feature representations, resulting in higher classification performance. The improved features are then combined, integrating further information from both networks.

iii The integrated feature set is further enhanced by the proposed Multiscale Feature Fusion Block (MFFB), which efficiently gathers multi-scale contextual information by examining the feature maps with filters of varying receptive fields. The output from the MFFB module undergoes adaptive average pooling, is then flattened, and is ultimately processed through fully connected layers to obtain the final classification. This HMCF is an innovative method for efficient feature extraction and classification. This proposed architectural integration is the novel approach and has several advantages.

2 Methodology

Different deep learning techniques have been employed for image classification, with the primary approach being CNN. This is due to CNN capability to extract patterns, classify, and detect problems in PV module images[52,53]. CNN is capable to solve various difficulties in computer vision and pattern recognition, including image classification, object detection, facial recognition, and natural language processing[54-55].

The proposed HMCF architecture as shown in Fig. 2 is a hybrid deep learning framework designed to improve the performance of image classification tasks by utilizing the advantages of several advanced components. The input image is concurrently processed by two pre-trained CNNs: EfficientNetB0 and ResNet50, both pre-trained on ImageNet dataset, serve as the foundational networks. The features extracted by these feature extractors are individually refined by CBAMs. These refined features are concatenated together. Now these features are processed by Multiscale feature fusion block (MFFB), which ensures that the model is able to identify faults of different sizes and shapes. EfficientNetB0 and ResNet50 are recognized for their capacity to capture significant spatial and contextual information from images. ResNet50, which facilitates deep residual learning to capture complicated hierarchical features with minimal degradation, and EfficientNetB0, recognized for its combined scaling method that attains perfect balance between precision and computational efficiency to get efficient, high-resolution features. Features from both models are concurrently retrieved

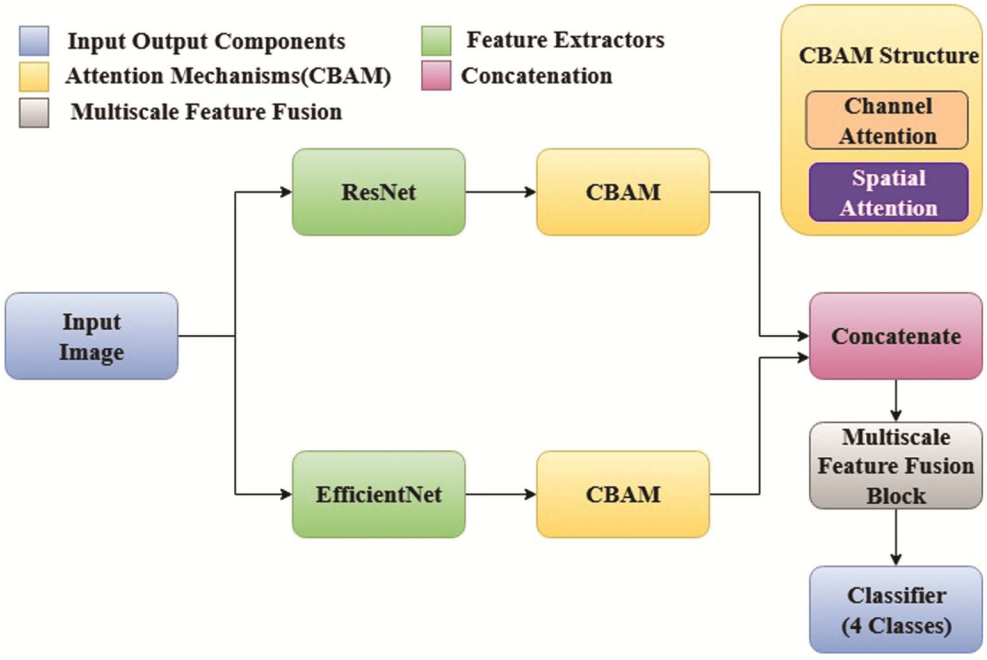

Fig. 2 — Proposed Hybrid CNN Model

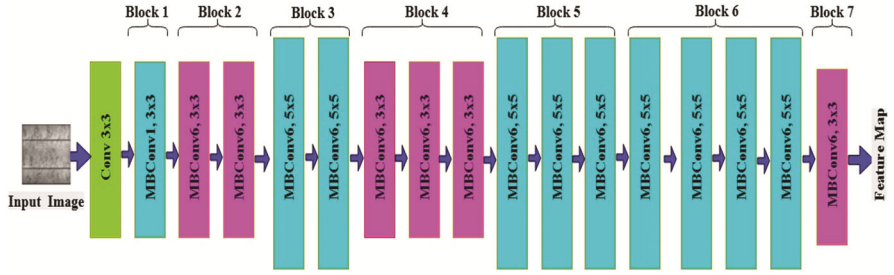

Fig. 3 — Architecture and internal components of EfficientNetB0

and subsequently integrated to form a robust, complete feature representation from the input image that builds on the advantages of both models.

2.1 EfficientNetB0

EfficientNetB0 signifies a notable improvement in CNN architectures, attaining an optimal balance between model efficacy and computational performance. The architecture is fundamentally based on the revolutionary compound scaling method, which uniformly changes the network's depth, width, and resolution through set scaling coefficients, hence ensuring optimal resource use and improved accuracy across various workloads. The architecture, shown in Fig. 3, includes Mobile Inverted Bottleneck Convolution (MBConv) blocks, derived from MobileNetV2, which integrate depth wise and point wise convolutions within an inverted residual framework. This approach improves computational performance and facilitates the extraction of complex information. Moreover, EfficientNetB0 incorporates Squeeze-and-Excitation (SE) optimization, which adaptively adjusts channel-wise feature responses by modelling relationships among channels. This method enables the network to concentrate on the most important features, hence enhancing representational capacity. The integration of these elements allows EfficientNetB0 to attain superior accuracy with markedly fewer parameters and reduced processing requirements relative to earlier CNN architectures, making it especially appropriate for use in resource-limited environments.

2.2 ResNet50

ResNet50 is a crucial architecture, especially recognized for its effectiveness as a feature extractor

