

Deep learning-based fully automatic system for segmentation and defect classification of the solar modules using electroluminescence images

Mustafa Yusuf Demirci^{a,*}, Nurettin Beşli^b, Abdulkadir Gümüşçü^b

^a T.C. Ministry of National Education, Mehmet Tuza Pakpen MTAL, Büyükkayacık OSB Mah, Yayıncık Cd. NO:12, 42280, Konya, Turkey
^b Harran University, Department of Electrical and Electronics Engineering, Osmancık Campus, Şanlıurfa - Mardin Highway 18. Km, Şanlıurfa, Turkey

ARTICLE INFO

Keywords:
 Defect detection
 Deep learning
 Solar cells
 Image segmentation
 Vision transformer

ABSTRACT

In order to ensure high efficiency operation of solar power plants, monitoring and maintenance of their components are of great importance. As PhotoVoltaic (PV) modules are the key components of the solar powered electricity production systems, determination of the module cell defects would prevent future malfunctions and power losses. Using ElectroLuminescence (EL) imaging, even the finest defects like micro-cracks could be detected precisely. Manual evaluation of the EL images is time-consuming, labor-intensive and prone to human errors while emerging automatic systems often focus on a single aspect of the problem. To address these issues, a complete fast and robust evaluation framework was proposed. In this work, we propose SolarELTester, a novel end-to-end fully automatic segmentation and classification system for segmentation and identification of the solar cell defects using the integration of state-of-the-art deep learning algorithms on module EL images. Proposed system composed of four stages: image preprocessing, segmentation using Region-based Convolutional Neural Networks (R-CNN), defect classification using modified lightweight CNN and Vision Transformer (ViT) models and finally defect ratio calculation. Proposed system was able to detect solar cells even from distorted module images at 99.73 % mean Average Precision (mAP) and classify five types of common solar cell defect types with 98.71 % accuracy for the validation dataset while maintaining real-time performance. Therefore, the proposed methodology represents a significant stride towards automation of the solar module analysis and offers a flexible framework for monitoring and quality assessment in solar-powered systems.

1. Introduction

In the global world, there is a huge demand for energy, which is increasing day by day. This huge demand, especially for fossil fuels, has led to a global energy crisis. Also, COVID-19 global pandemic and regional wars have contributed to energy crisis as well [1]. The main drawbacks of fossil fuels are that their resources are limited and expensive, and they also have a major impact on environmental pollution. According to the latest global energy outlook report by the International Energy Agency (IEA), the energy industry is the main cause of air pollution which affects more than 90 % of people living on the planet [2]. Renewable energy is of great importance in preventing environmental damage and overcoming the energy crisis. Global forecast reports state that renewable capacity of the whole world is expected to increase by 2,400 GW between 2022 and 2027, equivalent to the entire installed power capacity of China today. Solar PhotoVoltaic (PV) provides 4.5 % of the entire global electricity generation, making it the

third-largest renewable electricity technology after hydropower and wind power [3].

This huge capacity led to need for effective maintenance and monitoring of the PV components. PV module health monitoring and performance assessment are crucial for the operation of solar power units; early diagnosis of flaws in manufacturing process is also important to prevent potential power losses [4]. Solar modules may have various defects caused by manufacturing issues, improper transportation and installation or even extreme weather conditions in operation. Manufacturing defects may result from poor soldering or lamination, and these production issues could lead to material defects and solder faults. Defects like finger interruptions, breakages and micro cracks may happen in all these stages. Also, some module defects like hot spot and potential induced degradation could happen over longtime [5,6]. These defects not only cause significant power losses but can also cause issues like short-circuits, arcs and even fire. Effective and reliable testing mechanisms are crucial for reducing such potential risks [7].

* Corresponding author.

E-mail addresses: mfecidpc@gmail.com (M.Y. Demirci), nbesli@harran.edu.tr (N. Beşli), agumuscu@harran.edu.tr (A. Gümüşçü).



Measurements such as I-V Curve and thermal imaging are generally used for performance evaluation at the module level, while electroluminescence (EL) imaging technique is used to evaluate cell-level defects of PV modules, that are more difficult to detect using other methods. EL imaging is based on the principle that feeding direct current to the solar module and capturing the resulted radiation using IR-filtered digital camera. Using this method, breakages, finger interrupts, shunted cells, soldering defects and even finest defects like micro-cracks could be effectively detected [8]. Sample EL image of a PV module with defective cells is shown in Fig. 1. In the image, red marks indicate solder defect, yellow mark indicates breakage, blue marks indicate micro crack and green marks indicate finger interruptions. Inspection and assessment of the EL images are mostly carried out by trained specialists. Therefore, this process is labor-intensive and time-consuming. Also, manual assessment is expensive and not consistent due to human errors and tedious nature of work. All these reasons necessitated the development of automated systems that will quickly detect cell defects and evaluate module performance accurately. Due to the advances in artificial intelligence in recent years, machine learning-based applications in industrial inspection have become widespread. Especially, Convolutional Neural Network (CNN)-based deep learning algorithms achieved state-of-the-art performance in image classification. AlexNet [9], VGG [10], GoogleNet [11], ResNet [12] and EfficientNet [13] architectures pushed the boundaries for image recognition and classification in the ImageNet challenge [14]. In 2021, researchers adapted the transformer mechanism for computer vision tasks called Vision Transformer (ViT) [15], inspired from original transformer model for machine translation [16]. ViT models have achieved significant progress in image detection, pushing the boundaries even further from CNNs. Therefore, this article provides a robust end-to-end solution for segmentation and detection of defective PV module EL images using machine vision and deep learning techniques. Proposed system is a complete framework which implements all stages on EL images from image preprocessing to report generation and utilizes advanced processing algorithms such as CNN and ViT. The rest of this article is organized as follows: In section 2, a literature review of the works, focusing deep-learning based detection of PV-EL images is given in detail. Section 3 summarizes the technical contributions of the proposed article. Section 4 presents details of the proposed defect segmentation and classification methodology. In section 5, experimental results and calculated performance metrics are given in detail. Finally, in section 6, the proposed work is summarized

and future work is discussed.

2. Related work

Due to the superior performance on image classification and detection tasks, many researchers utilized deep learning-based methods for automatic inspection of solar EL images. Bartler et al. presented automatic detection of solar cell defects using adapted VGG-16 CNN architecture. They used a custom dataset containing 98,280 solar cell EL images and obtained 7.73 % balanced error rate for binary classification of labels named good and defected [17]. In Ref. [18], Deitsch et al. presented automatic classification of solar cells using Support Vector Machine (SVM) modified VGG-19. They used a benchmark EL dataset named ELPV provided by Buerhop-Lutz et al. [19] and obtained 88.42 % accuracy for binary classification. In Ref. [20], Akram et al. presented light CNN model for classification of solar cell EL images. They trained CNN model from scratch and obtained 93.02 % accuracy for the binary classification of ELPV dataset. In Ref. [6], a feature extraction and classification framework for solar cells is presented. First, image features are extracted via pre-trained CNNs, and then feature selection and classification are performed by using SVM on ELPV dataset. 94.52 % binary classification accuracy is obtained. In another study, K.-M. Lin et al. proposed segmentation and lightweighted CNN-based classifier models for monocrystal solar cells. They trained models on the custom EL image dataset consisting of 68,700 annotated solar cell images and obtained 99.40 % accuracy for the binary classification of the defective PV cells [21].

With the recent advances in computer vision and machine learning, deep learning models are used not only for classification but also for object detection and localization. State-of-the-art detector algorithms such as Faster R-CNN [22] and YOLO [23] have gained attention for high detection performance and real-time detection speed. In Ref. [24], Otamendi et al. presented the segmentation and detection model of cell-level defects on PV-EL images. They utilized Faster R-CNN for segmentation of PV cells, performed anomaly classification using EfficientNet-B1 model, segmented the defects through generative encoders and obtained 84 % detection accuracy. In another study, Meng et al. proposed YOLO-based detection and segmentation of EL images. They trained models on a custom EL dataset and obtained 94.55 % average precision on test dataset for anomaly detection [25]. In another study, H.-H. Lin et al. presented segmentation of mono crystalline PV cell

