

Computer vision-based method for quantifying iron-related defects in silicon solar cell

Oleg Olikh[‡], Oleksii Zavhorodnii, Yulia Perets

Taras Shevchenko National University of Kyiv, Kyiv 01601, Ukraine

E-mail: olegolikh@knu.ua

Abstract. This study demonstrates the feasibility of employing transfer learning from pre-trained computer vision (CV) models to predict iron concentration in silicon solar cells, even when the available training datasets are extremely limited. The predictions are based on the kinetic dependencies of the short-circuit current following FeB pair dissociation, which were converted into images using a wavelet transformation. The performance of various combinations of CV models and regression algorithms was systematically analyzed. Specifically, several state-of-the-art CV architectures, including EfficientNetB7, MobileNetV2, NASNetLarge, ResNet152V2, Xception, and YOLOv4, were utilized either as classifiers or as feature extractors. Regression models, namely Random Forest, Gradient Boosting, eXtreme Gradient Boosting, Support Vector Regression (SVR), and Deep Neural Networks (DNNs), were trained to predict iron concentration from the extracted features. Training and testing were performed using both simulated and experimental datasets. In both cases, EfficientNetB7 and NASNetLarge provided the most informative features for subsequent regression. Among the regression algorithms, SVR and DNNs were identified as the most effective. These models achieved MSE, MAPE, MedAPE, and R^2 values of up to 0.001, 6%, 4%, and 0.999, respectively, for the simulated data, and 0.008, 10%, 5%, and 0.996 for the experimental data.

Keywords: defect, Si solar cell, iron contamination, machine learning, computer vision

Submitted to: *Semicond. Sci. Technol.*

[‡] Author to whom any correspondence should be addressed.

1. Introduction

Due to the urgent need to address environmental challenges and the growing global demand for renewable energy, the deployment of photovoltaic (PV) systems has been rapidly increasing worldwide. In particular, solar PV generation exceeded 1,600 TWh in 2023 [1, 2], rose by about 30% in 2024 [3], and forecasts indicate that total installed capacity will surpass 6 TW by 2030 [1]. Meanwhile, crystalline silicon photovoltaics, which have benefited from decades of scientific advancement and continuous cost reductions, continued to dominate the market in 2024, accounting for approximately 98% of the global share [4, 5].

As in other semiconductor devices, defects play a decisive role in determining the operating parameters of solar cells. Therefore, diagnosing defects, particularly determining their concentrations, is critically important for maintaining the stable performance of PV systems. In recent years, researchers have increasingly complemented established defect-characterisation methods with machine learning (ML) approaches that improve the accuracy, speed, and cost efficiency of these analyses. The use of ML methods for analysing macroscopic defects (such as cracks, finger failures, hotspots, and scratches) and point defects, however, differs significantly. Researchers typically detect macroscopic defects in PV systems using two main approaches [6, 7]. The first approach, Electrical Testing Techniques, involves analysing characteristic electrical curves of parameters such as current, voltage, and power. The second approach, Imaging-Based Techniques, involves analysing electroluminescence (EL) [8] or photoluminescence [9] images of solar cells. Numerous review studies demonstrate extensive use of ML in both approaches [10, 11, 12, 13, 14, 15, 16].

Regarding point defects, although they represent one of the primary limitations of PV devices, researchers have developed considerably fewer ML techniques specifically for their analysis. Existing ML-based approaches include extracting recombination-active centre parameters from lifetime measurements [17, 18, 19], detecting radiation-induced defects using Raman spectroscopy [20, 21], estimating the concentration of contaminant impurities from the ideality factor of current–voltage characteristics [22], and analysing variations in photovoltaic conversion parameters [23]. Nevertheless, a relatively limited number of such methods are offered.

One of the main challenges in applying ML methods effectively is that training the models requires a large amount of labelled data [12]. In practice, researchers often cannot obtain such large volumes of experimental data; therefore, they commonly employ approaches such as simulations, in which hundreds of thousands of dependencies are computed [17, 18, 19,

24, 23]; Physics-Informed Neural Networks (PINNs), which incorporate physical laws into the loss function to generate synthetic data [25, 26]; or Transfer Learning, in which a model trained on one task is adapted to another related task [27, 28]. However, simulations can be highly demanding in terms of time and computational resources; PINNs are primarily suitable for phenomena described by partial differential equations, and pre-trained models are not available for all types of physical problems. At the same time, one of the most extensively studied tasks in machine learning is computer vision (CV), for which many pre-trained models have been publicly released. Moreover, these models are typically trained on extremely large standard datasets. For example, EfficientNetB7 was trained on approximately 1.2 million images from the ImageNet dataset.