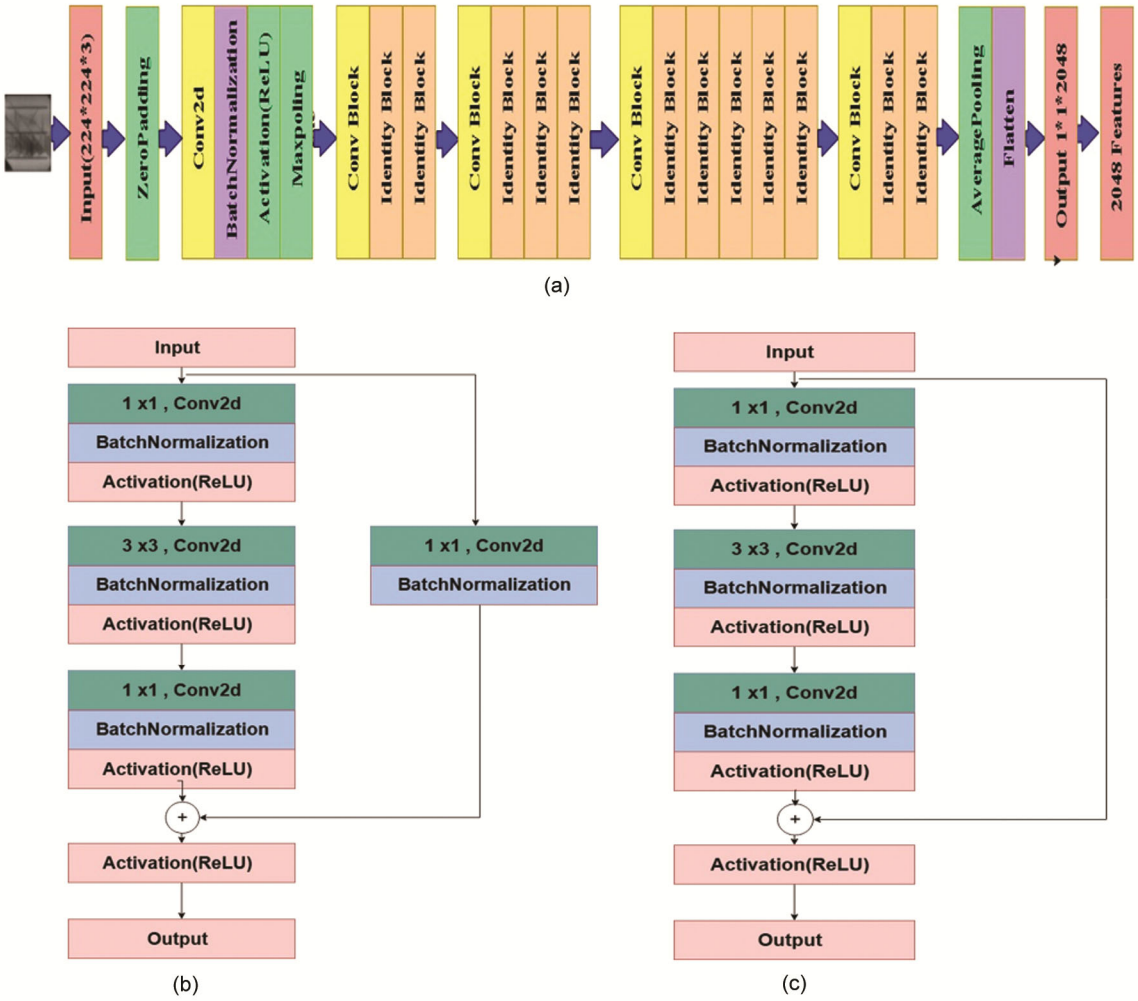

Fig. 4 — (a) The architecture and internal components of aResNet; (b) The architecture of the convolutional block; and (c) The architecture of the identity block

in complicated image analysis applications. This 50-layer deep CNN is characterized by its novel application of residual learning and skip connections, hence facilitating the training of considerably deeper networks. As a feature extractor, ResNet50's earliest layers' capture basic visual patterns, while the deeper levels incorporate higher-level representations, resulting in a comprehensive hierarchical feature collection. This hierarchical structure is especially beneficial for transfer learning; models pre-trained on large datasets can be fine-tuned for particular tasks with diminished processing demands and enhanced generalization.

The architecture and internal components of a deep CNN derived from the ResNet family is shown in Fig. 4 (a). It illustrates the comprehensive flow of the network, commencing with a 224×224×3 input image that is subjected to zero padding, followed by an initial convolutional layer, batch normalization, ReLU activation, and max pooling. The network's core consists of a series of convolutional and identity blocks that systematically extract hierarchical feature representations. The concluding layers consist of average pooling and flattening, resulting in a 2048-dimensional feature vector, appropriate forfurther classification. Figure 4 (b) and 4 (c) focus on the architecture of the convolutional block and identity block, respectively. The convolutional block comprises a sequence of 1×1, 3×3, and 1×1 convolutional layers, each succeeded by batch normalization and ReLU activation, along with a parallel shortcut path utilizing a 1×1 convolution to align dimensions. This block facilitates the acquisition of detailed features while mitigating the degradation

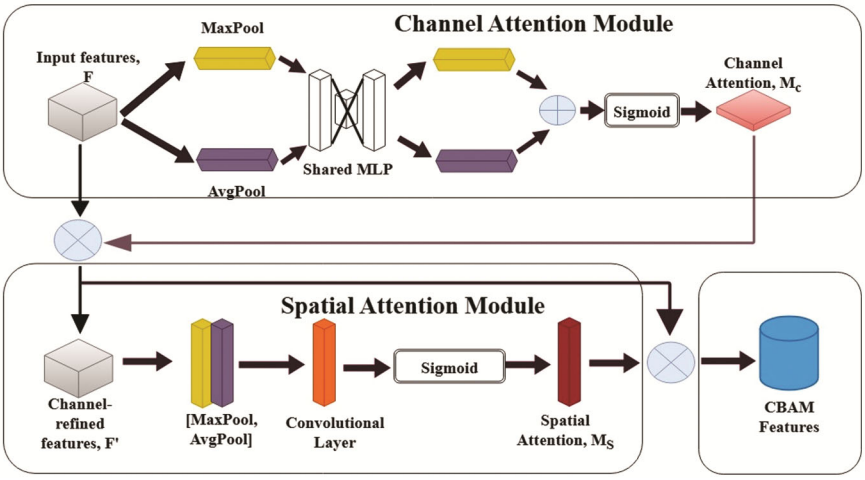

Fig. 5 — Convolutional Block Attention Module (CBAM), Channel Attention Module (CAM) out feed into the Spatial Attention Module (SAM)

issue in deep networks. Figure 4 (c) preserves the shortcut identity mapping and is employed when the input and output dimensions are identical. Collectively, these residual blocks enable the building of deeper networks while maintaining steady gradient flow, enhancing convergence, and improves representational capacity, rendering them exceptionally effective in intricate computer vision applications.

2.3 Convolutional Block Attention Module (CBAM)

A CBAM is independently applied to the output of each feature extractor to boost the quality of features extracted. The refined outputs from both networks are subsequently concatenated to create an integrated feature representation, thereby merging the mutually beneficial information acquired by each backbone. CBAM enhances the feature maps by progressively highlighting the most informative channels and spatial location, allowing the model to concentrate on relevant patterns while diminishing less valuable information. It has two sub-modules: Fig. 5 illustrates the Channel Attention Module (CAM) and the Spatial Attention Module (SAM), which are implemented in sequence[56-57].

The channel attention module emphasizes informative features by simulating the inter-channel relationships of feature maps. This is accomplished by consolidating spatial information using global average and max pooling procedures, subsequently employing a shared multi-layer perceptron (MLP), to generate channel-wise attention weights. The weights are subsequently employed to readjust the input feature map along the channel dimension. Given an intermediate feature map $F \in R^{C\times H\times W}$ where C, H, and W indicated the number of channels, height, and width respectively, the channel attention mechanism focuses on identifying important features. Channel attention is computed as:

$$Mc(F)=\sigma(MLP(AvgPool(F))+MLP(MaxPool(F)))\in R^{C\ X\ 1\ x\ 1} \quad \ldots (1)$$

In the given scenario, AvgPool and MaxPool execute global average pooling and maximum pooling across spatial dimensions respectively. The shared MLP comprises two completely connected layers. The sigmoid activation function 'σ' guarantees that the attention values range from 0 to 1. The input feature map is then refined using:

$$F' = Mc(F) \otimes F \quad \ldots (2)$$

where $\otimes$ signifies element-wise multiplication (broadcasted across spatial dimensions).

