Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

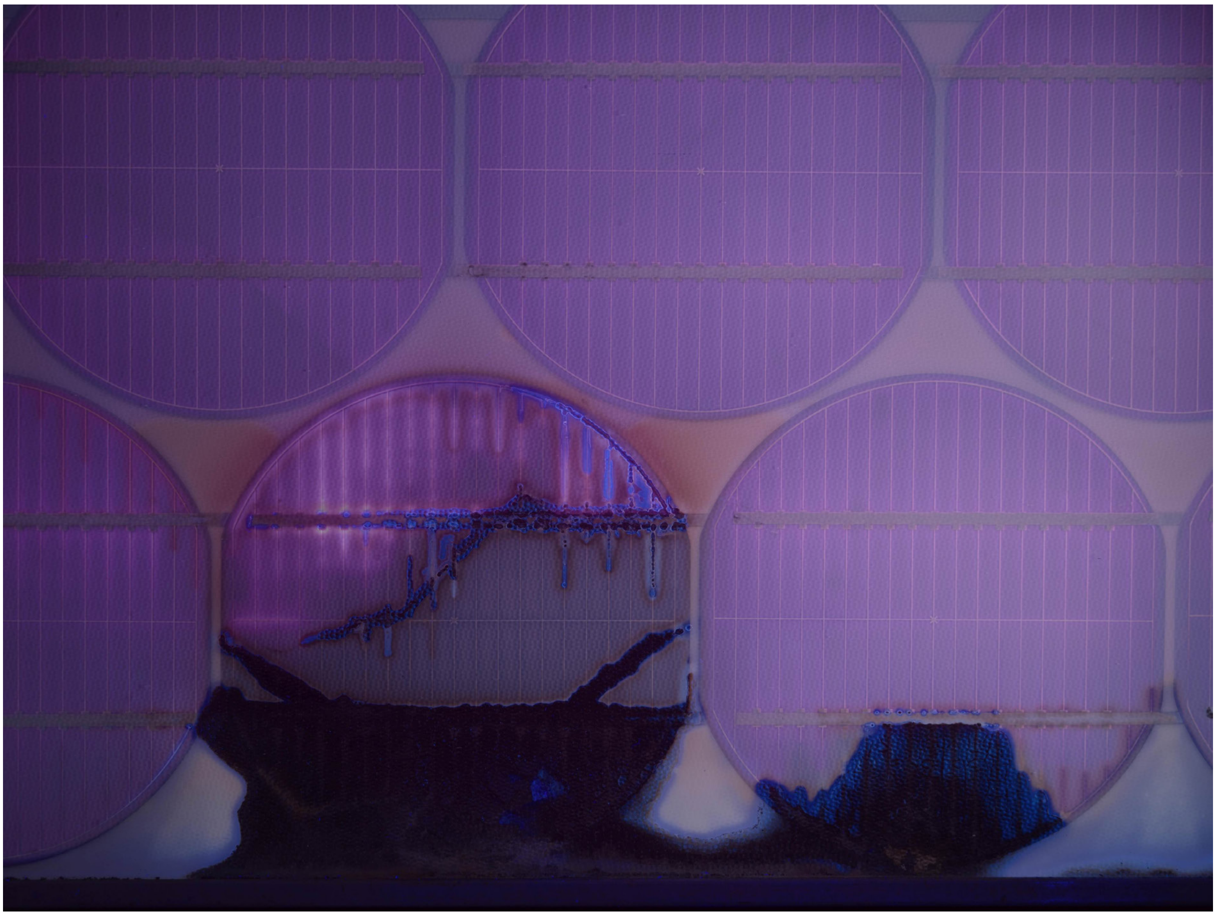
Federal Department of the Environment, Transport, Energy and Communications DETEC

Swiss Federal Office of Energy SFOE
Energy Research and Cleantech Division

Final report dated 28.01.2026.

EAGLE

A deep learning approach to photovoltaics reliability



Source: SUPSI ©

Date: 28.01.2026

Location: Bern

Publisher:
Swiss Federal Office of Energy SFOE
Energy Research and Cleantech
CH-3003 Bern
www.bfe.admin.ch

Subsidy recipients:
SUPSI-DACD-ISAAC
Campus Mendrisio
Via Flora Ruchat-Roncati 15
CH6850 Mendrisio
https://www.supsi.ch/isaac

Fernfachhochschule Schweiz FFHS
Zollstrasse 17
CH-8005 Zürich
https://www.ffhs.ch/de/forschung

Authors:
Mauro Caccivio, SUPSI, mauro.caccivio@supsi.ch
Ralf Jandl, FFHS, ralf.jandl@ffh.ch
Ebrar Özkalay, SUPSI, ebrar.oezkalay@supsi.ch
Danuta Paraficz, FFHS, danuta.paraficz@ffhs.ch
Natasa Sarafijanovic-Djukic, FFHS, natasa.sarafijanovic-djukic@ffhs.ch
Beatrice Paoli, FFHS, beatrice.paoli@ffhs.ch

SFOE project coordinators:
Stefan Oberholzer, stefan.oberholzer@bfe.admin.ch

SFOE contract number: SI/502555-01

Summary

Photovoltaic (PV) energy plays a key role in reducing global carbon emissions. As the number of PV systems installed continues to grow, ensuring the long-term reliability of PV modules becomes increasingly important. Failures can significantly reduce energy yield, yet many degradation mechanisms remain difficult to quantify using conventional inspection methods. Multispectral imaging enables the detection of such failures by capturing complementary information from different spectral domains.

The objective of this project was to develop a quantitative methodology for acquiring and automatically analysing multispectral images of PV modules. The proposed framework combines ultra-high-resolution multispectral imaging with artificial intelligence to detect defects visible at the solar cell level and correlate them with electrical performance losses. A single ultra-high-resolution multispectral camera was used to acquire visible (VI), electroluminescence (EL), and ultraviolet fluorescence (UVf) images.

During the project, a dataset of **241 PV modules corresponding to 11076 solar cells** was collected and processed. A dedicated segmentation pipeline automatically extracted individual cell images from module images. **More than 10% of the dataset was annotated by experts**, providing a high-quality training and validation dataset for machine learning models. The core component of the methodology is a multispectral defect classification model based on the **Channel Vision Transformer** architecture. The model was trained using a progressive transfer-learning strategy: pre-training on ImageNet, fine-tuning on large publicly available EL datasets, and final adaptation to the project's multispectral images. Despite the limited size of the proprietary dataset, this strategy enabled robust defect detection across different module technologies. The model achieved **>90% accuracy on labelled project data** and approximately **94% accuracy on an independent external dataset** (EL only). For unlabeled images, the system achieved **>70% correct failure identification**, meeting the project's performance targets.

The developed approach enables the detection of several PV failure modes, including **cell cracks, inactive regions, corrosion, delamination, and discolouration**. A key improvement over conventional methods is the ability to **detect multiple simultaneous defects**, reflecting the complex degradation patterns typically observed in aged modules. Combining multiple spectral channels also allows identification of defects not visible in EL imaging alone.

In addition to defect classification, the project demonstrated a data-driven method to estimate **module power loss directly from multispectral images**. A deep learning regression model based on **ResNet-18** was trained to predict the loss in maximum power using VI, EL, and UVf images. This establishes a quantitative link between degradation patterns in images and losses in electrical performance.

Some limitations remain. The methodology performs best for defects with clear optical signatures. Rare defects and complex combinations remain challenging due to limited training data. Furthermore, degradation mechanisms without visible optical signatures cannot be detected. The analysis also focused on cell-level visible defects, while module-level visible degradation patterns, such as potential-induced or ultraviolet-induced degradation-related EL patterns, were not included.

Overall, the project demonstrates that AI-based multispectral image analysis enables **scalable, quantitative, and automated detection of PV module defects** and links visual degradation patterns to **loss of electrical performance**. The results highlight the strong potential of AI-enhanced multispectral imaging as a next-generation tool for PV reliability assessment in research laboratories, industrial production, and operation and maintenance of PV systems.

Zusammenfassung

Die Photovoltaik (PV) spielt eine zentrale Rolle bei der Reduktion globaler CO_2-Emissionen. Mit der zunehmenden Anzahl installierter PV-Systeme gewinnt die langfristige Zuverlässigkeit von PV-Modulen immer mehr an Bedeutung. Defekte können den Energieertrag erheblich reduzieren, jedoch sind viele Degradationsmechanismen mit konventionellen Inspektionsmethoden nur schwer quantitativ zu erfassen. Multispektrale Bildgebung ermöglicht die Detektion solcher Fehler, indem sie komplementäre Informationen aus unterschiedlichen spektralen Bereichen erfasst.

Ziel dieses Projekts war die Entwicklung einer quantitativen Methodik zur Aufnahme und automatisierten Analyse multispektraler Bilder von PV-Modulen. Der entwickelte Ansatz kombiniert ultrascharfe multispektrale Bildgebung mit Methoden der künstlichen Intelligenz, um auf Solarzellebene sichtbare Defekte zu erkennen und mit elektrischen Leistungsverlusten zu korrelieren. Zur Aufnahme der Bilder wurde eine einzelne hochauflösende multispektrale Kamera verwendet, mit der Aufnahmen im sichtbaren Bereich (VI), Elektrolumineszenz (EL) sowie ultravioletter Fluoreszenz (UVf) erzeugt werden können.

Im Verlauf des Projekts wurde ein Datensatz von **241 PV-Modulen mit insgesamt 11.076 Solarzellen** erfasst und verarbeitet. Eine speziell entwickelte Segmentierungspipeline extrahierte automatisch einzelne Zellbilder aus den Modulaufnahmen. **Mehr als 10 % des Datensatzes wurden von Expertinnen und Experten annotiert**, wodurch ein hochwertiger Trainings- und Validierungsdatensatz für maschinelle Lernverfahren entstand. Der zentrale Bestandteil der Methodik ist ein multispektrales Klassifikationsmodell zur Defekterkennung auf Basis der **Channel Vision Transformer-Architektur**. Das Modell wurde mittels einer progressiven Transfer-Learning-Strategie trainiert: zunächst Vortraining auf ImageNet, anschließend Feinabstimmung auf großen öffentlich verfügbaren EL-Datensätzen und schließlich Anpassung an die multispektralen Bilder dieses Projekts. Trotz der begrenzten Größe des proprietären Datensatzes ermöglichte diese Strategie eine robuste Defekterkennung über verschiedene Modultechnologien hinweg. Das Modell erreichte eine **Genauigkeit von über 90 % auf den annotierten Projektdaten** sowie etwa **94 % Genauigkeit auf einem unabhängigen externen Datensatz (nur EL)**. Für nicht annotierte Bilder wurde eine **korrekte Fehlererkennung von über 70 %** erzielt und damit die im Projekt definierten Leistungsziele erreicht.

Der entwickelte Ansatz ermöglicht die Detektion verschiedener PV-Fehlermodi, darunter **Zellrisse, inaktive Zellbereiche, Korrosion, Delamination und Verfärbungen**. Ein wesentlicher Fortschritt gegenüber konventionellen Methoden besteht in der Fähigkeit, **mehrere gleichzeitig auftretende Defekte innerhalb einer einzelnen Zelle** zu erkennen, was die komplexen Degradationsmuster gealterter Module besser widerspiegelt. Durch die Kombination mehrerer spektraler Kanäle können zudem Defekte identifiziert werden, die in reinen EL-Aufnahmen nicht sichtbar sind.

Neben der Defektklassifikation wurde im Projekt auch eine datengetriebene Methode zur Abschätzung von **Leistungsverlusten von PV-Modulen direkt aus multispektralen Bildern** entwickelt. Hierfür wurde ein Deep-Learning-Regressionsmodell auf Basis von **ResNet-18** trainiert, um den Verlust der maximalen Leistung anhand von VI-, EL- und UVf-Bildern vorherzusagen. Dadurch wird eine quantitative Verbindung zwischen sichtbaren Degradationsmustern in den Bildern und dem Verlust elektrischer Leistung hergestellt.

Dennoch bestehen einige Einschränkungen. Die Methodik funktioniert am besten für Defekte mit klaren optischen Signaturen. Seltene Defekte sowie komplexe Defektkombinationen bleiben aufgrund begrenzter Trainingsdaten schwierig zu erkennen. Darüber hinaus können Degradationsmechanismen ohne sichtbare optische Signatur nicht detektiert werden. Die Analyse konzentrierte sich außerdem auf

sichtbare Defekte auf Zellebene, während **sichtbare Degradationsmuster auf Modulebene**, wie etwa durch potenzialinduzierte oder UV-induzierte Degradation verursachte EL-Muster, nicht berücksichtigt wurden.

Insgesamt zeigt das Projekt, dass KI-basierte multispektrale Bildanalyse eine **skalierbare, quantitative und automatisierte Erkennung von Defekten in PV-Modulen** ermöglicht und visuelle Degradationsmuster mit elektrischen Leistungsverlusten verknüpft. Die Ergebnisse verdeutlichen das große Potenzial KI-gestützter multispektraler Bildgebung als **Diagnosewerkzeug der nächsten Generation** für die Zuverlässigkeitsbewertung von PV-Systemen in Forschungslaboren, industrieller Produktion sowie im Betrieb und der Wartung von PV-Anlagen.

Main Findings

- A multispectral artificial intelligence (AI) model based on the **Channel Vision Transformer (ChannelViT)** architecture was developed to automatically detect and classify defects in photovoltaic (PV) modules three complementary imaging modalities: **visible (VI)**, **electroluminescence (EL)**, and **ultraviolet fluorescence (UVf)** images

- When trained and tested using **EL images** of crystalline silicon PV modules from publicly available datasets, the model achieved **94.4% classification accuracy on previously unseen images** for common EL-visible defect classes such as *good*, *cracks*, *cross cracks*, *dark regions*, and *corrosion*. This demonstrates strong performance for standard EL-based defect detection.

- The inclusion of **visible (VI)** and **ultraviolet fluorescence (UVf) images** enable the detection of additional defect classes that are typically not visible in EL images alone, such as ***discolouration*** and ***delamination***. However, larger labelled datasets are required to fully capture these defects across different PV module technologies.

- When combining all three image types (**VI, EL, and UVf**), the developed model achieved **>90% accuracy on expert-labelled multispectral data and ≥70% correct defect identification on unlabeled data**, demonstrating reliable performance for multispectral defect detection.

- A complementary **AI-based method for estimating PV module performance** directly from multispectral images was also developed. Using VI, EL, and UVf images together reduced the **mean absolute error (MAE) in predicted power loss from 5.5% (EL-only) to 4.8%**, demonstrating that combining multiple spectral imaging modalities improves the accuracy of performance loss estimation.

Contents

Abbreviations

AI	Artificial Intelligence
Al-BSF	Aluminum Back Surface Field
BIPV	Building-Integrated Photovoltaics
ChannelViT	Channel Vision Transformer
CNN	Convolutional Neural Network
COCO	Common Objects in Context
Duramat	Durable Module Materials Consortium
EL	Electroluminescence
EVA	Ethylene-Vinyl Acetate
FAIR	Facebook AI Research
FFHS	Fernfachhochschule Schweiz
FMEA	Failure Mode Effect Analysis
GB	Gigabyte
GPU	Graphics Processing Unit
HDBSCAN	Hierarchical Density-Based Spatial Clustering of Applications with Noise
HJT	Heterojunction Technology
IBC	Interdigitated Back Contact
IEC	International Electrotechnical Commission
IR	Infrared
LeTID	Light and Elevated Temperature Induced Degradation
MAE	Mean Absolute Error
MAST	Module Accelerated Sequential Testing
MedAE	Median Absolute Error
MB	Megabyte
MHSA	Multi-Head Self-Attention
Mpx	Megapixel
OFI	Österreichisches Forschungsinstitut für Chemie und Technik
PC	Personal Computer
PERC	Passivated Emitter Rear Cell
PID	Potential Induced Degradation
Pm	Maximum Power
Pmp	Maximum power point percentage
PV	Photovoltaic
PVLab	SUPSI Photovoltaics Laboratory

RAM	Random Access Memory
R&D	Research and Development
ResNet	Residual Neural Network
RGB	Red Green Blue
ROI	Region of Interest
SAM	Segment Anything Model
STC	Standard Test Conditions
t-SNE	t-Distributed Stochastic Neighbor Embedding
TOPCon	Tunnel Oxide Passivated Contact
TÜV	Technischer Überwachungsverein
UMAP	Uniform Manifold Approximation and Projection
US	United States
UV	Ultraviolet
UVf	Ultraviolet Fluorescence
UVID	Ultraviolet-Induced Degradation
VGG	Visual Geometry Group
VI	Visual Inspection
VPN	Virtual Private Network

1 Introduction

1.1 Background information and current situation

Photovoltaic (PV) energy is a crucial source for decarbonization: The continuous increment of temperature witnessed in the Alps (+2°C with respect to the reference average 1961-1990[1]) is the echo of the worldwide phenomenon of climate change caused mainly, if not exclusively, by greenhouse gases emissions. Among renewable sources, photovoltaic energy is crucial in order to reach zero net carbon emission, both in terms of cost competitive price and energy market penetration in the last years[2]. The explosive trend in price reduction of this technology has been possible thanks to a mix of constant improvement in cell efficiency, industrial optimization, and economies of scale, but important steps in terms of lifetime enhancement are needed in order to grant the lowest levelized cost of electricity. Presently, the typical power performance warranty for PV modules guarantees around 80% of the initial power output after 25 years, but an extension of this guaranteed power output time to 35 years and beyond is already envisioned by the European strategical plans[3].

The reliability of photovoltaic modules has been initially verified through accelerated testing based upon the ones identified by the Jet Propulsion Laboratory (JPL) in the late seventies, where five testing blocks of solar modules were used to lay the foundation for the international electrotechnical commission (IEC) approval testing standards, necessary to bring on the market safe and reliable products[4]. The IEC 61215 and 61730 standards remained basically unchanged for several years, until the explosive evolution of the market and the first massive failures of PV panels led to an important update in 2016, with more demanding testing sequences being developed to ensure the 25 years' lifetime frame, the standard for warranty[5]. The research of the optimal acceleration to detect failures at the earliest possible time is pursued further by the IEC technical committees, who are now working on a revision of the IEC 61215 and 61730 standards much faster than the previous update, to cover new defects coming from new technologies, such as ultraviolet-induced degradation (UVID)[6], potential-induced degradation (PID)[7], light and elevated temperature induced degradation (LeTID)[8], updated ultraviolet (UV) stress test to detect backsheet weaknesses[9]. The solar industry as well is looking for new formulas to grant defect free production and increase the warranty terms: some examples of the improvement of quality checks beyond IEC standards are the PV module reliability scorecard by PV Evolution Lab, the Thresher test sequence proposed by TÜV Süd (Technischer Überwachungsverein Süd), the MAST (Module accelerated sequential testing) approach from Dupont and the production approach by Sunpower, based upon FMEA (Failure mode effect analysis) for the prioritization of test development[10]. In September 2020 Violet Power, a US startup founded by the former R&D responsible of ARCO Solar, one of the pioneering PV manufacturers, announced for the first time a 50 years' warranty term, based upon a vertically integrated production, with all its steps under direct control and extended accelerated testing[11]. The new layouts and concepts subjected to standard and accelerated testing are showing some critical patterns, such as UV-induced degradation (UVID) of high-efficiency solar cells, and hotspot sensitivity on bifacial[12] and shingled cells[13]. On the other hand, even though the installed systems with high-efficiency modules are relatively young, the rapid pace of technological advancement meant they lack a long-term track record. various defects on module materials and solar cells could be thoroughly investigated through visual inspection (VI), electroluminescence (EL) and ultraviolet fluorescence (UVf) [15].

For all of the above reasons, the opportunity to use high-resolution images on different bandwidths (Visible VI, Ultraviolet UV, Infrared IR) to identify and quantify the appearance of failure modes represents a very promising way to achieve a cost wise and innovative instrument to monitor and compare the accelerated and/or not accelerated degradation of PV modules. The failure modes are automatically identified and classified by means of machine learning algorithms. Artificial intelligence (AI) algorithms have been used to successfully solve the task of classifying images in different application fields, and Vision Transformers (ViTs) have been proved to achieve very good performance[46]. ViTs use attention over the image and detecting certain patterns and spatial context. An additional advantage of ViTs is the transfer learning capability, meaning that ViTs which are trained for another image related task, i.e., already learned certain patterns, can be used, and adjusted to the problem at hand, which can reduce training time significantly[15]. Once deployed, the algorithms allow to identify the objects automatically in new data not yet seen by the model.

In the field of failure modes classification in photovoltaics, machine learning has been employed mainly for the analysis of EL images[18,19]. In these studies, EL images of the PV modules have been segmented to extract the single cells as a basic unit and these were classified into good functioning and defected ones. In some cases[18,19], the defected solar cells have been classified into more than one failure mode (e.g. cracked, corroded and so on), but every image could belong to just one class (e.g. a cracked cell could not be identified as also corroded). This is a limitation for PV modules that already experienced some aging, as the appearance of more than one failure mode per cell cannot be excluded. In the field of thermography, a Naive Bayes classifier has been trained on infrared (IR) images to carry out hot-spot diagnosis on full PV modules[22]. This approach, however, requires extensive feature extraction, does not provide information at the cell level and is not able to distinguish between different failure modes. The manual annotation of failure modes is known to be subjective. In this study[21], two different operators identified an average 1.4 and 2.8 defects per PV module, respectively.

