technologies

MDPI

Article

Cross-Technology Prediction of PV Cell Output Power Using a Convolutional Hierarchical Mixture of Experts Model

Héctor Felipe Mateo-Romero [1], Luis Hernández-Callejo [2,*], Miguel Ángel González Rebollo [3], Valentín Cardeñoso-Payo [4], Victor Alonso Gómez [5], Leonardo Cardinale-Villalobos [6], Jose Ignacio Morales Aragonés [7], Sara Gallardo Saavedra [2], Abel Méndez Porras [8] and Mario Carbonó dela Rosa [9]

[1] Departamento de Informática, Universidad de Valladolid, 47011 Valladolid, Spain; hectorfelipe.mateo@uva.es
[2] Departamento Ingeniería Agrícola y Forestal, Universidad de Valladolid, 42004 Soria, Spain; sara.gallardo@uva.es
[3] Departamento de Fisica de la Materia Condensada, Universidad de Valladolid, 47011 Valladolid, Spain; mrebollo@uva.es
[4] Departamento Informatica, Universidad de Valladolid, 47011 Valladolid, Spain; valentin.cardenoso@uva.es
[5] Departamento de Física, Universidad de Valladolid, 47002 Valladolid, Spain; victor.alonso.gomez@uva.es
[6] Departamento de Electronica, Campus San Carlos, Tecnológico de Costa Rica, Ciudad Quesada 21001, Costa Rica; lcardinale@itcr.ac.cr
[7] Instituto de Nanociencia y Materiales de Aragón, Consejo Superior de Investigaciones Científicas y Universidad de Zaragoza, 50009 Zaragoza, Spain; ziguratt@coit.es
[8] Escuela de Computación, Campus San Carlos, Tecnológico de Costa Rica, Ciudad Quesada 21001, Costa Rica; amendez@tec.ac.cr
[9] Escuela de Ciencias Básicas, Universidad Nacional Abierta y a Distancia, Santa Marta 470003, Colombia; ing.mariocarbono@gmail.com
* Correspondence: luis.hernandez.callejo@uva.es

Abstract

Accurate prediction of photovoltaic (PV) cell power from electroluminescence (EL) images is a key enabler for automated quality assessment and performance estimation in PV manufacturing and diagnostics. However, most existing image-based deep learning models are developed and evaluated for a single PV cell technology, limiting their ability to generalize across the wide variety of cell types used in practice. This work investigates the impact of PV cell technology on power prediction accuracy and proposes a novel *Convolutional Hierarchical Mixture of Experts (CHME)* architecture to overcome these generalization limitations. First, convolutional neural networks and feature-based machine learning models are evaluated on multiple PV cell technologies. While technology-specific convolutional models achieve low mean absolute errors (MAEs) of 0.02–0.04 when tested on the same cell type, their performance deteriorates substantially (MAEs of 0.08–0.23) when applied to different technologies. Feature-based models exhibit greater robustness across technologies but at the cost of lower prediction accuracy. To address these limitations, the proposed CHME model combines multiple pretrained technology-specific convolutional experts with a discriminator network that automatically identifies the PV cell technology and selects the most appropriate expert for power prediction. Experimental results demonstrate that CHME achieves the best overall performance, reducing the MAE to 0.0262 compared with 0.0298 for the best standalone convolutional model, while preserving adaptability to heterogeneous datasets. These results demonstrate that explicitly accounting for PV cell technology significantly improves image-based power prediction and that the proposed hierarchical mixture-of-experts framework provides an accurate, scalable, and easily retrainable solution for real-world PV diagnostic systems.

Check for updates

Academic Editor: Joshua M. Pearce

Received: 20 May 2026
Revised: 18 July 2026
Accepted: 7 August 2026
Published: 12 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Keywords: mixture of experts; convolutional neural networks; photovoltaic cells; machine learning; regression

1. Introduction

The global energy landscape has undergone significant changes due to inflationary pressures triggered by the COVID-19 pandemic and recent geopolitical conflicts, such as ongoing wars. These disruptions have caused a substantial increase in energy prices, with supply chain interruptions and rising demand compounding the situation [1,2]. This escalation, combined with the objectives outlined in the United Nations' Sustainable Development Goals (SDGs) [3], has accelerated the transition from non-renewable, polluting energy sources to sustainable and eco-friendly alternatives.

Among the diverse forms of renewable energy, solar energy stands out as particularly advantageous due to its widespread availability and the relative ease of installing photovoltaic (PV) modules [4]. However, maintaining solar installations presents considerable challenges, primarily due to the large number of PV modules, each consisting of numerous cells, and the diverse factors influencing their energy production, such as climate conditions, humidity, and temperature [5]. These complexities have spurred extensive research into optimizing PV system performance, with key areas including maximum power tracking, energy forecasting, parameter estimation, and defect detection [6].

Artificial Intelligence (AI) has emerged as a pivotal tool in addressing these challenges, demonstrating its versatility across numerous sectors, including renewable energy [7]. Techniques such as neural networks, fuzzy logic, and metaheuristics have proven effective in solving PV-related problems [6,7]. In photovoltaics, AI has been widely applied to power and energy forecasting, enabling more accurate predictions of PV generation for grid integration and energy management. It has also been extensively used for Maximum Power Point Tracking (MPPT), where intelligent optimization algorithms improve the tracking of the global maximum power point under dynamic environmental conditions and partial shading. In addition, AI has become a key tool for automated fault detection and diagnosis, allowing the identification, classification, and localization of defects to enhance system reliability while reducing inspection and maintenance costs. A critical focus of this research is the analysis of PV cells, which involves identifying surface defects using advanced imaging techniques such as electroluminescence (EL), photoluminescence (PL), and thermography. More information about this techniques can be found in the bibliography [8,9].

The innovative approach proposed in this paper goes beyond traditional defect detection by directly evaluating the impact of these defects on PV cell performance. Typically, the performance of PV modules is assessed using the intensity–voltage (I-V) curve, which can be measured either at the module level in operational settings or at the cell level under laboratory conditions.

Other studies have focused on automated defect detection and localization in photovoltaic cells, leveraging recent advances in deep learning and computer vision. These approaches employ techniques such as coordinate attention mechanisms to enhance the representation of spatial features, generative adversarial networks (GANs) to improve defect recognition through data augmentation and image enhancement, and YOLO-based object detection models for the real-time identification and localization of defects. Such methods have demonstrated improved detection accuracy and computational efficiency, facilitating the deployment of automated inspection systems for large-scale PV manufacturing and field monitoring. A complete review of the state of the art of AI application in PV systems was performed in [6], analyzing the different problems and solutions. The review identified

that most existing studies focus on defect detection and classification, whereas the direct estimation of the electrical performance or power output of photovoltaic cells and modules has been scarcely investigated. The present work addresses this research gap by estimating the normalized maximum power point directly from electroluminescence images.

Few studies have explored the use of machine learning for predicting the power output of PV modules or cells. Hoffmann et al. [10] proposed a deep learning approach to estimate the maximum power output of photovoltaic modules directly from electroluminescence (EL) images. Using a dataset of 719 modules with different degradation levels, they trained a ResNet-based model that achieved a mean absolute error of approximately 3.2%, outperforming conventional regression methods. In addition, they introduced a visualization technique to localize the contribution of individual cells to the predicted power loss.

In a subsequent study, Hoffmann et al. [11] extended this methodology to photoluminescence (PL) images, demonstrating that module power can also be estimated from this imaging modality using deep convolutional neural networks. Their model achieved a mean absolute error of 4.4%, while the generated regression maps enabled the localization of power losses and the identification of inactive regions, improving the interpretability of the predictions.

This paper presents a novel method for assessing PV cell performance using EL images. By integrating EL images with performance data derived from I-V curves, this approach enables accurate predictions across different PV cell technologies, including monocrystalline and polycrystalline cells. Moreover, a Convolutional Hierarchical Mixture of Experts (CHME) model is introduced, leveraging pretrained convolutional models to enhance predictive accuracy and reliability.

The remainder of the paper is structured as follows: Section 2 elaborates on the problem context, followed by details regarding data collection in Section 3.1. The preprocessing methodology is discussed in Section 3.3, while Sections 3.6 and 3.7 describe the evaluation metrics and machine learning models used. Section 4 presents the experimental results, and the conclusions are provided in Section 5.

2. Problem Definition

There are different ways of using the energy of the sun to produce electricity. The two most important are Solar Thermal Energy and Solar Photovoltaic Energy [12]. Their most important difference is that PV systems produce direct electricity converting the sunlight, while thermal systems produce thermal energy.

2.1. Structure of a Photovoltaic Cell

As previously noted, photovoltaic (PV) modules consist of numerous individual PV cells that convert sunlight directly into electricity. While the internal structure of these cells is highly complex [13], the front surface of a standard cell exhibits two primary grid components (see Figure 1):

- **Electrical Busbars (busbars):** Thin metallic rectangular strips printed on both the front and back sides of the PV cell to conduct the generated current. Their quantity is variable and depends on the specific cell design and manufacturer.
- **Fingers:** Ultra-thin grid lines oriented perpendicularly to the busbars. Their primary function is to collect the photo-generated current across the surface and deliver it to the busbars. These are typically screen-printed using a silver conductive paste.

Photovoltaic cells can be manufactured using various fabrication techniques [13]. The experiments in this study utilize the following cell technologies:

- **Monocrystalline Cells:** Silicon-based cells manufactured from a single continuous crystal structure. They exhibit high conversion efficiencies but incur higher production costs compared to alternative technologies.
- **Polycrystalline Cells:** Silicon-based cells formed by melting multiple silicon crystals together. While the resulting grain boundaries and surface irregularities render them less efficient, they offer a significantly more cost-effective manufacturing alternative.

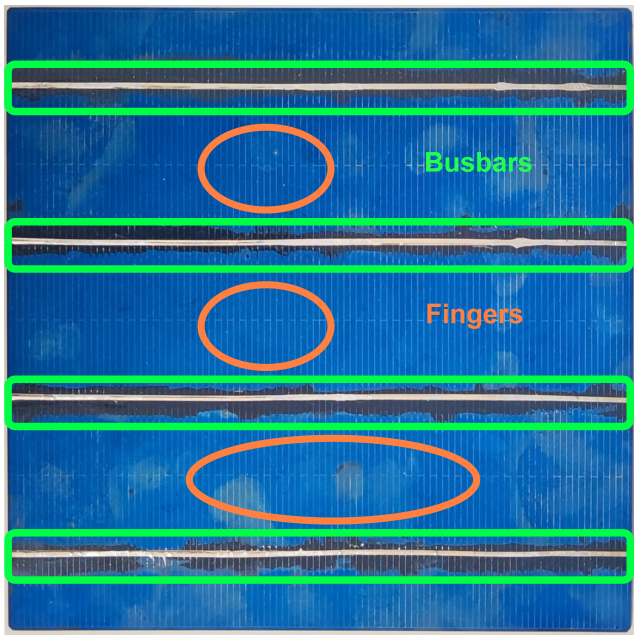


Figure 1. A polycrystalline photovoltaic cell, includes two distinct elements: the green lines represent the horizontal, wider bars called busbars, while the orange lines denote the thinner, perpendicular lines known as fingers.

The detection of problems in PV cells is not a trivial problem, since they can present problems that are not visible at a single glance. It is necessary to apply techniques to examine the surface of the cell. The most important techniques are the following:

2.2. Techniques to Capture the PV Module/Cells

- Thermography: This technique uses an infrared camera to detect the heat emissions from the module or cell.
- Electroluminescence: The EL technique consists of taking a photograph of the photons produced by the modules/cells when they are connected to an electric current. This technique is invasive since it is necessary to disconnect the module to inject current into it.