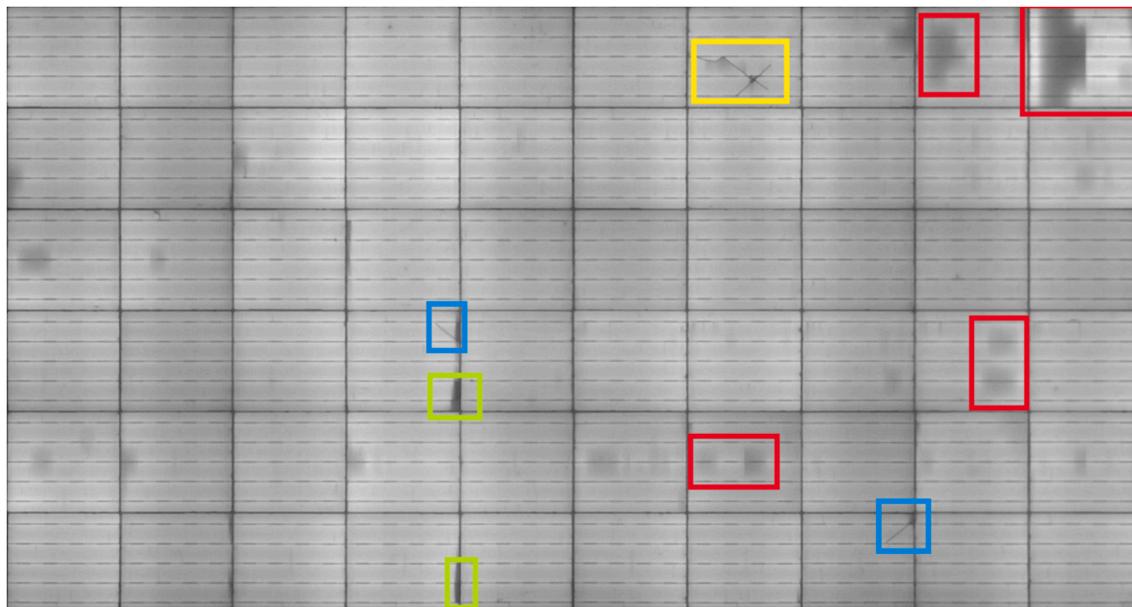


Fig. 1. EL image of a 60 cell PV module with various defective cells.

EL images by using YOLO-v4 model and pseudo-colorization with K-means clustering. They achieved 99.8 % accuracy on the custom EL dataset consisting of 7,140 solar cell images [26]. In another study, Korovin et al. proposed anomaly detection model SeMaCNN on EL images for heterojunction solar cells through the combination of distance-based feature maps and pre-trained CNN features. They trained models on the two different EL image datasets and achieved 92.5 % and 94.6 % accuracy for the custom test dataset and ELPV dataset, respectively [27].

Besides binary classification, researchers focused on the detection of multiple solar cell defects. In Ref. [28], authors presented a defect detection of solar cell images using custom CNN and Generative Adversarial Networks (GAN)-based data augmentation. They used two different datasets and detected four types of solar cells: defect-free, micro crack, finger interruption and break with 83 % accuracy. In another study, Zhao et al. presented multiple defect detection and segmentation on the custom EL image datasets through a mask R-CNN with ResNet-101 backbone. They obtained 70.2 % mAP for the 14 different defect types and top 97.3 % F1-score for the binary classification [29]. In Ref. [30], Wang et al. proposed EL image defect segmentation using feature fusion ResNet152-Xception model on two public datasets. They achieved 96.17 % accuracy for the binary classification and 92.13 % accuracy for the 9 types of defects. Semantic segmentation is another technique that was initially used to segment medical image via U-Net architecture [31]. Several methods also propose semantic segmentation of the solar EL images for defect detection [32,33]. Few recent studies have focused on utilizing transformers for EL image classification. In Ref. [34], researchers proposed transformer-based network for polycrystalline solar cell defect detection using DPiT, a method based on Swin Transformer model. They tested the models on the ELPV dataset and obtained the best accuracy of 91.7 % using 4-stage transformer model. In another work, researchers utilized ViT model for defect classification of monocrystalline PV cells. They used a custom dataset containing 5362 cell images and trained ViT- μ model from scratch. As a result, they obtained 97.8 % accuracy for the binary defect detection and 94.3 % accuracy for the 3-class (defect-free, crack, solder defect) [35].

3. Contributions

Based on the literature review, most of the related studies performed binary classification on the solar cell images i.e., functional or defective. However, in PV production units, the detection of various cell defects is essential for ensuring quality control, as each defective cell significantly contributes to the prediction of production costs and statistical analyses. In the literature, multi-class defect detection research is limited and the detection performance drops significantly in multi-class defect classification. In addition, most of these mentioned works have used public benchmark dataset which has serious class imbalance. However, some have achieved very high performance by using very large datasets, which are time-consuming and labor-intensive to generate. Furthermore, there are very few studies analyzing the use of vision transformers on PV cell defect detection problem. To address these above-mentioned issues, we proposed a novel segmentation and classification system for identifying cell-level defects of the solar module images. The proposed work is a fully automated monitoring and evaluation solution for the solar modules using cutting-edge deep learning techniques and provides robust, fast evaluation framework that can also be adaptable to another machine vision tasks. In this manner, technical contributions of the proposed work are summarized as follows:

1. We proposed a novel end-to-end solar module segmentation and multiple cell defect classification system named SolarELTester. Proposed system combines image preprocessing, cell segmentation, defect classification and report generation stages in a single framework.

2. Efficient cell segmentation of EL module images is realized by using modified Faster R-CNN. With the proposed method, even the distorted images of defected module can be segmented with high precision. Proposed model is capable of extracting PV cells with 99.73 % mAP on the test dataset.
3. For the defect classification stage, we proposed pre-trained custom CNN and Vision Transformer models. Both models achieved superior classification performance while keeping the detection time to a minimum. To our best knowledge, this is the first framework that integrates CNN and ViT-based end-to-end multi-class defect detection models for solar PV systems. While ViT model attained 98.70 % accuracy, the modified CNN reached 98.23 % accuracy on the test dataset with only 7 min of training time.
4. We introduced a novel multi-class EL image dataset consisting of 3,100 expert annotated EL images of monocrystalline solar cell. The proposed dataset is balanced and has diverse types of defects, which makes it suitable for academic use.

4. Methodology

The SolarELTester system essentially consists of four individual stages. In the first stage, a raw solar module EL image is fed to the image preprocessing unit for segmentation. In the second stage, module image is segmented into cells. In the third stage, image classification is performed by using CNN and ViT, and finally, in the fourth stage, classification statistics are calculated. General representation of the proposed framework is given in Fig. 2.

4.1. Datasets

In this study, a novel solar cell EL image dataset is proposed. Our dataset is composed of 3,100 EL images of monocrystalline PV cells with various defect types. In the formation of PV cell dataset, it was necessary to accurately extract each cell from module images. To address this issue, an independent PV module image dataset was first created. For the creation of this dataset, an extensive number of solar module EL images were collected. These module images were labeled by using bounding boxes and fed to the segmentation module which is detailed in Section 4.3. Afterwards, the collected module EL images were segmented by the developed detection system in order to divide module to individual cell images, thereby constructing the cell dataset. After detection and extraction, each of the obtained cell images was selected and labeled for five classes: functional (ok), solder defect (solder), finger defect (finger), material defect (material) and crack/breakage (crack). Each defect class contains 620 cell images. Sample images for the cell dataset are given in Fig. 3. Module EL images used for the segmentation and classification datasets were provided by the quality control unit of a national PV manufacturer. For the training of segmentation model, a total of 120 monocrystalline and polycrystalline module EL images were selected and labeled. PV module dataset was divided into 100 training and 20 test samples, whereas cell dataset was split into 80 % for training and 20 % for testing.

4.2. Image preprocessing

In the first stage, raw module EL image is taken and prepared for the image segmentation. Some of the module images were distorted and had shifted columns. This issue could be seen in the sample image given in Fig. 4. Owing to the abundance of such images in the collected data, the preprocessing stage was carried out through the development of a specialized image processing system, employing advanced techniques such as corner detection. For image processing, MATLAB image processing toolbox is used. Since EL images are grayscale whereas file system was RGB, original image was loaded and transformed from RGB to grayscale first. Then 7 x 7 median filter was applied to the image for emphasizing image edges. After median filtering, image binarization