At present, PV modules are guaranteed to retain 85% of the nominal power after 25 years of operation, but a longer lifetime of PV modules would help to reduce the price of the PV generated electricity. Understanding how the different failure modes affect the module performance is therefore of great importance. Several studies[22–24] describe the types of failure modes occurring in degraded PV modules. Some attempts have been done to analytically model the degradation of PV modules, and a review of them can be found in[23]. Such analytical models, however, rely on assumptions and simplifications that limit their employment in real settings and suffer from a lack of generalization capability. In[22] an equivalent-circuit model is used in combination with optoelectronic measurements to analyze the different contributions to the output power losses. The approach presented by the authors in[22] requires an extensive characterization of the PV modules, involving invasive procedures as the extraction of encapsulant samples for carrying out the optical measurements. The characterization has been done for two PV modules, and scaling up this methodology for a higher number of modules is unlikely.

1.2 Purpose of the project

The aim of this project was to define a novel methodology for the quantitative and timely analysis of defects and failures in PV modules by combining image analysis with artificial intelligence algorithms, and to correlate these observations with module performance losses.

A key feature of the methodology is the use of a unique, extremely high-resolution multi-spectral camera with sensitivity ranging from UV to IR, equipped with appropriate filters and lighting, and using a single camera for all measurements. This setup allowed rapid data acquisition, reducing uncertainties associated with multiple cameras while retaining the advantages of integrating several diagnostic tools for efficient defect evaluation.

The project aimed to advance the state of the art by developing an automated methodology for the identification of failure modes using a seven-channel multispectral image composed of visual inspection (VI-RGB), electroluminescence (EL-gray scale), and ultraviolet fluorescence (UVf-RGB) data. Combining multiple measurement techniques enables the identification of a wider spectrum of degradation modes, which is critical when analyzing modules exposed to real environmental aging. To achieve these goals, an automatic failure mode identification methodology at the cell level was implemented using Channel Vision Transformers (Channel ViTs [46–47]).

The methodology also enabled the identification of multiple degradation modes in a single solar cell, allowing quantitative analysis of the impact of different failure modes on module performance. Careful annotation of the training dataset reduced subjectivity and ensured high-quality model training.

The employment of machine learning instead of analytical models (such as in the studies[22–24] discussed in section 1.1) goes beyond the state of the art as it will enable to quantitatively analyze the impact of the different failure modes originating in field exposure and artificial aging experiments on the cell performance. Moreover, the approach presented in the present project overcomes the need of invasive procedures as it is easily scalable while providing quantitative insights on the degradation mechanisms of PV modules.

The results of the project will have a direct impact on the future methodology to trace defects, both in research laboratories and in production environment, allowing a quantitative and measurable way to record the evolution of module performances.

1.3 Objectives

The present project pursues the following objectives:

- Develop a methodology to enhance the standard visual inspection procedure included in IEC standards, in order to provide detailed and comprehensive documentation of the PV modules before, during intermediate steps and at the end of the accelerated aging procedures.
- Develop a methodology for the automatic identification of failure modes in solar cells and modules using artificial intelligence.
- Develop a data fusion strategy to combine images coming from different experimental techniques into one multichannel image.
- Quantitatively identify the failure modes affecting at most the performance of naturally and artificially aged modules, respectively.

In order to achieve these goals, the following research questions will be addressed:

- What are the optimal settings in order to achieve the best resolution to detect defects in the different wavelengths? What are the best settings to optimize signal-to-noise ratio?
- Which model based on neural networks achieves the better accuracy in identifying failure modes in crystalline silicon solar cells?
- What is the contribution of the different channels into identifying the failure modes?
- What is the relevant importance of the different failure modes in the degradation of the performance of the modules? How does this change in the different technologies?

The advancement of the project will be measured with the following measurable targets:

- Datasets of raw multispectral images of at least 230 modules.
- Datasets of annotated images. At least 30% of the images are labelled by experts.
- At least 90% accuracy in failure identification by machine learning prototypes for labelled data.
- At least 70% accuracy in failure identification by machine learning prototypes for unlabeled data (tested by experts).
- Identification of the top two failure modes that affect module performance.

2 Procedures and methodology

2.1 Data acquisition: hardware

The choice of image capture device has been focused on the need to have the highest possible resolution, in order to increase the ability to digitally zoom into the images, with the possibility of deepening the analysis in specific areas and avoiding the stitching of photos, thus reducing computational effort and digital artifacts.

In the first phase, a model was chosen that represents the highest resolution multispectral digital sensor on the market: the iXM-MV150, with a 150 Mpx sensor, equipped with a Linos 60 mm f/4 lens and manual focus (Figure 1). In a preliminary phase, before the project was awarded, Phase One offered a trial period with a similar 100 Mpx camera, the results of which were very satisfactory.

a) b)

Figure 1: Picture of the PhaseOne multispectral camera: (a) body and (b) lenses system.

Given the size of the files (870 MB), it has also been necessary to invest in a PC with an adequate graphic card, RAM and solid-state drive, so that the images can be easily opened, processed and stored.

2.2 Datasets

In the first year of the project, several datasets of electroluminescence (EL) images (see below) were used to set up and test the preprocessing pipeline, prior to the construction, calibration, and testing of the data acquisition system for multispectral images. Once the multispectral data acquisition system was operational, visual (VI), electroluminescence (EL), and ultraviolet-fluorescence (UVf) images of various old and new technologies were added to the dataset for further analysis.

Table 1 summarizes the EL datasets used to set up and test the preprocessing pipeline. Since all three imaging techniques and the electrical performance data are required for the project, these datasets could not be used for anything beyond the setup and testing of the preprocessing pipeline.

Table 1: Summary of the datasets used for setting up and testing the preprocessing pipeline.

Data Source	Data	Aging Method	VI	EL	UVf	Electrical data before aging	Electrical data after aging
French et al. EL Image Dataset of PV Module Under Step-wise Damp Heat Exposures[27]	1028 cell images	3000 hours of damp-heat cycling. Images were taken after 500 h, 1000 h, 1500 h, 2000 h, 2500 h and 3000 h of exposure	N	Y	N	N	N
Chen, Crack segmentation dataset[28]	29664 cell images	NA	N	Y	N	N	N
Infinity project[45]	2725 cell images	NA	N	Y	Y	Y	N

Dataset-1: French et al. EL Image Dataset of PV Module Under Step-wise Damp Heat Exposures[27]

Grayscale EL images, segmented at a cell level (250x250 pixels), manually labelled into *"good"*, *"cracked"* and *"corroded"*. An example from each class is shown in Figure 2. A total of 1028 images are available with a Creative Common license (80.4% good, 18.5% corroded, 1.1% cracked). The producer of the modules is not made public. The cells are monocrystalline-Si solar cells, and the modules underwent 3000 hours of damp-heat cycling. Images were taken after 500 h, 1000 h, 1500 h, 2000 h, 2500 h and 3000 h of exposure.

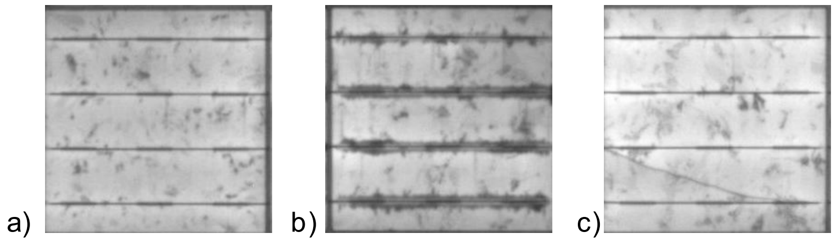
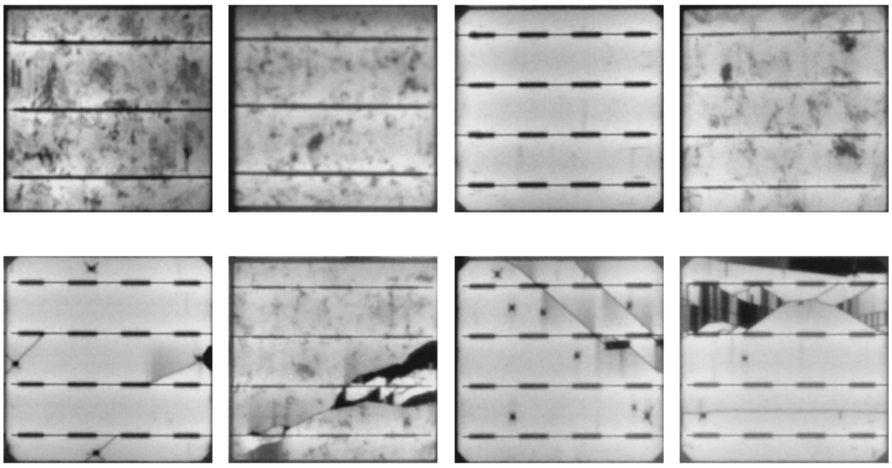


Figure 2: An example of (a) *"good"*, (b) *"corroded"*, and (c) *"cracked"* cell EL images[27].

Dataset-2: Chen, Crack segmentation dataset[28]

The full dataset contains 29664 EL images of different mono- and polycrystalline Si technology (s. Figure 3). They are segmented at a cell level (400x400 pixels) and consist of a grayscale image repeated three times to form a 3-channel image. A part of the cells (2159 images) is labelled as *"good"*, *"cracked"*, *"crossed"*, *"dark"* or a combination of the above.



a)

b)

Figure 3: Examples of (a) *"good"* and (b) defected solar cell images of different PV technologies[28].

Dataset-3: Infinity project [54]

The Österreichisches Forschungsinstitut für Chemie und Technik (OFI) has provided us with EL images of monocrystalline Si modules subjected to different cycles of artificial aging (Figure 4). The images were segmented into 2725 cells following the pipeline described in section 2.3. Also, in this case, a grayscale image is repeated three times to form a 3-channel image.

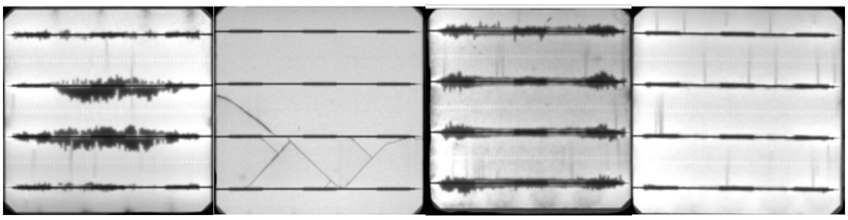


Figure 4: Examples of EL images acquired within the Infinity project.

2.3 Segmentation

Image segmentation refers to the task of automatically identifying objects within an image. There are different ways of performing this task, as depicted in Figure 5. In this project, instance segmentation has been used to identify the exact position of each cell in the modules.

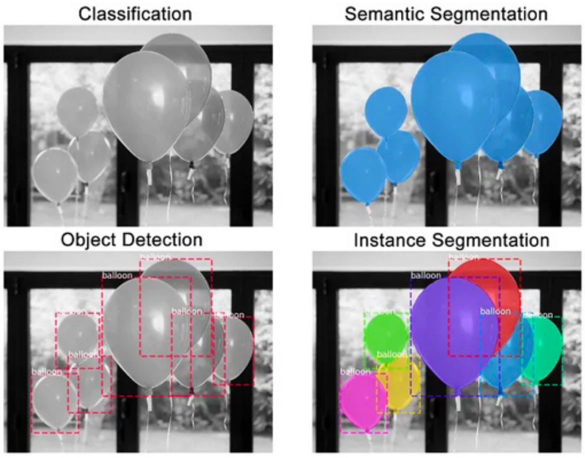


Figure 5: Classification: There is a balloon in this image. Semantic Segmentation: These are all the balloon pixels. Object Detection: There are 7 balloons in this image at these locations. The model accounts for objects that overlap. Instance Segmentation: There are 7 balloons at these locations, and these are the pixels that belong to each one of them[31].

In this project, Detectron2[32] has been used for the image segmentation task. Detectron2 is an open-source computer vision framework by Facebook AI Research (FAIR) based on PyTorch (Figure 6). It's designed for efficient object detection and segmentation model development, offering pre-trained models and modular components. It consists of several key components, including a backbone network (e.g., Residual Network ResNet), feature pyramid network, and task-specific heads for object detection and segmentation[33].

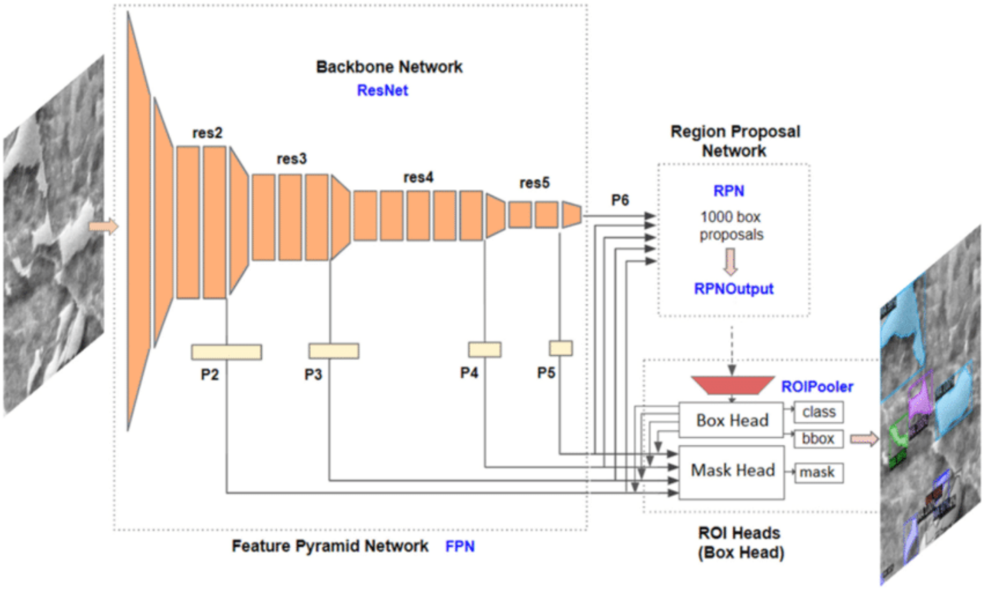


Figure 6: Schematic architecture of Detectron2. The backbone network provides feature maps (P1–P5) to the region proposal network. The ROI (Region of interest) head locates bounding boxes (bbox) and segments (mask) objects, together with the corresponding class (in our case just one class *cell*)[32].

Detectron2 is already pre-trained with the COCO[35] (Common Objects in Context) data set and is able to recognize various shapes and patterns. However, in order to use the model to segment solar cells, the model must be fine-tuned with images of solar cells. As different technologies of solar panels are being analyzed within the project, the model needs to be trained with them, as the shape and nature of the cells differs from one technology to the other. For each technology, at least one image that is as representative as possible must be provided, and the segmentation must be carried out manually by a person. The online tool VGG (Visual Geometry Group) Image Annotator has been employed for this manual masking task[43].

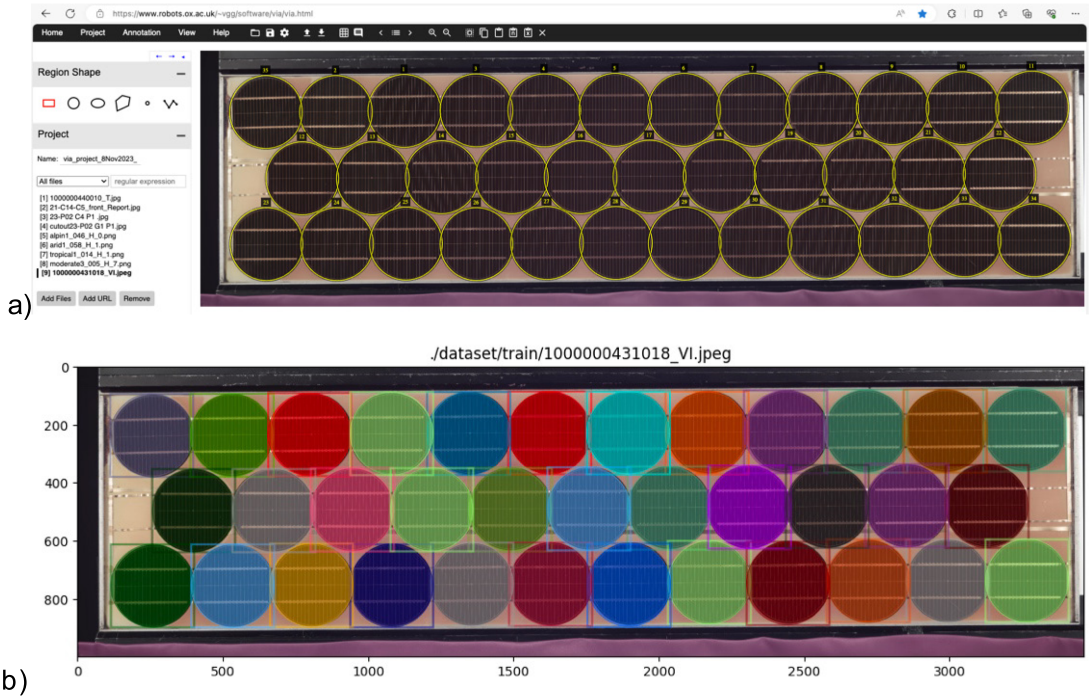


Figure 7: (a) Example of one manually masked TISO image. (b) Example of a new TISO image segmented with Detectron2.

The image and the mask evaluated manually are then used for training. Figure 7a shows one image from the TISO dataset taken as an example. Other technologies, such as the Infinity dataset, have been used for training. Before the segmentation, each TISO image is cut so that only the module is visible, resulting in 6925×2000 pixels images. The trained Detectron2 model can then be used for new TISO photos to automatically segment their cells (Figure 7b). The results of the segmentation and of the labelling (described in section 2.4) are displayed in a web application, described in section 2.5. In order to provide better navigation functionalities in the web application, an overview with all the segmented cells identified with progressive numbering is also produced during the segmentation step. An example for the TISO dataset is shown in Figure 8. In order to reduce the amount of data to be transferred, the overview is scaled by a factor 0.5. Each single segmented cell is then classified individually, which is why the corresponding segment image section for each individual cell is saved individually in the original resolution (approx. 1000 x 1000).

Figure 8: (a) Overview of a segmented TISO image. (b) Segmented individual TISO cell, ready for further analysis.

During the course of the project, an increasing variety of technologies were analyzed and segmented. Unfortunately, it had to be realized that Detectron2 is not as robust and generalizable as hoped. In some cases, not all cells were detected, or the model identified cells in areas of the image where there were none. Additionally, for Detectron2, the training dataset must be supplemented with several manually segmented examples of each new technology, and the model must be retrained. Due to these drawbacks, a newer technology was adopted, which is currently state-of-the-art in the field of image segmentation and is described below.

Segment Anything (SAM) is a new state-of-the-art image segmentation model developed by Meta AI and was published in April 2023. SAM is designed to generalize across a wide range of segmentation tasks with minimal or no task-specific training. Unlike traditional segmentation models such as Detectron2, which require task-specific fine-tuning and dataset augmentation for each new application, SAM emphasizes versatility and ease of use, providing a flexible, general-purpose solution. Segment Anything is a foundation model for image segmentation, capable of producing accurate masks for arbitrary objects in images. It can work with minimal input, such as points, bounding boxes, or textual prompts, to segment objects. This makes it highly suitable for scenarios where manual annotation is labor-intensive or infeasible.

SAM's architecture is built around three primary components[45]:

1. Image Encoder:
   a. A Vision Transformer (ViT) is used as the backbone for encoding the image.
   b. The input image is divided into patches, and their embeddings are extracted.
   c. This pre-trained encoder generates rich, high-dimensional representations of the entire image.
2. Prompt Encoder:
   a. Accepts interactive inputs like points, bounding boxes, or text prompts (via embeddings).
   b. Processes these inputs into embeddings that guide the segmentation process.
   c. This flexibility enables SAM to adapt to various task requirements.
3. Mask Decoder:
   a. Combines the image and prompt embeddings to generate high-quality segmentation masks.
   b. Outputs masks for all objects in the image or for specific objects based on the given prompt.

This modular architecture ensures that SAM can efficiently handle diverse segmentation tasks without fine-tuning.

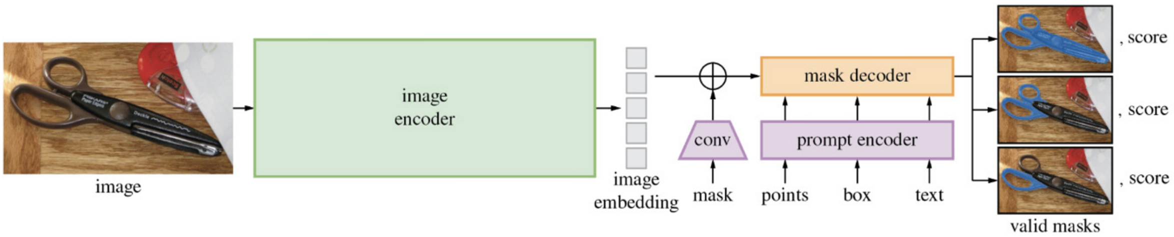


Figure 9: Architecture of Segment Anything Model with Encoder, Decoder and prompting[43].

Key benefits of SAM compared to Detectron2:

1. Generalization Across Tasks:
   a. SAM is trained on a massive dataset of over 1 billion masks, allowing it to generalize to new tasks and objects without additional training.
   b. Detectron2, in contrast, often struggles to generalize and requires fine-tuning on task-specific datasets for each new application.
2. Promptable Segmentation:

    a. SAM supports interactive prompts, such as a single click on an object, to refine segmentation dynamically.
    b. Detectron2 lacks this capability, requiring fully automated or predefined object classes.
3. No Fine-Tuning Needed:
    a. SAM performs zero-shot segmentation on new data, saving time and computational resources.
    b. Detectron2 requires fine-tuning with manually labeled datasets for each domain or object type.
4. Robustness:
    a. SAM is highly robust to variations in object appearance, image quality, and background clutter.
    b. Detectron2 often misidentifies or fails to detect objects under challenging conditions (as described above).
5. Ease of Use:
    a. SAM's ability to process arbitrary input prompts (e.g., points or bounding boxes) makes it more user-friendly.
    b. Detectron2 requires precise data preparation and pre-definition of object categories.
6. Scalability:
    a. SAM's foundation model approach enables it to handle large-scale datasets and applications without task-specific modifications.
    b. Detectron2's scalability is limited by its reliance on manual dataset preparation and retraining.