Both these techniques have their utility but electroluminescence is the most commonly used in works related to fault detection in the surface of the PV cell [6]. More information about these techniques can be found in the bibliography [8,9].

3. Data and Methodology

The objective of this section is to introduce the processes that were performed in the article: the collection of the data, the preprocessing of the data, the election of the output target, and the selection and training of the machine learning models for each experiment.

3.1. Data

This paper aims to analyze the differences in the performance of ML methods when they face different kinds of PV cells. For this objective, three datasets were prepared using

three different kinds of PV cells: polycrystalline with 4 busbars, monocrystalline with 3 busbars and monocrystalline with 4 busbars. Figure 2 presents the three kinds of PV cells.

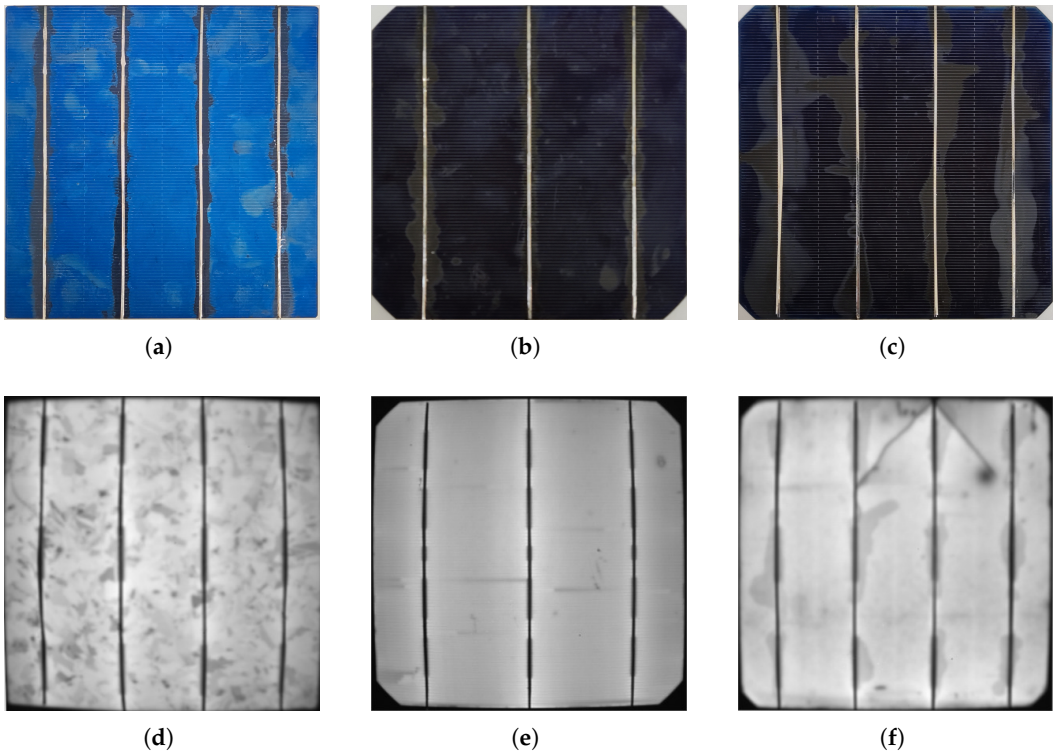


Figure 2. Different kinds of PV cells used in the experiments. (**a**) A polycrystalline photovoltaic (PV) cell equipped with four busbars, commonly referred to as Poly-4; (**b**) a monocrystalline photovoltaic (PV) cell featuring three busbars, commonly identified as Mono-3; (**c**) a monocrystalline photovoltaic (PV) cell with four busbars, typically referred to as Mono-4; (**d**) EL image of a Poly-4 cell; (**e**) EL image of a Mono-3 cell; (**f**) EL image of a Mono-4 cell.

3.2. Photovoltaic Cells

The three datasets provide the opportunity to analyze the differences in performance with cells of different technologies (Poly and Mono) that, as can be observed, change the aspect of the cell. They can also provide the opportunity to analyze the difference in performance due to the number of busbars since they change the structure of the cell.

Each dataset is composed of:

- Electroluminescence (EL) images of each PV cell were captured using an "InGaAs C12741-0" silicon detector camera from Hanamatsu Photonics, equipped with an 8 mm focal length lens and an aperture of f/1.4 (Figure 2). The current values used and their corresponding simulated irradiance levels are listed below:
  - **50% (500 W/m^2)**: Corresponds to a current of 1.27 A
  - **60% (600 W/m^2)**: Corresponds to a current of 1.52 A
  - **70% (700 W/m^2)**: Corresponds to a current of 1.78 A
  - **80% (800 W/m^2)**: Corresponds to a current of 2.03 A
  - **90% (900 W/m^2)**: Corresponds to a current of 2.28 A
  - **100% (1000 W/m^2)**: Corresponds to a current of 2.53 A

  In addition to the EL images, various masks have been designed to simulate different types of soiling commonly observed in real PV installations [14]. These masks are constructed using cardboard and foil paper, the latter being selected for its light emission properties. Figure 3 illustrates the different types of shadows cast when the

masks are positioned over various parts of the PV cells, as well as defects observed in the original cells in class 7.

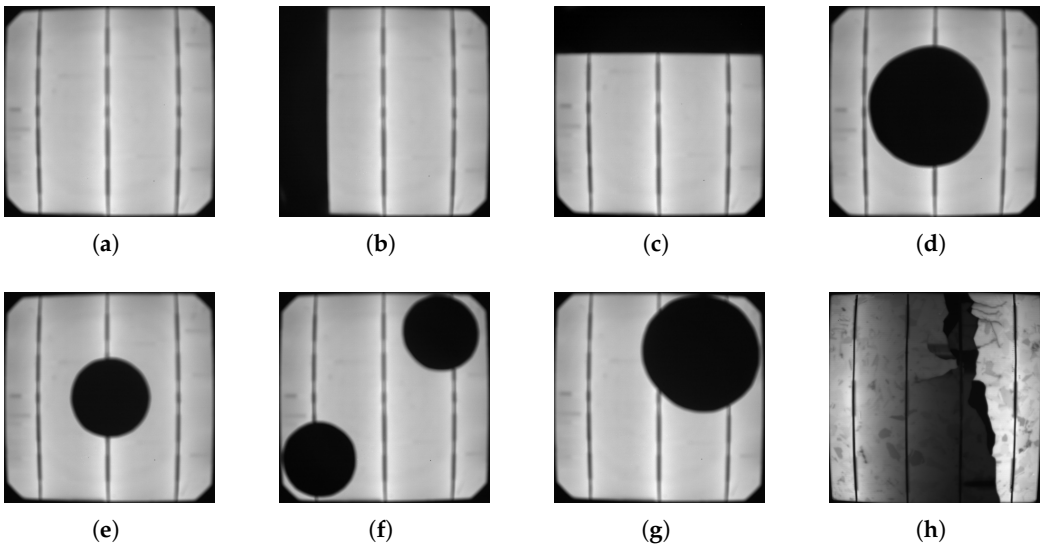


Figure 3. Different shadows presented in the datasets: 0: Original image, 1: Long horizontal line, 2: Long vertical line, 3: Big central circle, 4: Small central circle, 5: Two small circles in the corners, 6: Big circle in one corner. (**a**) Shadow 0; (**b**) Shadow 1; (**c**) Shadow 2; (**d**) Shadow 3; (**e**) Shadow 4; (**f**) Shadow 5; (**g**) Shadow 6; (**h**) Shadow 7.

- Current–Voltage (I-V) curve of each cell: The I-V curves of the cells were taken using an I-V tracer presented in [15]. This device is capable of measuring the I-V of a single PV cell. The illumination necessary for measuring the PV cell was provided by a customized LED array; more information about the array and the LEDs can be found in our previous work [16]. The temperature of the cells was monitored to avoid a decrease in performance due to overheating. The LED array was powered by a power supply, and by using different values of current, it simulated the different irradiances of the solar light; the values used were: 1.27 A, 1.78 A, 2.03 A, 2.28 A, 2.53 A.
  All I-V measurements were acquired under controlled laboratory conditions using the same I-V tracer, LED illumination system, acquisition protocol, and temperature monitoring procedure. As with any experimental measurement system, the acquired I-V curves contain a certain level of instrumental noise and variability. To reduce these effects, the current values were smoothed using a Savitzky–Golay filter before extracting the electrical parameters, as described in [15]. Furthermore, although the customized LED array incorporates a diffuser to improve irradiance uniformity, small spatial variations in illumination are unavoidable. Since all photovoltaic cells were measured using the same acquisition setup and under identical operating conditions, these sources of uncertainty affect the complete dataset consistently and do not introduce systematic differences between the evaluated models.
  As in the EL images, the masks were used to increase the number of defects present in the datasets (Figure 3). The shadows produced by the masks result in a reduction in the performance of the PV cells, which can be seen in their I-V curve, as shown in Figure 4.

The three original datasets were used to create four new datasets (Figure 5); these new datasets are vital to creating good comparisons between the different models, which are presented in the subsequent sections.

- **Mono-Combined**: A dataset composed of all the images from monocrystalline cells (Mono 3 + Mono 4).

- **Poly-4 Reduced**: A version of Poly-4 reduced to the same size as Mono-Combined. This dataset aims to facilitate a more valid comparison between models trained with the datasets by ensuring identical dimensions.
- **Mono + Poly**: A comprehensive dataset consisting of all available images, including both Poly-4 and Mono-Combined.
- **Mono + Poly Balanced**: A dataset made up of all the images from Mono-Combined and an equal number of images from Poly-4 (Poly-4 Reduced).

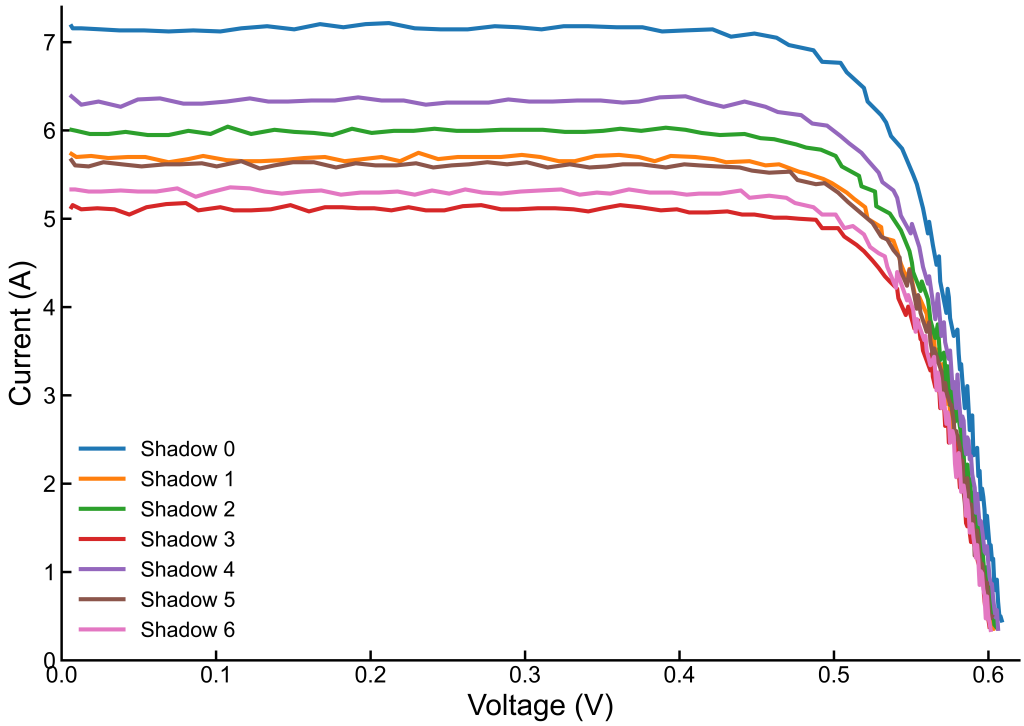


Figure 4. Different I-V curves of a cell with different shadows produced by the masks. Adapted from [17].