This work primarily aimed to apply standard pre-trained CV models to analyse electrophysical measurement results related to point defects. In particular, we focused on quantifying iron in boron-doped crystalline silicon solar cells by examining the short-circuit current (I_{sc}) relaxation following intense illumination. Iron represents one of the most prevalent, ubiquitous, and efficiency-limiting metallic impurities in such structures [29, 30], which, as noted earlier, form the foundation of the photovoltaic market. In Si:B, iron tends to form iron–boron pairs under equilibrium conditions, which intense illumination can dissociate [31, 32]. In fact, the aforementioned I_{sc} variations directly reflect the recovery process of iron–boron pairs [33].

It should be noted that the use of well-established computer vision benchmark architectures such as YOLO, MobileNetV2, EfficientNet, ResNet, Xception, GoogleNet, and other convolutional neural networks (CNNs) is a common approach for identifying macrodefects from EL measurements [8, 34, 6, 35, 36, 37, 38, 39]. However, in this case, the measurement result is an image, which makes the approach relatively straightforward. In our case, it was necessary to transform the time dependence into an image representation. Standard approaches to solving such problems involve the use of Fourier or wavelet transforms, and last were applied in this study. In photovoltaics, wavelet transforms are typically used for processing solar cell images to enhance the detection of macrodefects [40, 41], but they can also be employed to convert one-dimensional non-stationary signals into two-dimensional spectrograms and thereby enable the effective extraction of subtle features [42].

By applying computer vision models to wavelet spectrograms represented as images, we generated high-dimensional feature vectors and used them as inputs for traditional regression models. The $I_{sc}(t)$ de-

pendencies for training the regression models were obtained through both simulation and experimental measurements. In both cases, the hybrid Transfer Learning approach produced predictions with sufficiently high accuracy (within a few percent), even when training models on small datasets containing fewer than 30 samples. Importantly, the proposed approach is highly versatile and can be extended to a wide range of tasks related to defect characterization and other applications.

2. Methodology

2.1. General outline of the method

Figure 1 illustrates the workflow of the ML pipeline used to extract iron contamination from $I_{\text{sc}}(t)$ dependencies. The process consists of three main blocks: Data Acquisition, CNN Feature Processing, and Predictive Regression. The first stage involves either simulating or experimentally measuring the time dependence of the short-circuit current in a solar cell after the induced decay of FeB pairs. These procedures are described in more detail in Subsections 2.2 and 2.3. For the experimental curves, the data were smoothed using a Savitzky–Golay filter [43]. Subsequently, a continuous wavelet transform [44] was applied to convert the one-dimensional time dependencies into two-dimensional spectrograms represented as images, where each point corresponds to the amplitude of the wavelet coefficient at a specific time and frequency. The Morlet wavelet was employed, and the procedure was implemented using the Python package PyWavelets. Examples of the resulting images are shown in Figure 2b and Figure 2c. Data augmentation was then performed by flipping the images along the x- and y-axes and rotating them by 90°, 180°, and 270°. This procedure is known to improve the accuracy of ML model predictions, particularly when only small datasets are available [45].

During the CNN Feature Processing stage, all images (both original and augmented) were processed using one of the standard CV models to extract a feature set for each image. The selected models and feature extraction settings are described in Subsection 2.4. No CNN fine-tuning was performed; the models were used in their pre-trained form as downloaded. In general, the dimensionality of the feature vectors obtained from CNN outputs substantially exceeds the number of available samples, implying a high degree of redundancy. Therefore, to enable comparison and mitigate this effect, Principal Component Analysis (PCA) was applied in some cases to reduce the feature dimensionality with negligible loss of total variance.

The obtained feature sets served as inputs to regression models based on one of the standard

algorithms described in Subsection 2.5, which aimed to predict the iron concentration (N_{Fe}) in the solar cell. In the first case, the regression models were trained on a simulated training dataset and tested on both the simulated test dataset and experimental data. In the second case, a portion of the experimental results was used for training, while the remaining part was reserved for testing the corresponding models. During training, feature sets derived from the original wavelet spectrograms and their augmented versions were treated as separate samples. During testing, the median of the predicted values obtained from the original and augmented images was used as the final prediction. Model performance was evaluated using the metrics described in Subsection 2.6.