The Figure 10 shows a comparison of the segmentation quality using Detectron2 and SAM, with SAM demonstrating significantly better accuracy which is essential for the subsequent classification of the cells.

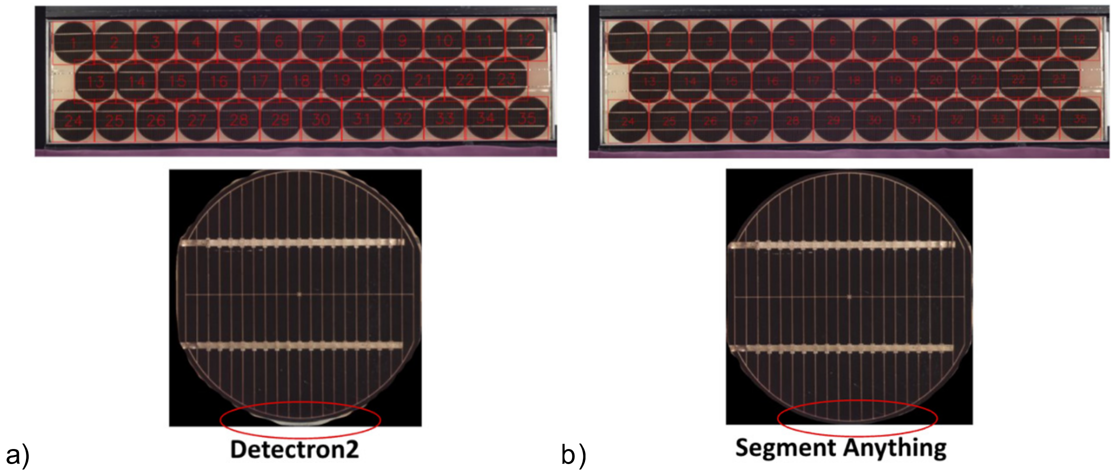


Figure 10: Segmentation using (a) Detectron2 and (b) SAM.

It is appreciated that SAM has been developed as a new, more robust, and simpler method for segmenting PV cells. For the project, SAM is now used for all technologies, with Detectron2 being applied only if SAM does not work for certain technologies (which is currently not the case).

2.4 Unsupervised labelling with UMAP

The first step in the preprocessing of the data is a semi-automatic labelling of the data with the algorithm UMAP (Uniform manifold approximation and projection)[34]. UMAP is a dimensionality reduction technique based on Riemannian geometry and algebraic topology. According to the authors, the algorithm has a good visualization quality and has better performance while preserving the global structure of the data better than other comparable algorithms such as t-SNE (t-distributed stochastic neighbor embedding)[37]. The data are preprocessed as follows:

- Pixels are scaled between 0 and 1.
- A pre-trained VGG16[38] network is used to extract relevant features from the images. VGG16 is a relatively shallow network based on convolutional neural networks with 16 trainable layers. It achieves very good performance on the ImageNet[39] classification task and with 4.2 ms per inference step on a GPU it is a relatively fast model. The model is loaded with the ImageNet weights, and the top classification layer is removed, a common practice when this pre-trained network is used as a feature extractor.
- Subsequently, a global pooling layer is applied to the VGG16 features to reduce their number to 512.
- The resulting 512 features are given to UMAP as an input. UMAP produces unsupervised embeddings in 10 or 20 dimensions, depending on the input dataset. A two-dimensional embedding is also evaluated for visualization purposes.
- HDBSCAN (Hierarchical density-based spatial clustering)[40] is used to cluster the high dimensions UMAP embeddings. HDBSCAN provides density-based hierarchical clustering, and it is used to find groups of solar cells images with similar features. This process is also completely unsupervised. The results of the clustering are projected in 2D for visualization purposes.

A Dash[39] application has been programmed to visualize the results of the unsupervised labelling and to provide each cluster with a label. The following requirements had to be met:

- Quick development and easy maintenance.
- Integration of interactive plots.
- Good integration with existing Python code and Jupyter notebooks.

2.5 Rendering of the results and expert labelling

As mentioned, three photos (Visual, EL, UVf) were taken for each solar panel, each approximately 1 GB in size. This results in very large amounts of data that require a suitable storage location, accessible quickly and easily by all project participants. The decision was made to utilize the FFHS SharePoint, with VPN (Virtual private network) access set up for SUPSI. The underlying OneDrive synchronization of the SharePoint enables automatic syncing of data across the PCs and servers of all project participants.

Additionally, a software solution was needed for the labelling of panels and their cells by experts. This software should allow convenient display of panels and cells, as well as filtering based on various criteria (e.g., technology, error category, etc.). To ensure continuous accessibility for all project participants, a Mac Studio Server was set up at FFHS, reachable via VPN by SUPSI researchers. In order to display the segmented cells and their classification predicted by the model, a simple web interface was developed using Flask. Flask[42] is a popular Python web framework used to develop web applications. It is simple, lightweight, and flexible, making it an excellent choice for building small to medium-sized web applications. Flask can be used to create a basic web service by writing a few lines of code in Python. The web interface offers the user a quick and easy way to view the different solar panels and navigate to the single cells in order to check, or if necessary correct, the classifications made by the model. The user can look at entire modules, can zoom into certain image areas to take a closer look at parts of the modules, and can also look at individual segmented cells and classify them.

2.6 Classification

The Vision Transformer (ViT) (Figure 11: The image is divided into fixed-size patches, linearly embedded, and augmented with position embeddings. A learnable "classification token" is added, and the sequence is processed by a standard Transformer encoder for classification.) is a cutting-edge architecture for image classification, including the domain of damage classification in photovoltaic (PV) cells. Its superiority lies in its ability to model long-range dependencies and understand complex patterns effectively, which are essential for detecting subtle damage in PV cells, such as cracks, hotspots, or delamination. Traditional Convolutional Neural Networks (CNNs) rely on localized receptive fields, which may miss global contextual information, while ViT excels in this domain. ViT captures global patterns through self-attention mechanisms, which are crucial for identifying dispersed or subtle damage across PV cells. It performs well with large-scale datasets, which can be advantageous for training robust damage classifiers if substantial labeled data is available. The ViT architecture is inspired by the Transformer model used in natural language processing (NLP). The key components of the ViT architecture are[46]:

1. Patch Partitioning:
    - The input image is divided into fixed-size patches (e.g., 16x16 pixels).
    - Each patch is flattened into a 1D vector and linearly projected into a patch embedding.
2. Positional Encoding:
    - Positional encodings are added to the patch embeddings to retain spatial information, which is absent in the standard Transformer.
3. Transformer Encoder:
    - The patch embeddings are fed into a Transformer encoder composed of layers of:
        - Multi-Head Self-Attention (MHSA): Captures dependencies between patches by computing attention scores across all patches.
        - Feedforward Networks (FFN): Non-linear transformations for feature enhancement.
        - Layer Normalization and Skip Connections: Stabilize training and facilitate gradient flow.
4. Classification Token:
    - A learnable "class token" is prepended to the input sequence, and its final embedding is used as the image representation for classification.
5. Output:
    - The final embedding is passed through a fully connected layer for damage classification into the desired number of categories.

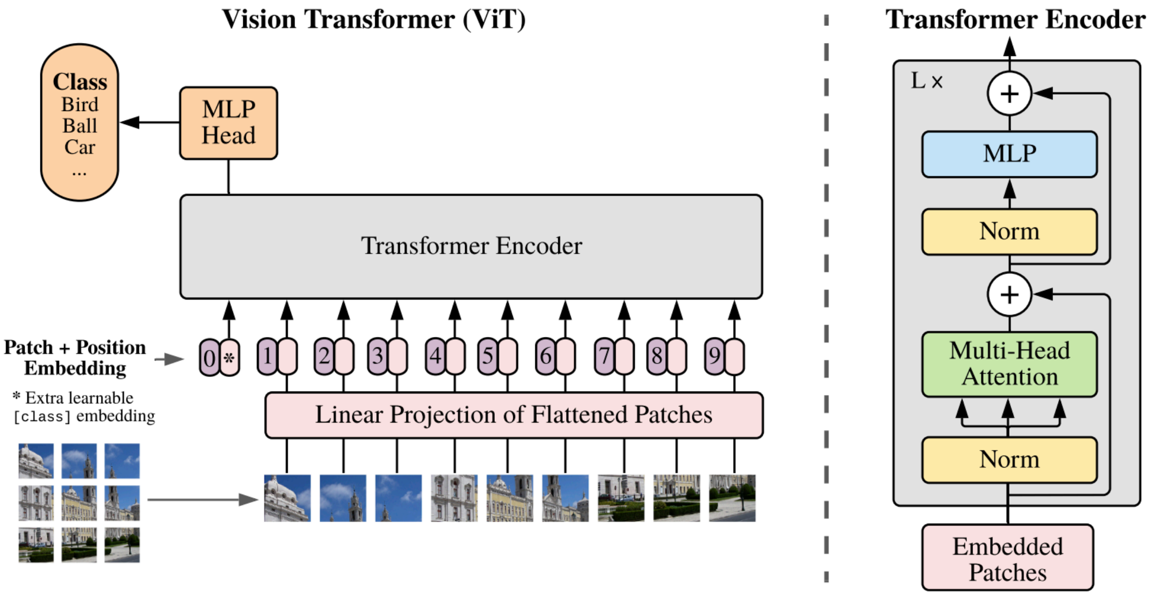


Figure 11: The image is divided into fixed-size patches, linearly embedded, and augmented with position embeddings. A learnable "classification token" is added, and the sequence is processed by a standard Transformer encoder for classification.

ChannelViT (Figure 12) is a specialized extension of the ViT architecture designed to handle images with multiple channels, such as the 7-channel images in this project (3 channels for VI, 3 channels for UVf and 1 channel for EL, simply gray scale, since each RGB channel in EL are identical), where each channel may contain different types of information[45].

The key extensions of the ChannelViT architecture are:

1. Channel-wise Tokenization:
   - Instead of treating the input image as a single entity, ChannelViT processes each channel separately.
   - For each channel, patch embeddings are computed independently.
2. Channel Integration:
   - A specialized attention mechanism integrates information across channels, ensuring that inter-channel relationships (e.g., between thermal, optical, and spectral channels) are effectively captured.
3. Multi-Head Cross-Attention:
   - Enhances the self-attention mechanism to include interactions between patches across different channels, enriching the feature representations.
4. Channel-Aware Positional Encoding:
   - Positional encodings are adapted to incorporate channel-specific spatial information, crucial for preserving the integrity of multi-channel data.
5. Hierarchical Feature Fusion:
   - ChannelViT combines features from different channels at multiple levels of abstraction, allowing for a more comprehensive understanding of the multi-channel input.

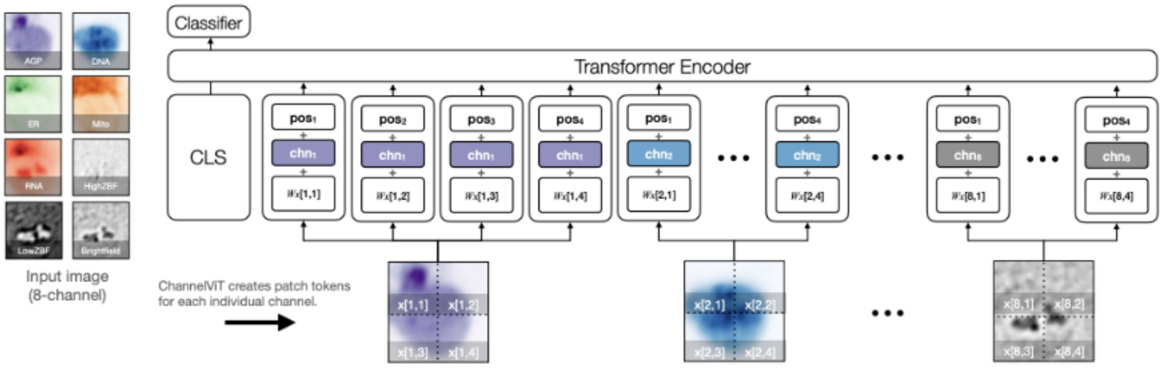


Figure 12 Example of ChannelViT processes cell image with five channels

Unlike standard ViT, ChannelViT explicitly accounts for the unique information contained in each channel and integrates it intelligently. By leveraging inter-channel dependencies, ChannelViT often achieves higher accuracy in tasks where different channels provide complementary information, such as detecting PV-cell damage using combined thermal, optical, and electroluminescence images. ChannelViT can handle images with varying numbers of channels, making it flexible for your use case of combined 3-image, 3-channel data (total 7 channels). The attention mechanism enables ChannelViT to focus on relevant patterns, even in scenarios where some channels might have noise or irrelevant details.

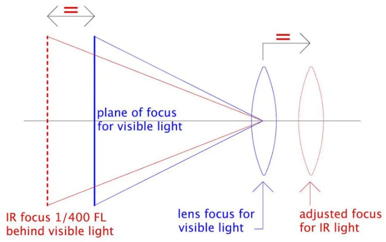
3 Activities and results

3.1 WP2 Instrument setup and calibration and data acquisition

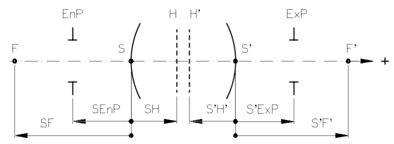
3.1.1 Instrument Setup and Calibration

Regarding the location of the imaging setup, the first solution considered has been to use the existing visual inspection setup used in the SUPSI photovoltaics laboratory (PVLab), consisting of a horizontal table with 1000 lux illumination with the camera placed above it, but the interference with ongoing lab services, the stray light and the variable environmental conditions, have led to the multispectral setup being placed in a dedicated room, which was finally identified in the rear part of the dark room, with the advantage of a controlled environment and the absence of reflections thanks to the special black paint used to reduce them even in the IR range. The camera distance can be varied according to the dimensions of the module, in order to completely fill the sensor and make the most of the camera's high resolution.

As previously mentioned in Section 2.1, the Phase One iXM-MV150 camera, with a 150 Mpx sensor, equipped with a Linos 60 mm f/4 lens and manual focus is used. Particular attention has been paid to creating an effective and reproducible setup. For visual inspection (VI) and UV fluorescence (UVf), the focus of the Linos lens has been precisely calibrated and fixed at a common point, while for the electroluminescence (EL) image, i.e. infrared light, a different set point has been verified, to optimize accuracy: the two settings were fixed as points on the focus ring to be able to quickly reproduce the same conditions for each module. As shown in Figure 13, this small difference in focus is mainly due to the phenomenon of refraction.



Specification		ON	7721-9011
image circle max. (mm)	70.9	working distance (mm)	334 ... ∞
focal length f' (mm)	60.6	interface	M45x0.75
magnification β' [range]	-0.033 [-0.2 ... 0]	filter thread	M49x0.75 / M40.5x0.5
spectral range λ (nm)	400 - 750	weight (g)	240
schematic diagram	*) in air	design includes CCD cover glass:	yes / 2mm N-BK7



SF (mm)	-33.6	f-stop	Ø EnP	Ø ExP
S'F' (mm) *)	42.3	4	14.4	14.5
HH' (mm) *)	8.5	5.6	10.8	10.9
SH (mm)	26.9	8	7.6	7.6
S'H' (mm) *)	-18.3	11	5.5	5.5
SEnP (mm)	26.6	16	3.8	3.8
S'ExP (mm) *)	-18.6	22	2.8	2.8

Figure 13: near IR focus shift vs VIS and UVf[41] and specs of the Qioptiq LINOS lens, optimized for 400-750nm (VIS).

For the visual inspection setup, two flash units have been purchased with dedicated rectangular soft boxes to evenly distribute the light over the wide surfaces of the modules. The flashes have been precisely positioned at a 45° angle to the surface of the module to avoid glass reflections (Figure 14a). A color checker table is positioned on top of the modules (Figure 14: Schematic view of (a) visual inspection (VI) and (b) UV-fluorescence (UVf) setups.), to grant full traceability of the color space, so to be able to quantify properly e.g. the yellowing rate of materials or the oxidation of silver-plated busbars.

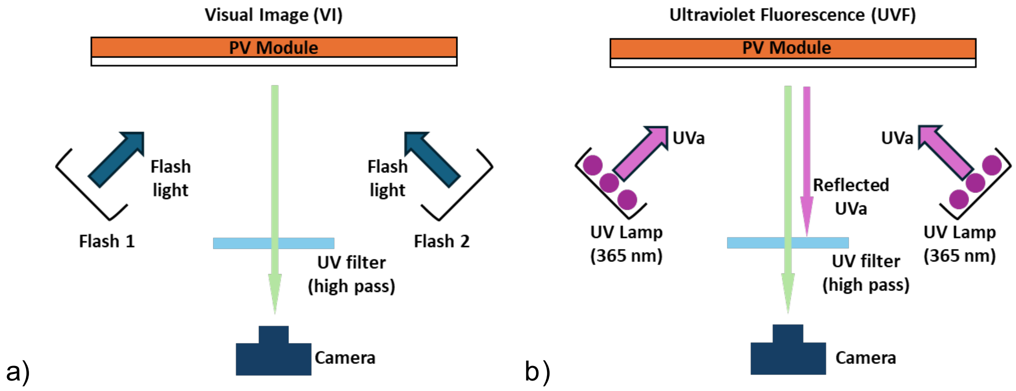


Figure 14: Schematic view of (a) visual inspection (VI) and (b) UV-fluorescence (UVf) setups.

For the UV-fluorescence (UVf) imaging setup, a similar approach was employed: four UV lamps (emitting at 365 nm) were positioned at a 45° angle, and a UV filter was actively used to eliminate UV reflections (Figure 14b). The UV filter optimizes the image quality to filter the UV radiation from the spots at 365 nm with a high pass filter that cuts at 385 nm, as can be seen in Figure 15. UVf imaging setup is crucial because UV light can be absorbed by degradation by-products of polymers, such as encapsulants, which may exhibit fluorescence effects. When UV light (high energy) is absorbed by these materials, they emit photons in the visible range (lower energy), enabling the detection of historically hot areas and defects, such as moisture penetration, cracks in the solar cells (thanks to bleaching due to oxygen ingress especially for glass/backsheet modules), and other potential issues. This type of measurement depends on the polymer type, as not all polymers produce degradation by-products with fluorescence effects (i.e. polyvinyl butyral PVB).

a) b)

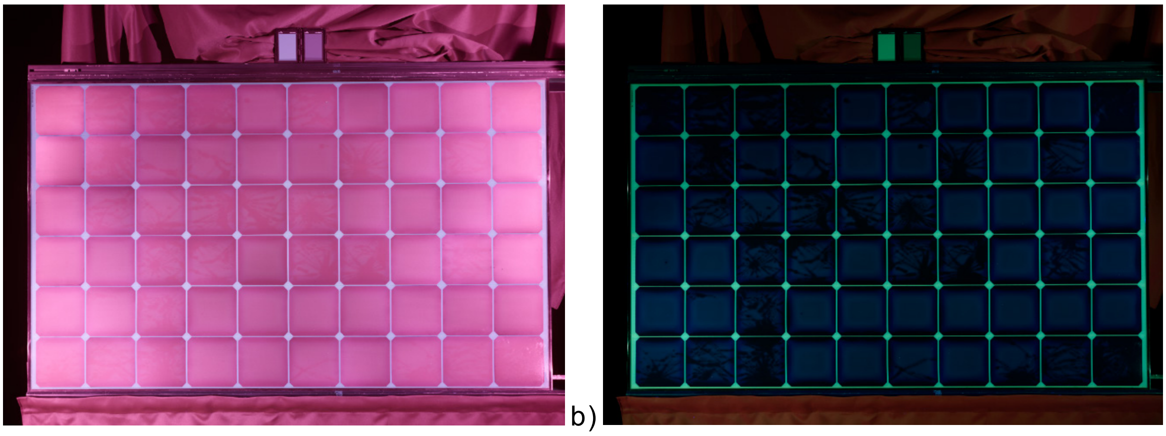


Figure 15: UV-fluorescence (UVf) image (a) without and (b) with UV filter (high pass).

The electroluminescence phenomenon is based on the impression of the digital sensor by near IR photons coming from the use of the solar cells as IR LEDs: a power supply is connected to the PV module under inspection (positive to positive, negative to negative) so to bias the circuit in order to allow current flow and the consequent IR emission from the cells. The setup, illustrated in Figure 16 for what concerns the camera and filtering means, adds a power supply to the other two acquisition methodologies, being electroluminescence the only one using direct light coming from the samples and not reflected or re-emitted irradiation. The accurate focus of electroluminescence images is far more complicated than the visible ones, due to the low level of light and higher noise in the so called “live view” mode for long-term exposures, where a digital magnified area is used to fix the optimal sharpness.