In addition, the spatial attention module identifies focal points within the feature map by bringing out important spatial locations. It employs max pooling and average pooling operations along the channel axis, concatenates the resultant two-dimensional feature maps, and processes them via a convolutional layer to produce spatial attention weights. These are subsequently employed to enhance the feature map in the spatial dimension.

Spatial attention is calculated from the channel-refined feature map *F'* as follows:

$$Ms(F') = (f^{7X7}([AvgPool(F'); MaxPool(F')])), Ms(F') \in R^{1\ X\ H\ X\ W} \quad \ldots (3)$$

Here, (;) signifies channel-wise concatenation, while $f^{7\times7}$ refers to a convolutional layer containing a

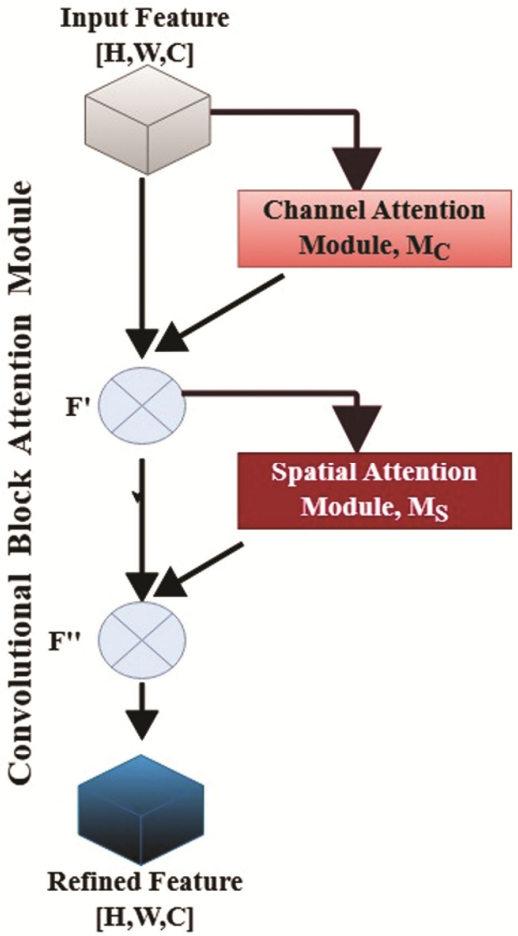

Fig. 6 — Flow diagram of Convolutional Block Attention Module (CBAM)

7×7 kernel. A sigmoid activation signifies output values within the range of 0 to 1. The final refined feature map finally gets obtained as follow:

$$F'' = Ms(F') \otimes F' \quad \ldots (4)$$

CBAM integrates both attention methods to adaptively modify feature representations, enabling the network to concentrate on significant regions while eliminating less important information. CBAM introduces negligible computational overhead and may be effortlessly integrated into current CNN architectures, resulting in enhanced performance across diverse computer vision tasks as shown in Fig. 6

2.4. Multiscale Feature Fusion Block (MFFB)

This block acquires multiscale information with the use of diverse dilation rates in simultaneous convolutional layers. This work proposed an enhanced version of atrous spatial pyramid pooling named as MFFB module to enhance the extraction of multi-scale contextual information while preserving computational efficiency by incorporating at attention mechanisms and efficient convolutions into the conventional ASPP structure as shown in Fig. 7. The MFFB comprises several parallel atrous (dilated) convolution branches with varying dilation rates r∈{1,6,12,18,24}, enabling the module to capture a range of receptive fields. The MFFB includes both local and global contexts, allowing the model to identify fine and broad structural patterns in PV images, which are crucial for identifying faults. Each

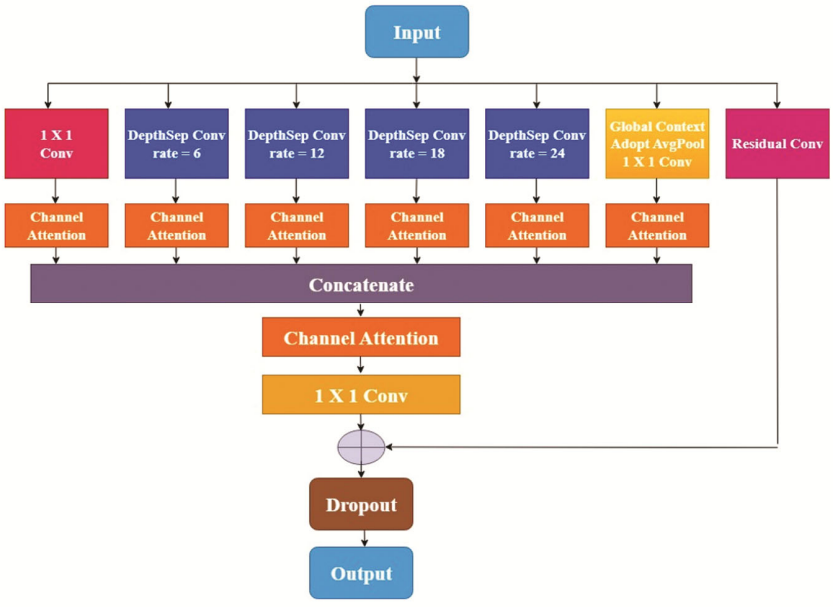

Fig. 7 — Enhanced Atrous Spatial Pyramid Pooling (EASPP) architecture named as Multiscale Feature Fusion Block (MFFB)

branch processes the input feature map $F \in R^{C\times H\times W}$ utilizing a depth wise separable convolution, which breaks down standard convolution into a depth wise convolution succeeded by a point wise convolution. For a specified dilation rate r, the branch output is delineated as:

$$F_r = \text{ReLU}(\text{BN}(Conv^{1X1}(\text{ReLU}\ (\text{BN}\ (Conv_{3X3}^{(r)}(\text{F})))))) \quad \ldots (5)$$

where $Conv_{3X3}^{(r)}$represent a depthwise convolution with dilation rate r, and BN and ReLU denote Batch Normalization and the ReLU activation function, respectively.Furthermore, a global context branch calculates:

$$F_{gc} = \text{ReLU}(\text{BN}(Conv_{1x1}\ (\text{GAP}(\text{F})))) \quad \ldots (6)$$

where GAP denotes global average pooling. The resultant feature map is subsequently bi-linearly up-sampled to correspond with the spatial dimensions of F. To recalibrate feature responses, each output branch, including global context, is adjusted using a Squeeze-and-Excitation (SE) block.

$$F_r' = F_r \cdot \sigma(Conv_{1x1}(\text{ReLU}(Conv_{1x1}\ (\text{GAP}(F_r))))) \quad \ldots (7)$$

where 'σ' denotes the sigmoid activation function. All enhanced outputs are concatenated:

$$F_{cat} = \text{Concat}(F_{r1}', F_{r2}', F_{r3}', \ldots\ldots\ldots, F_{rn}', F_{gc}') \quad \ldots (8)$$

Subsequently, the channel attention method is implemented on *Fcat*:

$$F_{att} = F_{cat} \cdot \sigma(Conv_{1x1}(\text{ReLU}(Conv_{1x1}\ (\text{GAP}(F_{cat}))))) \quad \ldots (9)$$

The attended features are ultimately merged via a 1×1 convolution and integrated with a residual connection:

$$F_{out} = \text{Dropout}(\text{ReLU}(\text{BN}(Conv_{1x1}\ (F_{att}))) + F_{res}) \quad \ldots (10)$$

where F_{res} is either an identity mapping or a 1×1 projection of the input, based upon the channel dimensions. This approach allows the MFFB module to adaptively concentrate on useful multi-scale features with no extra overhead, which is highly effective for intensive prediction applications like semantic segmentation.