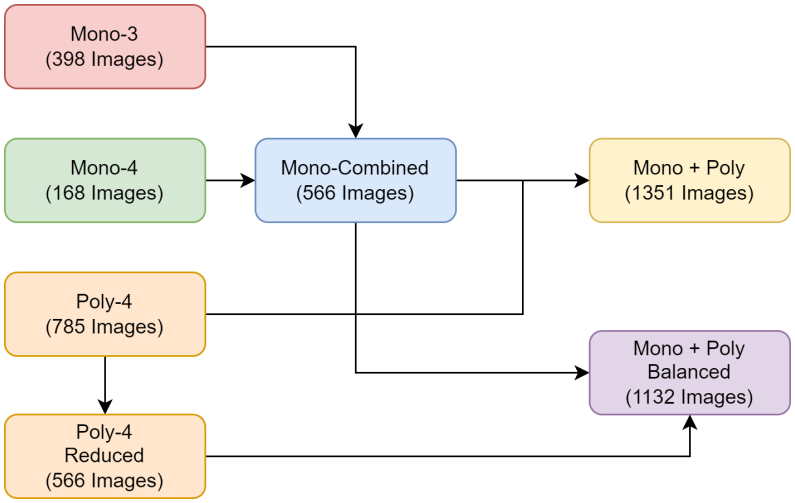


Figure 5. Datasets used in the experiments. Adapted from [17].

Figure 5 summarizes the datasets generated. The original datasets, Mono-3, Mono-4, and Poly-4, contain 398, 168, and 785 images, respectively. The Mono-Combined dataset was created by merging Mono-3 and Mono-4, resulting in 566 images. Subsequently, the Mono + Poly dataset combined Mono-Combined with Poly-4, yielding a total of 1351 images. Finally, to reduce class imbalance, the Poly-4 Reduced dataset was cre-

ated with 566 images, leading to the balanced Mono + Poly Balanced dataset containing 1132 images.

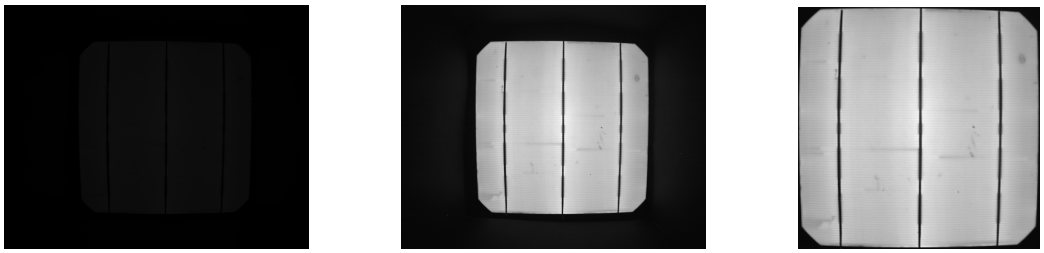
More information about the methodology used to take the measurements can be found in our previous work [16,18]. The data are publically available to download from the Zenodo Repository [18].

3.3. Preprocessing

The captured EL images needed to be processed to be suitable for the ML methods since they presented some problems:

- Presence of white pixels due to dead pixels of the camera lens or luminous noise. This problem was solved by removing these pixels from the images.
- Low brightness and contrast in the captured image. This problem was solved by a min-max normalization of the images, improving their visibility.
- Removing the dark contours of the images. This problem was solved by automatically finding the corners of the cells and cropping the image to the area occupied by the cell.

Figure 6 presents a sample in different stages of the preprocessing. A more detailed explanation of these processes can be found in [18].



(**a**) (**b**) (**c**)

Figure 6. An electroluminescence image of a photovoltaic cell during different stages of preprocessing. (**a**) Original image; (**b**) standardized image; (**c**) cropped image.

3.4. Labeling

The I-V Curve of a PV cell can be used as a reference for the performance of that cell, as it has been shown in Figure 4. There are different methods for expressing the performance of a PV cell but one of the most important is finding the Maximum Power Point (MPP) of the cell [19]. The MPP can be found as the absolute maximum of a Power–Voltage (P–V) curve of a cell which can be computed using the Current and Voltage values of the I–V curve. Figure 7 presents a sample with its I–V curve and P–V curve, indicating the MPP of the curve.

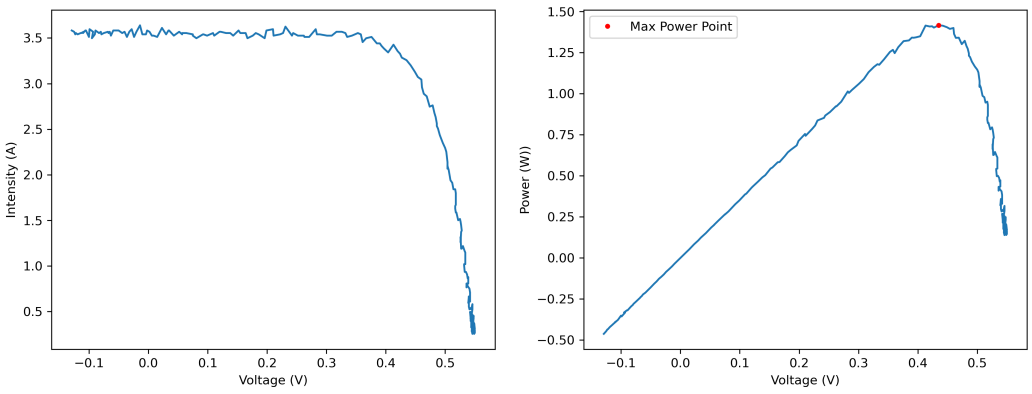


Figure 7. Intensity–Voltage curve and Power–Voltage curve of a polycrystalline PV cell.

The MPP of a cell depends on the irradiance of the light, so different irradiances will produce different MPPs. To solve this problem, it is necessary to standardize the value, making it independent of the value of the irradiance. The following processes are applied toward that objective:

- The measurements of the I-V curves are grouped by the current (Irradiance) used to obtain them. Each dataset is processed separately from the other datasets.
- The MPP of each cell is calculated, using the measured values of Current and Voltage.
- The mean value and standard deviation are computed for each group. These values are used to perform a Z-score normalization.
- The resulting values of the Z-score normalization present a mean of 0 and a standard deviation of 1. This obtained value will represent the MPP of a theoretical PV cell in a good state.
- A min-max normalization is performed to obtain a variable with a domain of [0, 1]. This value will provide a measure of the performance of the cell that is independent of the irradiance that was used to take the measurement.

The resulting value measures the relative performance of a cell, providing a value between $[0, 1]$. Low values correspond with cells that are not performing well, and high values with cells that are performing near their theoretical optimal condition.

3.5. Feature Extraction

The extraction of features in the original images are important due to different factors: Traditional machine learning algorithms are not suitable for dealing directly with images, so it is necessary to extract the most important aspects of the images before using these methods. Another important factor is that features provide an abstraction of the original images, reducing their complexity. This property can be extremely useful in improving the generalization capacities of the models.

After the manual selection process, features were automatically extracted from each image. The features that were selected include:

- **The mean, median, mode, and variance** of all pixels in the image. These statistical measures provide insights into the distribution of pixel values within the image.
- **The roughness of the image histogram**. This is determined by identifying the inflection points within the histogram. Inflection points are regions in the histogram where the slope changes direction, often indicating significant transitions in intensity. The roughness measure helps to characterize the texture and overall distribution of pixel intensities in the image.
- **The number of significant peaks** in the histogram. This measure counts the important peaks in the histogram, where each peak represents a dominant intensity value. Only the prominent peaks are considered; if the distance between two peaks is too small, the lower peak is discarded. This feature helps to quantify the distinct intensity levels present in the image, which may correlate with specific image features or objects.
- **The distance between the two highest peaks** in the histogram. This feature measures the separation between the two most significant intensity peaks in the image histogram, which can indicate the contrast or differentiation between the major intensity regions in the image.
- **The height and width of the two highest peaks**. These measurements give an idea of the prominence and spread of the two highest intensity peaks in the histogram. The height of the peak reflects the frequency of the intensity, while the width indicates how dispersed the intensity values around the peak are.
- **The count of dark, gray, and white pixels** in the image. Each pixel is classified into one of three categories based on its intensity value. Dark pixels are those with values

less than 0.55, gray pixels are those with values between 0.55 and 0.8, and white pixels have values greater than 0.8. These intervals were chosen after analyzing the distribution of histogram values across all images. The mean positions of the dark and gray peaks in the histogram were used as the thresholds to define these categories. The values for each category are expressed as proportions of the total number of pixels in the image, giving a relative measure of how much of the image falls into each category.

3.6. Metrics

For the evaluation of the performance of the AI models, two distinct metrics were selected:

- Mean Absolute Error (MAE):

$$\text{MAE} = \sum_{i=1}^{D} |x_{\text{real},i} - x_{\text{pred},i}|$$

The Mean Absolute Error calculates the average magnitude of the errors between predicted and actual values, where the error is measured as the absolute difference between the predicted value, $x_{\text{pred},i}$, and the real value, $x_{\text{real},i}$, for each data point i. This metric provides a straightforward evaluation of prediction accuracy, giving equal weight to all errors regardless of their size.

- Mean Squared Error (MSE):

$$\text{MSE} = \sum_{i=1}^{D} (x_{\text{real},i} - x_{\text{pred},i})^2$$

The Mean Squared Error computes the average of the squared differences between the predicted and actual values. The error for each data point is calculated as the square of the difference between $x_{\text{pred},i}$ and $x_{\text{real},i}$. This metric penalizes larger discrepancies more heavily than smaller ones, making it sensitive to outliers in the data.

Additional details regarding these metrics can be found in the relevant literature [20], as well as in our previous work [16].

These metrics were selected because they offer straightforward and effective methods for directly comparing the real values with the predicted values. Both the MAE and MSE provide clear insights into the accuracy of the model's predictions. Additionally, these metrics are easily computable using the tools and libraries employed in this study, making them practical for implementation in the evaluation process.

3.7. Models

Different AI methods were employed to assess the performance differences across each dataset. Each method was fine-tuned to optimize its task while maintaining a capacity for generalization.

The datasets used were partitioned into three distinct sets: Training (60%), Validation (20%), and Test (20%). The Training and Validation datasets are the only ones utilized during the training process in order to avoid bias for the Test set. Table 1 presents the distribution of each dataset into these subsets.

Table 1. Dataset split into training, validation, and test subsets.

Dataset	Total	Training (60%)	Validation (20%)	Test (20%)
Mono-3	398	239	79	80
Mono-4	168	101	33	34
Mono-Combined	566	340	113	113
Poly-4	785	471	157	157
Poly-4 Reduced	566	340	113	113
Mono + Poly	1351	811	270	270
Mono + Poly Balanced	1132	680	226	226

3.7.1. Convolutional Neural Network Model

Convolutional Neural Networks (CNNs) are the state-of-the-art method for dealing with images [21], even in the field of PVs [6]. CNNs do not need a previous feature extraction, since they can deal directly with the input image. The experiments were carried out in Python V3.10, using the Tensorflow V2.19.0 library [22].