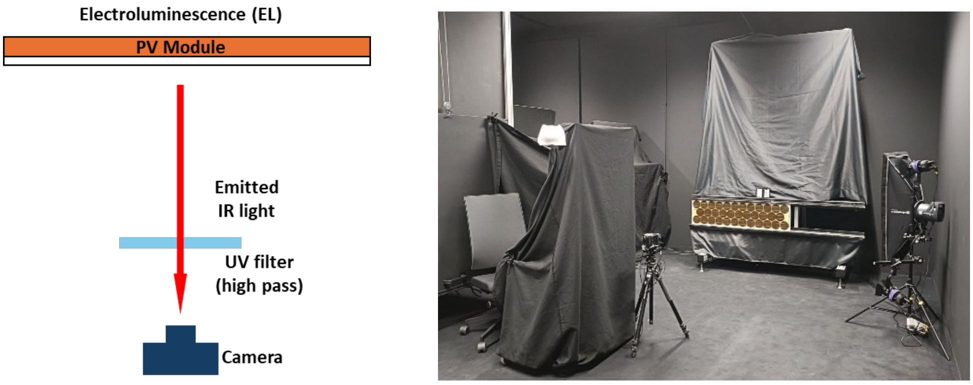


Figure 16: Schematic view of electroluminescence (EL) setup, overall view of the setup in the dark room.

The logistics have also been carefully managed, to optimize the handling of hundreds of samples and to maintain the best repeatability of the images.

A set of remotely controlled switches implemented (Figure 17), in order to efficiently semi-automate the process, activating separately the different shooting setups, that is:

1) The two flash units for VI images.
2) The power supply for EL images.
3) The 4xUV lamps for UVf images.
4) A spotlight to change modules.

Figure 17: Remotely controlled electric switches for rapid selection of the light sources during data acquisition.

3.1.2 Data Acquisition

Data acquisition – including VI, EL, and UVf imaging, as well as electrical performance measurements of the modules at standard test conditions (STC, 1000 W/m² and 25 °C) measured in SUPSI PVLab (ISO17025 accredited for PV module testing) using a class A+A+A+ solar simulator (Pasan IIIb) – was carried out on aged modules originating from other activities. Each dataset includes visual (VI), electroluminescence (EL), and ultraviolet fluorescence (UVf) images, as well as electrical performance measurements after ageing. Before ageing, either measured electrical values or datasheet values were used. The two primary datasets were the TISO and Catamarano. Additional datasets were later created and incorporated into the project, as described in Section 3.2.

Dataset-1: TISO (172 modules) [49, 50]

The first set of modules measured with the multispectral imaging setup consists of the TISO modules. A total of 172 modules were measured across all three spectral ranges (EL, UVf, and VI; see Figure 18). Each module contains 35 cells, resulting in a total of 6020 cells analyzed. Electrical performance measurements were also conducted for each module

a)

b)

c)

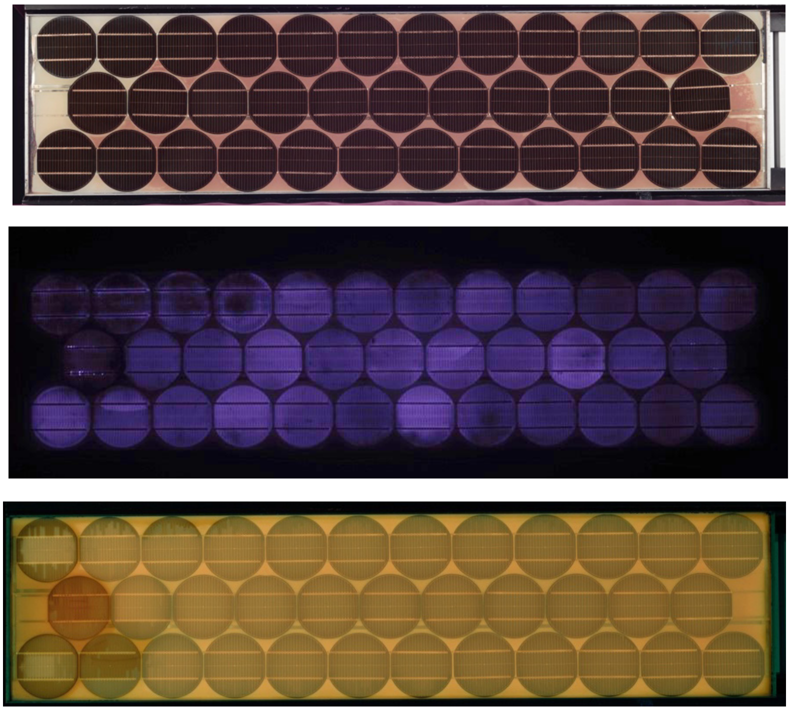


Figure 18: (a) Visible (VI), (b) electroluminescence (EL), and (c) ultraviolet fluorescence (UVf) images of a TISO module.

Dataset-2: Catamarano (8 modules) [53]

Several IBC modules, originally installed on a catamaran that completed a world tour, were measured after being dismantled. Their multispectral images (see Figure 19) and power measurement data were also incorporated into the project.

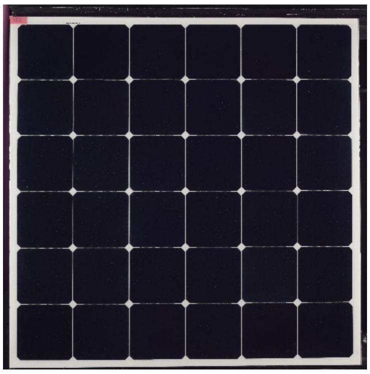
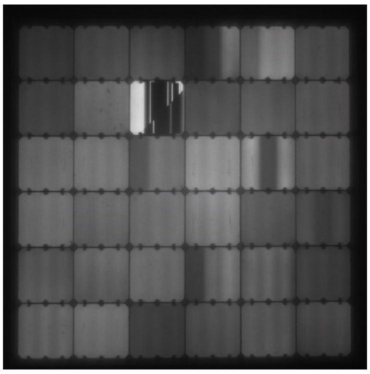
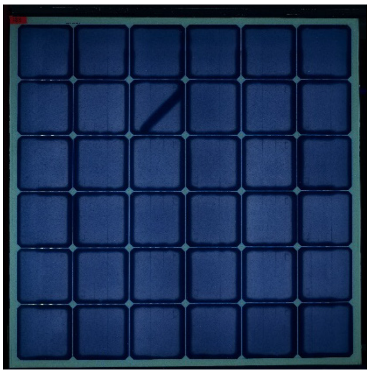
a)  b)  c) 

Figure 19: (a) Visible (VI), (b) electroluminescence (EL), and (c) ultraviolet fluorescence (UVf) images of a Catamarano module.

3.2 WP3 Accelerated Testing of New Technologies

Measurements of various PERC half-cut, PERC third-cut, PERC shingled, HJT, IBC, and TOPCon modules were conducted and added to the project's dataset pool (Table 2). Data collected from the ACHILLES and CICLO 14 (ATTRACT and REBIPV) projects was particularly valuable. Within these projects, several tests – including thermal cycling, hot-spot endurance, hail testing, and outdoor accelerated aging using heat blankets and shadow masks – were performed. In addition, further modules from PVVintage (within PVDECTECT project) were incorporated into the dataset. Below there are more details on the datasets.

Table 2: Summary of the datasets which have multispectral images and electrical performance data.

Data Source	Data	Ageing Method	VI	EL	UVf	Electrical data before ageing	Electrical data after ageing
TISO	172 modules	Outdoor	Y	Y	Y	Y	Y
Catamarano	8 modules	Outdoor	Y	Y	Y	Y	Y
Achilles	25 modules	Outdoor + Indoor (Hail test)	Y	Y	Y	Y	Y
Ciclo 14	32 modules	Outdoor (Open-rack, BIPV roof and BIPV facade) + Indoor Accelerated Ageing (Standard and Accelerated Thermal cycling and Standard and Accelerated Hot spot)	Y	Y	Y	Y	Y
PVVintage	8 modules	Outdoor	Y	Y	Y	Y	Y
23-089/A/1	1 module	Outdoor	Y	Y	Y	Y	Y
24-128-A	3 modules	Outdoor	Y	Y	Y	Y	Y
24-P10-A	6 modules	Outdoor	Y	Y	Y	Y	Y
25-019-A	3 modules	Outdoor	Y	Y	Y	Y	Y
25-P03-A	10 modules	Outdoor	Y	Y	Y	Y	Y

Dataset-3: Achilles (25 modules)

PV modules of different technologies have been damaged within the framework of the project ACHILLES. Hail tests have been performed with different ice ball diameters and energies. A total of 25 modules have been measured at different testing stages. In addition, several modules were collected directly from the field after the hailstorm of the 25th August 2023 in Locarno. Figure 20 shows an example of multispectral images.

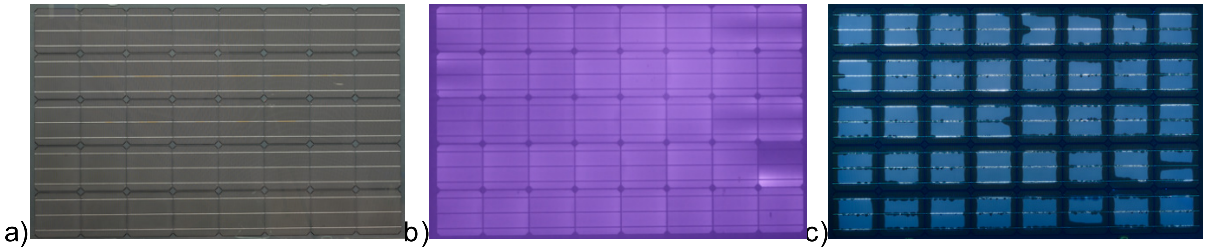


Figure 20: (a) Visible (VI), (b) electroluminescence (EL), and (c) ultraviolet fluorescence (UVf) images of an Achilles module.

Dataset-4: Ciclo 14 (ATTRACT and REBIPV) (32 modules) [49]

PV modules of various technologies (Al-BSF, PERC half-cut, PERC shingled, PERC third-cut, HJT, IBC, TOPCon, and coloured PERC full-cell) underwent a series of indoor accelerated aging tests, including thermal cycling and hot-spot endurance, as well as long-term outdoor monitoring in different mounting configurations (open-rack, BIPV-insulated, and BIPV-ventilated), within the framework of the Ciclo 14, which includes the ATTRACT and REBIPV projects. The indoor thermal cycling and hot-spot endurance tests were conducted at elevated temperatures, beyond the standard test conditions, in addition to the standard test conditions. Additionally, some modules were installed in outdoor accelerated aging setups, using heat blankets and shadow masks in addition to outdoor monitoring at open-rack and BIPV mounting configurations. Several reference modules for each type were also kept indoors as non-weathered samples for comparison. Figure 21 shows an example of the multispectral images of a module which was installed outdoors with a shadow mask in BIPV insulated configuration (no rear side ventilation).

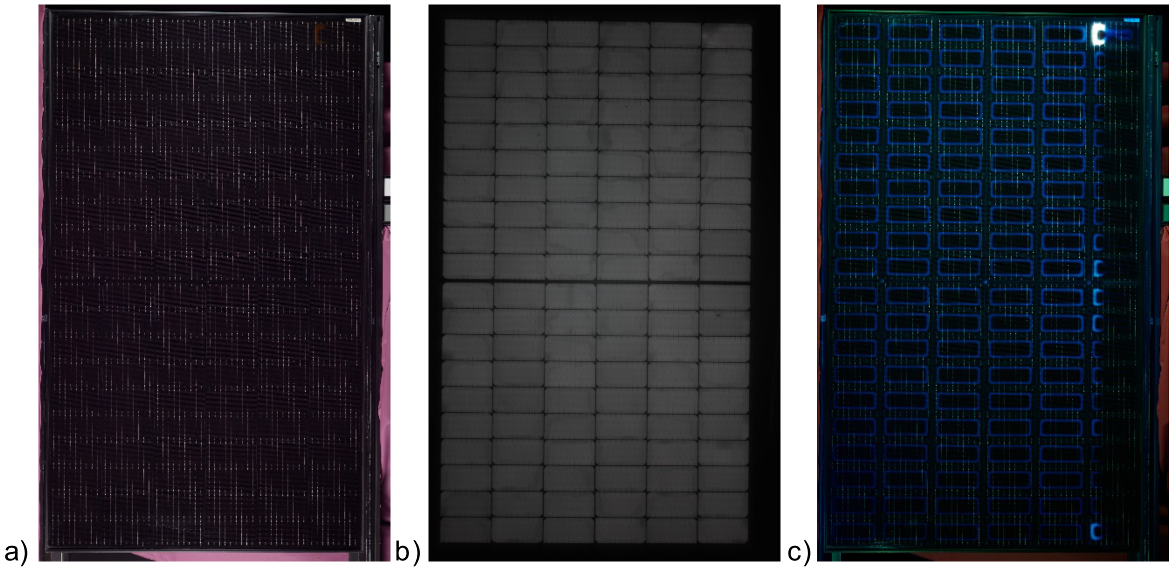


Figure 21: (a) Visible (VI), (b) electroluminescence (EL), and (c) ultraviolet fluorescence (UVf) images of a Ciclo 14 module.

Dataset-5: PVVintage (within PVDETECT project) (8 modules) [52]

Several ARCO and Siemens M55 modules (Al-BSF), which had been installed at various altitudes across Switzerland, were measured after approximately 30 years of outdoor operation. Figure 22 shows the images of one such module.

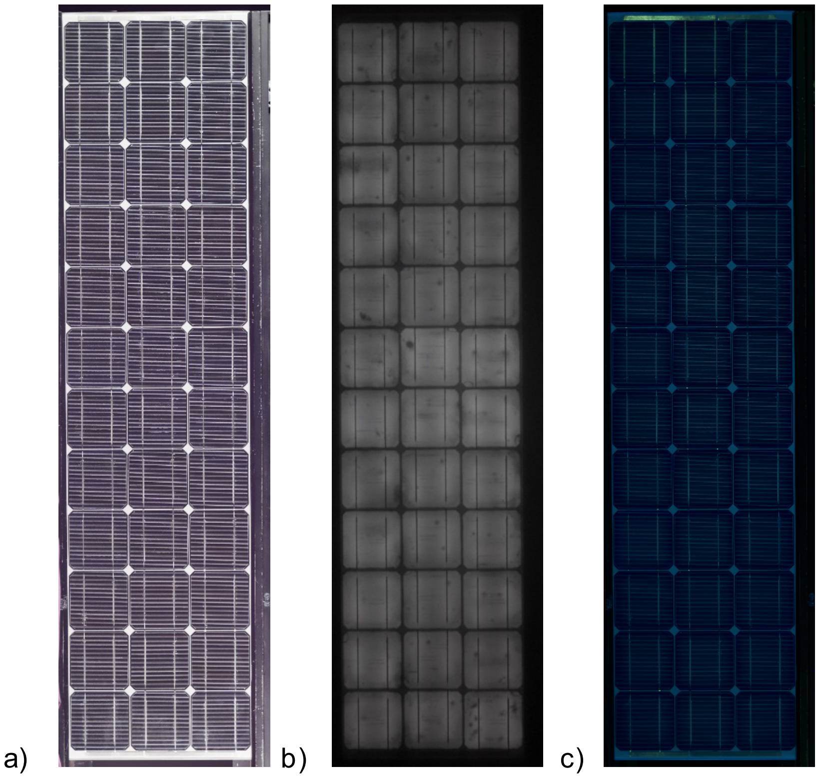


Figure 22: (a) Visible (VI), (b) electroluminescence (EL), and (c) ultraviolet fluorescence (UVf) images of a PVVintage module.

Other datasets (23 modules)

These datasets of the project consisted of five different module technologies, each of which is presented separately in Table 2. All modules are field-aged with varying outdoor operating durations, and most of them are based on PERC technology. Figure 23 shows the images of one such module.

a) b) c)

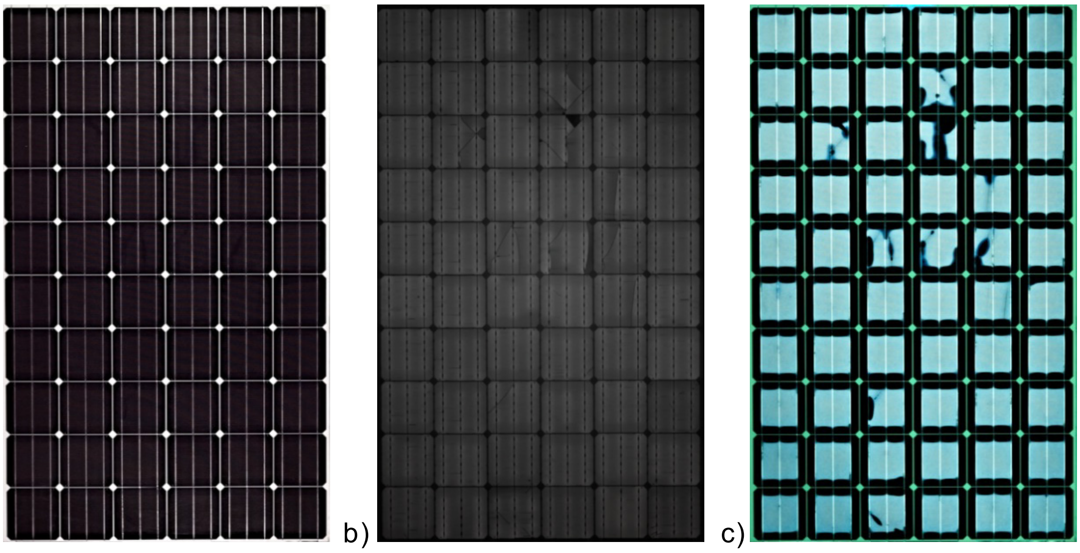
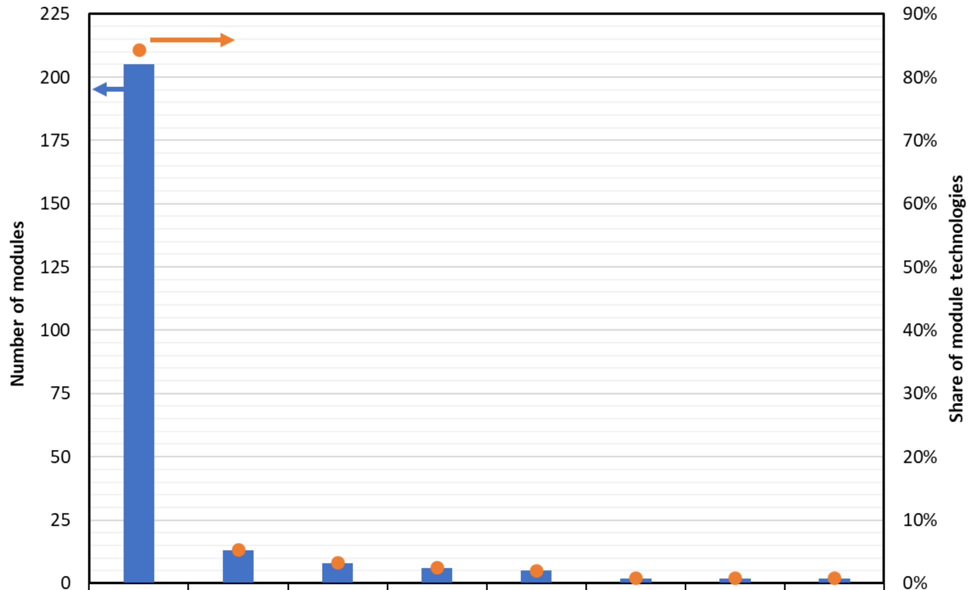


Figure 23: (a) Visible (VI), (b) electroluminescence (EL), and (c) ultraviolet fluorescence (UVf) images of a module from other datasets.

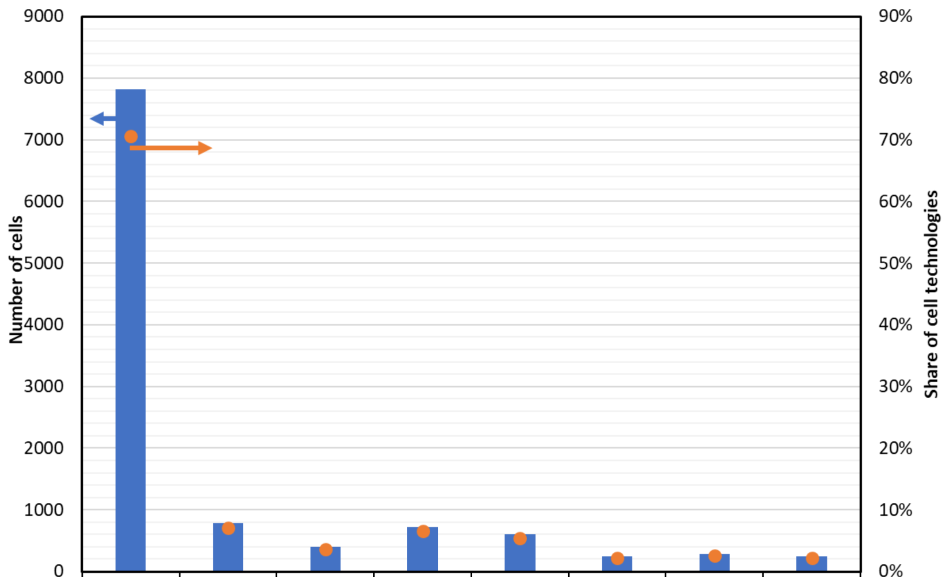
Figure 24 shows the distribution of module technologies within the dataset pool (Table 2). In total, 241 modules were included. Given the requirement for complete data across all modalities (VI, EL, UVf, and electrical performance), assembling such a large dataset was challenging. Overall, data from 377 modules was available, including modules from previous years; however, modules missing at least one measurement mode could not be used. Statistically, the majority of the modules in the dataset are Al-BSF (84%, 205 modules). The second most common technology in the dataset is PERC, accounting for 7% (18 modules), although in Figure 24 this category is further divided by wafer format (i.e. full-cell, half-cut, 3-cut) or contacting technology (i.e. shingled). Since the adoption of TOPCon technology has progressed very rapidly, only two TOPCon modules were available for this project. End of this work package, raw dataset, including VI, EL and UVf images and electrical performance, of both old and new technologies were delivered for activities in WP4 and WP5.