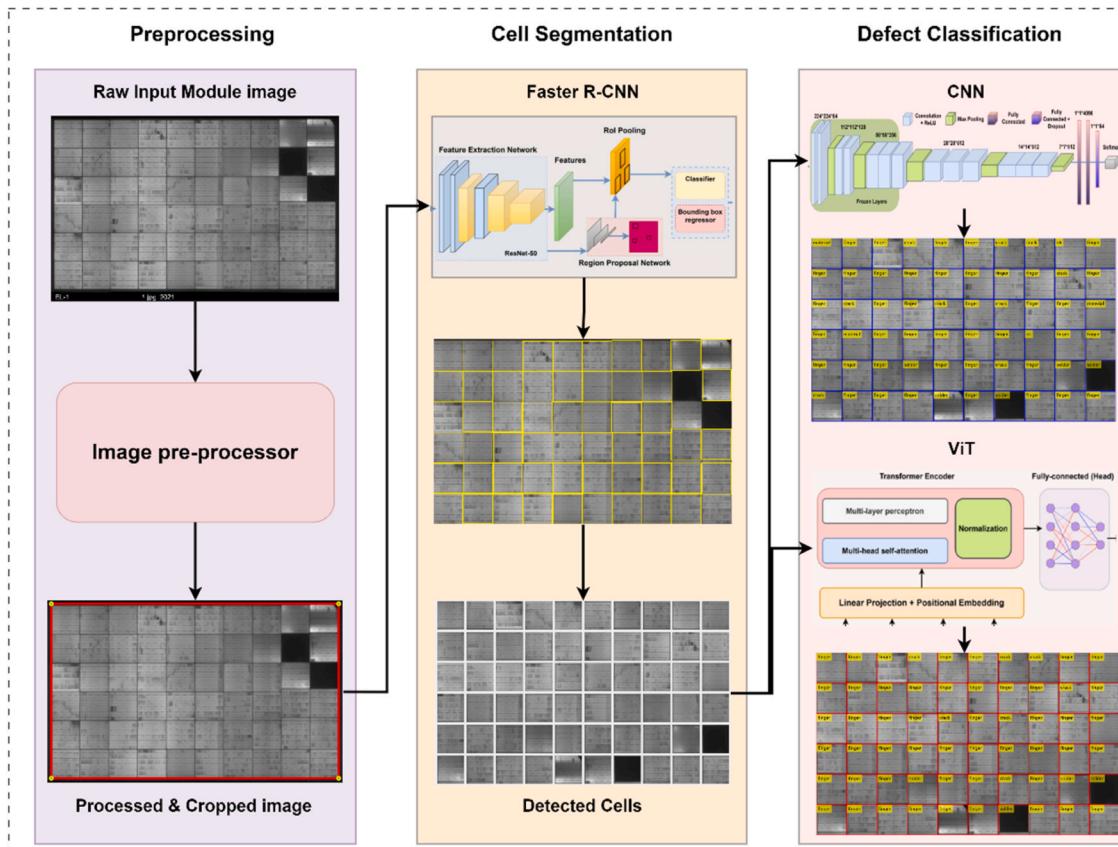


Fig. 2. Schematic of the proposed framework.

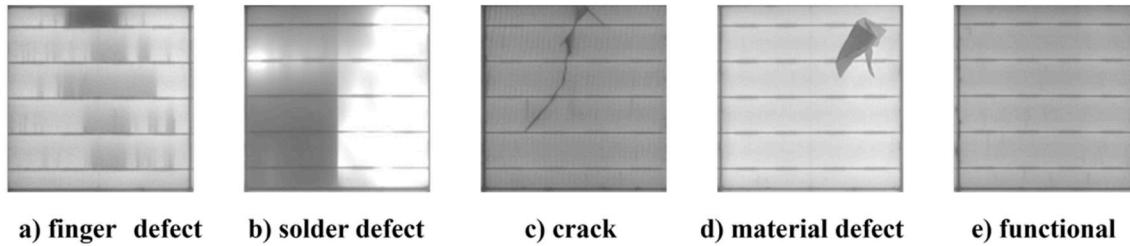


Fig. 3. Sample images for the proposed PV cell EL image dataset.

was performed by using adaptive local thresholding for background subtraction. In adaptive local thresholding, a threshold value is calculated for each pixel individually using the mean intensity of its neighborhood pixels [36]. In the next stage, binary image mask was created by using convex hull algorithm. Then, corners were detected by using Harris method, and the 50 strongest corner points were determined for x and y axes. Next, minimum and maximum values of these points were found and these four points (x_0, x_1, y_0, y_1) were determined as module corner points. Lines are fitted along x and y axes, and finally, image is cropped using these points. Fig. 5 presents the visualization of the proposed image processing algorithm applied to a module EL image. The original unprocessed image is shown in Fig. 5(a), while Fig. 5(b) depicts the binarized module image. The Convex Hull algorithm was subsequently applied to the binarized image, as illustrated in Fig. 5(c). In Fig. 5(d), the edge lines detected by the algorithm are highlighted in red, and the corners identified are marked with yellow dots.

4.3. Image segmentation

After preprocessing, segmentation of the module image into single

cells was performed. For the segmentation stage, conventional image processing techniques such as contour tracing and line detection were initially considered, but since the collected field data contained degraded and shifted images as shown in Fig. 4, it was decided to extract cells using deep learning methods to obtain more cell data. Deep learning-based object detectors are used to locate objects in images also videos. Two main types of detectors are used in object detection: single-stage and two-stage detectors. Single-stage detectors rely on the bounding box predictions over the entire image using CNN backbones. YOLO [23] and RetinaNet [37] architectures are popular state-of-the-art single-stage detectors. Two-stage detectors, on the other hand, solve the detection problem in two steps: region proposal and classification. Region proposals are fragments of images that may contain objects. After generating region proposals, classification is performed on the extracted features using CNN and bounding box regression. Region based CNN (R-CNN) [38] and its successors are widely used two-stage object detector algorithms for their high detection precision. In general, single-stage detectors are faster than two-stage detectors while two-stage detectors are more accurate [39]. In the proposed detector, the latest variant of R-CNN, Faster R-CNN detector, is adapted. In the

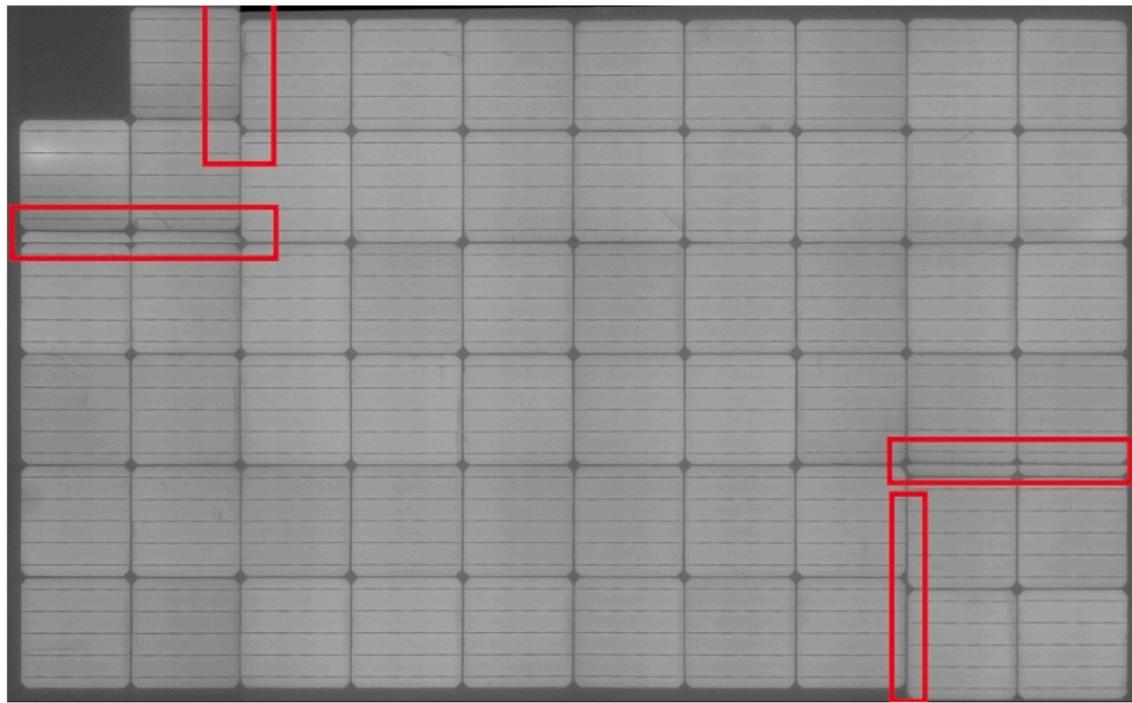


Fig. 4. Distorted PV module EL image sample.