2.2. Simulation details

To obtain the $I_{\text{sc}}(t)$ dependencies, I - V curve simulations were performed for a silicon n^+ - p - p^+ structure under monochromatic illumination using SCAPS-1D version 3.3.11. The SCAPS-1D software [46] is a widely used tool for modeling solar cells while accounting for defect states [47, 48, 49, 50, 51, 52]. I_{sc} values were extracted from the simulated I - V curves using a standard procedure [53].

During the simulations, the base thickness of the structure was set to 380 μm , and boron was used as the doping element with a concentration of $N_B = 1.36 \times 10^{15} \text{ cm}^{-3}$. The temperature was maintained at 340 K, and monochromatic illumination with a wavelength of 940 nm and an intensity of 5 W/m² was applied, corresponding to the experimental conditions (see Subsection 2.3). One of the modeling parameters was the total concentration of iron impurity atoms, N_{Fe} . It was assumed that Fe atoms were uniformly distributed throughout the base and p^+ layer of the solar cell and could exist either in interstitial positions, with a concentration N_{Fe_i} , or as FeB pairs, with a concentration N_{FeB} . The time dependence of N_{Fe_i} after pair dissociation follows the well-known expression [54, 55]:

$$N_{\text{Fe}_i}(t) = (N_{\text{Fe}_i,0} - N_{\text{Fe}_i,\text{eq}}) \times \exp(-t/\tau_{\text{ass}}) + N_{\text{Fe}_i,\text{eq}}, \quad (1)$$

where $N_{\text{Fe}_i,0}$ is the concentration of interstitial iron atoms formed due to FeB pair dissociation, $N_{\text{Fe}_i,0} = N_{\text{Fe}_i}(t=0) = N_{\text{Fe}}$; $N_{\text{Fe}_i,\text{eq}}$ is the portion of interstitial iron atoms that remain unpaired in the equilibrium state $N_{\text{Fe}_i,\text{eq}} = N_{\text{Fe}_i}(t \rightarrow \infty)$, according to [54, 55]

$$N_{\text{Fe}_i,\text{eq}} = \frac{N_{\text{Fe}}}{\left[1 + N_B \cdot A_z \cdot \exp\left(\frac{E_b}{kT}\right)\right] \left[1 + \exp\left(\frac{E_F - E_{\text{Fe}_i}}{kT}\right)\right]}, \quad (2)$$

E_b is the binding energy of the FeB pairs (taken as 0.582 eV [55]), A_z depends on the number of possible orientations of the pair and lattice site density (taken