	Al-BSF	IBC	PERC Full-cell	PERC Half-cut	HJT	PERC 3-cut	TOPCon	PERC Shingled
Number of modules	205	13	8	6	5	2	2	2
Proportion of modules	84%	5%	3%	2%	2%	1%	1%	1%

Figure 24: Statistical summary of the PV module technologies used in the project.

Since 172 of 205 Al-BSF modules are TISO modules, each containing only 35 cells, the proportion of Al-BSF cells relative to the total number of cells is a little lower. Modern commercial modules typically contain at least 120 cells, which changes the weighting when calculating technology share based on cell count rather than module count. As shown in Figure 25, when considering the number of cells, the share of Al-BSF technology drops to 71% (1,964 cells). Although Al-BSF remains the most represented technology, PERC cells constitute 14% (1,596 cells) of all data.



	Al-BSF	IBC	PERC Full-cell	PERC Half-cut	HJT	PERC 3-cut	TOPCon	PERC Shingled
Number of cells	7812	780	396	720	600	240	288	240
Proportion of cells	71%	7%	4%	7%	5%	2%	3%	2%

Figure 25: Statistical summary of the solar cell technologies used in the project.

3.3 WP4 Data Preprocessing and annotation

3.3.1 Segmentation

The segmentation of images from new technologies should be automated as much as possible using Segment Anything. PV panels differ significantly in structure and size depending on the technology. Therefore, the PV panel must first be identified in the image to focus only on the relevant image section. For this purpose, an algorithm analyzes the electroluminescence image to detect the brightest and darkest pixels, in order to determine the contours of the PV panel. Our pipeline then utilizes a two-stage approach. First, we use SAM's automatic mask generation to infer the panel's grid layout. This step leverages a novel peak-detection method on the centroid coordinates to accurately determine the number of rows and columns, thereby eliminating manual input.

A known limitation of using a generalized segmentation model like SAM in an automated mode is its tendency to produce redundant and overlapping masks for the same object. For a structured grid like a photovoltaic panel, this leads to inefficient and unreliable segmentation. To address this challenge, we developed a novel, two-stage pipeline that combines SAM’s powerful zero-shot capabilities with a classical signal processing approach.

Instead of directly using SAM for cell segmentation, we first leverage its automatic mode to identify the centroids of potential cells. By analyzing the spatial distribution of these centroids, we were able to detect the periodic peaks in their coordinates using SciPy find_peaks. This method allowed us to accurately and robustly infer the precise row and column structure of the PV panel, a critical step that eliminates the need for manual grid estimation.

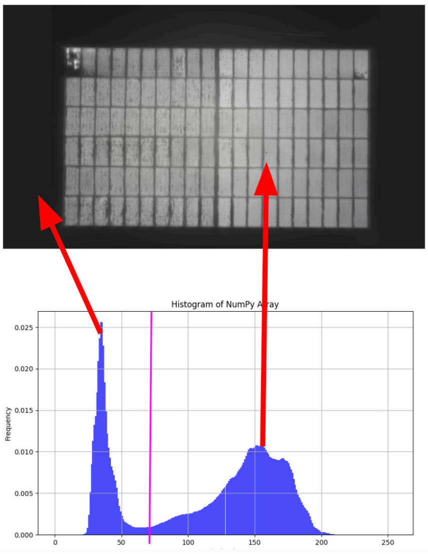


Figure 26: Determine the contour of the PV-Panel based on the brightness histogram.

Following segmentation, the identified regions are post-processed to enhance the quality and usability of the results. This includes the alignment of segmented images and the adjustment of the bounding boxes to account for the variability in panel size, perspective distortion, and slight misalignment. These refinements ensure that the extracted cell regions are properly standardized for subsequent analysis or model training.

Second, SAM's predictor is applied in a guided manner, using the inferred grid coordinates to constrain the segmentation process. This approach improves precision by constraining the segmentation process within the expected cell boundaries, resulting in cleaner and more consistent mask generation.

This refined approach transforms a potential weakness of the base model into a key advantage. By guiding SAM with the empirically determined grid coordinates, we ensure that the subsequent segmentation produces precise, non-overlapping masks that are constrained to the actual cell boundaries.

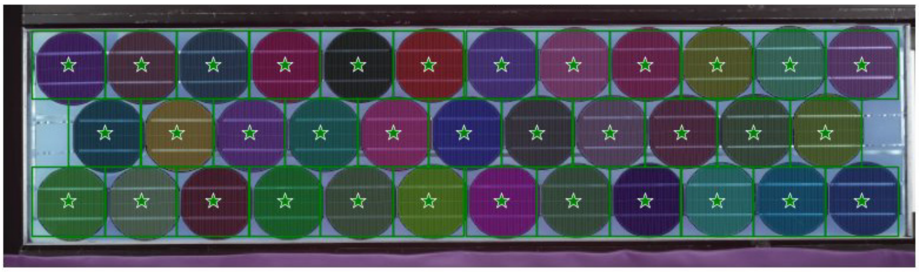
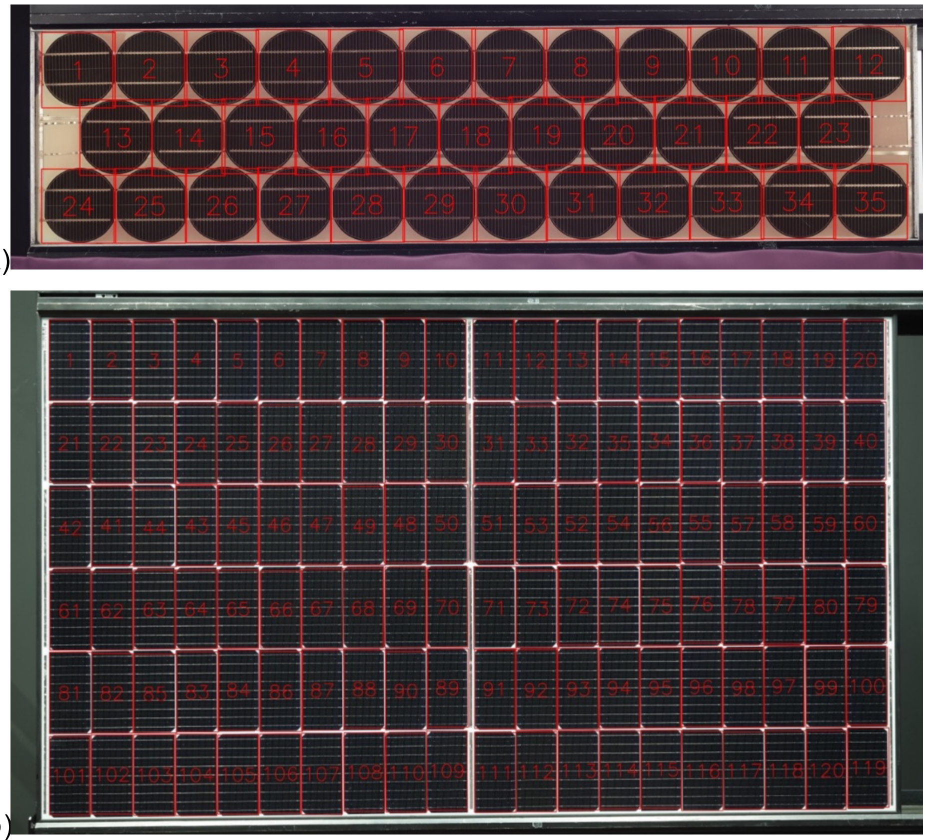


Figure 27: Automatic detected mask for Segment Anything.

This mask, together with the image section, is then passed to Segment Anything to segment the exact positions of the cells. shows two example panels whose segmentation was performed using this model. It can be observed that the cells for both the TISO panel (Figure 28) and the Achilles panel (Figure 28b) were segmented relatively accurately.

a)



b)

Figure 28: (a) Overview of a segmented TISO image. (b) Overview of a segmented Achilles image

3.3.2 Unsupervised labelling

French et al. Dataset

The results of the pipeline described in section 2.4 for the first dataset is depicted in Figure 29. Plots show that a completely unsupervised pipeline is effective in separating good cells from corroded ones. Due to the very small number of cracked cells, this class cannot be separated from the one containing good cells. HDBSCAN is effective in recognizing the two separate clusters.

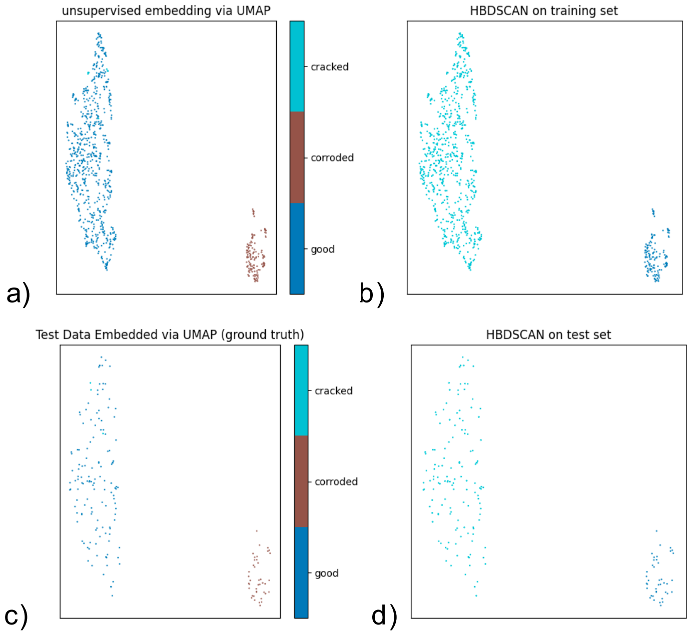


Figure 29: (a) two-dimensional representation of the UMAP embeddings of the French dataset. The different colors represent the different labels. (b) HDBSCAN clusters. (c) Test data embedded in the same space as (a). (d) HDBSCAN clusters on test data.

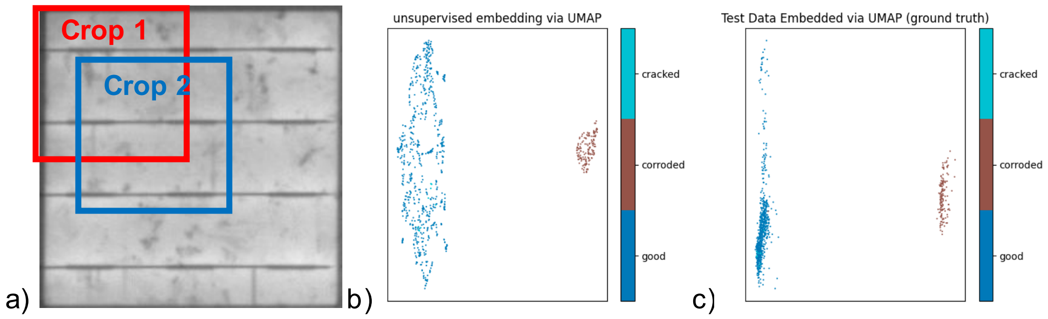


Figure 30: (a) Representation of the cropped images. (b) UMAP embeddings for the first set of cropped images. (c) Representation of the second set of cropped images projected onto the UMAP embeddings.

In order to test the robustness of the method with respect to translation, a test was performed with different crops of the images. All the crops have the same lateral sizes but are shifted with respect to one another (Figure 29**).** The embeddings are evaluated on the first set of crops Figure 31**:** (a) Two-dimensional projection of the 20-dimensional embeddings and clusters found on the full dataset (global average pooling). (b-g) Examples of cells coming from the different clusters. (h) Examples of cells not assigned to any cluster.0b**)** and the second set of crops is projected onto the same space (Figure 30c). The images of the second set classes (crop 2) belonging to different classes are still well separated, proving that the method is robust with respect to translational shifts, which is to be expected due to the employment of convolutional neural networks in the preprocessing. The test has been repeated for different offsets of the crops, as well as with crops coming from a different set of images (test set). In all the experiments, the good and corroded images are well separated.

2. Chen, Crack segmentation dataset

The unsupervised labelling routine has been applied to the second dataset of EL images. As several technologies are represented in the dataset, a first round of UMAP effectively separates the different technologies into six different clusters (Figure 31).

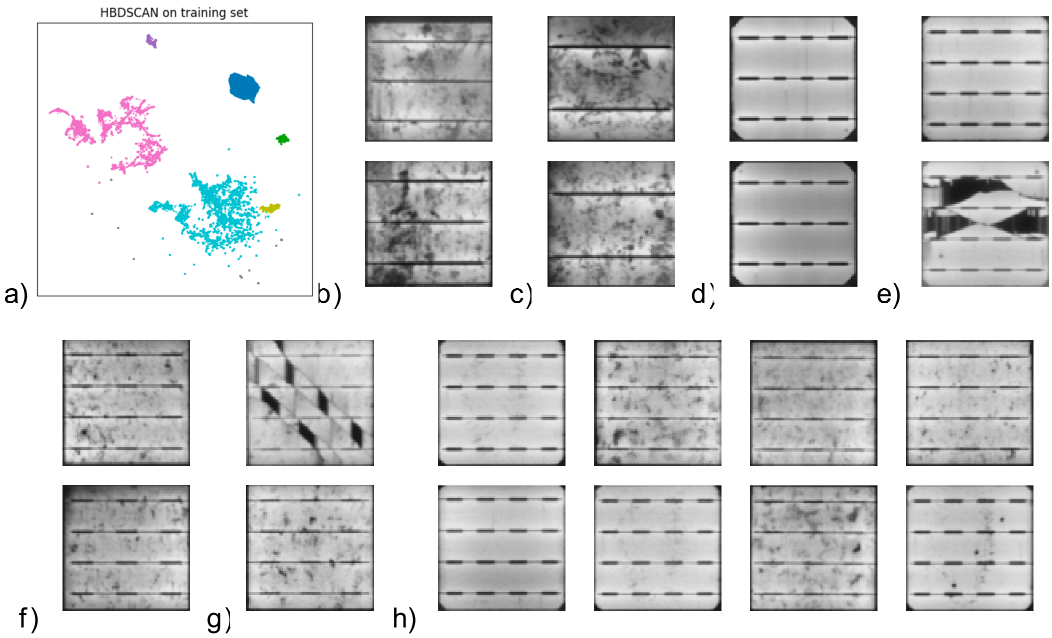


Figure 31: (a) Two-dimensional projection of the 20-dimensional embeddings and clusters found on the full dataset (global average pooling). (b-g) Examples of cells coming from the different clusters. (h) Examples of cells not assigned to any cluster.

When global average pooling was used during preprocessing, 318 images (approximately 1% of this dataset) were not assigned to any cluster. This is due to the fact that a tradeoff has to be found when choosing the HDBSCAN parameters in order not to have a significant number of very small clusters. Thanks to the Dash application, it is possible to visualize the small groups of images not assigned to any cluster, and it is therefore possible to assign them to a specific technology. Figure 30a-g show two images from each technology cluster, and Figure 31h depicts as some examples of images not assigned to any cluster.

The same analysis has been performed using global max pooling in the preprocessing. The results are shown in Figure 31a. 1607 images (around 5.4% of the dataset) were not assigned to any cluster. However, almost all of these images could be assigned to the same HDBSCAN cluster. 90 images formed a separate unidentified subcluster, and they could also be assigned to one cluster. Using global max pooling leads to the formation of eleven clusters in the UMAP embeddings. Some of the different clusters belong to the same technology, but different failures have already been separated from each other. (e.g. Figure 32).

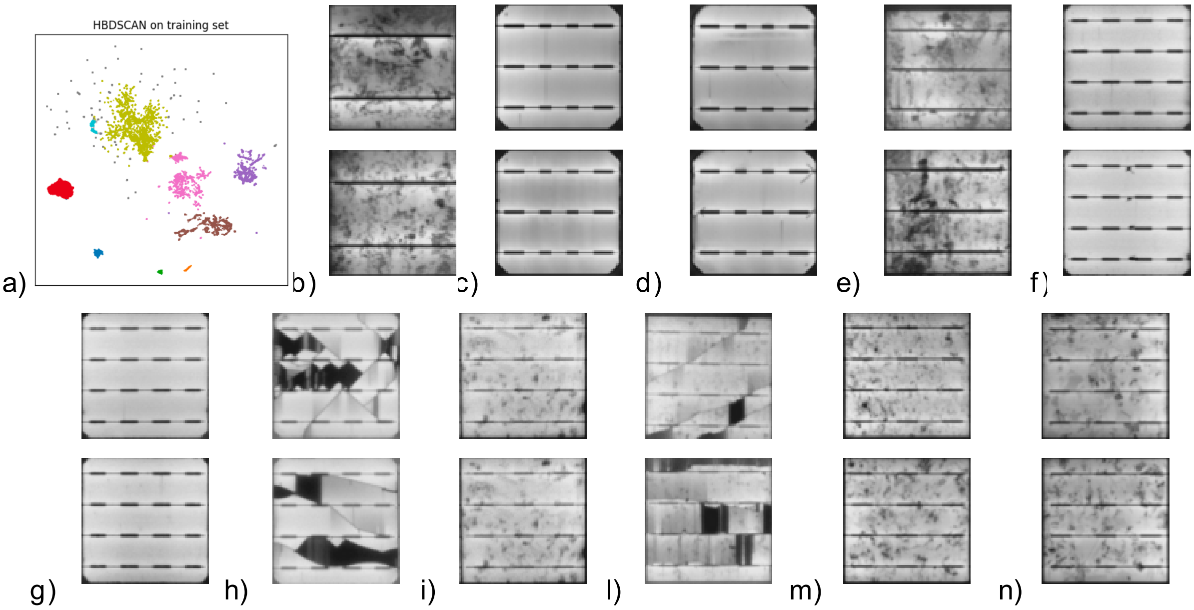


Figure 32: (a) Two-dimensional projection of the 20-dimensional embeddings and clusters found on the full dataset (global max pooling, n_neighbors=50). (b-n) Examples of cells coming from the different clusters.

The analysis routine has been repeated using the images of the technology represented by Figure 32f-h (10043 images) to evaluate 10-dimensional embeddings and to evaluate if the routine is effective in separating good and defected cells. 460 of these images also have labels, which are only used for the evaluation. Label 0 (blue) corresponds to good cells, while labels 2-6 correspond to different failures such as cracks, crosses and dark areas on the cells, as well as combinations of them. Figure 33a shows that good cells are quite well separated from defected cells, apart from a small number of cells presenting cracks or crosses. Some defects or combinations of defects are clustered together, as it can be seen from Figure 33b. The orange cluster in Figure 33b has been analyzed in Figure 34 with the Dash application. After zooming into this region, the cells belonging to this cluster are selected and visualized. The screenshot only shows some of the cells visualized. All cells of this cluster present the cross failure, and some of them also present cracks and darker regions. No intact cell is represented in this cluster.

It is worth noting that the pipeline fails on multi-crystalline Si technologies, which however will not be the focus of the present project.

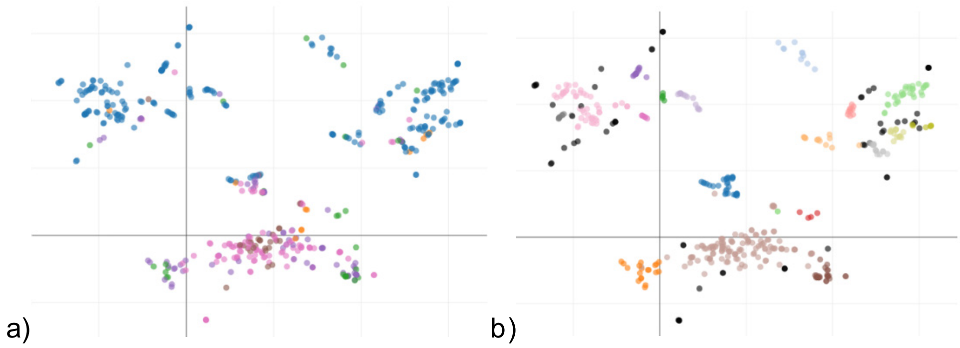


Figure 33: (a) Two-dimensional projection of the 10-dimensional embeddings found for one technology (n_neighbors=50). Only labelled images are represented. The color blue represents good images, while other colors represent cells with cracks, crosses, dark areas or a combination of those. (b) HDBSCAN clustering applied to the embeddings.

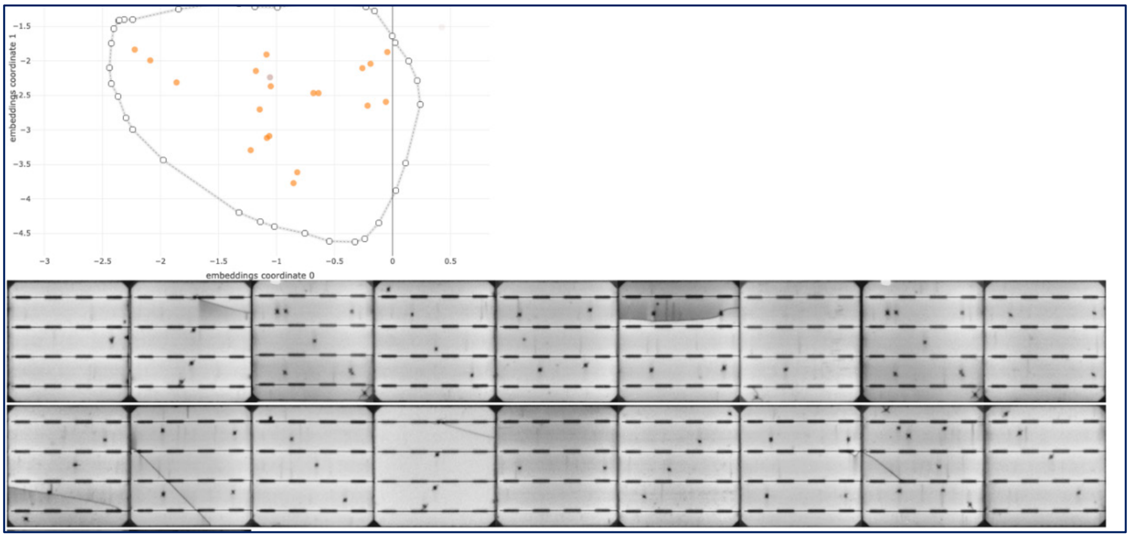


Figure 34: Screenshot of the Dash application. The area relative to the orange cluster (bottom links in Figure 33b) has been zoomed, and the cells selected and visualized.