2.5 Classification

The resulting feature map, encompassing both detailed and multiscale data, is transmitted through fully interconnected layers to execute the final classification, detecting various defects in solar images, including micro-cracks, finger, hot spot, short circuit etc.

This architecture proposes a comprehensive solution for fault detection in solar panels through deep learning by integrating efficient and deep feature extraction techniques with multiscale contextual framework. The combined use of Efficient NetB0 and ResNet50, along with the included multiscale dilation block, guarantees that the model can precisely and effectively identify and categorize defects.

2.6 Loss Function

Cross-entropy loss is a commonly employed objective function in problems relating to classification, especially in deep learning models like CNNs. It measures the divergence between the expected probability distribution and the actual label distribution. Represent the true label for a specific sample as a one-hot encoded vector, $y = [y_1, y_2, y_3 \ldots\ldots\ldots y_{CT}]$, where $y_1 \in \{0,1\}$ and CT is the total number of classes. Let $\hat{y} = [\hat{y}_1, \hat{y}_2, \hat{y}_3 \ldots\ldots\ldots \hat{y}_C]$ represent predicted probability vector produced by the model via a soft max activation. The cross-entropy loss LS_{CE} for a single sample is defined as:

$$Ls_{CE} = -\sum_{i=0}^{C} y_i \log(\hat{y}_i) \quad \ldots (11)$$

This loss function penalizes incorrect predictions more heavily as the confidence of the incorrect class increases, effectively encouraging the model to assign high probability to the correct class. In the case of a batch of N samples, the average loss is computed as:

$$Ls_{CE}^{BATCH} = -\frac{1}{N}\sum_{n=1}^{N}\sum_{i=1}^{C} y_i^{(n)} \log(\hat{y}_i^{(n)}) \quad \ldots (12)$$

Cross-entropy loss is differentiable and aligns closely with the maximum likelihood estimation under a categorical distribution assumption, making it especially suitable for training models using gradient descent techniques.

3 Results

This section offers an in-depth discussion of the experimental process, including evaluation metrics, hardware and software utilized, dataset construction, augmentation techniques, comparison with existing

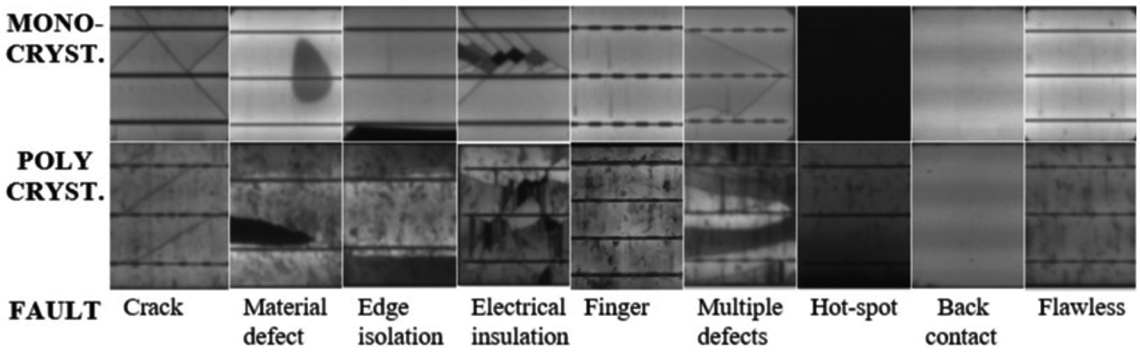

Fig. 8 — Different types of faults found in elpv-dataset

standard CNN models, and an assessment of the proposed HMCF compared to the other author's work. Presenting the experimental findings in tabular and graphical format enhances intuitive understanding.

3.1 Evaluation Metrics

Four standard metrics have been employed to assess the classification performance of the proposed method: recall, precision F1 score and accuracy (ACC).

$$\text{Recall} = T\_P/(T\_P + F\_N) \quad \ldots (13)$$

$$\text{Precision} = T\_P/(T\_P + F\_P) \quad \ldots (14)$$

$$\text{F1 Score} = 2*(\text{Precision} * \text{Recall})/(\text{Precision} + \text{Recall}) \quad \ldots (15)$$

$$\text{Accuracy} = (T\_P + T\_N)/(T\_P + T\_N + F\_P + F\_N) \quad \ldots (16)$$

True positives are represented as T_P, false positives as F_P, true negatives as T_N, and false negatives as F_N.

3.2 Hardware and Software

The hardware and software implementation utilized the Py Torch framework, with experiments conducted on an Nvidia GeForce RTX 3090 24 GB GPU, supported by 128 GB of Corsair Vengeance RGB PRO DDR4 3200 MHz RAM and a 3.2 GHz Intel i9 processor with 12 cores and 24 threads.

3.3 Dataset Construction

In this work, we use *zae-bayern/elpv* dataset[58,59]. The dataset comprises 2,624 grayscale images, each measuring 300 by 300 pixels, representing good and faulty solar cells with varied types of defects. The defects in these images are categorized as monocrystalline-good, monocrystalline-defective, polycrystalline-good, and polycrystalline-defective. Monocrystalline and polycrystalline solar panels image with various types of defects with flawless images is shown in Fig. 8. It illustrates eight types of defects found in the available *elpv-dataset*: crack, material defect, edge isolation, electrical insulation, finger, multiple defects, hot spot and back contact. For experiments, we have divided the dataset into three parts: training, validation, and testing. The partitioning ratio for this experiment is 60:20:20, allocating 60% to the training dataset and 20% each to testing and validation. Table 1 presents the partitioning details of the dataset, including the total count of training, testing, and validation images. For training, we use an initial learning rate of 1e-4 with cosine decay scheduling and implemented dropout rate of 0.5 in the decoder and classification layers. The values were chosen based on existing research and limited validation trials that exhibited consistent performance throughout training iterations.