Preprocessing

The images are processed as explained in Section 3.3. Subsequently, their size is reduced according to a scale factor, which is treated as a hyperparameter. This process adapts the images to the model input while improving computational efficiency and reducing memory consumption and training time, thereby enabling the algorithm to operate within the available hardware limitations.

Architecture

The process of designing the architecture of the model was performed as a manual process, increasing or reducing the complexity of the model depending on the metrics in the training process. The final architecture is composed of 13 different layers, using Convolutional 2D, Max Pooling, Flatten, and Dense layers. Figure 8 presents its architecture. This process was performed using the Mono + Poly dataset as input dataset, since it provides samples of the different kinds of PV cells.

Hyperparameter Optimization

Neural networks in general and CNNs have an important drawback since they are extremely dependent on their hyperparameters. However, the vast number of possible configurations makes the process of finding the best configuration manually almost impossible. For this reason, the Keras Tuner [23] library was used. It provides different methods to ensure the optimization of the hyperparameters of the NN models. The chosen algorithm, Bayesian Optimization Tuner [24], takes into account the results from previous trials to choose the next values; this informed search improves the exploitation of the search. The process was performed using the Mono + Poly dataset as the input dataset due to same reasons as before.

A summary of the hyperparameters for each model can be found in Figure 9.

Training

The seven different datasets were used to train seven different networks using the architecture and hyperparameters presented in the previous sections.

The training process was performed for 500 epochs, saving the model with lower MAE as the final model. Figure 10 shows the evolution of MAE in the training set and validation set for each model. Herein, the difference in performance across the different datasets can be observed, which will be discussed in Section 4.

Traditional machine learning methods have been used for regression tasks in different fields [25] due to their low computing requirements and good performance.

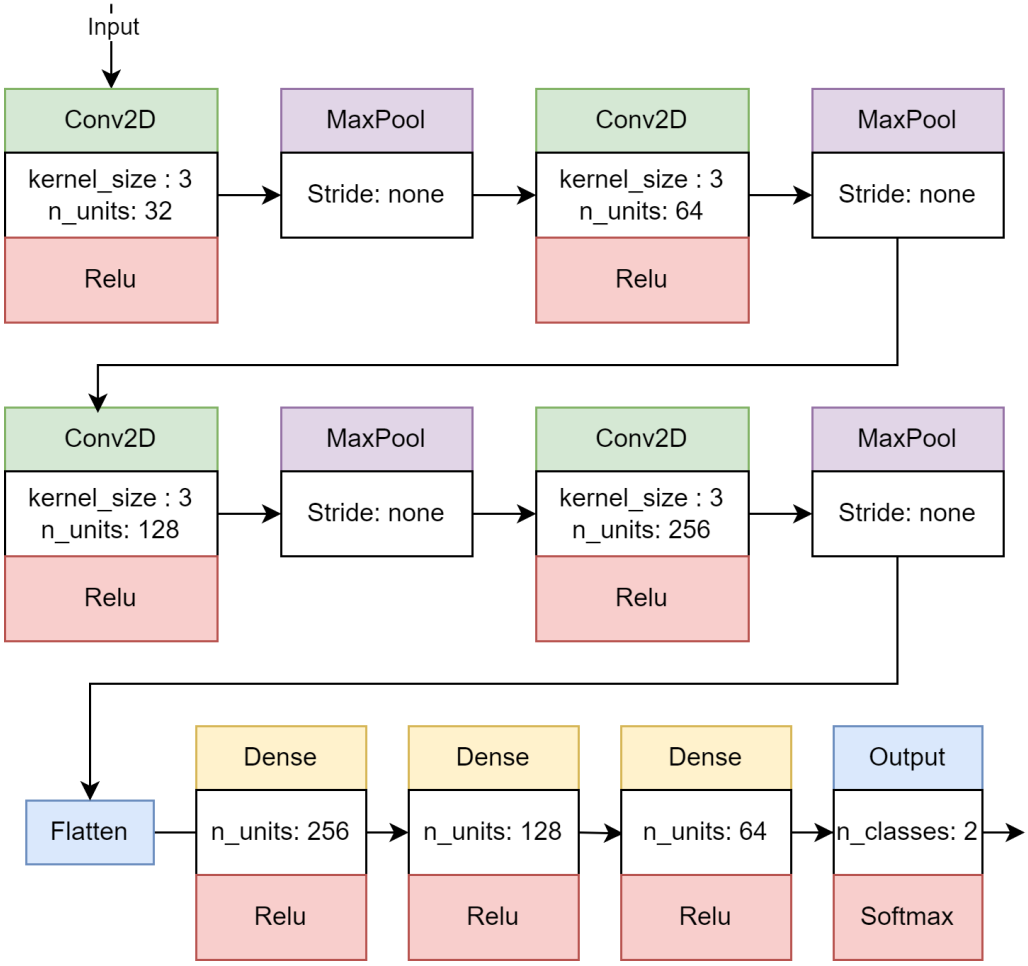


Figure 8. The optimal architecture found for the convolutional model. Adapted from [17].

Mono model	
batch size	8
learning rate	0.00025
scale factor	0.66
optimizer	Adam

Mono-3 model	
batch size	8
learning rate	0.0005
scale factor	0.66
optimizer	Adam

Mono-4 model	
batch size	8
learning rate	0.00025
scale factor	0.66
optimizer	Adam

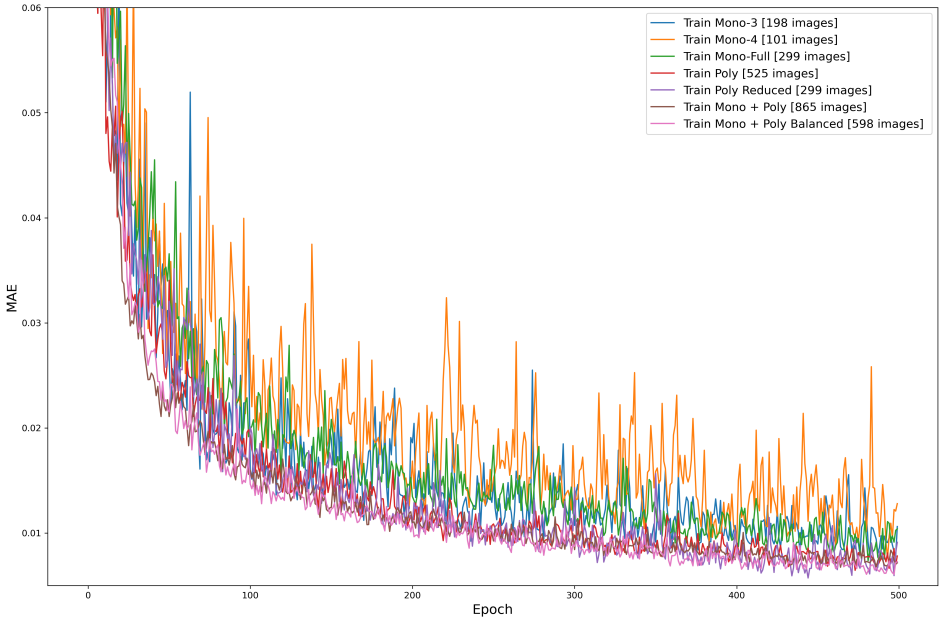
Poly-Complete model	
batch size	8
learning rate	0.00025
scale factor	0.66
optimizer	Adam

Poly-Reduced model	
batch size	8
learning rate	0.00025
scale factor	0.66
optimizer	Adam

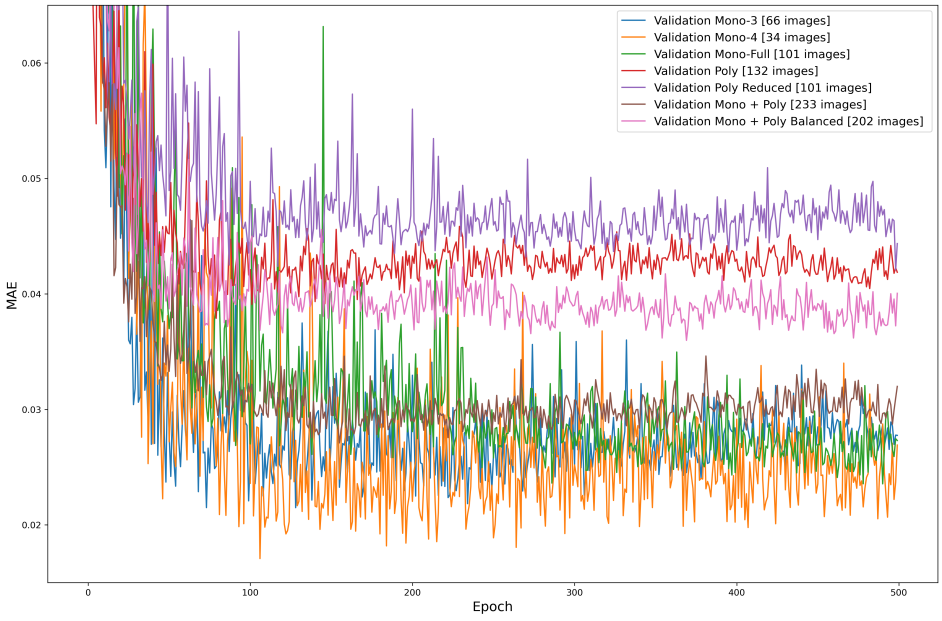
Mono + Poly model	
batch size	8
learning rate	0.001
scale factor	0.66
optimizer	Adam

Mono + Poly Balanced model	
batch size	8
learning rate	0.001
scale factor	0.66
optimizer	Adam

Figure 9. Optimal hyperparameters identified for Convolutional Neural Networks include batch sizes $\in [8, 16, 32, 64]$, learning rates $\in [0.01, 0.0001]$, scale factors $\in [0.5, 1]$, and optimizers such as Adam, Nadam, and SGD. Adapted from [17].



(**a**)



(**b**)

Figure 10. Progression of the mean absolute error during the training process of the CNN models. (**a**) Mean absolute error on the training set; (**b**) mean absolute error on the validation set.

Algorithms

The different algorithms explained below were implemented in Matlab 2024A using the Regression Learner application.

- Gaussian Process Regression (GPR) [26]: A supervised machine learning technique employed for regression and probabilistic classification tasks. Predictions are probabilistic in nature, making it particularly effective for small datasets. The method supports various kernels, such as Exponential, Matern 5/2, Rational Quadratic, and others.
- Ensemble Bagged Tree [27]: A machine learning technique that employs an ensemble of regression decision trees combined using the bagging (Bootstrap Aggregating) method. This approach enhances the model's predictive performance by reducing

variance and mitigating overfitting. By training multiple trees on different subsets of the data and averaging their predictions, the ensemble achieves greater stability and accuracy compared to individual decision trees.

- Support Vector Machine (SVM) [28]: A robust and versatile supervised machine learning algorithm commonly used for classification and regression tasks. Its flexibility comes from the wide range of kernel functions available which allow it to model complex, non-linear relationships in the data. SVMs are known for their ability to handle high-dimensional datasets efficiently while maintaining robustness against overfitting, particularly when regularization is applied. The method aims to find the hyperplane that best separates the classes in a dataset, maximizing the margin between data points of different classes. This makes it highly effective for tasks involving complex decision boundaries.
- Neural Network (NN) [29]. NNs provide extremely high performance, but they require high amounts of data and can have high computation demand depending on the number of layers.