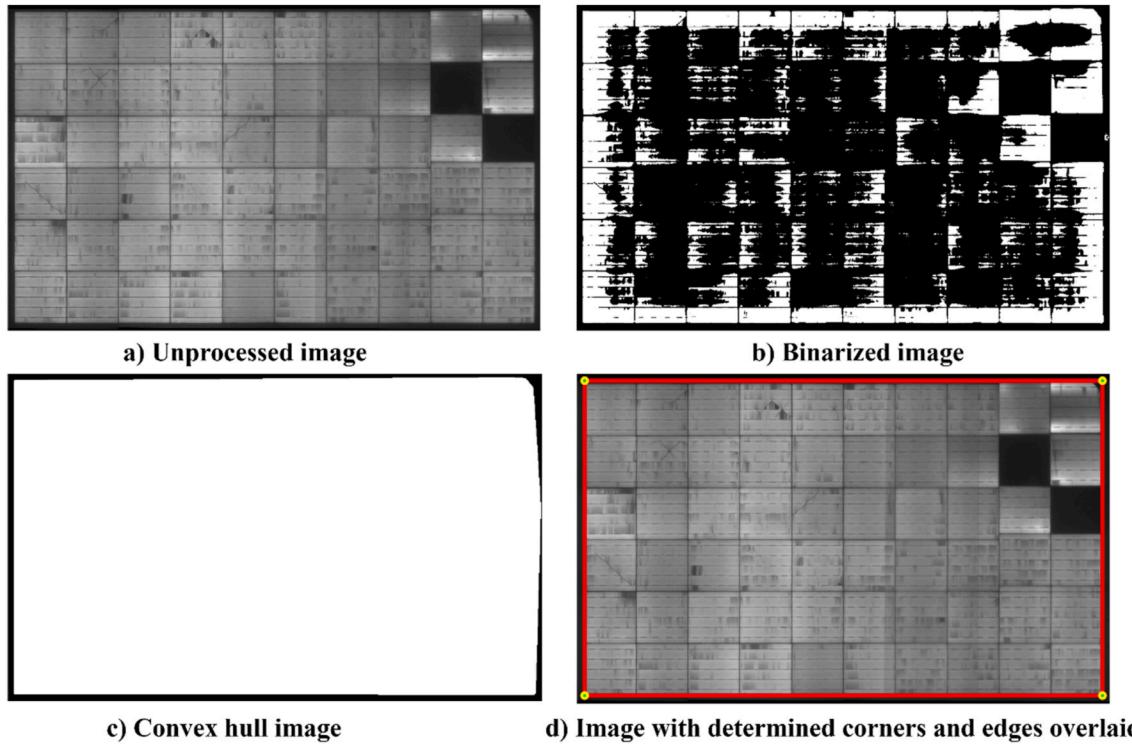


Fig. 5. Visualization of the proposed image preprocessor module.

Faster R-CNN, region proposals are computed by using internal Region Proposal Network (RPN). RPN generates proposals and bounding box coordinates simultaneously. Therefore, detection speed is improved while accuracy remains high [22]. In the proposed detector, we used ResNet-50 as feature extraction network. Due to its skip connections and deeper structure, ResNet architecture reduces detection time and performs well in feature extraction and classification tasks [12]. Input

images are resized to the 1120 x 2100 pixels and the last classification layer is changed to one-class for cell detector. Schematic of adapted Faster R- CNN network is given in Fig. 6. Training parameters are presented in the next section.

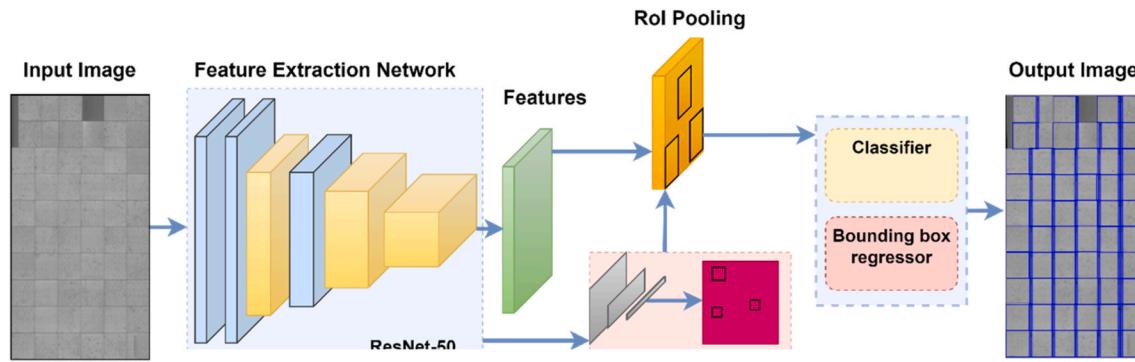


Fig. 6. Faster R-CNN model for module image segmentation.

4.4. Defect classification

In the SolarELTester, multi-class defect classification is performed after segmentation stage. In the proposed methodology, two different defect classification models are utilized: CNN and Vision Transformer. Both CNN and ViT models were pre-trained with ImageNet dataset. Details for the proposed classification pipelines are given in the following sections.

4.4.1. Classification using CNN

Convolutional Neural Networks are widely used for image classification problems. There are two main strategies for training CNNs: training from scratch and transfer learning (fine tuning). In the first method, network is trained with random initial weights and image features are learned from scratch during training session. This is usually time consuming and needs large amount of data due to the convolutional structure of CNNs [6]. For addressing these issues, transfer learning is commonly used approach for CNN training. In the transfer learning method, network is initially trained with large datasets. After that, network is trained by using target dataset, which is called fine-tuning. Transfer learning method is effective when dealing with limited data, and is more robust because of the knowledge transfer, and also a cost-effective solution as it reduces training time dramatically [4,40]. In the proposed method, VGG-16 pre-trained network was adapted and customized for the created dataset. First, the last fully connected layer was changed for five classes. Second, a fully connected layer with 64-dimension was added for increasing non-linearity and reducing overfitting. At the end of the network, additional 10 % dropout layer is added. Finally, layer freezing is applied to the initial layers of the network in order to reduce training time. First 14 layers of the network are kept frozen. Thus, training time was reduced significantly. Proposed CNN schematic was given in Fig. 7. Network hyperparameters are presented in the experimental results section.

4.4.2. Classification using vision transformer

Transformer structure become popular soon after the presentation of

state-of-the-art studies [16]. Due to its superior performance, it has become a new standard for machine translation and language modelling [41]. Although transformers were designed for machine translation and language modelling originally, researchers tried to adapt it to other domains such as image processing. Dosovitskiy et al. proposed Vision Transformer (ViT), a new learning model that relies on transformers entirely to process images. ViT model has shown that image classification tasks could be performed with the highest accuracy even without convolutional structures [15]. ViT is basically a modified transformer encoder structure which is fed with identical patches of image instead of word tokens. Image patches are flattened into a sequence of vectors using linear projection. Patch embedding and positional embedding are given to the transformer encoder network. Basic schematic of ViT model is given in Fig. 8.

ViT models are performed best when trained on large datasets and fine-tuned on target datasets. Therefore, transfer learning approach is chosen for the optimal performance and resource trade-off. In the proposed work, ViT-based model is adapted and trained with ImageNet-2012 dataset [14]. Base model has $384 \times 384 \times 3$ input size and 86.8 million parameters total. For transfer learning, only self-attention layer is trained and other layers are kept frozen as mentioned in Ref. [42]. Next, head network is replaced with five-dimension fully connected layer. Network hyperparameters are presented in the experimental results section.

4.5. Statistical analysis

It is crucial to keep statistics of defective PV modules, especially in production units. Since the quality control and production planning will be directly affected by the records kept, the fault statistics of the module cells are calculated at this stage of the SolarELTester system. After classification stage, defect rates of each module are calculated by using simple weighted percentage. For instance, defect rate of cracked cells were calculated by dividing cracked cell number with total detected cell number. All defect rates were calculated with the same weight value and total module defect rate is found calculated by dividing number of all

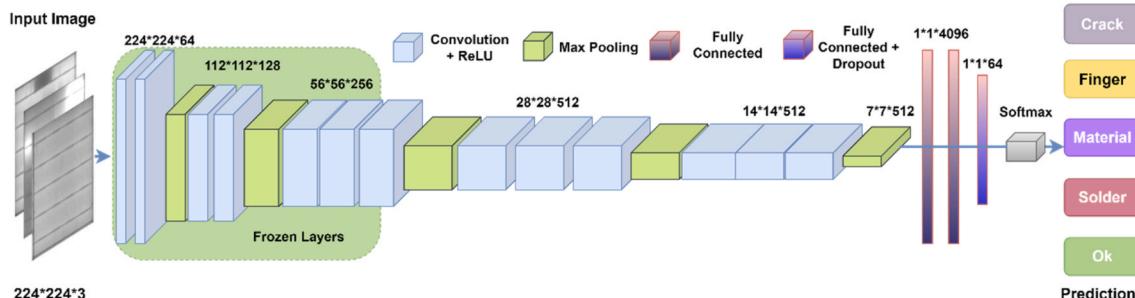


Fig. 7. Proposed custom VGG-16 network for defect classification system.

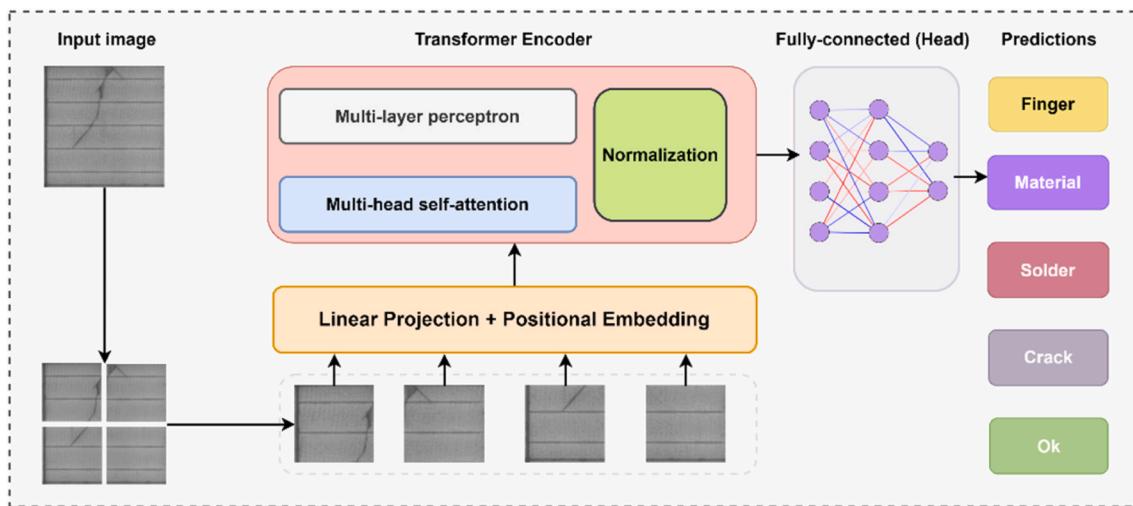


Fig. 8. Schematic diagram of the Vision Transformer model.