3.4 Augmentation Techniques

To enhance the generalization capacity of the proposed HMCF framework, various data augmentation techniques were utilized during training. Random rotations of ±15°, horizontal and vertical flipping, random cropping with a variance of up to 10%, and mild intensity jittering were specifically applied to the input photos. The selected augmentations were chosen for their suitability for Electroluminescence (EL) images, where minor fluctuations in orientation, illumination, or local contrast commonly appear in practice. Augmentation operations were conducted stochastically during each training period to enhance dataset variety and mitigate the risk of over fitting. A quantitative assessment of model performance, both with and without augmentation, is presented in Table 2. The proposed HMCF model trained without data augmentation, obtained a precision of 0.85, a recall of

0.84, F1-score of 0.85, and an AUC of 85.45%. After implementing data augmentation, the proposed model demonstrated a significant enhancement in performance, achieving precision, recall, and F1-scores of 0.93, with an AUC of 93.18% (Table 2).

This substantial improvement verifies that data augmentation is essential for enhancing model robustness by introducing variability, hence enabling the model to more effectively identify the complex patterns of various types of faults in different PV cell types. This study attempts to enhance the precision of identifying solar cell defects via EL images.

However, implementing typical data augmentations, such as horizontal flipping, small rotations, random cropping, and intensity jitter to enhance generalization, a quantitative ablation analysis for each augmentation is not included in this work.

3.5 Comparison with Existing Standard CNN Models

This section provides a comparative examination of different CNN models for detecting faults in solar PV cells utilizing EL images. The models were evaluated on their capacity to categorize both flawless and faulty solar cells for monocrystalline and polycrystalline types. Performance parameters such as precision, recall, F1-score, and accuracy were calculated to evaluate the durability and reliability of each model.

These experiments were performed on a dataset including EL images that were appropriately labelled, and the evaluation was performed independently for each cell type.

Table 1 — Partitioning details of used datasets

Dataset	Training	Testing	Validation	Total
Monocrystalline- good	286	95	95	476
Monocrystalline-defective	402	134	131	667
Polycrystalline-good	471	156	155	782
Polycrystalline-defective	419	140	140	699
Total overall				2624

Table 2 — Performance of proposed HMCF model with and without augmentation

	Precision	Recall	F1 Score	AUC (%)
Without data augmentation	0.85	0.84	0.85	85.45
With data augmentation	0.93	0.93	0.93	93.18

3.5.1 Monocrystalline Silicon Cell Classification

Table 3 presents the results for the classification of monocrystalline cells. The proposed HMCF had the superior performance among all models, attaining a precision and F1 score of 0.95 and 0.91 for flawless cells and 0.94 and 0.97 for faulty cells respectively. This illustrates a robust capacity to differentiate among the four categories, especially in recognizing monocrystalline solar cell faults with high recall rate of 0.99.

ResNet50 and EfficientNetB0 demonstrated commendable performance after proposed hybrid model, attaining F1-scores of 0.93/0.96 and 0.94/0.94, for flawless and faulty cells respectively. In this table, highest performance metrics is represented by bold letter and subsequent performance model is represented by underline. VGG19 demonstrated marginally superior performance relative to VGG16, with F1-scores of 0.88 and 0.92, in contrast to 0.84 and 0.90 for flawless and faulty cells respectively. Advanced models such as ResNeXt101, Dense Net121, and DenseNet161 gives comparatively inferior performance, especially in flawless cell detection.

3.5.2 Polycrystalline Silicon Cell Classification:

Table 4 depicting the results for the classification of polycrystalline solar cells. The proposed HMCF

Table 3 — Performance of CNN models in Monocrystalline Silicon Cell detection3

S. No	CNN Model	Flawless monocrystalline cell			Faulty monocrystalline cell		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score
1	ResNet50	0.93	0.90	0.93	0.93	0.98	0.96
2	EfficientNetB0	0.92	0.93	0.94	0.96	0.97	0.94
3	VGG16	0.83	0.86	0.84	0.91	0.89	0.90
4	VGG19	0.88	0.87	0.88	0.92	0.92	0.92
5	ResNext101	0.83	0.76	0.80	0.86	0.90	0.88
6	DenseNet121	0.83	0.81	0.82	0.88	0.90	0.89
7	DenseNet161	0.89	0.80	0.84	0.88	0.94	0.91
8	MobileNet	0.85	0.77	0.81	0.85	0.90	0.88
9	Proposed HMCF	0.95	0.94	0.91	0.94	0.99	0.97

Table 4 — Performance of CNN models in Polycrystalline Silicon Cell detection

S. No	CNN Model	Flawless polycrystalline cell			Faulty polycrystalline cell		
		Precision	Recall	F1-Score	Precision	Recall	F1-Score
1	ResNet50	0.86	0.93	0.89	0.92	0.82	0.87
2	EfficientNetB0	0.90	0.92	0.91	0.91	0.88	0.89
3	VGG16	0.91	0.96	0.93	0.93	0.87	0.90
4	VGG19	0.91	0.96	0.94	0.92	0.87	0.90
5	ResNext101	0.89	0.90	0.89	0.86	0.84	0.85
6	DenseNet121	0.85	0.91	0.88	0.85	0.76	0.81
7	DenseNet161	0.92	0.88	0.90	0.83	0.90	0.86
8	MobileNet	0.83	0.87	0.85	0.81	0.75	0.78
9	Proposed HMCF	0.93	0.96	0.94	0.93	0.89	0.91

Table 5 — Performance Comparison of proposed hybrid model with other existing standard models

S. No	CNN Model	Precision	Recall	F1-Score	Overall Accuracy
1	ResNet50	0.91	0.91	0.91	90.82
2	EfficientNetB0	0.92	0.92	0.92	92.23
3	VGG16	0.91	0.91	0.91	90.62
4	VGG19	0.92	0.92	0.92	91.58
5	ResNext101	0.87	0.87	0.87	86.67
6	DenseNet121	0.85	0.85	0.85	85.29
7	DenseNet161	0.89	0.88	0.88	88.38
8	MobileNet	0.83	0.83	0.83	83.05
9	Proposed HMCF	0.93	0.93	0.93	93.18

outperformed all standard CNN models, attaining F1-scores of 0.94 for flawless cells and 0.91 for faulty cells. VGG19 and VGG16 attained similar outcomes with F1-scores of 0.94/0.93 for flawless and 0.90/0.90 for faulty cells, respectively. EfficientNetB0 shows uniform execution across both cases, with F1-scores of 0.91 and 0.89. Alternative models like ResNeXt101, DenseNet121, and MobileNet exhibited diminished efficacy, especially in identifying faulty polycrystalline cells. The performance gap is due to noise and uneven textures in polycrystalline cell images, challenging precise defect detection for models with fewer parameters or fewer feature extraction abilities. In this table, highest performance metrics are represented by bold letter and subsequent performance model is represented by underline. The suggested model outperforms all other techniques having the highest value of precision, recall and F1-score.

From Tables 3 and 4, the slightly lower performance of polycrystalline cells compared to monocrystalline cells can be attributed to the inherent characteristics of polycrystalline modules. Polycrystalline solar cells typically exhibit more heterogeneous surface textures, irregular grain boundaries, and higher levels of background noise in EL images. These factors reduce the contrast between defective and non-defective regions, making defect boundaries less distinct and consequently more challenging for the model to classify. In contrast, monocrystalline cells possess a more uniform crystal structure and cleaner EL patterns, which allow defects to appear more sharply, resulting in higher classification accuracy. Despite these challenges, our proposed hybrid model still achieved competitive results on polycrystalline cells, outperforming other standard CNN architectures.