Preprocessing

The images are processed as explained in Section 3.3. After that, the features of the images are extracted as explained in Section 3.5.

Model Selection and Hyperparameter Optimization

The model selection process was conducted using the Mono + Poly Dataset, as it includes samples representing all types of PV systems under consideration. All the described algorithms were trained with their hyperparameters optimized using a Bayesian optimization approach, which is integrated directly into the application. Figure 11 illustrates the results of the five best-performing models along with their corresponding hyperparameters. Among these, the Exponential GPR emerged as the top-performing algorithm in terms of both mean absolute error and mean squared error.

Bagged Trees	
MAE	0.0729
MSE	0.0110
Maximum leaf size	8
Number of learners	30

Rational Quadratic GPR	
MAE	0.718
MSE	0.0100
Basic Function	Constant
Kernel Function	Rational Quadratic

Exponential GPR	
MAE	0.0698
MSE	0.0095
Basic Function	Constant
Kernel Function	Exponential

Medium Gaussian SVM	
MAE	0.0822
MSE	0.0119
Kenel Scale	3.9
Kernel Function	Gaussian

Trilayered NN	
MAE	0.0856
MSE	0.0149
Number and size of layers	3 layers: [10,10,10]
Activation Function	Relu

Figure 11. Results of various machine learning methods applied to the Mono + Poly dataset. GPR: Gaussian Process Regression, SVM: Support Vector Machine, NN: Neural Network, MAE: Mean Absolute Error, MSE: Mean Squared Error. Adapted from [17].

Feature Selection

Parameter selection is extremely useful to find the optimal configuration of the input features. Table 2 presents all the features after applying the F-test algorithm [30] to rank them. Figure 12 shows the performance of the chosen model with different amounts of

features; it can be seen how removing features improves performance, with eleven features being the best-performing experiment. Given this result, only these eleven features were considered in the subsequent experiments.

Table 2. Features sorted according to their F-score. The table displays the ranking of various features based on their F-scores, with higher values indicating more important features for the analysis. Adapted from [17].

Feature	F-Type Score	Feature	F-Type Score
Blacks	440.83	Peak 1 Width	29.47
Grays	405.62	Whites	21.88
Mean	302.85	Mode	21.06
Kurtosis	245.48	Peaks Distance	18.73
Peak 0 Height	190.62	Peak 1 Height	13.96
Roughness	189.67	Peak 1 Height	13.96
Variance	138.90	Peak 0 Width	6.71
Median	60.22	Peaks Number	1.61

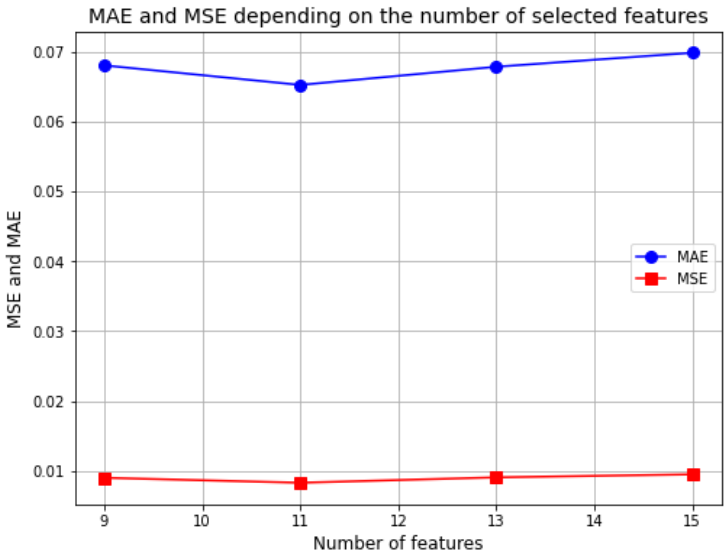


Figure 12. Validation set results for the GPR-Exponential method with Feature Selection applied to remove varying amounts of features, based on their F-score, eliminating the least important first. The red line represents the mean squared error, while the blue line represents the mean absolute error. Adapted from [17].

3.7.2. Convolutional Hierarchical Mixture of Experts

The Mixture of Experts approach, as outlined by [31], is a machine learning technique designed to decompose a complex problem into a series of smaller, more manageable subproblems. This methodology is particularly valuable for the problem at hand, as it allows each type of photovoltaic cell to be treated as a distinct subproblem. Among the various forms of MoE, the Hierarchical Mixture of Experts model is specifically employed,

as discussed in [32]. Given that the models in question are based on Convolutional Neural Networks (CNNs), the proposed architecture is referred to as the Convolutional Hierarchical Mixture of Experts (CHME).

Unlike conventional Mixture-of-Experts architectures, where a gating network learns a probabilistic routing policy and may suffer from expert collapse by assigning most samples to a single expert, the proposed CHME employs a supervised discriminator trained to classify the photovoltaic cell technology of the input electroluminescence image. The predicted cell type deterministically selects the corresponding expert (Mono-3, Mono-4, or Poly), eliminating the need for probabilistic routing during both training and inference. Consequently, each expert is trained exclusively using the samples belonging to its associated cell technology, preventing competition among experts and inherently avoiding the expert collapse problem. Since the routing mechanism is explicitly supervised rather than learned through optimization of a gating distribution, no additional load-balancing regularization loss is required in the proposed architecture.

Preprocessing

The images are processed, as in the case of the simple CNN models: improving the quality of the images (Section 3.3) and reducing their size. This model directly uses the images, since it is based on the convolutional models.

Architecture

This approach presents a Tree-Structured architecture (see Figure 13), the highest level is composed of a Discriminator model, capable of identifying which type is the input. After that, in the next level, a different regression model is chosen depending in the decision of the discriminator model. The architecture of the discriminator model was chosen using a manual process, changing the architecture to optimize the performance of the model. The final architecture can be seen in Figure 14. The regression models were obtained from Section 3.7.1, maintaining their architecture and hyperparameters.

Hyperparameter Optimization

The process of optimizing the parameters of networks was performed using Keras Tuner as in Section 3.7.1. Figure 15 presents the hyperparameters for all the networks, with the regression networks maintaining their hyperparameters.

Training

The training of the four networks was performed individually. The three regression experts were trained independently using only the samples corresponding to their respective photovoltaic technologies, while the discriminator was trained separately as a supervised multi-class classifier. Consequently, the routing mechanism was fixed after training and did not require joint optimization with the regression experts.

The details of the trainings M3, M4, and MP can be found in Section 3.7.1. The discriminator network was trained for 20 epochs using Binary Cross Entropy [33] as the loss function. Accuracy was used as an additional metric. The evolution of both metrics during the training process can be observed in Figure 16. It can be seen that by epoch 20, the model can classify the different types of PV cells with 100% accuracy and high confidence, as indicated by its low Binary Cross Entropy. The training process requires a lower amount of epochs than the regressor models since recognizing the kind of PV cell is a different, less complex problem than estimating the output power. It can also be observed how the model does not suffer from overfitting, since the training and validation values are extremely similar in the last epochs.

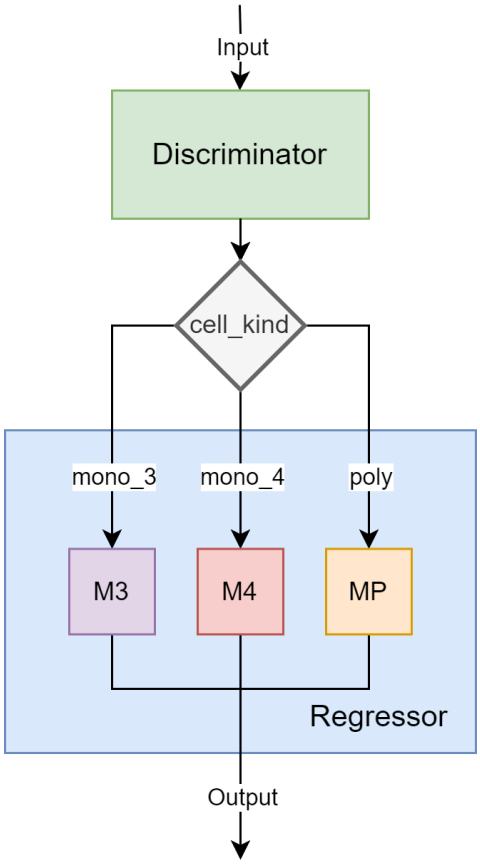


Figure 13. Architecture. of the Convolutional Hierarchical Mixture of Experts model. The `Cell_kind` represents the output of the discriminator network, which is responsible for selecting the appropriate expert for a given input. The `M3` network is a Convolutional neural network trained exclusively with the Mono-3 dataset, focusing on one type of photovoltaic system. Similarly, the `M4` network is trained solely with the Mono-4 dataset, and the `MP` network is trained with the Poly dataset. Each of these networks is specialized in processing data corresponding to their respective datasets. The final output is determined by the mixture of experts mechanism, where the discriminator directs the input to the most relevant network based on the characteristics of the input data. Adapted from [17].

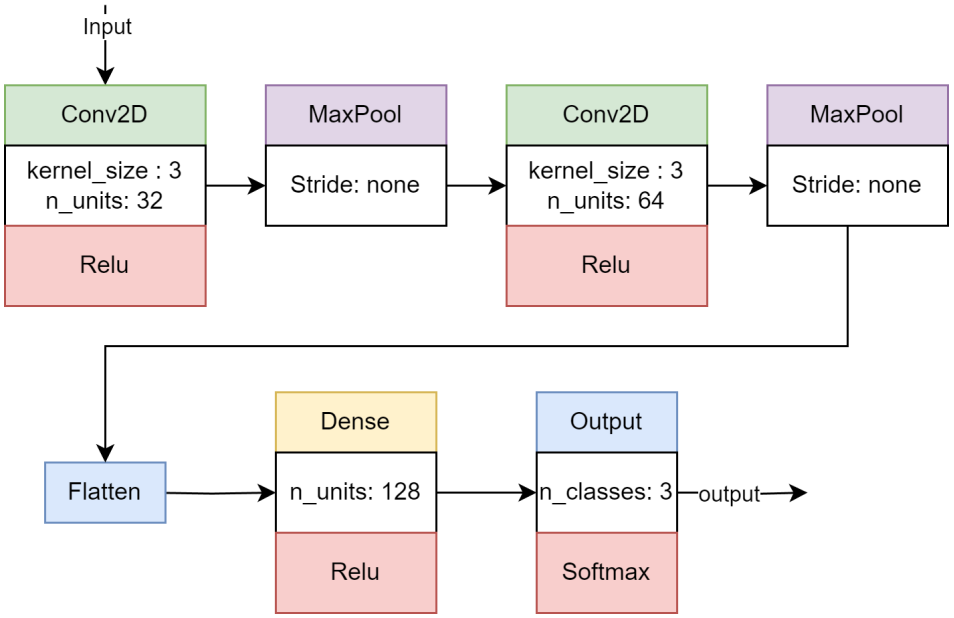


Figure 14. Architecture of the discriminator model.