EL test summary:		Defect Rate (%)
Num.of detected cells	:	60
Num.of functional cells (ok)	:	2
Num.of defected cells (finger interruption)	:	41
Num.of defected cells (cracked)	:	9
Num.of defected cells (solder defected)	:	5
Num.of defected cells (material defect)	:	3
 Total module defect rate (%)	:	 96.67

Fig. 9. SolarELTester statistical report example.

defected cells to the detected cells. A sample report for a 60-cell module is given in Fig. 9.

5. Experimental results

In this section, training parameters of the proposed SolarELTester models, results of detection and classification stages are given in detail along with performance metrics.

5.1. Experimental settings

All coding and experiments were carried out using MATLAB R2023b along with Deep Learning Toolbox, Statistics and Machine Learning Toolbox. Training and testing stages for detector and classifier networks were carried out on a workstation with Intel Core i7-11700K CPU @ 3.70 GHz, NVIDIA RTX 3080Ti GPU with 12 GB memory and main memory of 32 GB RAM.

5.2. Evaluation metrics

For the classification networks, confusion matrices were plotted for each model and essential evaluation metrics of accuracy, precision, recall and F-score were used to measure performance. To calculate these metrics, True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN) values are counted for each class. Confusion matrix structure for binary-class is given in Fig. 10.

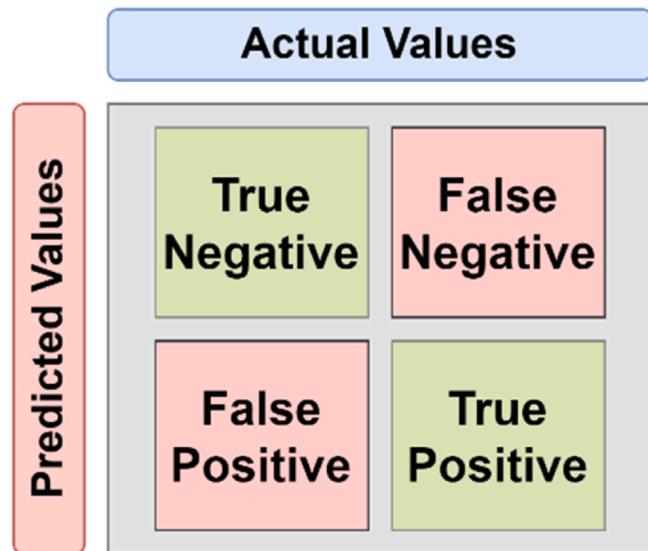


Fig. 10. Binary confusion matrix.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$F\text{-Score} = 2 * \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4)$$

On the other hand, the performance assessment of segmentation stage is different from that of classification stage. Proposed PV module segmentation is basically an object detection task. For assessing the performance of the detector networks, mostly calculated metric is average precision (AP) [43]. Calculation of AP value requires Intersection over Union (IoU) as well as precision and recall rates. IoU is a type of similarity measurement which is used to calculate accuracy of the detected bounding boxes according to the ground truth boxes in object detection. Calculation of the IoU is given in Equation (5):

$$\text{IoU} = \frac{X_{pr} \cap X_{gt}}{X_{pr} \cup X_{gt}} \quad (5)$$

Where X_{pr} represents the predicted bounding box area while X_{gt} represents ground truth bounding box area. The calculated IoU is compared with threshold value to determine whether object is detected or not. It is worth noting that true negative (TN) is not used in the object detection because there may have been infinite possibilities of “no object” bounding boxes in single image. Therefore, precision and recall are calculated in object detection as follows:

$$\text{Detection}_{\text{Precision}} = \frac{TP}{\sum \text{Pred}} \quad (6)$$

$$\text{Detection}_{\text{Recall}} = \frac{TP}{\sum Gt} \quad (7)$$

$\text{Detection}_{\text{Precision}}$ is true detection rate in the predictions while $\text{Detection}_{\text{Recall}}$ is the detection rate of the ground truth data. Therefore, Average precision is calculated using $\text{Precision} \times \text{Recall}$ area under curve. Ideally, there should be balance between these two metrics across all points. A common method to calculate AP value is interpolation of 11-points: maximum precision values are interpolated to 11 equal recall intervals and averaged. Let j is a set of equally spaced values (e.g. $[0.1, 0.2 \dots 0.9, 1]$). Calculation for 11-point method is formulated as follows:

$$AP = \frac{1}{11} \sum_{\text{Recall}_j} \text{Precision} (\text{Recall}_j) \quad (8)$$

5.3. Solar cell detection results

For image segmentation stage, a modified Faster R-CNN model was utilized. Proposed network was trained on 100 training module EL images which were selected randomly from segmentation dataset of 120 EL images. Remaining 20 images were reserved for test. Data augmentation

was applied to the training dataset in the form of random flipping on the X-Y axes. The “activation_40_relu” layer of the ResNet-50 network was selected as feature extraction layer. Training hyperparameters of the proposed detection model is given in Table 1.

Training time was measured as 76 min for the proposed Faster R-CNN network with the above-mentioned hyperparameters. After training stage, test images were detected and segmented by the trained detector network. Average precision of 0.9973 with IoU threshold of 0.75 for total 100 training images were measured. Precision-recall curve is given in Fig. 11. Detection results for a sample PV module EL image are given in Fig. 12. As can be seen from the figure, image is heavily distorted, but proposed Faster R-CNN model was able to detect module images accurately.

5.4. Defect detection – classification results

Detected images were cropped by using predicted bounding box coordinates after segmentation stage. Training and test datasets were created by using proposed method, and classifier networks (CNN and ViT) were trained with training PV cell dataset. To prevent overfitting and increase diversity, data augmentation was applied to the training dataset for both CNN and ViT models with same parameters. Online data augmentation was applied to the training dataset with the following settings: random reflection over X-Y axis, 180-degree rotation, random scaling in $\pm 5\%$ range, random X-Y axis shearing in $\pm 5\%$ range. Training hyperparameters for the proposed classification networks are given in Table 2.

Classification metrics are given in Table 3. As can be seen from the table, both proposed methods reached high performance values. While CNN model reached 98.23 % accuracy, ViT model reached 98.71 % accuracy. In terms of performance, ViT-based model outperformed CNN-based model in all metrics. Both models achieved over 98 % accuracy and over 98 % recall rate which indicates the robustness of the proposed methods. Confusion matrices for the proposed methods were given in Fig. 13.

When confusion matrices were analyzed, it could be inferred that ViT model performed better for detection of finger interruption, solder defects and material defects better, while CNN model performed better for detection of functional cells and cracked cells. Despite the small gap in overall classification performance, CNN gained the advantage when training and evaluation times were measured. Training and testing time for both models were given in Table 4. When both models were trained on the same dataset of 2480 cell images, training time was measured as 187 min for the ViT model and 7 min 15 s for the CNN model. In terms of

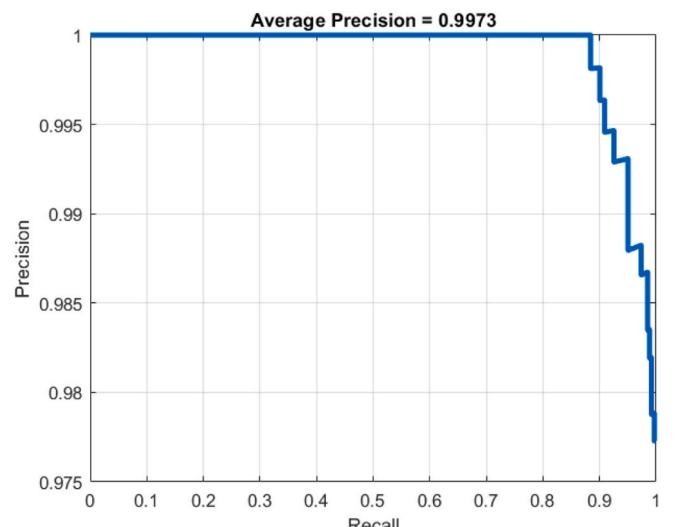


Fig. 11. Precision - Recall curve for the cell detection network.