3.5.3 Overall Performance Comparison

Table 5 shows the weighted average precision, recall, F1-score, and overall accuracy for both types of cells: monocrystalline and polycrystalline. In both the cases, proposed HMCF had the best overall performance, with an F1-score of 0.93 and a classification accuracy of 93.18%, which is about 1% higher than most efficient model and about 10% higher than lowest performing model, EfficienctNetB0 and MobileNet respectively. In this table, highest performance metrics are represented by bold letter and subsequent performance model is represented by underline. This shows that proposed hybrid model can generalize accurately and give reliable outcomes across all fault categories. Models like ResNeXt101, DenseNet121, and MobileNet did not do as well, with MobileNet having the lowest total accuracy at 83.05%. This shows that there is a trade-

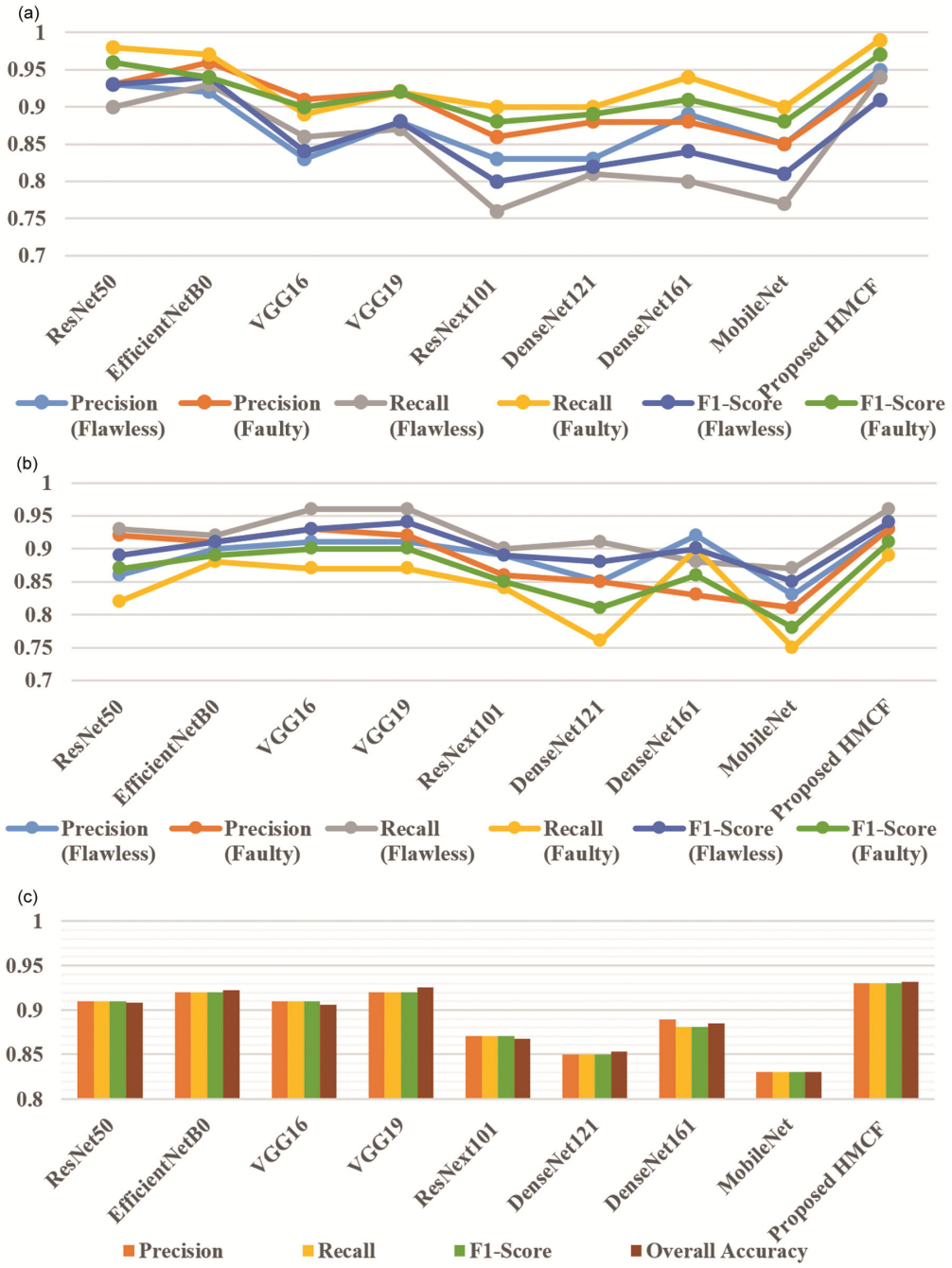

Fig. 9 — Performance of proposed Model in Flawless and Faulty (a) Monocrystalline Cells; (b) Polycrystalline cells; and (c) Performance Evaluation using weighted Average Value.

off between how complicated a model is and how well it can classify things. Fig. 9 shows the performance of flawless and faulty monocrystalline and polycrystalline cell with overall performance evaluation using average weightage values. Deeper or hybrid models are better at finding faults, but they also require more computing capability.

In spite of the suggested HMCF framework attaining high precision and recall, a qualitative analysis of errors indicates that certain limits persist. false positives are predominantly linked to categories that display significant visual similarity, such as edge isolation and micro-cracks, which may have fine textural or luminance patterns. False negatives commonly occur in

Table 6 — Performance comparison of proposed hybrid model with other authors'

S. No	Authors Work	Precision	Recall	F1-Score	Accuracy
1	Fan T *et al*. [60]	0.90	0.87	0.88	89.23
2	Eesaar H *et al*. [61]	0.92	0.90	0.92	92.12
3	Rahman M R *et al*. [62]	0.91	0.91	0.88	91.34
4	Karimi A M *et al*. [3]	0.92	0.81	0.85	86.34
5	Tan H *et al*. [19]	0.77	0.79	0.78	76.30
6	Proposed HMCF	0.93	0.93	0.93	93.18

cases where cracks are small or partially concealed by noise, rendering them challenging to identify even for skilled observers. Although HMCF diminishes these errors relative to baseline architectures via its multiscale contextual learning, further misclassifications remain in difficult scenarios.

Proposed HMCF model accurately captures both low-level and high-level information which enhances its precision and recall, especially in detecting multiple faults. CBAM enhances the feature maps by focusing on the most important channels and spatial location.MFFB module enhance the extraction of multi-scale contextual information while maintaining computational efficiency, which make this proposed model better than existing ones. Although lightweight models such as MobileNet provide faster prediction but lower accuracy making them less appropriate for high-precision industrial inspection applications. The findings indicate that implementing hybrid deep learning architectures can markedly enhance defect classification in PV cells, hence improving the reliability of solar panels and the quality control of production.

3.6 Comparison with Existing Work

In this section, performance of the proposed model is evaluated against different state-of-the-art methodologies documented in the literature to assess its effectiveness on the same dataset (*zae-bayern/elpv* dataset[50,51]) with same data partitioning strategies. In this paper, 60:20:20 partitioning scheme (training: validation: test) and applied augmentation strategies as described in section dataset construction is adopted for proposed methodology as well as for the existing work presented by other authors. Table 6 presents a comparison of the proposed model with other techniques by different authors in their research work in terms of evaluation metrics: precision, recall, F1-score and accuracy.