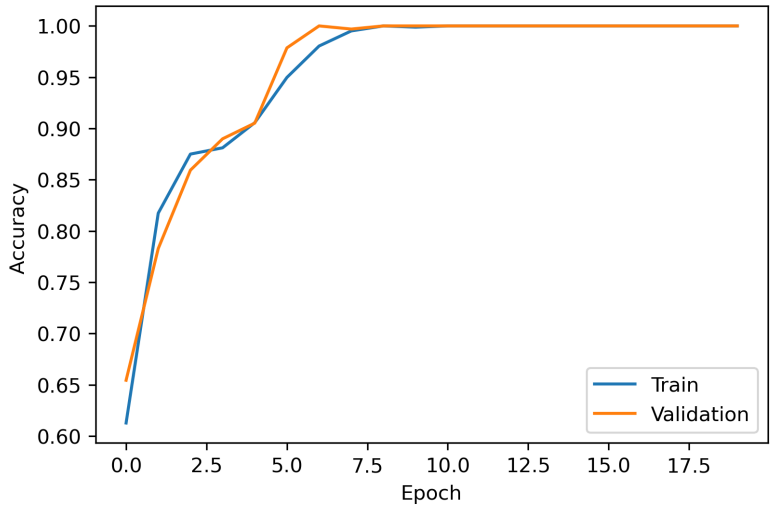
Discriminator	
batch size	8
learning rate	0.00025
scale factor	0.66
optimizer	Adam

Mono-3 model	
batch size	8
learning rate	0.0005
scale factor	0.66
optimizer	Adam

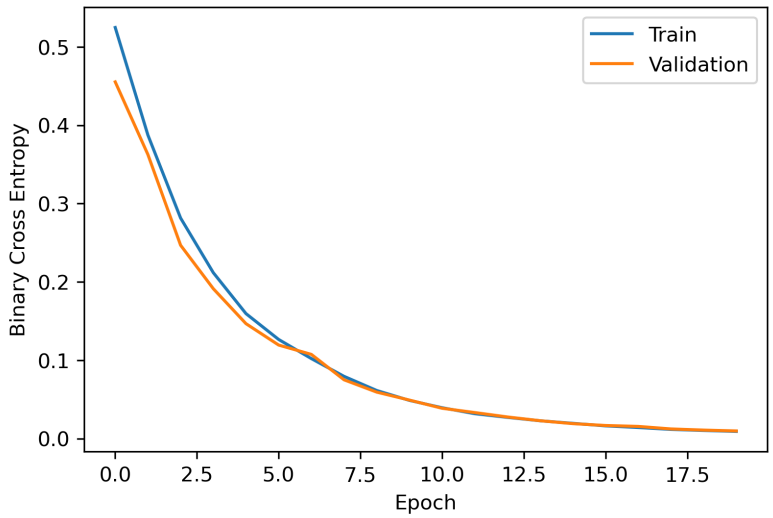
Mono-4 model	
batch size	8
learning rate	0.00025
scale factor	0.66
optimizer	Adam

Poly-Complete model	
batch size	8
learning rate	0.00025
scale factor	0.66
optimizer	Adam

Figure 15. Optimal hyperparameters found for each of the networks of the Convolutional Mixture of Experts: batch_size $\in [8, 16, 32, 64]$, learning_rate $\in [0.01, 0.00001]$, scale_factor in $[0.5, 1]$, optimizer: Adam, Nadam, SGD. Adapted from [17].



(**a**)



(**b**)

Figure 16. Progression of key metrics used to evaluate the performance of the discriminator network during its training. The accuracy and binary cross entropy provide insights into how well the discriminator network learns to classify and minimize errors. Adapted from [17]. (**a**) Evolution of the accuracy metric during the training process of the discriminator network; (**b**) evolution of the binary cross entropy loss during the training process of the discriminator network.

4. Results

This section analyses the performance of the proposed models for each dataset. First, the CNN models will be addressed. After that, the feature-based methods are tested. Finally, the combination of experts based on CNN is compared with the other methods.

4.1. Comparison of the CNN Models Trained with Different Datasets

The experiments were designed as explained in Section 3.7. Each model was tested with various sets, ensuring that any sample used in the training set and validation set appears in them. The following facts can be observed from their results (Table 3).

- As it can be observed, the results in the Test sets are extremely similar to the Validation sets, this ensures the generalization capacities of the models.
- The Mono Complete model is capable of dealing with any kind of monocrystalline image presented in the dataset with an MAE of 0.0194, which shows that it can find the most important patterns for dealing with these kinds of images in the training phase. However, these patterns do not seem to be suitable for dealing with polycrystalline images, which can be seen in the MAE obtained when testing the model with the Poly dataset (0.2169); this value indicates that the model is not capable of predicting the output power for these kinds of PV cells.
- The Mono-3 is capable of dealing with monocrystalline images with 3 busbars (MAE = 0.0211). The performance when dealing with the images from Mono-4 is quite worse (MAE = 0.1283), but it could be considered enough in some situations. The results demonstrate that this model is not optimal for dealing with every kind of monocrystalline PV cell presented in the datasets. However, it can still provide valid performance (Mono Complete MAE = 0.0758). The performance with the polycrystalline dataset is similar to that of the previous model.
- The Mono-4 model has a similar behavior to the previous one, showing good performance for Mono-4 (MAE = 0.0216) and a valid performance in Mono-3 (MAE = 0.1190). This model could be used to deal with any kind of mono PV cell but it would not lead to optimal results. The performance with polycrystalline cells is similar to the previous models.
- The Poly Complete model shows a great performance in its test set with an MAE of 0.0337; however, this result is lower than the Mono Complete result. This difference could be explained given that polycrystalline PV cells present more irregularities in their surface, as can be observed in Figure 2. The situation with the Mono-Complete dataset is similar to the previous cases. The model is not capable of dealing with this kind of PV cell since they present different patterns than the polycrystalline cells (MAE = 0.2228). This behavior happens with both kinds of monocrystalline PV cells (MAE = Mono-3 0.2030 and Mono-4 0.2628).
- The performance of Poly Reduced is similar to that of the Complete model (MAE = 0.0339). The performance in the mono datasets is quite similar to that of the Complete model (MAE = Mono Complete 0.2167, Mono-3 0.1919 and Mono-4 0.2690).
- The combination of monocrystalline and polycrystalline datasets results in an MAE of 0.0288. The model achieves an MAE of 0.0265 for the monocrystalline dataset, with 0.0259 for Mono-3 and 0.0281 for Mono-4. For the polycrystalline dataset, the model performs with an MAE of 0.0288. This demonstrates that the performance across all datasets is remarkably similar, indicating that training the model with both datasets enables it to learn the diverse patterns necessary to handle all types of PV cells effectively.
- The results of the Mono + Poly balanced are similar to those of the not-balanced dataset, with a Test MAE of 0.298. The performance in mono corresponds to an MAE

of 0.0312, with 0.0295 for Mono-3 and 0.0351 for Mono-4. The MAE for poly is 0.0288. These results are quite similar to the previous model, but they are still a bit lower due to the reduction in samples of polycrystalline cells in the training phase.

Table 3. Results of CNN-based models, comparing the performance across various configurations. The metrics used are mean absolute error (MAE) and mean squared error (MSE), showing the evaluation results for different models, including Mono Complete, Mono Busbar-3, Mono Busbar-4, Poly Complete, Poly Reduced, Mono + Poly, and Mono + Poly Balanced. The table highlights the best-performing configurations in terms of MAE and MSE for the validation and test datasets. Bold items are the best performing values. Adapted from [17].

	Metrics	Validation	Test	Mono Complete	Mono Busbar-3	Mono Busbar-4	Poly	Mono + Poly
Poly Complete	MAE	0.0313	0.0337	0.2228	0.2030	0.2628	0.0341	0.1211
	MSE	0.0019	0.0022	0.0728	0.0617	0.0960	0.0024	0.0351
Poly Reduced	MAE	0.0435	0.0339	0.2167	0.1919	0.2688	0.0339	0.1290
	MSE	0.0042	0.0021	0.0700	0.0551	0.1014	0.0021	0.0345
Mono Complete	MAE	0.0207	0.0194	0.0194	0.0221	0.0255	0.2159	0.1195
	MSE	0.0010	0.0006	0.0006	0.0009	0.0010	0.0639	0.0340
Mono Busbar-3	MAE	0.0184	0.0214	0.0758	0.0211	0.1283	0.1668	0.1488
	MSE	0.0008	0.0008	0.0124	0.0008	0.0239	0.0439	0.0307
Mono Busbar-4	MAE	0.0216	0.0215	0.0753	0.1190	0.0215	0.2351	0.1913
	MSE	0.0008	0.0007	0.0122	0.0223	0.0007	0.0786	0.0586
Mono + Poly	MAE	0.0281	0.0288	0.0265	0.0259	0.0281	0.0301	**0.0288**
	MSE	0.0015	0.0015	0.0012	0.0011	0.0013	0.0018	**0.0015**
Mono + Poly Balanced	MAE	0.0246	0.0298	0.0312	0.0285	0.0351	0.0305	0.0298
	MSE	0.0012	0.0018	0.0016	0.0012	0.0019	0.0019	0.0018

As can be seen, the performance of the models decreases with other kinds of PV cells. This can be explained by the structural differences between the different cells. As shown in Figure 2, the number of busbars changes the basic structure of the cell. Moreover, the surfaces of monocrystalline cells and polycrystalline cells are also completely different, being that the monocrystalline cells are more homogeneous than the polycrystalline cells.

4.2. Comparison of the Feature-Based Models Trained with Different Datasets

The main objective of the feature-based models is to reduce the effects of the structure of the cells in the generalization capacities of the models. These features, as explained, were obtained by analyzing the most important characteristics of the images and their histograms. The models were trained and tested exactly with the same samples as in the previous section. Table 4 presents the results of these experiments.

It can be seen how, in general, the models obtain worse results than the CNN models, but they still provide results that can be considered good (Mean MAE of 0.06); however, it can also be seen how there is a major difference. The Mono-3 and Mono-4 models have extremely similar results, which shows how the patterns found in their training phase are useful for classifying the other kinds of PV cells. These results shows how this feature-based regression can reduce the effects of the structure of the PV cells on the performance. However, there is still a difference between the Mono models and Poly models, which shows how the feature-based proposal cannot override the effects of the surface features of the PV cells.

Table 4. Results of the best feature-based method, comparing the performance across various configurations. The metrics used are mean absolute error (MAE) and mean squared error (MSE), with the evaluation results shown for different models, including Mono Complete, Mono Busbar-3, Mono Busbar-4, Poly Complete, Poly Reduced, Mono + Poly, and Mono + Poly Balanced. The table highlights the best-performing configurations in terms of MAE and MSE for validation and test datasets. Bold items are the best performing values. Adapted from [17].

	Metrics	Validation	Test	Mono Complete	Mono Busbar-3	Mono Busbar-4	Poly	Mono + Poly
Poly Complete	MAE	0.0540	0.0521	0.1881	0.1406	0.2515	0.0540	0.1534
	MSE	0.0056	0.0051	0.0536	0.0330	0.0813	0.0056	0.0416
Poly Reduced	MAE	0.0524	0.0530	0.1897	0.2523	0.2688	0.0524	0.1442
	MSE	0.0058	0.0059	0.0543	0.0806	0.1014	0.0058	0.0375
Mono Complete	MAE	0.0662	0.0648	0.0648	0.0699	0.0531	0.2114	0.1724
	MSE	0.0081	0.0071	0.0071	0.0081	0.0037	0.0634	0.0476
Mono Busbar-3	MAE	0.0651	0.0581	0.0591	0.0581	0.0604	0.2092	0.1771
	MSE	0.0075	0.0058	0.0068	0.0055	0.0053	0.0639	0.0532
Mono Busbar-4	MAE	0.0668	0.0458	0.0654	0.0853	0.0458	0.2151	0.1846
	MSE	0.0044	0.0036	0.0076	0.0124	0.0036	0.0706	0.0526
Mono + Poly	MAE	0.0698	0.0748	0.0864	0.0777	0.1000	0.0698	0.0748
	MSE	0.0095	0.0108	0.0119	0.0093	0.0162	0.0095	0.0108
Mono + Poly Balanced	MAE	0.0778	0.0681	0.0677	0.0664	0.0712	0.0686	**0.0681**
	MSE	0.0113	0.0092	0.0083	0.0078	0.0095	0.0103	**0.0092**

4.3. Performance of the Convolutional Hierarchical Mixture of Experts

As seen previously, training models with images of all the kinds of PV cells is the only way to obtain good performance. Using an abstraction method like feature extraction lightly reduced this behavior; however, it also considerably reduced the performance of the models. The high specialization of the models becomes a problem, since they are not capable of dealing with other kinds of PV cells. This property can be used to obtain a complex model that is more capable of dealing with different kinds of PV cells and a model trained by combining images of various PV kinds.