Table 1
Hyperparameters of the proposed Faster R-CNN detector.

Hyperparameter	Value
Mini-batch size (training)	1
Mini-batch size (testing)	4
Max. epochs	30
Initial learning rate	1×10^{-3}
Number of anchor boxes	2
Negative overlap range	[0–0.5]
Positive overlap range	[0.8–1]

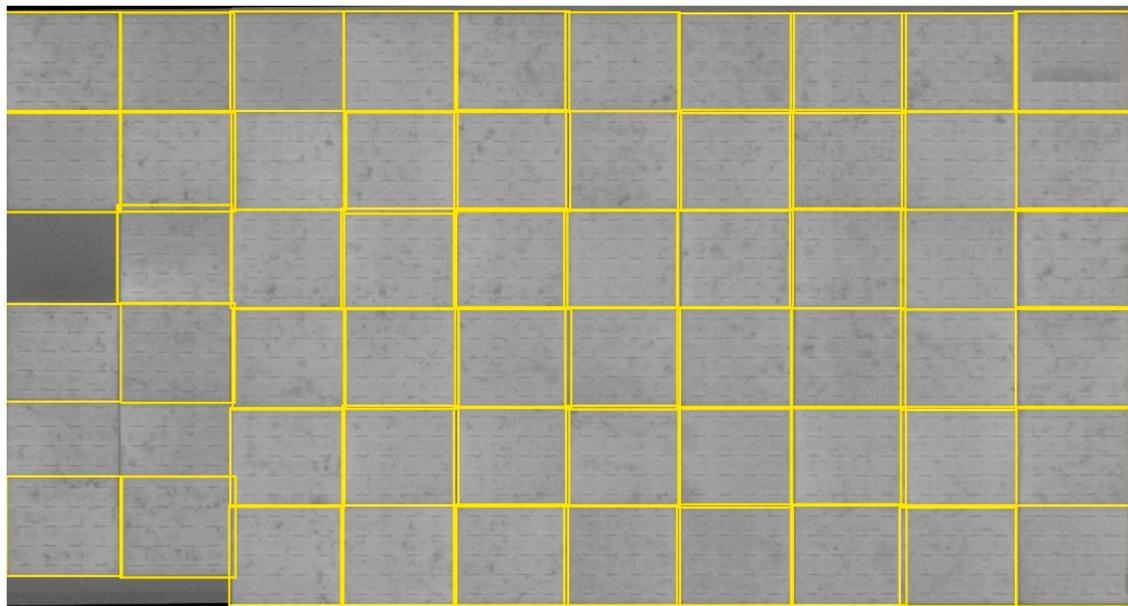


Fig. 12. Visualization of the cell detector results on a distorted EL image of PV module.

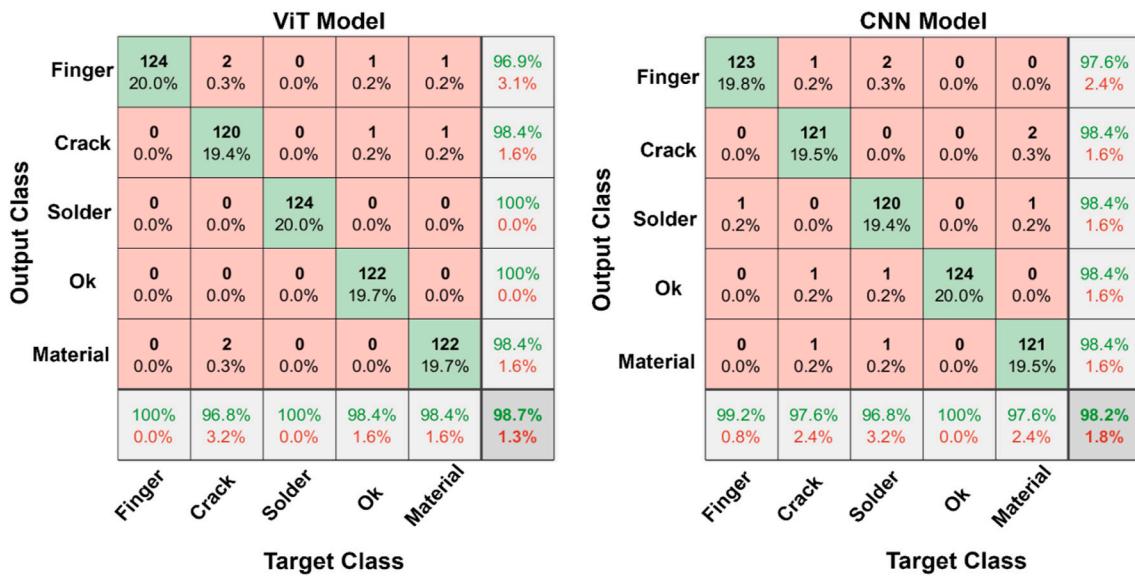


Fig. 13. Confusion matrices for the proposed classifiers.

Table 2
Training hyperparameters for the classification networks.

Hyperparameter	Value (CNN)	Value (ViT)
Max. epochs	40	8
Mini-batch size	64	12
Solver algorithm	SGD	ADAM
Initial learning rate	1×10^{-4}	1×10^{-4}
L2-Regularization	1×10^{-4}	none

Table 3
Classification results.

Method	Accuracy (%)	Precision (%)	Recall (%)	F-Score (%)
CNN (Custom VGG-16)	98.23	98.23	98.23	98.22
Vision Transformer (ViT-based)	98.71	98.72	98.71	98.71

runtime performance, CNN model clearly outperformed ViT model. CNN model processed a single cell image in 5.62 ms and ViT model processed a single cell image in 43.38 ms. For 60-cell standard PV module, the processing time of CNN model is approximately 337 ms while the processing time of ViT model is approximately 2.6 s. Therefore, CNN model performed faster and hold real-time performance.

Table 4
Runtime performances of the proposed classifier models.

Method	Training Time (minutes for whole Dataset)	Testing Time (seconds for whole dataset)	Testing Time (milliseconds for single image)	Testing Time (seconds for 60-cell PV module)
CNN (Custom VGG-16)	7 min	3.49	5.62	0.337
Vision Transformer (ViT-based)	187 min	26.90	43.38	2.6

Classification Results - CNN										
material	finger	finger	crack	finger	finger	crack	crack	ok	finger	
finger	finger	finger	finger	finger	finger	finger	finger	crack	finger	
finger	crack	finger	finger	crack	finger	crack	finger	finger	material	
finger	material	finger	finger	finger	finger	finger	ok	finger	finger	
finger	finger	finger	solder	finger	finger	crack	finger	solder	solder	
crack	finger	finger	finger	solder	finger	solder	finger	finger	finger	

Fig. 14. Visualization of the classification stage - CNN model.

5.5. Visualization of the classification results

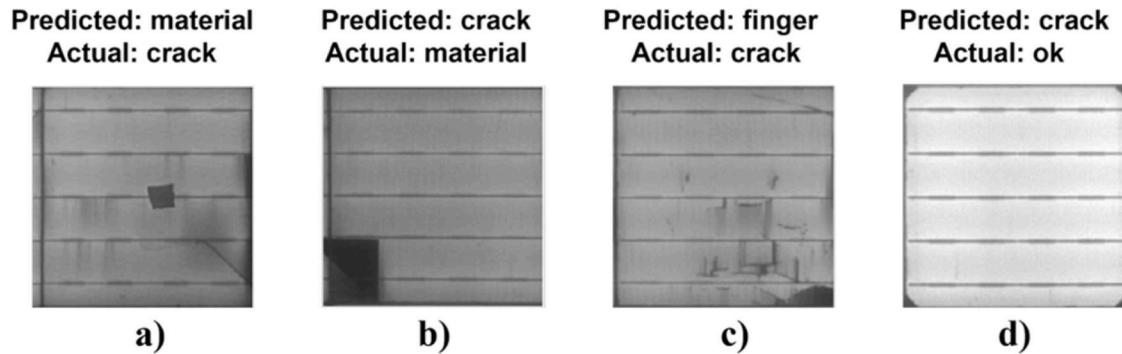
Visualization of the proposed CNN and ViT classification models is given in Fig. 14 and Fig. 15., respectively. As can be seen from the

figures, the models made different predictions for some cell images but similar predictions for the majority. The results of the models were consistent. To evaluate and analyze classification results in more detail, examples of misclassification by both models were given in Fig. 16.

Classification Results - Vision Transformer ViT-base model										
finger	finger	finger	crack	finger	finger	crack	crack	finger	finger	
finger	finger	finger	finger	finger	finger	finger	finger	crack	finger	
finger	finger	finger	finger	crack	finger	crack	finger	finger	finger	
finger	finger	finger	finger	finger	finger	finger	finger	finger	finger	
finger	finger	finger	solder	finger	finger	crack	finger	solder	solder	
finger	finger	finger	finger	finger	finger	solder	finger	finger	finger	

Fig. 15. Visualization of the classification stage - ViT model.