3. *Infinity project*

The preprocessing pipeline has been applied to the Infinity dataset as well. The embeddings have been evaluated on 90% of the dataset, and the rest of the images have been projected in the same embedding space. Several clusters of images have been identified, as shown in Figure 35a. Some defects have been well separated (e.g. Figure 35b and e), while some contamination is present in other clusters (e.g. Figure 35f and l). It is however worth noting that images with similar features tend to be close to one another in the embedding space, which could reduce the labelling effort even if some manual intervention is needed. Figure 36 shows how the test images (not used for evaluating the embeddings) are separated into clusters. Even if some contamination is present, the different defects are well separated.

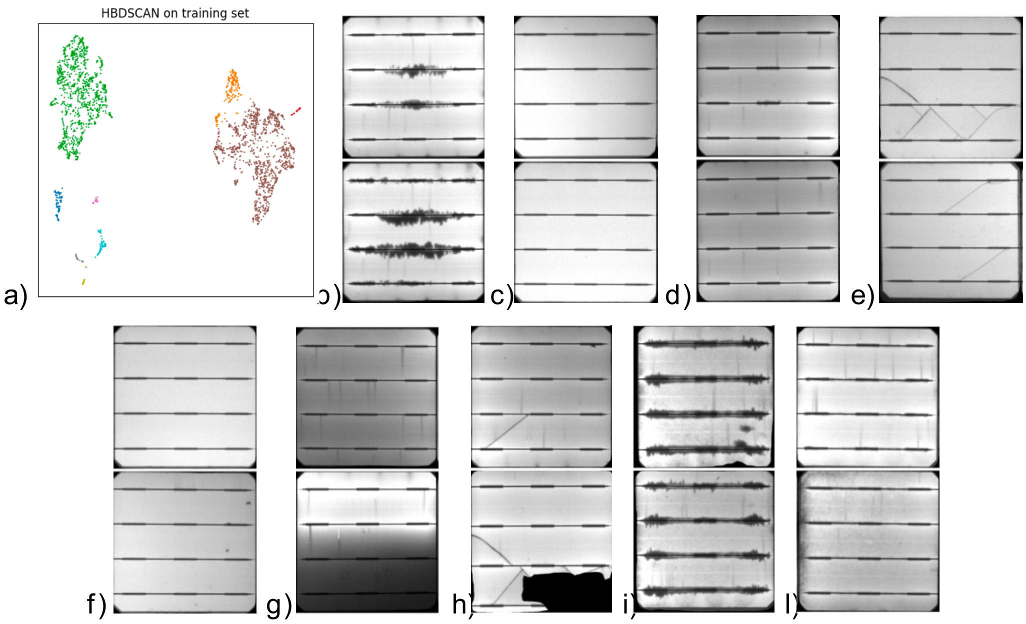


Figure 35: (a) 2-dimensional embeddings of the EL images of the infinity project evaluated on 90% of the dataset. (b – l) Examples of cells coming from different clusters.

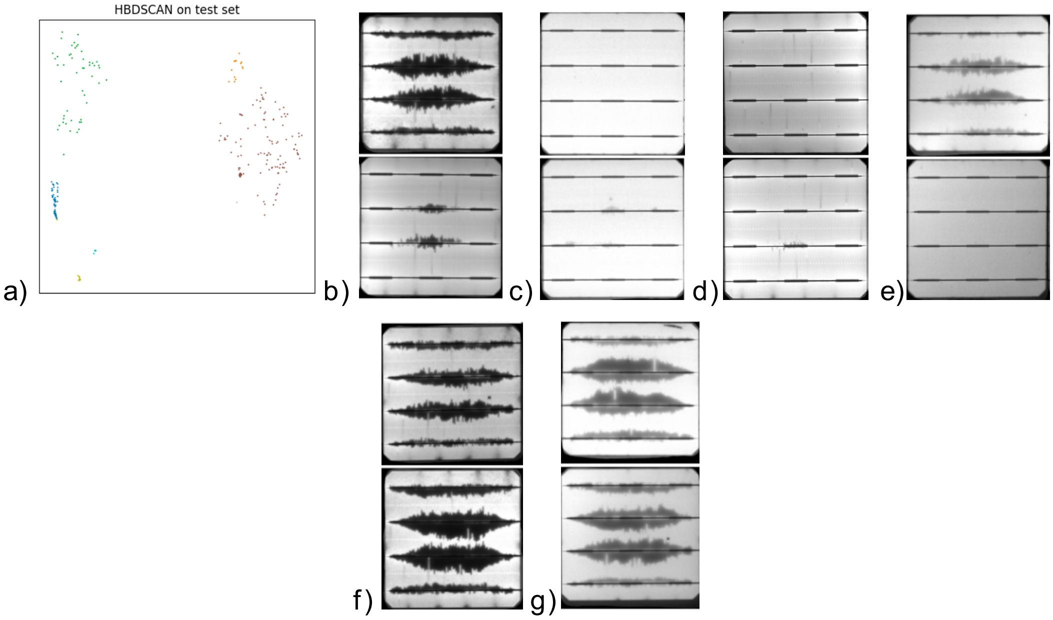


Figure 36: (a) Representation of the test images (10% of the dataset) on the 2-dimensional embeddings. (b – g) Examples of cells coming from different clusters.

4. TISO Dataset

A preliminary analysis has been carried out on the TISO multispectral images. This dataset is much more complicated than the previous one for the following reasons:

- The modules have aged naturally in outdoor conditions for many years and present therefore several types of defects.
- All three channels consist of RGB images, which are more complex than single channel gray-scale EL images.

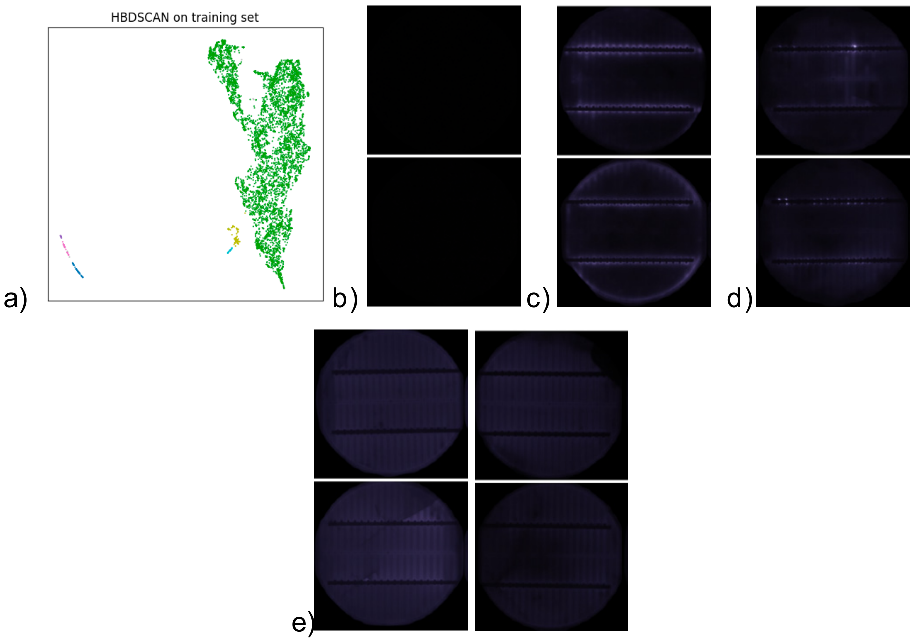


Figure 37: 2-dimensional embeddings of the TISO EL images. (b – e) Examples of cells coming from different clusters.

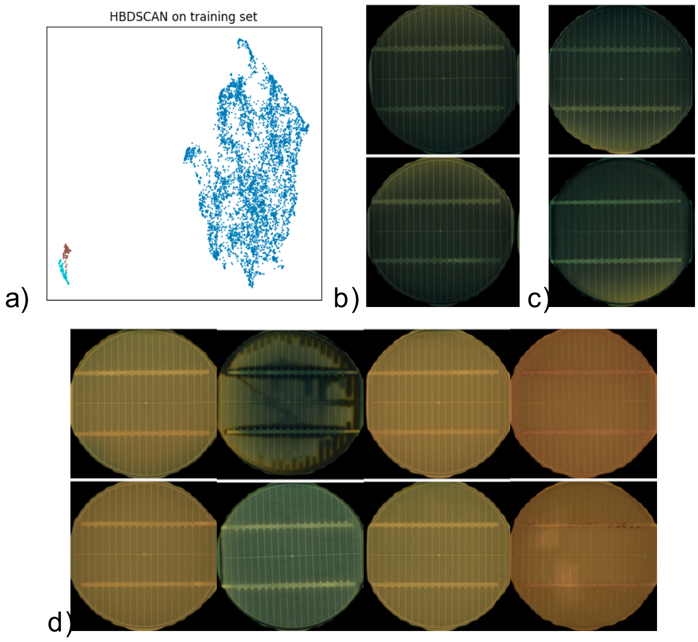


Figure 38: 2-dimensional embeddings of the TISO UV images. (b – e) Examples of cells coming from different clusters.

The segmented TISO images have been downscaled in resolution to 224x224 pixels for this preliminary analysis to reduce the computational effort. Each spectral channel has been considered separately. Figure 38a shows **Error! Bookmark not defined.**d).

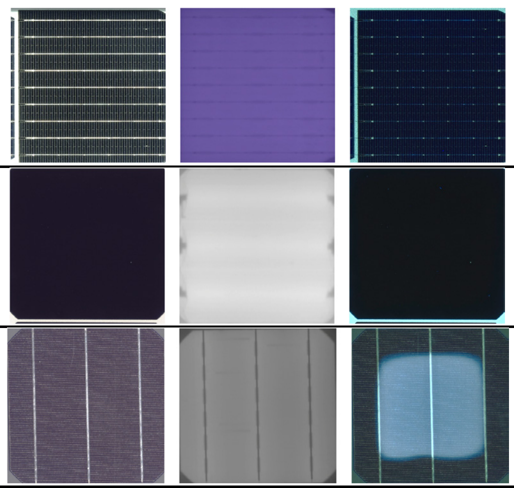
3.3.3 Annotation Frame

An annotation framework was defined prior to labeling the multispectral images. We created seven labels: *"good," "crack," "dark," "corrosion," "discoloration"* and *"delamination."*

Label-1: "good"

This label indicates that the solar cell shows no visible defects. Table 3 presents some examples.

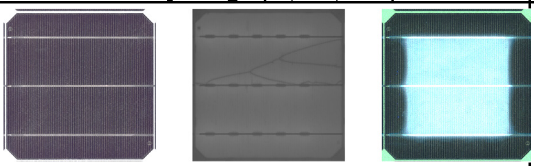
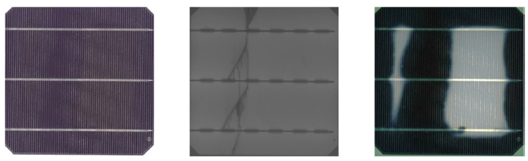
Table 3: Examples for *"good"* label.

Label	Early Stage (VI, EL, UVf)	Later Stages (VI, EL, UVf)
Good		NA

Label-2: "crack"

This label indicates that a solar cell crack is present. Table 4 shows representative examples. As expected, EL imaging reveals the crack most clearly. Depending on the module materials and layer structure, the crack may also become visible in the UVf image at later stages. In a few cases, cracks can even appear in the VI image, particularly when encapsulant discoloration is present in the non-cracked areas.

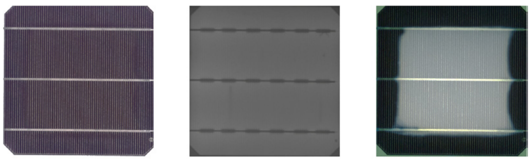
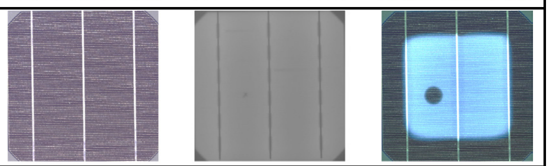
Table 4: Examples for *"crack"* label.

Label	Early Stage (VI, EL, UVf)	Later Stages (VI, EL, UVf)
Crack		

Label-3: "cross"

This label refers to small cross-shaped cracks on solar cells. These cracks are typically very small and can easily be missed during EL image inspection by person. They are directly visible in the EL image, while over time and depending on the module materials and layer structure, they may also become visible in the UVf image due to oxygen bleaching (Table 5).

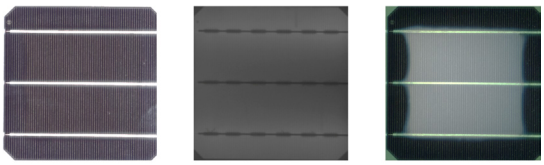
Table 5: Examples for *"cross"* label.

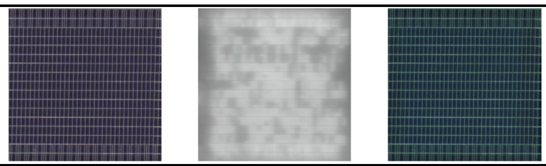
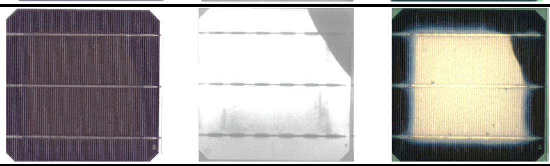
Label	Early Stage (VI, EL, UVf)	Later Stages (VI, EL, UVf)
Cross		
		

Label-4: "dark"

This label indicates the presence of a dark region in the EL image of solar cells. Since this phenomenon appears only under electroluminescence, it is visible exclusively in EL images. The dark area may be caused by a disconnected cell region, damaged busbars, broken fingers, or other similar defects (Table 6).

Table 6: Examples for *"dark"* label.

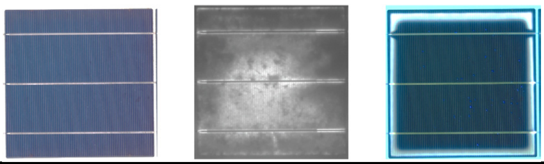
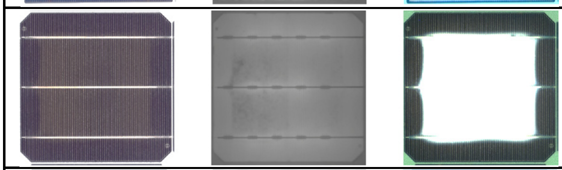
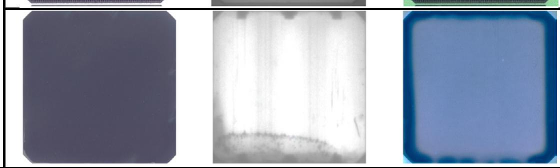
Label	Early Stage (VI, EL, UVf)	Later Stages (VI, EL, UVf)
Dark		NA

Label-5: *"corrosion"*

This label indicates the presence of corrosion on the solar cell. In some cases, corrosion can be identified in both the EL and VI (less likely) images, while in others it is visible in only one of them, depending on where it happens. UVf imaging can provide additional insights into the underlying corrosion mechanisms. For example, in Table 7, the difference between the corrosion example in the second row and the one in the first row is that the second case is caused by degradation by-products of encapsulant. This is confirmed by the presence of discoloration in the VI image and strong fluorescence signals in the UVf image. In contrast, the first-row example shows neither discoloration nor strong fluorescence (which may depend on the encapsulant material), and the darkening originates from the cell edges. These observations suggest that corrosion in that case is more likely due to moisture ingress through the backsheet and/or module edges. Thus, while VI and UVf images may not always directly indicate the presence of corrosion, they provide important complementary information that helps to better understand the underlying degradation mechanisms compared to using EL alone.

Table 7: Examples for *"corrosion"* label.

Label	**Early Stage (VI, EL, UVf)**	**Later Stages (VI, EL, UVf)**
Corrosion	  	NA

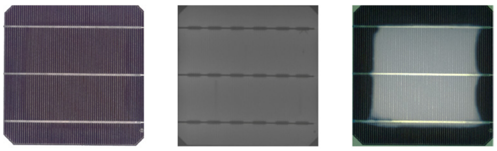
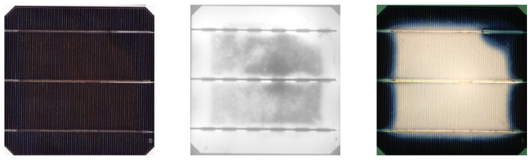
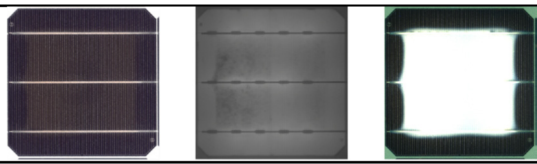
Label-6: " discoloration"

This label indicates the presence of discoloration, which is directly visible in the VI images. Since discoloration is mostly caused by degradation by-products, these compounds may exhibit fluorescence; therefore, discolored areas can appear brighter in the UVf images. Additionally, if these degradation by-products lead to corrosion – for example, acetic acid formed during EVA degradation – corrosion features may also become visible in the EL images. However, this typically occurs at later stages rather than immediately (Table 8).

When the discoloration is localized to a specific region of a cell (rather than covering the entire cell), as shown in the second and third early-stage examples in Table 8, UVf imaging is helpful for identifying the affected areas. In contrast, in cases of uniform discoloration (as in the first early-stage example), the UVf image simply appears brighter than that of non-discolored cells (not shown). Because our evaluation is performed at the cell level, this brightness difference alone is not sufficient for classification, making UVf less effective in such uniformly discolored cases.

Degradation by-products that cause discoloration – and the associated brightness increase in UVf images – can, over time, also lead to corrosion, especially if they cannot diffuse out of the module through the backsheet or edges. This corrosion may eventually become visible in the EL images, as illustrated in the later-stage examples.

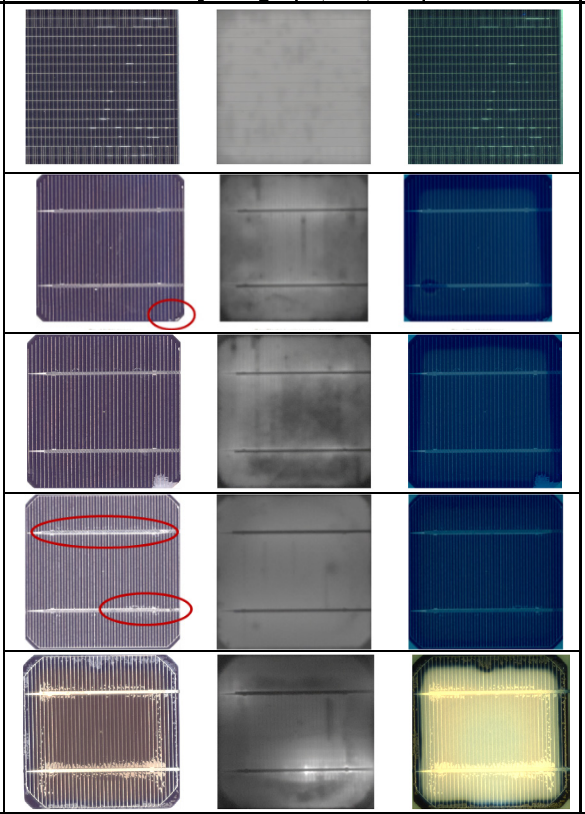
Table 8: Examples for *"discoloration"* label.

Label	Early Stage (VI, EL, UVf)	Later Stages (VI, EL, UVf)
Discol-ouration		
		
		

Label-7: "delamination"

This label indicates the presence of delamination between at least two layers of the PV module stack. At early stages, delamination is typically detectable only in the VI images (Table 9). At later stages, if the delaminated areas lead to secondary defects – such as corrosion caused by moisture ingress – they may also become visible in other imaging modalities.

Table 9: Examples for “delamination” label.

Label	Early Stage (VI, EL, UVf)	Later Stages (VI, EL, UVf)
Delami-nation		NA

As can be seen in many of the examples in the tables above, a single solar cell may exhibit more than one defect. For this reason, each cell was classified with up to three defects using the web interface described in Section 3.3.4. For example, in cases where discoloration is more severe, it may also lead to oxidation of the front metallization, which becomes visible in the EL image (red frames in Figure 39). In such situations, cells were labelled for multiple failure modes. In rare cases where a cell exhibited more than three defects, the labels were added manually in the code rather than through the web interface.