The proposed model, Convolutional Hierarchical Mixture of Experts (CHME), was explained in Section 3.7.2. Each specialized network was trained with the training set of its respective dataset, using the same methodology as the convolutional models. The model was tested using all the Test sets from the used datasets (Mono-3, Mono-4, and Poly Completed) and obtained an MAE of 0.0262 and an MSE of 0.0012. These values are a bit lower than the results obtained by the Mono + Poly model in Table 3.

Figure 17 presents the Box Plot model of all the CNN-based models and the CHME model. It can be seen how the amount of outliers is less than the those from the Mono + Poly and Mono + Poly balanced datasets, and in general, they are of lower value (with a sole exception). These values are additional evidence of the better performance of the model. This model is capable of obtaining the best performance of each individual model, since they are only used when the discriminator decides that a cell is suitable for that network. It also has other important advantages, since it is not necessary to completely retrain the model if new data is available. The retraining would only affect the discriminator network—which has quick training as shown previously—and the corresponding network. A new network would have to be created if a new kind of PV cell were presented.

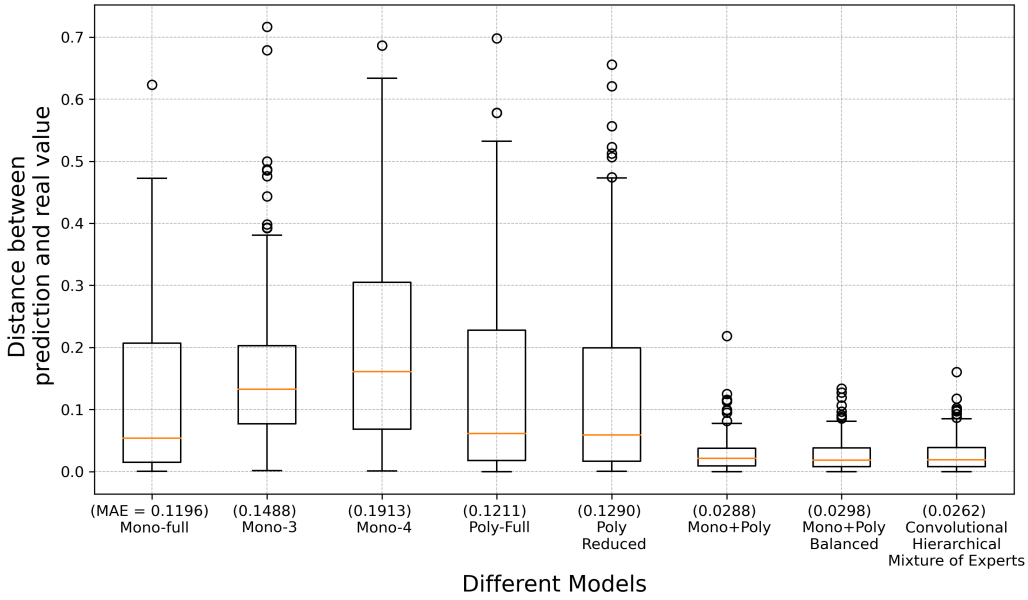


Figure 17. Boxplot comparing the distance between predictions and actual values for all CNN-based models, as well as for the CHME model. The plot visually represents the distribution of errors for each model, highlighting the median, quartiles, and potential outliers in the predictions. Additionally, the mean absolute error (MAE) for each model is provided, offering a quantitative measure of the average absolute difference between the predicted and true values. This comparison allows for a clear understanding of the model's performance in terms of prediction accuracy across different models. Adapted from [17].

5. Conclusions

The prediction of the output power of PV cells is a tricky problem due to different factors, such as the limitations of having to measure new cells when training models or the vast amount of different technologies used to build PV cells. Different models were tested in order to examine the effects of various technologies on the performance of the models. The results show how the Convolutional Models are able to obtain low error values if they deal with the same kinds of PV cells used for their training, but they are unable to obtain great results when they deal with new kinds of PV cells. This limitation also appears in the Feature-based methods. It can be concluded that the patterns presented in one kind of PV cell are notable enough that it becomes difficult to understand the patterns presented in the other kinds of PV cells, making it necessary to train the models with the kinds of PV cells that are going to be used.

The combined method, called the Convolutional Hierarchical Mixture of Experts, uses various Convolutional Networks that are trained with different kinds of PV cells. This method is able to obtain better results than a single Convolutional Network trained with all the different kinds of PV cells. This method takes advantage of the high specialization observed in the Convolutional Models, using the most appropriate model for each kind of PV cell. The CHME model is also easier to retrain than the single Convolutional, since only the discriminator (which is fast to train) and the corresponding network need to be retrained when data from a new kind of PV cell becomes available.

Regarding its practical applicability, the CHME model is highly recommended for automated quality control lines in PV manufacturing facilities and high-throughput laboratory settings where electroluminescence (EL) imaging is standard. Implementing this model allows production plants to automatically estimate the individual maximum power output of different cell lines (e.g., mono- and polycrystalline) right after fabrication, bypassing the temporal cost of continuous flash testing. In large-scale operational solar plants, this

model can be integrated into drone-based or mobile EL inspection workflows to diagnose and quantify severe localized degradation or mechanical defects, mapping specific cell degradation directly to power loss estimations.

All in all, the results have shown how it is possible to create models for different kinds of PV cells with low error, with the CHME being the best model in terms of error and facility to retrain. However, these models have the limitation of needing to be trained with data of a PV kind in order to have a good performance.

5.1. Hardware Requirements and Economic Feasibility

The practical implementation of the proposed CHME model does not require substantial extra equipment or prohibitive costs, as it seamlessly integrates into standard PV inspection workflows.

For a **new or existing utility-scale power system**, EL testing is already widely adopted as a standard benchmark for quality assurance during commissioning or periodic maintenance. The model utilizes standard EL images, which are typically captured using an InGaAs camera already owned by maintenance contractors or integrated into automated manufacturing lines.

The primary components and associated cost factors for deploying the CHME model are detailed below:

- **Data Acquisition (Existing Equipment):** No extra hardware investment is required on-site if EL testing infrastructure (such as standard digital or InGaAs cameras and custom power supplies to inject current into the PV strings) is already available.
- **Computational Infrastructure (Low-to-Medium Cost):** The trained CHME framework can execute inference locally on a standard industrial PC equipped with a consumer-grade Graphical Processing Unit, or remotely via a cloud computing platform. Because model inference takes only milliseconds per cell image, high-end enterprise servers are unnecessary.
- **Integration and Maintenance (Software Cost):** The deployment expenses are limited to software engineering for API integration—linking the image capture software to the Python/TensorFlow environment hosting the CHME architecture. Furthermore, the modular nature of the CHME minimizes long-term software maintenance costs; if a new PV cell technology is adopted, only a specific sub-network and the quick-training discriminator need updating, avoiding costly system-wide re-development.

5.2. Limitations

The first limitation concerns the experimental generation of the degradation labels. The different degradation levels were simulated by introducing controlled shading using opaque materials such as cardboard and aluminum foil. Although this approach provides a repeatable and controlled environment for acquiring paired electroluminescence images and I-V measurements, it does not fully reproduce the wide variety of degradation mechanisms found in real photovoltaic cells, such as cracks, delamination, PID, corrosion, or soldering failures. In particular, the artificial masks exhibit regular geometric patterns with well-defined edges that differ from the irregular spatial structures typically observed in naturally occurring defects. Consequently, there is a risk that deep learning models may partially learn these artificial geometric characteristics rather than the underlying physicochemical features associated with real degradation processes. Although the diversity of mask shapes and positions was designed to reduce this effect, additional validation using electroluminescence images of naturally degraded PV cells is required to assess the true generalization capability of the proposed framework.

Furthermore, the proposed methodology requires representative training data for each photovoltaic technology considered. As shown in the experimental results, convolutional models are highly specialized and their performance decreases significantly when they are applied to cell technologies that were not included during training. Although the CHME architecture mitigates this limitation by assigning a dedicated expert to each cell type, a new expert must still be trained whenever a previously unseen PV technology is introduced.

The experiments were conducted under controlled laboratory conditions using electroluminescence images acquired with the same imaging system and illumination setup. Consequently, the robustness of the proposed approach under different acquisition systems or environmental conditions has yet to be evaluated.

The proposed framework also relies on experimentally measured I-V curves, which inherently contain measurement uncertainty associated with the electrical acquisition system and the illumination setup. Although all measurements were acquired under controlled laboratory conditions using the same I-V tracer, LED illumination system, acquisition protocol, and temperature monitoring procedure, small variations due to instrumental noise and illumination non-uniformity are unavoidable. To mitigate these effects, all samples were acquired under identical conditions and the measured current signals were smoothed using a Savitzky–Golay filter before extracting the electrical parameters. Nevertheless, measurement uncertainty remains an inherent limitation of the generated dataset and may introduce a small amount of variability into the regression targets.

Finally, the performance labels are obtained from normalized MPP values. While this normalization makes the labels independent of irradiance and reduces the influence of measurement variability, the proposed framework estimates the relative electrical performance of a PV cell rather than its absolute output power under arbitrary operating conditions.

5.3. Future Work

Future research will focus on validating the proposed methodology using electroluminescence images of naturally degraded photovoltaic cells acquired under real operating conditions. Particular attention will be devoted to evaluating whether the models trained using controlled artificial degradation are able to generalize to realistic degradation mechanisms, including PID, microcracks, corrosion, dust accumulation, and other naturally occurring defects with irregular spatial distributions. In particular, transfer learning techniques will be investigated to adapt the models trained on controlled degradation scenarios to real degradation conditions while minimizing the amount of additional labeled data required.

Another important research direction is the development of more general feature extraction strategies capable of learning technology-independent representations. Such approaches could reduce the need to train a dedicated expert for each photovoltaic technology and improve the generalization capability of the proposed framework when new PV cell types are introduced.

Future work will also extend the dataset by incorporating additional photovoltaic technologies, manufacturing processes, and degradation mechanisms. Increasing the diversity of the training data will enable a more comprehensive evaluation of the scalability and robustness of the proposed Convolutional Hierarchical Mixture of Experts architecture.

Furthermore, the proposed methodology will be evaluated under different acquisition systems and imaging conditions to assess its robustness against variations in camera characteristics, illumination conditions, and measurement setups. This evaluation will help determine the applicability of the proposed framework in practical inspection environments.

Finally, future work will investigate alternative deep learning architectures, including Vision Transformers and self-supervised learning methods, to improve prediction accuracy while reducing the dependence on labeled data. Model optimization and compression

techniques will also be explored to facilitate the deployment of the proposed framework on embedded and edge-computing devices for real-time photovoltaic inspection.