ViT model predictions



CNN model predictions

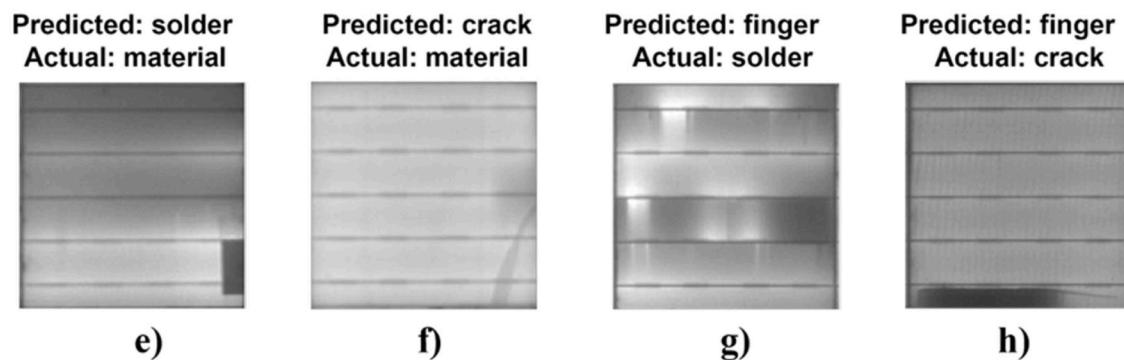


Fig. 16. Misclassification samples for the proposed CNN and ViT models.

As can be seen from Fig. 16, misclassified samples were mostly multi-defected cell images. Since the proposed method assigned cells into a single class, deep networks generally tried to classify cells according to the most dominant defect. This situation can be seen especially when looking at the predictions made by CNN in Fig. 16(e)–(g) and (h). When the predictions of ViT model are examined, the results appear to differ from CNN evaluations. ViT model performed better at detecting finer details. As seen in Fig. 16(d), the cell image is originally labeled as functional, however ViT model classified it as cracked because of the uneven corners. This is primarily due to the small patch-based input and attention mechanism of the transformers. Additionally, larger input size of the ViT model also contributes to this effect.

For better understanding of internal mechanisms of the proposed classifier networks, Class Activation Maps (CAM) were utilized for the sample cell images classified by both models. In the proposed method,

Gradient-Weighted CAM (Grad-CAM) is used. Grad-CAM provides visualization for deep neural networks using gradients of the predicted labels and generally last-convolutional layer of the network to generate heatmaps [44]. Fig. 17 shows the Grad-CAM overlaid heatmaps of the classified sample images of both models. When Fig. 17 is examined, it could be seen that the models had produced highly differentiated activation maps. It can be concluded that the ViT model focused more on finer details such as dendritic patterns and small cracks or fingers, while the CNN model mainly focused on only defected area itself, as shown in Fig. 17(a), (b), (c) and (d). Another notable difference is observed in the functional cell's heatmaps. As illustrated in Fig. 17(e), CNN model mostly focused on central part of the image whereas ViT model focused more on the cell edges. Nevertheless, both models performed well enough and detected defective PV cell regions with very high accuracy.

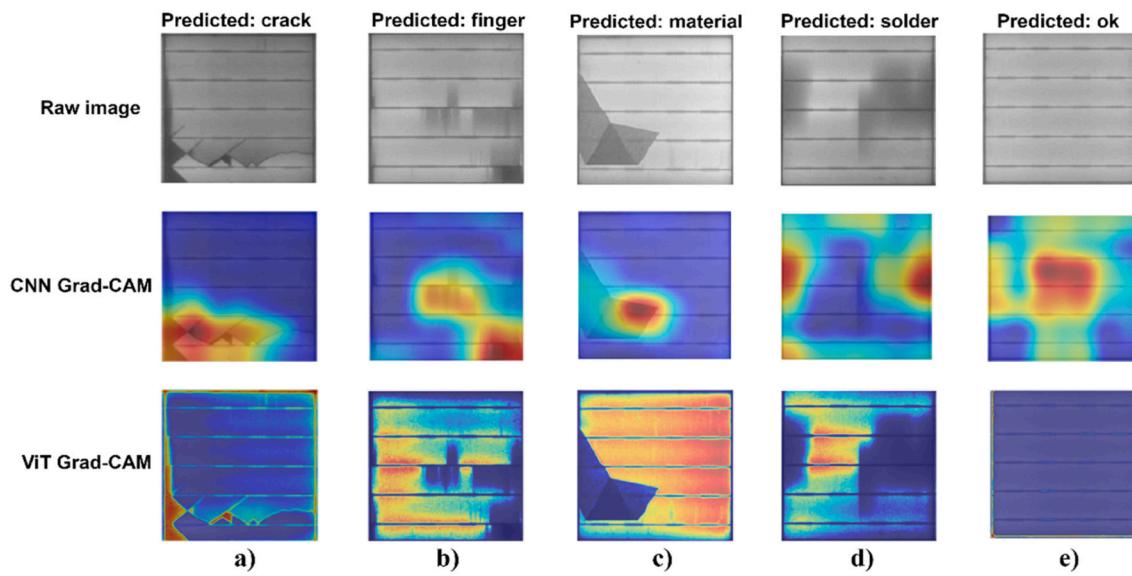


Fig. 17. Class activation maps for the proposed CNN and ViT based models.

6. Conclusions and discussions

In this work, we proposed SolarELTester, end-to-end robust segmentation, classification and evaluation framework for the solar cell EL images. A novel balanced multi-class EL image dataset was created and labeled for this work. Using state-of-the-art AI-Techniques, pre-trained Faster R-CNN-based detector network and CNN and ViT based classifier networks are proposed for PV-EL image defect detection problem. Proposed models were trained and tested on the created EL dataset. With the proposed methods, superior detection and classification performance was obtained for both cell detection and multi-class defect classification. In the proposed classification system, the optimized pre-trained ViT and CNN models were implemented with fast training and detection time. Our proposed segmentation model was able to detect even distorted EL images of PV modules and segment them into individual cells with over 99 % accuracy. Both classifier models reached beyond 98 % detection accuracy. CNN based model hold real-time detection performance rate with only 5.62 ms for a single image detection. ViT based model offered better accuracy and better detection performance in finer defects due to its attention-based mechanism. Proposed methods effectiveness was proven with the performance metrics and visualizations.

For future work, semi-supervised and self-supervised learning methods could be considered for reducing the time and labor for manual labeling of the images. Also, a module-based segmentation system can be utilized for better detection of the PV cells with multiple defects.

CRediT authorship contribution statement

Mustafa Yusuf Demirci: Writing – review & editing, Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Nurettin Beşli:** Writing – review & editing, Validation, Supervision, Project administration, Formal analysis, Conceptualization. **Abdulkadir Gümüşü:** Writing – review & editing, Supervision, Resources, Project administration, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data sets generated during the current study are available from the corresponding author on reasonable request.