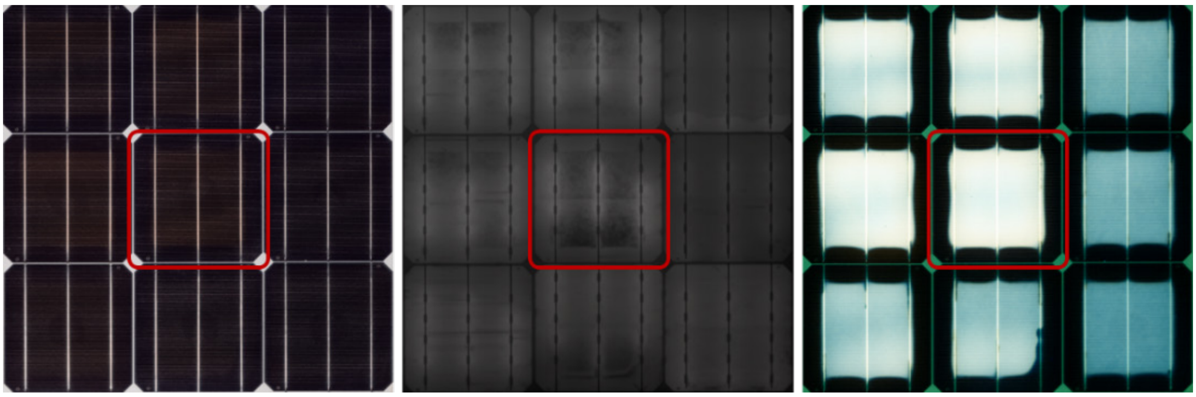


Figure 39: (left) VI, (middle) EL, and (right) UVf images of a section of a module after 10 years of operation under frequent shading. The cell within the red frame is labelled as “discoloration” and “corrosion”.

3.3.4 Web Interface

Figure 40 shows a screenshot of the Flask web application. On the left, the solar panels are listed in a tree structure, grouped by technology. The user can filter this list according to his/her wishes, for example to only look at certain technologies or only certain classifications (e.g. good, cracked cells and so on). By clicking on the name of the panel, the corresponding images (VI, UVf, EL) are displayed on the top-right area.

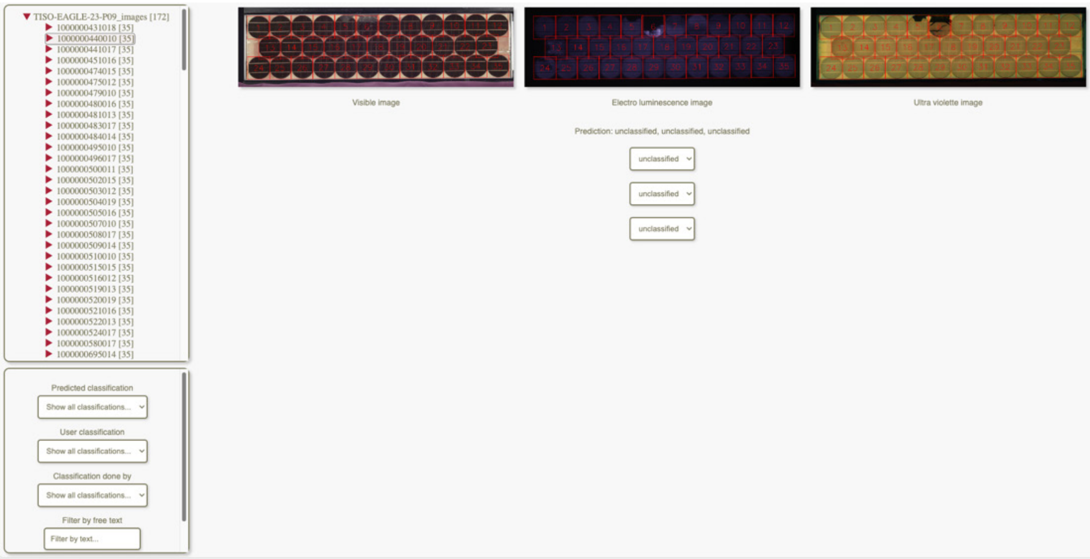


Figure 40: Screenshot of the initial view of the web application.

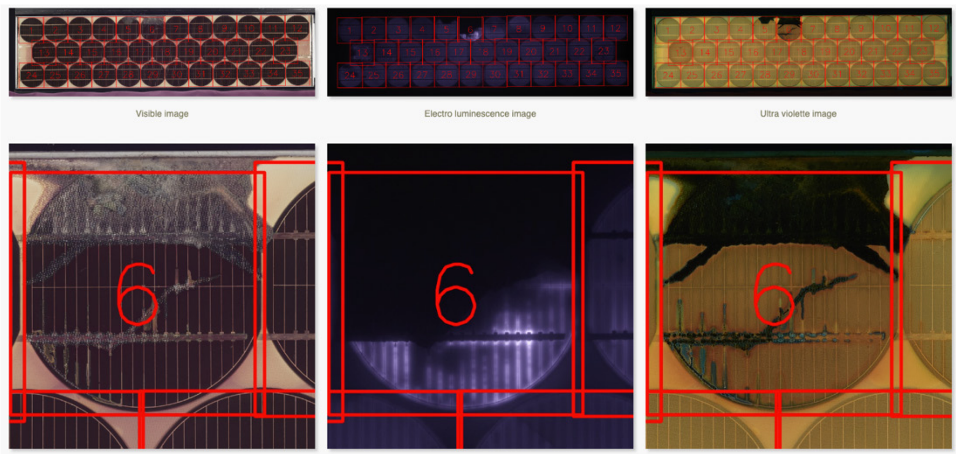


Figure 41: Example of a zoomed TISO cell.

Moving the mouse cursor over one of the three images results in the corresponding area being displayed larger in its full resolution. This makes it easier to identify the defects in individual cells. Each cell is uniquely identified by the module name and cell number. Figure 41 shows an example where cell number six is zoomed and shows visible burns.

The tree structure on the left of Figure 40 allows to easily navigate to each single cell of a module and to either classify it or change the label provided by the model. The user has to identify himself before being able to perform a classification. As cells can present more than one failure, up to three defects can be attributed to one cell, as shown in Figure 42.

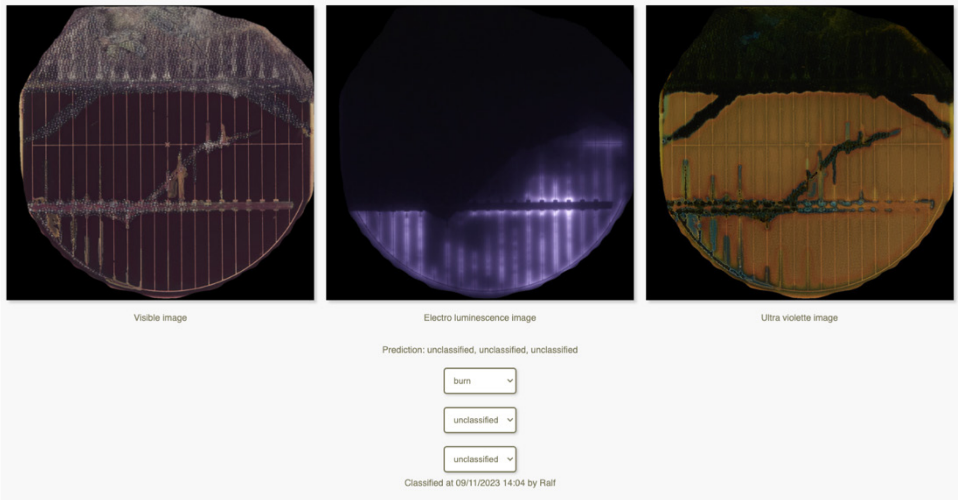


Figure 42: Example of an unclassified TISO image. Up to three defects can be attributed to the cell.

3.4 WP5 Data analysis and modelling

Identification of Failure Modes

Finally, we automatically identify and categorize PV module defects at the cell level. As mentioned before, the automatic identification and classification of failure modes are achieved by employing Vision Transformers (ViTs), specifically, the Channel Vision Transformer (ChannelViT).

The annotated multispectral images serve as the training dataset for the classification task, enabling the assignment of one or more failure modes to each cell. This approach aims to surpass the state of the art by developing a methodology for automatic failure identification using multispectral images (Visible (VI), Electroluminescence (EL), UV-Fluorescence (UVf)), allowing for the identification of a wider spectrum of degradation modes critical for analyzing modules exposed to real environmental aging.

The methodology enables the identification of multiple degradation modes in a single solar cell, a significant improvement over previous studies where typically only one failure per cell was considered. Up to three defects can be attributed to one cell through the developed web interface.

Because the ChannelViT model processes 7-channel multispectral images with an input size of 224 × 224 pixels image, the architecture contains approximately 21 million trainable parameters. Training such a large model directly on our proprietary dataset was not feasible due to the limited number of expert-labeled samples. For this reason, we applied a progressive training approach. First, we initialized the model using ImageNet weights. ImageNet is a large general-purpose dataset (over 14 million images) that allows the model to learn basic visual features before adapting to the PV domain.

In the second step, we fine-tuned the model on PV-specific datasets from Duramat and Infinity. The Duramat (from the U.S. Department of Energy's Durable Module Materials Consortium) dataset contains several thousand EL images of modules and cells with well-defined annotations for cracks, inactive regions, and dark patterns. While the Infinity dataset consists of 2725 labeled cells and provides consistent, high-quality labels for typical failure modes. These datasets are crucial because they contain far more labeled PV defects than we could collect ourselves, allowing the model to learn domain-specific EL characteristics before training on our own multispectral data.

As an initial example, the model was trained on the Duramat dataset but tested on the Infinity dataset. This means the model was trained on different technologies than those used for classification prediction.

Although the model never "saw" the images from the Infinity dataset during training, it achieves an impressive accuracy of 94.4% in prediction. As illustrated in the following image, the results demonstrate excellent performance.

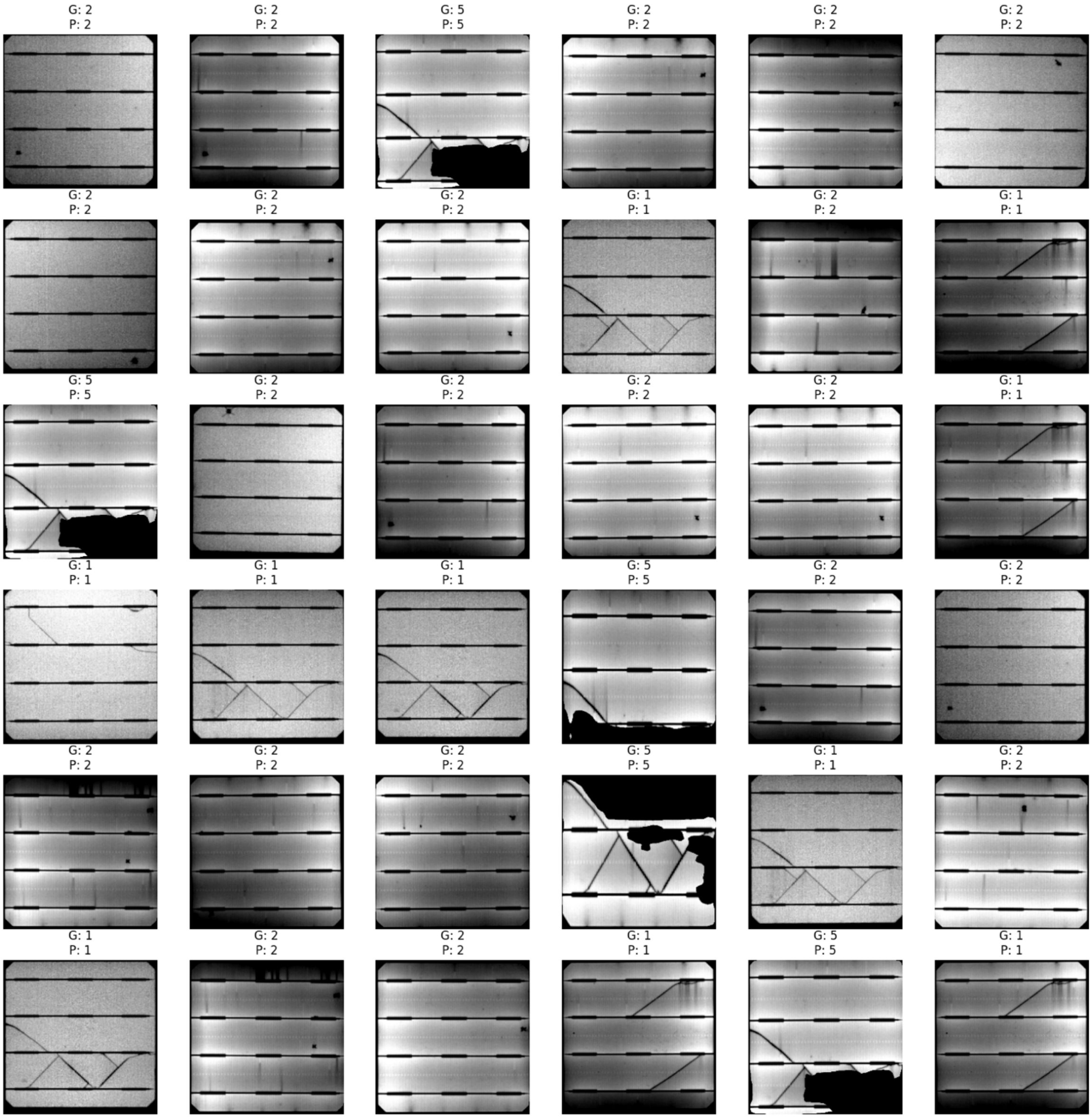


Figure 43: Examples of correct Vision Transformer predictions for the Infinity dataset, where the ground truth (G) is the same as the prediction (P).

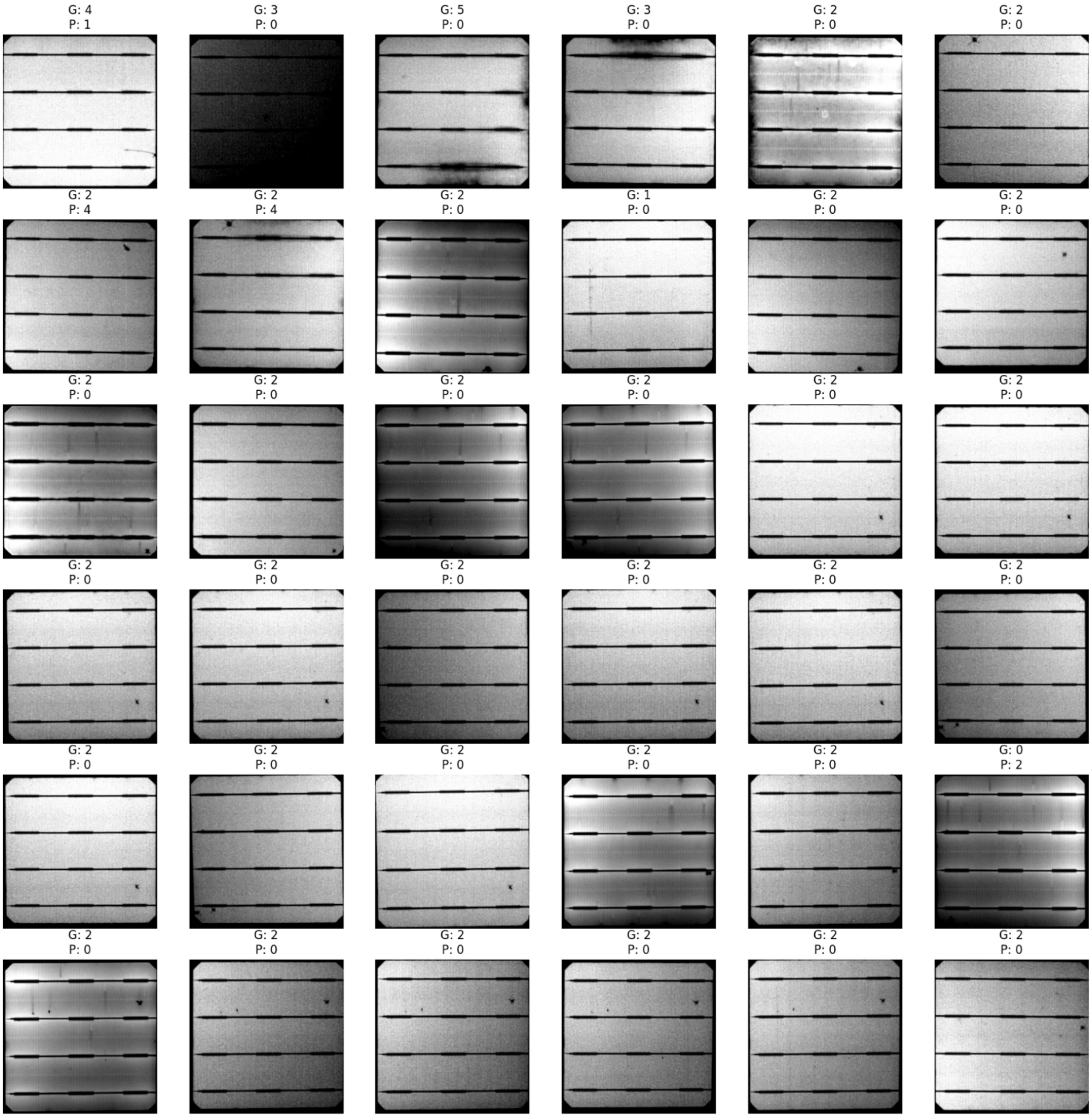


Figure 44: Examples of wrong Vision Transformer predictions, where the ground truth (G) and the prediction (P) are different.

When examining the misclassified examples, it becomes evident that none of them represent clear or obvious errors. In most cases, the model classified the cells as *good*, meaning that no defect was detected. These samples are borderline cases where even human experts would have difficulty assigning a definitive label, as the defects, typically small cross-pattern artefacts, are extremely subtle and barely visible. The ambiguity in these samples highlights the inherent challenge of distinguishing between truly defect-free cells and those with very minor or early-stage degradation.

Our machine learning model at this stage is designed to classify various distinct and combined failure modes observed in PV modules: good, crack, cross, dark and corrosion, which are all mainly visible in EL images. Other defects (delamination and discoloration) rely on other bands like VI and UVf images.

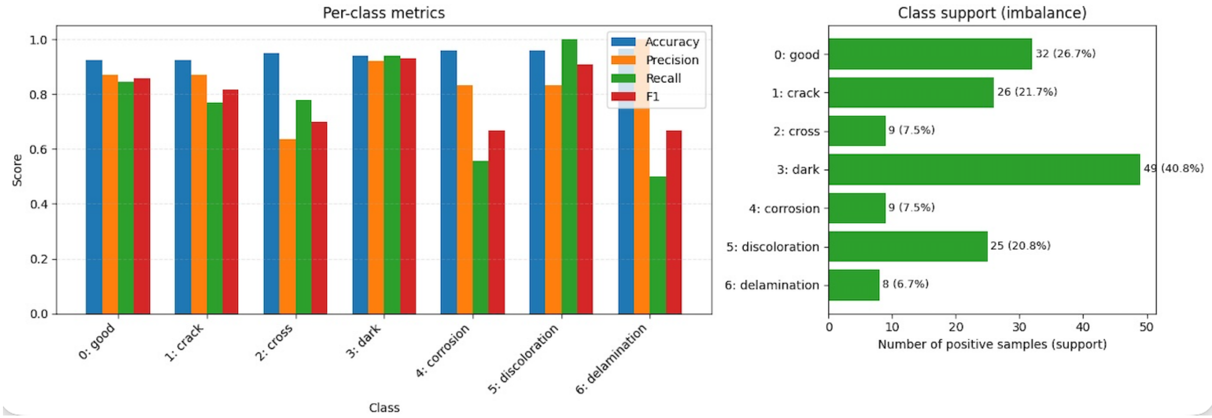


Figure 45: Per-class performance metrics (Accuracy, Precision, Recall, F1) and class support distribution for the failure-mode classifier. The left plot shows model performance across the seven failure-mode categories, while the right plot illustrates the class

After completing this intermediate adaptation phase, we trained the model on our proprietary 7-channel dataset. The results are highly encouraging (Figure 45); however, the current number of expert-labeled samples and the strong heterogeneity of the dataset, covering multiple PV module technologies, still pose a limitation. While the model reliably detects the defined failure modes, achieving robust generalization to entirely new technologies will require a larger and more balanced training set.

The long-term goal is to develop a model that can be applied directly to new images, regardless of whether they originate from known or entirely new PV technologies. For EL-based classification, it has already been demonstrated that this level of generalization is achievable: the model generalizes very well on unseen data because EL-related defects are well represented in the current training dataset. For VI and UVf images, the situation is also very promising. Although the current number of labeled samples does not yet fully capture the diversity and subtleties of these defect types across all technologies, the existing results show that the model can already extract useful patterns from the multi-spectral inputs. This provides a strong foundation for further improvement.

Performance Prediction

Next, we investigate whether PV module performance can be accurately predicted using images acquired in VI, EL and UVf domains. In this analysis, we evaluate PV module performance using the peak maximum power (Pm) as the primary metric. To achieve this, we employ a state-of-the-art deep learning model tailored for image-based analysis.

Dataset

Only the TISO dataset is used to perform the experiments. The reason is that the other datasets contain very few samples and exhibit a highly skewed distribution of Pm values. Using such datasets would likely cause the model to learn that a particular type of PV module corresponds to a specific Pm value, rather than learning which image features indicate performance degradation. The TISO dataset contains data of 157 solar panels, and it consists of the two components:

- Image Data: Preprocessed images of solar panels in the EL, UV, and VI modalities. EL images are single-channel, while UV and VI images are three-channel color images. All images were resized to 224x224 pixels to match the input size required by the model architecture.
- Target Variable: Instead of predicting the absolute Pm value, we use the Pm value relative to the nominal Pm value, which we refer to as the peak maximum power point percentage (Pmp).

We perform two sets of experiments: one using all three modalities (EL, UV, VI) as input, and another using only EL images.

Model Architecture

A ResNet machine learning model [55], pretrained on the ImageNet dataset, is used to process image data. ResNet (Residual Network) architectures employ convolutional layers with residual connections, which facilitate optimization during training and enable the use of deeper neural networks. The ResNet family includes several variants with different depths. In this study, we experiment with ResNet-18, ResNet-34, and ResNet-50, consisting of 18, 34, and 50 layers, respectively.