Author Contributions: Conceptualization, H.F.M.-R.; methodology, H.F.M.-R.; software, H.F.M.-R.; validation, H.F.M.-R., L.H.-C. and M.Á.G.R.; formal analysis, H.F.M.-R.; investigation, H.F.M.-R., V.C.-P., V.A.G., L.C.-V., J.I.M.A. and S.G.S.; resources, L.H.-C., M.Á.G.R. and A.M.P.; data curation, H.F.M.-R.; writing—original draft preparation, H.F.M.-R.; writing—review and editing, H.F.M.-R., L.H.-C., M.Á.G.R., V.C.-P., V.A.G., L.C.-V., J.I.M.A., S.G.S., A.M.P. and M.C.d.R.; visualization, H.F.M.-R.; supervision, L.H.-C. and M.Á.G.R.; project administration, L.H.-C.; funding acquisition, L.H.-C. and M.Á.G.R. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Universidad de Valladolid with the 2020 predoctoral contracts, co-funded by Santander Bank. This work was also supported by the Spanish Ministry of Science, Innovation, and Universities through the "Plan Estatal de Investigación Científica, Técnica y de Innovación" (project ID: PID2023-148369OB-C43). This study was also supported by the Universidad de Valladolid with ERASMUS+ KA-107 and with ERASMUS+ PRACTICAS. We also appreciate the help of the project "Identificación de Fallas en Instalaciones Solares Fotovoltaicas (cod: 1360051) of Tecnológico de Costa Rica" for hosting Hector Felipe Mateo Romero during their stay.

Data Availability Statement: The data can be downloaded from https://doi.org/10.5281/zenodo.14439369 [34].

Conflicts of Interest: The authors declare no potential conflicts ofinterest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
CHME	Convolutional Hierarchical Mixture of Experts
CNN	Convolutional Neural Network
EL	Electroluminescence
GAN	Generative Adversarial Network
GPR	Gaussian Process Regression
I-V	Current–Voltage
LED	Light Emitting Diode
M3	Mono-3 expert network
M4	Mono-4 expert network
MP	Poly-4 expert network
MAE	Mean Absolute Error
ML	Machine Learning
MoE	Mixture of Experts
MPP	Maximum Power Point
MPPT	Maximum Power Point Tracking
MSE	Mean Squared Error
NN	Neural Network
P–V	Power–Voltage
PID	Potential-Induced Degradation
PL	Photoluminescence
PV	Photovoltaic
SDGs	Sustainable Development Goals
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
YOLO	You Only Look Once

References

1. Brookings Institution. *How the COVID-19 Pandemic and Geopolitical Conflicts Affected Global Energy Prices*; Technical Report; Brookings Institution: Washington, DC, USA, 2023.
2. Stanford Institute for Economic Policy Research. *Inflation and Energy Price Shocks: A Global Perspective*; Technical Report; Stanford Institute for Economic Policy Research: Stanford, CA, USA, 2023.
3. Rey, J.; Sachs, J.D. From Millennium Development Goals to Sustainable Development Goals. *Lancet* **2012**, *379*, 2206–2217. [CrossRef] [PubMed]
4. International Renewable Energy Agency. *Future of Solar Photovoltaic: Deployment, Investment, Technology, Grid Integration and Socio-Economic Aspects*; Technical Report; International Renewable Energy Agency: Abu Dhabi, United Arab Emirates, 2019.
5. Jathar, L.D.; Ganesan, S.; Awasarmol, U.; Nikam, K.; Shahapurkar, K.; Soudagar, M.E.M.; Fayaz, H.; El-Shafay, A.S.; Kalam, M.A.; Bouadila, S.; et al. Comprehensive Review of Environmental Factors Influencing the Performance of Photovoltaic Panels: Concern over Emissions at Various Phases Throughout the Lifecycle. *Environ. Pollut.* **2023**, *326*, 121474. [CrossRef] [PubMed]
6. Mateo Romero, H.F.; Gonzalez Rebollo, M.A.; Cardenoso Payo, V.; Alonso Gomez, V.; Redondo Plaza, A.; Moyo, R.T.; Hernandez Callejo, L. Applications of Artificial Intelligence to Photovoltaic Systems: A Review. *Appl. Sci.* **2022**, *12*, 10056. [CrossRef]
7. Gaviria, J.F.; Narvaez, G.; Guillen, C.; Giraldo, L.; Bressan, M. Machine Learning in Photovoltaic Systems: A Review. *Renew. Energy* **2022**, *196*, 298–318. [CrossRef]
8. Koester, L.; Lindig, S.; Louwen, A.; Astigarraga, A.; Manzolini, G.; Moser, D. Review of Photovoltaic Module Degradation, Field Inspection Techniques and Techno-Economic Assessment. *Renew. Sustain. Energy Rev.* **2022**, *165*, 112616. [CrossRef]
9. Abdulla, H.; Sleptchenko, A.; Nayfeh, A. Photovoltaic Systems Operation and Maintenance: A Review and Future Directions. *Renew. Sustain. Energy Rev.* **2024**, *195*, 114342. [CrossRef]
10. Hoffmann, M.; Buerhop-Lutz, C.; Reeb, L.; Pickel, T.; Winkler, T.; Doll, B.; Würfl, T.; Peters, I.M.; Brabec, C.; Maier, A.; et al. Deep-Learning-Based Pipeline for Module Power Prediction from Electroluminescence Measurements. *Prog. Photovolt. Res. Appl.* **2021**, *29*, 920–935. [CrossRef]
11. Hoffmann, M.; Hepp, J.; Doll, B.; Buerhop-Lutz, C.; Peters, I.M.; Brabec, C.; Maier, A.; Christlein, V. Module-Power Prediction from PL Measurements Using Deep Learning. In *Proceedings of the Conference Record of the IEEE Photovoltaic Specialists Conference (PVSC)*; IEEE: Piscataway, NJ, USA, 2021; pp. 827–830. [CrossRef]
12. Maddileti, T. Review on Types of Solar Power Systems. *J. Eng. Sci.* **2019**, *10*, 499–502.
13. Al-Ezzi, A.S.; Ansari, M.N.M. Photovoltaic Solar Cells: A Review. *Appl. Syst. Innov.* **2022**, *5*, 67. [CrossRef]
14. Köntges, M.; Kurtz, S.; Packard, C.E.; Jahn, U.; Berger, K.A.; Kato, K.; Friesen, T.; Liu, H.; Van Iseghem, M.; Wohlgemuth, J.; et al. *Review of Failures of Photovoltaic Modules*; Technical Report; IEA International Energy Agency: Paris, France, 2014.
15. Morales-Aragonés, J.I.; Gómez, V.A.; Gallardo-Saavedra, S.; Redondo-Plaza, A.; Fernández-Martínez, D.; Hernández-Callejo, L. Low-Cost Three-Quadrant Single Solar Cell I-V Tracer. *Appl. Sci.* **2022**, *12*, 6623. [CrossRef]
16. Mateo Romero, H.F.; Hernández-Callejo, L.; González Rebollo, M.Á.; Cardeñoso-Payo, V.; Alonso Gómez, V.; Morales Aragonés, J.I.; Moyo, R.T. Optimized Estimator of the Output Power of PV Cells Using EL Images and I-V Curves. *Sol. Energy* **2023**, *265*, 112089. [CrossRef]
17. Mateo Romero, H.F. Employing Artificial Intelligence Techniques for the Estimation of Energy Production in Photovoltaic Solar Cells Based on Electroluminescence Images. Ph.D. Thesis, Universidad de Valladolid, Valladolid, Spain, 2024. [CrossRef]
18. Mateo-Romero, H.F.; Hernandez-Callejo, L.; Gonzalez Rebollo, M.A.; Cardeñoso-Payo, V.; Gomez, V.A.; Bello, H.J.; Moyo, R.T.; Morales-Aragonés, J.I. Synthetic Dataset of Electroluminescence Images of Photovoltaic Cells by Deep Convolutional Generative Adversarial Networks. *Sustainability* **2023**, *15*, 7175. [CrossRef]
19. Derbeli, M.; Napole, C.; Barambones, O.; Sanchez, J.; Calvo, I.; Fernández-Bustamante, P. Maximum Power Point Tracking Techniques for Photovoltaic Panel: A Review and Experimental Applications. *Energies* **2021**, *14*, 7806. [CrossRef]
20. Botchkarev, A. A New Typology Design of Performance Metrics to Measure Errors in Machine Learning Regression Algorithms. *Interdiscip. J. Inf. Knowl. Manag.* **2019**, *14*, 45–76. [CrossRef] [PubMed]
21. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future Directions. *J. Big Data* **2021**, *8*, 53. [CrossRef] [PubMed]
22. Abadi, M.; Agarwal, A.; Barham, P.; Brevdo, E.; Chen, Z.; Citro, C.; Corrado, G.S.; Davis, A.; Dean, J.; Devin, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. 2015. Available online: https://www.tensorflow.org (accessed on 6 August 2026).
23. O'Malley, T.; Bursztein, E.; Long, J.; Chollet, F.; Jin, H.; Invernizzi, L. Keras Tuner. 2019. Available online: https://keras.io/keras_tuner/ (accessed on 6 August 2026).
24. Garnett, R. *Bayesian Optimization*; Cambridge University Press: Cambridge, UK, 2022.

25. Géron, A. *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, 1st ed.; O'Reilly Media: Sebastopol, CA, USA, 2017.
26. Williams, C.K.I.; Rasmussen, C.E. Gaussian Processes for Regression. In *Proceedings of the Advances in Neural Information Processing Systems*; Touretzky, D., Mozer, M.C., Hasselmo, M., Eds.; MIT Press: Cambridge, MA, USA, 1995; Volume 8.
27. Plaia, A.; Buscemi, S.; Fürnkranz, J.; Loza Mencía, E. Comparing Boosting and Bagging for Decision Trees of Rankings. *J. Classif.* **2022**, *39*, 78–99. [CrossRef]
28. Hearst, M.A.; Dumais, S.T.; Osuna, E.; Platt, J.; Schölkopf, B. Support Vector Machines. *IEEE Intell. Syst. Their Appl.* **1998**, *13*, 18–28. [CrossRef]
29. Schmidhuber, J. Deep Learning in Neural Networks: An Overview. *arXiv* **2014**, arXiv:1404.7828. [CrossRef]
30. Pudjihartono, N.; Fadason, T.; Kempa-Liehr, A.W.; O'Sullivan, J.M. A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction. *Front. Bioinform.* **2022**, *2*, 927312. [CrossRef] [PubMed]
31. Masoudnia, S.; Ebrahimpour, R. Mixture of Experts: A Literature Survey. *Artif. Intell. Rev.* **2014**, *42*, 275–293. [CrossRef]
32. Jordan, M.I.; Jacobs, R.A. Hierarchical Mixtures of Experts and the EM Algorithm. In *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*; IEEE: Piscataway, NJ, USA, 1993; Volume 2, pp. 1339–1344. [CrossRef]
33. Zhang, Z.; Sabuncu, M.R. Generalized Cross Entropy Loss for Training Deep Neural Networks with Noisy Labels. *arXiv* **2018**, arXiv:1805.07836. [CrossRef]
34. Mateo Romero, H.F.; Hernández Callejo, L.; Alonso Gómez, V.; Morales-Aragonés, J.I. SIPV-EL. 2024. Available online: https://zenodo.org/records/14439369 (accessed on 6 August 2026).

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.