The proposed model outperformed all other techniques across all evaluation metrics. The proposed model exhibited a precision, recall and F1 score of 0.93, 0.93, and 0.93 respectively with an overall accuracy of 93.18%. T. Fan *et al*.[60]achieved a precision, recall and F1-score of 0.90, 0.87 and 0.88 respectively with accuracy of 89.23%. The methodology employed by H. Eesaar *et al*.[61] achieved a precision and F1-score of 0.92, alongside an accuracy of 92.12%, indicating reduced sensitivity and a more balanced performance. M.R. Rahman *et al.*[62] achieved a recall score of 0.91, comparable to ours; however, their F1-score diminished to 0.88, indicating an imbalance in their predictions. A.M. Karimi *et al*.[3] achieved a precision score of 0.92, although their recall score was significantly lower at 0.81, resulting in an F1-score of 0.85 and an accuracy of 86.34%. Similarly, H. Tan *et al*.[19] achieved an architecture F1 score of 0.78 with an accuracy of 76.30, the lowest among all.

In this table, bold showing the best effective model and underline values illustrating second-best. This comparative study demonstrates that the proposed model maintains a robust equilibrium between precision and recall, yielding an elevated F1-score and overall classification accuracy. The model's constant performance across all metrics indicates its robustness and suitability for diversified situations, making it an excellent option for real-world fault detection applications.

4 Limitation

A detailed comparison of parameter counts between conventional models and the proposed HMCF framework is given in Table 7. This comparison highlights the trade-off between precision and efficiency. Although lightweight models facilitate real-time inference on edge devices, they frequently exhibit inferior performance in tough scenarios where defect signs are subtle or distorted by noise. Conversely, massive networks attain significant accuracy; nevertheless, their substantial computational and energy requirements limit scalability in real-world solar farms. Through the integration of advanced multiscale contextual learning, fault detection accuracy is continuously boosted across all defect categories,

Table 7 — Model wise number of parameters in millions

S. No	Model	Number of parameters(millions)
1.	ResNet50	25.6
2.	EfficientNetB0	5.3
3.	VGG16	138.4
4.	VGG19	143.7
5.	ResNeXt101	84.3
6.	DenseNet121	8.0
7.	DenseNet161	28.7
8	MobileNetV2	3.5
9.	Proposed HMCF	43.9

thereby minimizing false detections and consequent maintenance expenses. Consequently, the additional computational burden is justified by the tangible operational advantages in photovoltaic defect investigation.

The proposed HMCF framework demonstrates great overall precision and recall; nevertheless, an examination of misclassified data exposes several limits. False positives occur more frequently in fault categories exhibiting significant visual similarities, such as hot-spot and back-contact defects, or edge isolation and micro-cracks. False negatives are frequently linked to faults that are barely visible, partially obscured, or compromised by image noise. Despite the hybrid architecture with multiscale feature fusion mitigating these mistakes relative to conventional CNN baselines, residual misclassifications persist in more difficult scenarios. The next development of this research will incorporate a confusion matrix-based error analysis to quantify misclassifications on a per-class basis and investigate solutions such as uncertainty-aware modeling and targeted augmentation that can reduce these errors.

5 Conclusion and Discussion

This study presents a novel Hybrid Multiscale Contextual Framework (HMCF) model that combines ResNet50 and EfficientNetB0 models for efficient fault detection in solar panels utilizing EL images. This hybrid approach utilizes the extensive depth of ResNet50 and the computational efficiency of EfficientNetB0 to subtle defects in EL images and improved reliability of fault classification. In this novel approach, we also introduce CBAM and proposed MFFB to enhance the accuracy from previous CNN models. CBAM ensures that model is able to focus on important features and MFFB augmented the model's ability to capture multi-scale contextual information, leading in an overall accuracy of 93.18%, which is 1% higher than the available conventional CNN model (EfficientNetB0). In this work although, overall efficiency is high but at the cost of increase computational complexity. The hybrid model provides a reliable, scalable, and efficient approach for automated identification of solar panel faults. For future work, the methodology can be extended to larger-scale PV modules, incorporating real-time monitoring and adaptive learning to further enhance predictive accuracy while keeping computational complexity at check.

References

1 Prime J, Ahmed I A, Akande D, Elhassan N, Melnikov Y & Whiteman A, IRENA, *Renew capacity statistics,*(Abu Dhabi), 2025 p.21.

2 Sharma P & Mishra R K, *Mater Renew Sustain Energy*, 14 (2025) 21.

3 Karimi A M, Fada J S, Hossain M A, Yang S, Peshek T J, Braid J L & French R H, *IEEE J Photovolt,* 9 (2019) 1324.

4 Lowe D G, *Int J Comput Vis,* 60 (2004) 91.

5 Bay H, Ess A, Tuytelaars T & Van Gool L, *Comput Vis Image Understand*, 110 (2008) 346.

6 Dalal N & Triggs B, *IEEE Comput Soc Conf Comput Vis Pattern Recognit (CVPR)*, San Diego, CA, USA, 2005, p. 886.

7 Ojala T, Pietikäinen M &Mäenpää T, *IEEE Trans Pattern Anal Mach Intell,* 24 (2002) 971.

8 Jégou H, Perronnin F, Douze M, Sánchez J, Pérez P &Schmid C, *IEEE Trans Pattern Anal Mach Intell,* 34 (2012) 1704.

9 Aha D W, Kibler D F & Albert M K, *Mach Learn*, 6 (1991) 37.

10 Mantel C, Villebro F, Reis Benatto G A, Parikh H R, Wendlandt S, Hossain K, Poulsen P, Spataru S, Sera D & Forchhammer S, Machine learning prediction of defect types for electroluminescence images of photovoltaic panels, paper presented to SPIE International Society for Optics and Photonics, San Diego, United States,August 2019, 1113904.

11 Breiman L, *Mach Learn*, 45 (2001) 5.

12 Ahmad A, Jin Y, Zhu C, Javed I, Maqsood A & Akram M W, *IET Renew Power Gener*, 14 (2020) 2693.

13 Tella H, Hussein A, Rehman S, Liu B, Balghonaim A & Mohandes M, *Case Stud ThermEng,* 66 (2025) 105749.

14 Lu S D, Wang M H, Wei S E, Liu H D & Wu C, *Processes*, 9 (2021) 1635.

15 Chen X, Karin T & Jain A, *Sol Energy*, 242 (2022) 20.

16 Abdelsattar M, Abdelmoety A, Ismeil M A & Emad-Eldeen A, *IEEE Access*, 13 (2025) 4136.

17 Trupke T, Bardos R A, Schubert M C & Warta W, *ApplPhys Lett*, 89 (2006) 1854.

18 Johnston S, Guthrey H, Yan F, Zaunbrecher K, Al-Jassim M, Rakotoniaina P &Kaes M, *IEEE J Photovolt*, 4 (2014) 348.

19 Tan H, Zhou H, Duan C, Zhang T & Zhang L, *IEEE Access*, 12 (2024) 61512.

20 Qian X, Li J, Cao J, Wu Y & Wang W, *Neural Netw*, 127 (2020) 132.

21 Abdullah-Vetter Z, Buratti Y, Dwivedi P, Sowmya A, Trupke T & Hameiri Z, Localization of defects in solar cells using luminescence images and deep learning, paper presented to *IEEE* 48th Photovolt Spec Conf, Fort Lauderdale, USA,0745-0749 June 2021.