In the experiment using all image types (EL, UV, VI), each modality is processed by an identical ResNet model to extract modality-specific feature representations. Since EL images contain only a single channel, we duplicate this channel to form a three-channel input compatible with the ResNet architecture. The resulting feature vectors from all modalities are then concatenated to form a joint representation, which is subsequently passed through a fully connected neural network to produce the final prediction. In contrast, in the experiment where only EL images are used, the (duplicated) three-channel EL input is processed by a single ResNet model, and the resulting feature representation is directly fed into the fully connected network to generate the prediction.

Model Training and Evaluation

Due to the relatively small dataset size, model evaluation is repeated 50 times using different random train-validation-test splits. The final results are averaged across all runs.

In each run, the dataset is randomly split into training (80%), validation (15%), and test (5%) subsets. The training subset is used to train the ResNet-based model, while the validation subset is used to prevent overfitting via early stopping with a patience of 10 epochs. Training is performed with the Adam optimizer and the Mean Squared Error (MSE) loss function. The test subset is used to evaluate the model.

The evaluation metrics commonly used for regression problems, Mean Absolute Error (MAE), Median Absolute Error (MedAE), Mean Squared Error (MSE), and the coefficient of determination (R^2), are reported. Table 10 shows the average metrics across the 50 runs for the experiment using only EL image, while Table 11 reports the results for the experiment using all three modalities.

We observe that incorporating UV and VI images alongside EL images leads to improved performance compared to using EL images alone. This indicates that the UV and VI modalities contain complementary degradation-related information that EL imaging alone does not capture. This is a key result of our study: multispectral inputs systematically improve model performance across all evaluated architectures.

Table 10 shows the performance of the ResNet models when trained on EL-only input, while Table 11 presents the results when combining EL, UV, and VI modalities. Across all metrics (MAE, MedAE, MSE, and R^2), the models benefit from the additional spectral information. In particular, ResNet-18 shows a substantial improvement in MAE (from 5.46 to 4.79) and R^2 (from 0.24 to 0.39), confirming that combining modalities enhances the model's ability to capture relevant degradation features. This demonstrates that multispectral imaging is not only useful for qualitative inspection but also quantitatively improves the prediction of module performance.

Furthermore, the shallower ResNet variants outperform the deeper ResNet-50 model, suggesting that increased depth does not necessarily yield performance gains for this task. The reason is that more complex models tend to overfit the training data set and hence do not generalize well.

The best results are achieved with ResNet-18 using all three input modalities, with an average median absolute error of 3.81% and a mean absolute error of 4.79%.

To further illustrate the model's behavior, Figure 46 shows a scatter plot of predicted versus true Pmp values across all 50 runs for the best-performing model. The predictions show a strong correlation with the ground truth, with most points lying close to the diagonal. However, performance declines for panels with Pmp values below 65% or above 85%, due to the limited number of samples in these regions of the dataset.

Table 10: ResNet model performance – input consists of images in EL domain

	ResNet-18	ResNet-34	ResNet-50
MAE	5.46	6.25	10.05
MedAE	4.42	5.48	8.43
MSE	47.11	59.81	153.2
R^2	0.24	0.15	-1.51

Table 11: ResNet model performance – input consists of images in EL, UV and VIS domain.

	ResNet-18	ResNet-34	ResNet-50
MAE	4.79	5.63	9.30
MedAE	3.81	4.82	7.02
MSE	37.90	50.04	151.6
R^2	0.39	0.19	-1.45

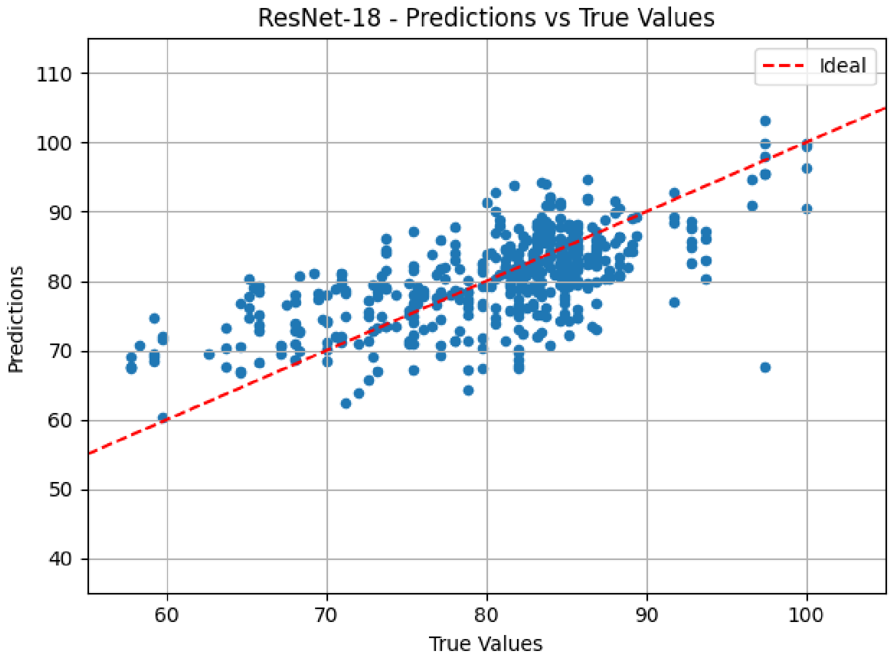


Figure 46: Prediction versus true values of Pmp for ResNet-18 model with the input of images in EL, UV, and VI domains.

4 Evaluation of the results

The first phase of the project focused on setting up the test bench and establishing an efficient and repeatable multispectral imaging procedure. At the beginning, 172 TISO modules were processed across the complete multispectral range. Additionally, multispectral measurements were carried out on 71 pre-stressed c-Si PV modules, obtained through various acquisition methods from multiple projects (i.e., ACHILLES and PVDETECT), and these were also incorporated into the dataset. The addition of new data was stopped before half of 2025. In addition to the multispectral imaging, electrical performance of the modules was measured at STC. In parallel with the multispectral measurements, a dedicated defect-labeling structure was developed considering the multispectral images, allowing multiple labels per cell. Labeling continued throughout the measurement phase and was finalized once the necessary amount of data had been annotated. In total, >10% of the available data was labeled by experts.

The segmentation pipeline demonstrates reliable performance. The time needed for preprocessing the full TISO dataset and uploading the segmented images within minutes, which means that the images are ready for further processing on the same day.

The developed web interface is user-friendly and allows efficient navigation through different technologies, modules, and individual cells. The different users can identify themselves to keep track of who created or modified the labeling of each specific image. The implementation with Flask allows for the rapid development of new features if additional requirements arise during the project.

The ChannelViT-based classifier represents a central element of the project goals. Due to the limited size of the proprietary multispectral dataset, a progressive training strategy was adopted: initialization on ImageNet, fine-tuning on external PV-specific EL datasets (Duramat and Infinity) and finally adaptation to the project's 7-channel images.

Even in its prototype stage, the model delivers very promising results. When trained on Duramat and evaluated on the Infinity dataset (i.e. on unseen module technologies), the classifier achieves an accuracy of about 94 %, demonstrating strong cross-dataset generalization for EL-visible failure modes. On the project's multispectral data, the per-class metrics (Figure 45) show that well-represented classes such as "good", "crack" and "dark" reach high precision and recall values, while performance is naturally lower for rare classes such as "delamination" and for complex combinations of defects. Misclassified examples are typically borderline cases where experts even find it difficult to decide whether a subtle structure should already be regarded as a defect, which highlights both the intrinsic ambiguity of some labels and the high sensitivity of the model.

Overall, the current results indicate that ChannelViT is capable of reliably detecting multiple simultaneous failure modes at cell level and of leveraging complementary information from VI, EL and UVf channels. Further improvements are expected as the labelled dataset grows and becomes more balanced across technologies and defect types.

As mentioned previously, the advancement of the project is measured with the following measurable targets:

- Datasets of raw multispectral images of at least 230 modules.
    - We have collected images of 241 modules, which make 11076 cells in total (Figure 24 and Figure 25).
- Datasets of annotated images. At least 30% of the images are labelled by experts.
    - >10% of data was labeled by experts, Expert labeling required significantly more effort than initially anticipated. Many images demanded careful inspection and repeated verification to ensure accuracy, which slowed the pace of annotation. Nevertheless, the expert-labeled portion of the dataset remains of high quality and provides a valuable foundation for training and evaluating models.
- At least 90% accuracy in failure identification by machine learning prototypes for labelled data.
    - >90% accuracy is achived for labelled data.
- At least 70% accuracy in failure identification by machine learning prototypes for unlabeled data (tested by experts).
    - > also 70% is archived for unlabeled data, this is especially valid for all the defects visible in well studied EL visible damages, in the damages type that require UVf and VI we aim for better accuracy in the next steps.
- Identification of the top two failure modes that affect module performance.
    - Regarding the measurable target to identify the top two failure modes, the project achieved a more integrated solution than originally envisioned. While Task 5.2 (the generation of explicit cell-defect ratios per module) was not the primary focus, we successfully fulfilled the core objective of Task 5.3 by developing a robust deep learning methodology for performance-loss estimation.
    - Using a ResNet-18 model trained on multispectral data, we demonstrated that overall module performance (PMPP) can be accurately predicted directly from VI, EL, and UVf images. This approach provides a significant advancement toward automated reliability assessment, as it correlates visual degradation patterns to electrical power loss without requiring manual cell-level statistical intermediates

5 Outlook

In a follow-up project, we plan to further expand and balance the labelled multispectral dataset by integrating additional PV module technologies, performing targeted accelerated ageing experiments, and systematically annotating newly acquired multispectral images. This will improve the representation of rare defects and enable the models to better capture characteristic signatures of defects visible in the VI and UVf channels. By enlarging the training dataset and refining the model through transfer learning, the ChannelViT classifier is expected to achieve robust performance not only for well-known EL-based failure modes but also for complex multispectral defect combinations across diverse PV technologies.

Future work will also extend the analysis to module-level degradation patterns. For instance, EL patterns related to PID and UVID will be investigated and incorporated into the classification framework. This will allow the methodology to capture degradation mechanisms that manifest at the module scale, rather than solely at the individual cell level.

The methodology will be implemented indoors, in laboratories and manufacturing environments. Indoor implementation will enable faster and more accurate defect detection and classification, reducing technician bias and supporting quality assurance. Combined with automated analysis, this approach will increase throughput and ensure consistent inspection standards.

Beyond the scientific development of improved models, a key objective of the continuation project is to transfer the methodology from controlled laboratory environments to operational PV systems. The ultimate goal is to enable quantitative assessment of PV systems in the field using multispectral imaging. This approach will allow operators to identify defective modules directly in installed systems without dismantling or individually testing each module. From a practical perspective, such a methodology could significantly reduce the effort required for system diagnostics and maintenance. Instead of testing modules one by one, multispectral images of entire module rows or systems can be automatically analysed to identify modules with potential defects or degradation patterns. This capability is particularly valuable for large-scale or difficult-to-reach PV installations, where locating problematic modules would otherwise require substantial time and manual inspection.

Another practical advantage is the ability to estimate performance losses directly from imaging data. If the predicted power loss of a module exceeds warranty limits after accounting for methodological uncertainties, this information can support faster, more informed maintenance or replacement decisions. In this way, multispectral imaging combined with AI-based analysis provides an efficient tool for rapid system diagnostics, targeted maintenance, and improved operational decision-making.

To validate this approach under realistic conditions, the planned continuation project will include studies on outdoor PV modules and installations, where multispectral imaging will be evaluated as a tool for defect identification, classification, and performance loss estimation.

References

1. MeteoSchweiz. *Klimabulletin Sommer 2020*. (Zürich, 2020).
2. IRENA. *Renewable Power Generation Costs in 2019*. (2020).
3. EU Commision. *SET-Plan - Declaration on Strategic Targets in the Context of an Initiative for Global Leadership in Photovoltaics (PV)*. https://setis.ec.europa.eu/system/files/2021-04/declaration_of_intent_pv.pdf (2021).
4. Osterwald, C. R. & McMahon, T. J. History of accelerated and qualification testing of terrestrial photovoltaic modules: A literature review. *Progress in Photovoltaics: Research and Applications* 17, 11–33 (2009).
5. Sample, T. Standards for PV - Overview of IEC Related PV Standards and How They Contribute to Reduced Costs of Energy. *Proc. 35th European Photovoltaic Solar Energy Conference and Exhibition* 5DP.1.1, 991–996 (2018).
6. Gebhardt P., Fokuhl, E. et. al. Stabilization procedures for TOPCon PV modules after UV-induced degradation,Solar Energy Materials and Solar Cells, Volume 294, 2026, 113885, ISSN 0927-0248.
7. Virtuani, A., Annigoni, E. & Ballif, C. One-type-fits-all-systems: Strategies for preventing potential-induced degradation in crystalline silicon solar photovoltaic modules. *Progress in Photovoltaics: Research and Applications* 27, 13–21 (2019).
8. Ramspeck, K. *et al.* Light Induced Degradation of Rear Passivated mc-Si Solar Cells. *Proc. 27th European Photovoltaic Solar Energy Conference and Exhibition* 2DO.3.4, 861–865 (2012).
9. Gebhardt, P., Kräling, U., Fokuhl, E., Hädrich, I., & Philipp, D. (2024). Reliability of Commercial TOPCon PV Modules—An Extensive Comparative Study. *Progress in Photovoltaics: Research and Applications*.
10. Liu, F., Jiang, L. & Yang, S. Ultra-violet degradation behavior of polymeric backsheets for photovoltaic modules. *Solar Energy* 108, 88–100 (2014).
11. Sample, T. *New Combined Cycles for Future Test Standards in IEC TC82 WG2*. http://sophia-pv.fe.uni-lj.si/ (2018).
12. Osborne, M. Game changer: Violet Power to offer 50-year solar panel warranty with US-made IBC technology. *PV-Tech* (2020).
13. Monokroussos, C. *Supplementary Power Rating and Type Approval for Bifacial PV Modules. TÜV Forum - PV Module Technology & Application (presentation)* (2019).
14. Clement, C. E., Singh, J. P., Birgersson, E., Wang, Y. & Khoo, Y. S. Hotspot development and shading response of shingled PV modules. *Solar Energy* 207, 729–735 (2020).
15. Jordan, D. C. *et al.* High Efficiency Module Degradation – from Atoms to Systems. *Proc. 37th European Photovoltaic Solar Energy Conference and Exhibition* 4BO.14.2, 828–833 (2020).
16. Krizhevsky, A., Sutskever, I. & Hinton, G. E. ImageNet Classification with Deep Convolutional Neural Networks. *Adv Neural Inf Process Syst* 25, (2012).
17. Hussain, M., Bird, J. J. & Faria, D. R. A study on CNN transfer learning for image classification. *Advances in Intelligent Systems and Computing* 840, 191–202 (2019).
18. Fada, J. S. *et al.* Electroluminescent Image Processing and Cell Degradation Type Classification via Computer Vision and Statistical Learning Methodologies. in *2017 IEEE 44th Photovoltaic Specialist Conference (PVSC)* 3456–3461 (2017). doi:10.1109/PVSC.2017.8366291.
19. Deitsch, S. *et al.* Automatic classification of defective photovoltaic module cells in electroluminescence images. *Solar Energy* 185, 455–468 (2019).
20. Karimi, A. M. *et al.* Automated Pipeline for Photovoltaic Module Electroluminescence Image Processing and Degradation Feature Classification. *IEEE J Photovolt* 9, 1324–1335 (2019).

21. Karimi, A. M. *et al.* Feature Extraction, Supervised and Unsupervised Machine Learning Classification of PV Cell Electroluminescence Images. in *2018 IEEE 7th World Conference on Photovoltaic Energy Conversion (WCPEC) (A Joint Conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC)* 418–424 (2018). doi:10.1109/PVSC.2018.8547739.
22. Niazi, K. A. K., Akhtar, W., Khan, H. A., Yang, Y. & Athar, S. Hotspot diagnosis for solar photovoltaic modules using a Naive Bayes classifier. *Solar Energy* 190, 34–43 (2019).
23. Hamzavy, B. T. *et al.* Study of PV module degradation rate prediction through correlation of field-aged and accelerated-aged module degradation data. in *2017 IEEE 44th Photovoltaic Specialist Conference (PVSC)* 2618–2621 (2017). doi:10.1109/PVSC.2017.8366127.
24. Liu, Z. *et al.* Quantitative analysis of degradation mechanisms in 30-year-old PV modules. *Solar Energy Materials and Solar Cells* 200, 110019 (2019).
25. Ndiaye, A. *et al.* Degradations of silicon photovoltaic modules: A literature review. *Solar Energy* 96, 140–151 (2013).
26. Sharma, V. & Chandel, S. S. Performance and degradation analysis for long term reliability of solar photovoltaic systems: A review. *Renewable and Sustainable Energy Reviews* 27, 753–767 (2013).
27. Virtuani, A. *et al.* 35 years of photovoltaics: Analysis of the TISO-10-kW solar plant, lessons learnt in safety and performance—Part 1. *Progress in Photovoltaics: Research and Applications* 27, 328–339 (2019).
28. Annigoni, E. *et al.* 35 years of photovoltaics: Analysis of the TISO-10-kW solar plant, lessons learnt in safety and performance—Part 2. *Progress in Photovoltaics: Research and Applications* 27, 760–778 (2019).
29. French, R. H., Karimi, A. M. & Braid, J. L. Electroluminescent (EL) Image Dataset of PV Module Under Step-wise Damp Heat Exposures. (2019) doi:10.17605/OSF.IO/4QRTV.
30. Chen, X. Electroluminescence Image Analysis - crack segmentation. https://datahub.duramat.org/dataset/metadata/crack-segmentation (2022) doi:10.21948/1871275.
31. Abdulla, W. Splash of Color: Instance Segmentation with Mask R-CNN and TensorFlow. *Matterport Engineering Techblog* https://engineering.matterport.com/splash-of-color-instance-segmentation-with-mask-r-cnn-and-tensorflow-7c761e238b46 (2018).
32. Wu, Y., Kirillov, A., Massa, F., Lo, W.-Y. & Girshick, R. *Detectron2*. https://github.com/facebookresearch/detectron2 (2019).
33. Lin, T.-Y. *et al.* Feature Pyramid Networks for Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 2117–2125 (2017).
34. Ackermann, M., Iren, D., Wesselmecking, S., Shetty, D. & Krupp, U. Automated segmentation of martensite-austenite islands in bainitic steel. *Mater Charact* 191, 112091 (2022).
35. Lin, T. Y. *et al.* Microsoft COCO: Common Objects in Context. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8693 LNCS, 740–755 (2014).
36. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *ArXiv* 1802.03426, (2018).
37. Van Der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* 9, 2579–2605 (2008).
38. Simonyan, K. & Zisserman, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings* (2014).
39. Deng, J. *et al.* ImageNet: A large-scale hierarchical image database. 248–255 (2010) doi:10.1109/CVPR.2009.5206848.

40. Campello, R. J. G. B., Moulavi, D. & Sander, J. Density-based clustering based on hierarchical density estimates. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 7819 LNAI, 160–172 (2013).
41. Dash. https://dash.plotly.com/.
42. Miguel Grinberg. Flask Web Development. *O'Reilly* 1–314 (2018).
43. VGG Image Annotator. https://www.robots.ox.ac.uk/~vgg/software/via/via.html
44. Ravi, N. et al. SAM 2: Segment Anything in Images and Videos. doi: 10.48550/arXiv.2408.00714
45. Kirillov, A. et al. Segment Anything. arXiv:2304.02643 (2023).
46. Dosovitskiy A., et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. Conference paper at ICLR 2021 (2021).
47. Y. Bao, S. Sivanandan, and T. Karaletsos. Channel Vision Transformers: An Image Is Worth 1 x 16 x 16 Words. Apr. 19, 2024, *arXiv*: arXiv:2309.16108. doi: 10.48550/arXiv.2309.16108.
48. Warren, T. Focus Adjustment and Other Considerations when Shooting Infrared. *35mmc* Focus Adjustment and Other Considerations when Shooting Infrared (2023).
49. Özkalay, E., Virtuani, A., Eder, G., Voronko, Y., Bonomo, P., Caccivio, M., Ballif, C., & Friesen, G. *Correlating long-term performance and aging behaviour of building integrated PV modules.* Energy and Buildings, 316, 114252 (2024)
50. Virtuani, A., Caccivio, M., Annigoni, E., Friesen, G., Chianese, D., Ballif, C., & Sample, T. *35 years of photovoltaics: Analysis of the TISO-10-kW solar plant, lessons learnt in safety and performance—Part 1.* Progress in Photovoltaics: Research and Applications, 27(4), 328–339 (2019).
51. Annigoni, E., Virtuani, A., Caccivio, M., Friesen, G., Chianese, D., & Ballif, C. *35 years of photovoltaics: Analysis of the TISO-10-kW solar plant, lessons learnt in safety and performance—Part 2.* Progress in Photovoltaics: Research and Applications, 27(9), 760–778 (2019).
52. Özkalay, E., Quest, H., Gassner, A., Virtuani, A., Eder, G. C., Vorstoffel, S., Buerhop-Lutz, C., Friesen, G., Ballif, C., Burri, M., & Bucher, C. *Three decades, three climates: environmental and material impacts on the long-term reliability of photovoltaic modules.* EES Solar, Issue 4 (2025).
53. PlanetSolar - Catamarano, https://www.planetsolar.org/en/world-premiere/history/
54. AIT Austrian Institute of Technology. (n.d.). *INFINITY – In-depth analysis of climatic impacts on PV systems*. https://www.ait.ac.at/infinity
55. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).