References

- [1] P.K. Ozili, E. Ozen, Global energy crisis: impact on the global economy, *SSRN Electron. J.* 8 (4) (2023) R50, <https://doi.org/10.2139/ssrn.4309828>.
- [2] IEA - International Energy Agency, World Energy Outlook 2023, OECD, 2023, <https://doi.org/10.1787/827374a6-en>.
- [3] IEA - International Energy Agency, Renewables 2022 - Analysis and Forecasts to 2027, OECD, 2022, <https://doi.org/10.1787/96bc279a-en>.
- [4] H. Munawer Al-Otum, Deep learning-based automated defect classification in Electroluminescence images of solar panels, *Adv. Eng. Inform.* 58 (2023) 102147, <https://doi.org/10.1016/j.aei.2023.102147>, August.
- [5] M. Kötges, et al., IEA-PVPS T13-01 2014 Review of Failures of Photovoltaic Modules Final, 1911, July, 2014 doi: 978-3-906042-16-9.
- [6] M.Y. Demirci, N. Beşli, A. Güümüşü, Efficient deep feature extraction and classification for identifying defective photovoltaic module cells in Electroluminescence images, *Expert Syst. Appl.* 175 (2021), <https://doi.org/10.1016/j.eswa.2021.114810>, March.
- [7] M. Waqar Akram, G. Li, Y. Jin, X. Chen, Failures of photovoltaic modules and their detection: a review, *Appl. Energy* 313 (May 2022) 118822, <https://doi.org/10.1016/j.apenergy.2022.118822>, February.
- [8] T. Fuyuki, H. Kondo, Y. Kaji, T. Yamazaki, Y. Takahashi, Y. Uraoka, One Shot Mapping of Minority Carrier Diffusion Length in Polycrystalline Silicon Solar Cells Using Electroluminescence, 2005, pp. 1343–1345, <https://doi.org/10.1109/pvsc.2005.1488390>.
- [9] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, *Adv. Neural Inf. Process. Syst.* 2 (2012) 1097–1105.
- [10] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc., 2015, pp. 1–14.
- [11] C. Szegedy, et al., Going deeper with convolutions, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2015, pp. 1–9, <https://doi.org/10.1109/CVPR.2015.7298594>, 07-12-June.
- [12] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2016, pp. 770–778, <https://doi.org/10.1109/CVPR.2016.90>, 2016-Decem.
- [13] M. Tan, Q.V. Le, EfficientNet: rethinking model scaling for convolutional neural networks, in: 36th Int. Conf. Mach. Learn. ICML, 2019, pp. 10691–10700, 2019, vol. 2019-June.
- [14] O. Russakovsky, et al., ImageNet large scale visual recognition challenge, *Int. J. Comput. Vis.* 115 (3) (2015) 211–252, <https://doi.org/10.1007/s11263-015-0816-y>.
- [15] A. Dosovitskiy, et al., An image is worth 16X16 words: transformers for image recognition at scale, in: ICLR 2021 - 9th Int. Conf. Learn. Represent., 2021, <https://doi.org/10.48550/arXiv.2010.11929>.
- [16] A. Vaswani, et al., Attention is all you need, *Adv. Neural Inf. Process. Syst.* (2017) 5999–6009, <https://doi.org/10.48550/arXiv.1706.03762>, 2017-Decem, no. Nips.
- [17] A. Bartler, L. Mauch, B. Yang, M. Reuter, L. Stoicescu, Automated detection of solar cell defects with deep learning, *Eur. Signal Process. Conf.* (2018) 2035–2039, <https://doi.org/10.2391/EUSIPCO.2018.8553025>, 2018-Septe.
- [18] S. Deitsch, et al., Automatic classification of defective photovoltaic module cells in electroluminescence images, *Sol. Energy* 185 (2019) 455–468, <https://doi.org/10.1016/j.solener.2019.02.067>, February.
- [19] C. Buerhop-Lutz, S. Deitsch, A. Maier, F. Gallwitz, C.J. Brabec, A benchmark for visual identification of defective solar cells in electroluminescence imagery, in: 35th Eur. PV Sol. Energy Conf. Exhib., 2018, pp. 1287–1289.
- [20] M.W. Akram, et al., CNN based automatic detection of photovoltaic cell defects in electroluminescence images, *Energy* 189 (2019) 116319, <https://doi.org/10.1016/j.energy.2019.116319>.
- [21] K.M. Lin, H.H. Lin, Y.T. Lin, Development of a CNN-based hierarchical inspection system for detecting defects on electroluminescence images of single-crystal silicon photovoltaic modules, *Mater. Today Commun.* 31 (2022) 103796, <https://doi.org/10.1016/j.mtcomm.2022.103796>, June.
- [22] S. Ren, K. He, R. Girshick, J. Sun, Faster R-CNN: towards real-time object detection with region proposal networks, *IEEE Trans. Pattern Anal. Mach. Intell.* 39 (6) (2017) 1137–1149, <https://doi.org/10.1109/TPAMI.2016.2577031>.
- [23] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2016, pp. 779–788, <https://doi.org/10.1109/CVPR.2016.91>.
- [24] U. Otamendi, I. Martínez, M. Quartulli, I.G. Olaizola, E. Viles, W. Cambarau, Segmentation of cell-level anomalies in electroluminescence images of photovoltaic modules, *Sol. Energy* 220 (May 2021) 914–926, <https://doi.org/10.1016/j.solener.2021.03.058>, October 2020.
- [25] Z. Meng, S. Xu, L. Wang, Y. Gong, X. Zhang, Y. Zhao, Defect object detection algorithm for electroluminescence image defects of photovoltaic modules based on deep learning, *Energy Sci. Eng.* 10 (3) (2022) 800–813, <https://doi.org/10.1002/ese.3.1056>.
- [26] H.H. Lin, H.K. Dandage, K.M. Lin, Y.T. Lin, Y.J. Chen, Efficient cell segmentation from electroluminescent images of single-crystalline silicon photovoltaic modules and cell-based defect identification using deep learning with pseudo-colorization, *Sensors* 21 (13) (2021), <https://doi.org/10.3390/s21134292>.
- [27] A. Korovin, et al., Anomaly detection in electroluminescence images of heterojunction solar cells, *Sol. Energy* 259 (2023) 130–136, <https://doi.org/10.1016/j.solener.2023.04.059>, May.
- [28] W. Tang, Q. Yang, K. Xiong, W. Yan, Deep learning based automatic defect identification of photovoltaic module using electroluminescence images, *Sol. Energy* 201 (2020) 453–460, <https://doi.org/10.1016/j.solener.2020.03.049>, November 2019.
- [29] Y. Zhao, K. Zhan, Z. Wang, W. Shen, Deep learning-based automatic detection of multitype defects in photovoltaic modules and application in real production line, *Prog. Photovoltaics Res. Appl.* 29 (4) (2021) 471–484, <https://doi.org/10.1002/pip.3395>.
- [30] J. Wang, et al., Deep-learning-based automatic detection of photovoltaic cell defects in electroluminescence images, *Sensors* 23 (1) (2023), <https://doi.org/10.3390/s23010297>.
- [31] W. Weng, X. Zhu, U-net: convolutional networks for Biomedical image segmentation, *IEEE Access* 9 (2021) 16591–16603, <https://doi.org/10.1109/ACCESS.2021.3053408>.
- [32] M.R.U. Rahman, H. Chen, Defects inspection in polycrystalline solar cells electroluminescence images using deep learning, *IEEE Access* 8 (2020) 40547–40558, <https://doi.org/10.1109/ACCESS.2020.2976843>.
- [33] L. Pratt, D. Govender, R. Klein, Defect detection and quantification in electroluminescence images of solar PV modules using U-net semantic segmentation, *Renew. Energy* 178 (2021) 1211–1222, <https://doi.org/10.1016/j.renene.2021.06.086>.
- [34] X. Xie, H. Liu, Z. Na, X. Luo, D. Wang, B. Leng, DPiT: detecting defects of photovoltaic solar cells with image transformers, *IEEE Access* 9 (2021) 154292–154303, <https://doi.org/10.1109/ACCESS.2021.3119631>.
- [35] J.P.C. Barnabé, L.P. Jiménez, G. Fraidenraich, E.R. De Lima, H.F. Dos Santos, Quantification of damages and classification of flaws in mono-crystalline photovoltaic cells through the application of vision transformers, *IEEE Access* 11 (2023) 112334–112347, <https://doi.org/10.1109/ACCESS.2023.3322653>.
- [36] D. Bradley, G. Roth, Adaptive thresholding using the integral image, *J. Graph. Tool.* 12 (2) (2007) 13–21, <https://doi.org/10.1080/2151237x.2007.10129236>.
- [37] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, Focal Loss for dense object detection, *IEEE Trans. Pattern Anal. Mach. Intell.* 42 (2) (2020) 318–327, <https://doi.org/10.1109/TPAMI.2018.2858826>.
- [38] R. Girshick, J. Donahue, T. Darrell, J. Malik, Rich feature hierarchies for accurate object detection and semantic segmentation, in: Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., 2014, pp. 580–587, <https://doi.org/10.1109/CVPR.2014.81>.
- [39] M. Carranza-García, J. Torres-Mateo, P. Lara-Benítez, J. García-Gutiérrez, On the performance of one-stage and two-stage object detectors in autonomous vehicles using camera data, *Remote Sens.* 13 (1) (2021) 1–23, <https://doi.org/10.3390/rs13010089>.
- [40] F. Demir, V. Bajaj, M.C. Ince, S. Taran, A. Şengür, Surface EMG signals and deep transfer learning-based physical action classification, *Neural Comput. Appl.* 31 (12) (Dec. 2019) 8455–8462, <https://doi.org/10.1007/s00521-019-04553-7>.
- [41] T. Wolf, et al., Transformers: state-of-the-art natural language processing, in: EMNLP 2020 - Conf. Empir. Methods Nat. Lang. Process. Proc. Syst. Demonstr., 2020, pp. 38–45, <https://doi.org/10.18653/v1/2020.emnlp-demos.6>, January 2020.
- [42] H. Touvron, M. Cord, A. El-Nouby, J. Verbeek, H. Jégou, Three things everyone should know about vision transformers, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2022, pp. 497–515, https://doi.org/10.1007/978-3-031-20053-3_29, 13684 LNCS.
- [43] R. Padilla, S.L. Netto, E.A.B. da Silva, A Survey on performance metrics for object-detection algorithms, in: 2020 International Conference on Systems, Signals and Image Processing (IWSSIP), Jul. 2020, pp. 237–242, <https://doi.org/10.1109/IWSSIP48289.2020.9145130>, 2020-July.
- [44] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization, in: Proc. IEEE Int. Conf. Comput. Vis., 2017, pp. 618–626, <https://doi.org/10.1109/ICCV.2017.74>, 2017-October.