22 Ge C, Liu Z, Fang L, Ling H, Zhang A & Yin C, *IEEE Trans Parallel Distrib Syst*, 32 (2021) 1653.

23 Su B, Zhou Z & Chen H, *IEEE Trans IndInformat*, 19 (2023) 404.

24 Su B, Chen H, Chen P, Bian G, Liu K & Liu W, *IEEE Trans IndInformat*, 17 (2021) 4084.

25 Su B, Chen H & Zhou Z, *IEEE Trans Ind Electron*, 69 (2022) 3161.

26 Acikgoz H, Korkmaz D & Budak U, *Expert Syst Appl*, 229 (2023) 120546.

27 Wang J, Bi L, Sun P, Jiao X, Ma X, Lei X & Luo Y, *Sensors*, 23 (2022) 297.

28 Fu H & Cheng G, *Energy Sour Part A Recovery Utilization Environ Effects*, 45 (2023) 8686.

29 Lu S, Wu K & Chen J, *IEEE Access*, 11 (2023) 71026.

30 Li J, Wu W & Chen H, GCSC detector: A detector for photovoltaic cell defect based on deep learning, paper presented to 42nd Chin Control Conf, Tianjin, China, 6913–6917 July 2023.

31 Liu Y, Xu J & Wu Y, *IEEE Trans Ind Electron*, 69 (2022) 8452.

32 Wang Y, Li L, Sun Y, Xu J, Jia Y, Hong J, Hu X, Weng G, Luo X, Chen S, Zhu Z, Chu J & Akiyama H, *Energy*, 229 (2021) 120606.

33 Yadav S, Thakur V, Kaul A, Das A, Rewari S & Nand D, Gallium Nitride Gate-All-Around Macaroni Field Effect Transistor (GaN-GAA-MCFET): Biosensor for DNA detection, paper presented to *IEEE Int Conf Devices for Integrated Circuit,* Kalyani, India, 348-352 April 2025.

34 Das A, Bhardwaj A, Das K, Yadav S, Kaul A, Goyal P, Sharma S, Rewari S, Kanaujia B K & Gupta R S, *Indian J Pure Appl Phys*, 63 (4) (2025) 281.

35 Das A, Rewari S, Kanaujia B K, Deswal S S & Gupta R S, *Micro Nano Struct*, 204 (2025) 208152.

36 Das A, Singh P, Sharma S, Goyal P, Joshi O, Nakra P & Gupta R S, *ECS J Solid State Sci Technol*, 14 (2025) 053006.

37 Bhardwaj A, Das A, Kumar P, Raj B & Yadav R, *Ind J Pure Appl Phys*, 63 (2025) 617.

38 Bartler A, Mauch L, Yang B, Reuter M & Stoicescu L, Auto mated detection of solar cell defects with deep learning, paper presented to 26thEur Signal Process Conf, Rome, Italy, 2035-2039 December 2018.

39 Zhu J, Zhou D, Lu R, Liu X & Wan D, *Nondestruct Test Eval*, 40 (2024) 309.

40 Abdelsattar M, AbdelMoety A & Emad-Eldeen A, *Sci Rep* 15 (2025) 24356.

41 Masita K, Hasan A, Shongwe T & Abu H, *Solar Energy Advances*, 5 (2025) 2667.

42 Dunderdale C, Brettenny W, Clohessy C & Van Dyk E E, *Prog Photo volt Res Appl*, 28 (2020) 177.

43 Sungeetha D & Gomathi S, Maximizing the Performance of Solar Panels through the Use of Artificial Intelligence for Dust Detection in Renewable Systems, paper presented to Int Conf Electron Renewable Syst, Tuticorin, India 96-99 February 2025.

44 Wang J, Du H & Zeng Y, *IEEE Trans Consum Electron*, 71 (2025) 6602.

45 Rudro R A M, Nur K, Sohan M F A A, Mridha M F, Alfarhood S, Safran M & Kanagarathinam K, *Energy Rep*, 12 (2024) 1580.

46 Cao Y, Pang D, Zhao Q, Yan Y, Jiang Y, Tian C, Wang F & Li J, *Eng Appl ArtifIntell*, 131 (2024) 107866.

47 Bhardwaj A, Das A, Rai A, Krishna R, Upadhyay A K, *Semiconductors*, 59 (5) (2025) 474–482.

48 Das A, Bhardwaj A, Gupta A, Raj G, *Semiconductors*, 59 (5) (2025) 427.

49 Bhardwaj A, Das A, Sharma U, Gupta A, Roy S, *Semiconductors*, 59 (5) (2025) 382.

50 Bhardwaj A, Das A, Roy S, Khushboo, Aman A, Raj G, *Semiconductors*, 59 (6) (2025) 521.

51 Bhardwaj A, Das A, Yadav R, Kumar P, *Micro electron Reliab*, 172 (2025) 115840.

52 Henrique R, Alves F, Antero G, Júnior D D & Marra E G, *Renew Energy*, 179 (2021) 502.

53 Cipriani G, Amico A D, Guarino S, Manno D, Traverso M & Dio V D, *Energies*, 13 (2020) 17.

54 Zyout I & Oatawneh A, Detection of PV solar panel surface defects using transfer learning of the deep convolutional neural networks, paper presented to Adv Sci Eng Technol Int Conf, Dubai, United Arab Emirates, 1-4 February 2020.

55 Husain AA, Maity T & Yadav R K, *IET Image Process*, 14 (2020) 10.

56 Woo S, Park J, Lee J Y &Kweon I S, *Computer Vision and Pattern Recognition (ECCV),* (2018) 3. https://doi.org/ 10.48550/ arXiv.1807.06521

57 Saxena P & Bhandari A K, *IEEE Trans ArtifIntell*, 5 (2024) 3510.

58 Buerhop-Lutz C, Deitsch S, Maier A, Gallwitz F, Berger S, Doll B, Hauch J, Camus C & Brabec C, paper presented to 35th Eur Photovolt Sol Energy Conf (EUPVSEC), Brussels, 1287-1289 2018.

59 Deitsch S, Christlein V, Berger S, Buerhop-Lutz C, Maier A, Gallwitz F &Riess C, *Sol Energy*, 185 (2019) 455.

60 Fan T, Sun T, Xie X, Liu H & Na Z, *IEEE Access*, 10 (2022) 16269.

61 Eesaar H, Joe S, Rehman M, Jang Y & Chong K, *Energies*, 16 (2023) 7726.

62 Rahman M R, Tabassum S, Haque E, Nishat M M, Faisal F & Hossain E, paper presented in 3rdIntConf Sustainable Technol Ind 4.0 (STI), Dhaka, Bangladesh, 1-6 December 2021.