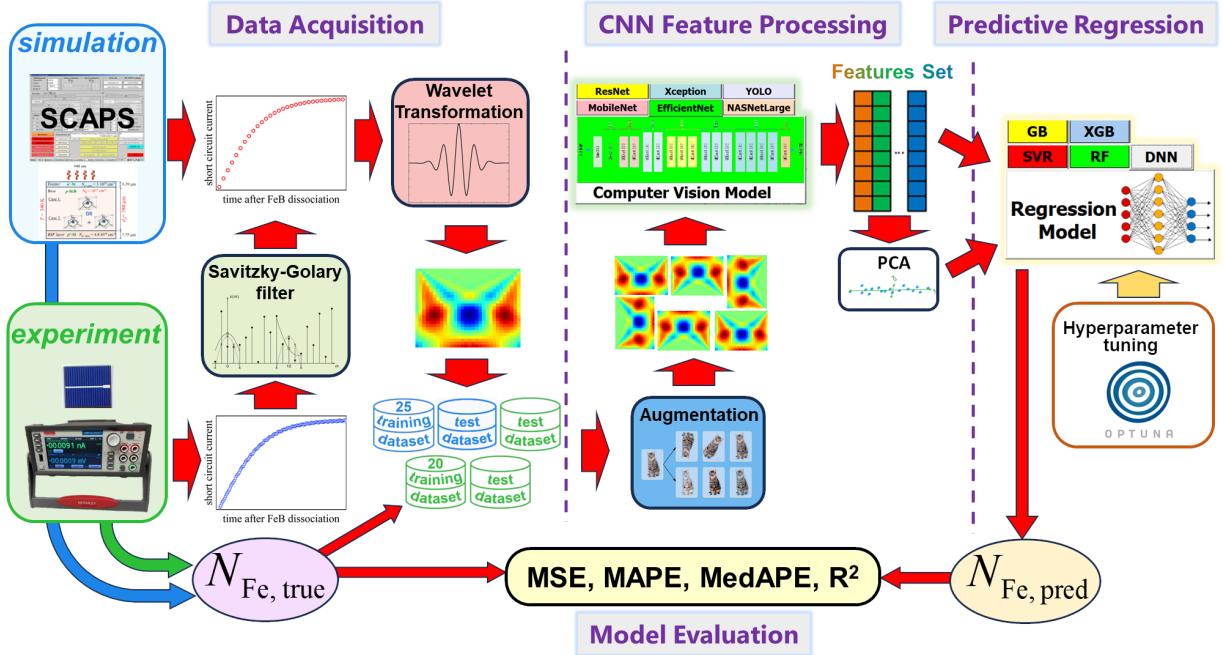


Figure 1. The workflow of the ML pipeline

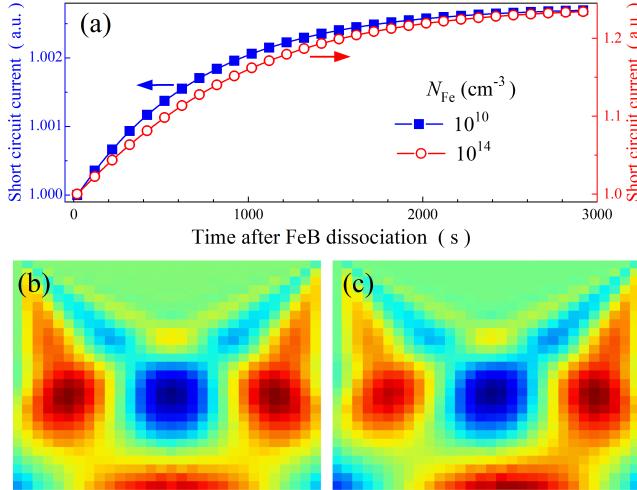


Figure 2. Simulated time dependencies of the short-circuit current (a) and the corresponding wavelet spectrograms for iron concentrations of 10^{10} cm^{-3} (b) and 10^{14} cm^{-3} (c). The data in panel a are shown with filled squares for the concentration corresponding to panel b and with open circles for that corresponding to panel c.

as 10^{-23} cm^3 [54]), E_F is the Fermi level, E_{Fe_i} is the position of the donor Fe_i level relative to the valence band maximum (taken as 0.394 eV [32]); τ_{ass} is the characteristic time of the complex association, according to [56, 32, 57]

$$\tau_{\text{ass}} = A \times \frac{T}{N_A} \exp \left(\frac{E_m}{kT} \right), \quad (3)$$

E_m is the energy of Fe_i^+ migration (taken as 0.66 eV

[32, 56, 57]), A is the constant (taken as $5.7 \times 10^5 \frac{\text{s}}{\text{K cm}^3}$ [57]). In its turn, the iron–boron pair concentration N_{FeB} , was estimated from

$$N_{\text{FeB}}(t) + N_{\text{Fe}_i}(t) = N_{\text{Fe}}. \quad (4)$$

Overall, the concentrations of iron-related defects depended not only on time but also on their spatial position within the structure, reflecting the non-uniformity of the $(E_F - E_{\text{Fe}_i})$ difference.

A detailed description of the modeling approach, including the values of silicon and defect parameters employed, is provided elsewhere [24, 58].

To create the training dataset, 25 N_{Fe} values were selected, evenly distributed on a logarithmic scale from 10^{10} cm^{-3} to 10^{14} cm^{-3} . Examples of the resulting dependencies are shown in Figure 2, along with the corresponding wavelet spectrograms. The simulated test dataset consisted of 10 dependencies calculated for 10 N_{Fe} values not included in the training dataset.

2.3. Experiment details

The proposed method was validated using measurements of $n^+ \text{-} p^+$ silicon solar cells fabricated on $380 \mu\text{m}$ thick p -type boron-doped Cz wafers ($N_B = 1.36 \times 10^{15} \text{ cm}^{-3}$). The n^+ ($0.7 \mu\text{m}$, $20\text{--}30 \Omega/\square$) and p^+ ($0.6 \mu\text{m}$, $10\text{--}20 \Omega/\square$) layers were formed by phosphorus and boron diffusion, respectively. Details of the fabrication procedure are provided elsewhere [33].

Current–voltage curves and I_{sc} kinetics were measured with a Keithley 2450 source meter. A 940 nm SN-HPIR940nm-1W LED (intensity 5 W/m^2)

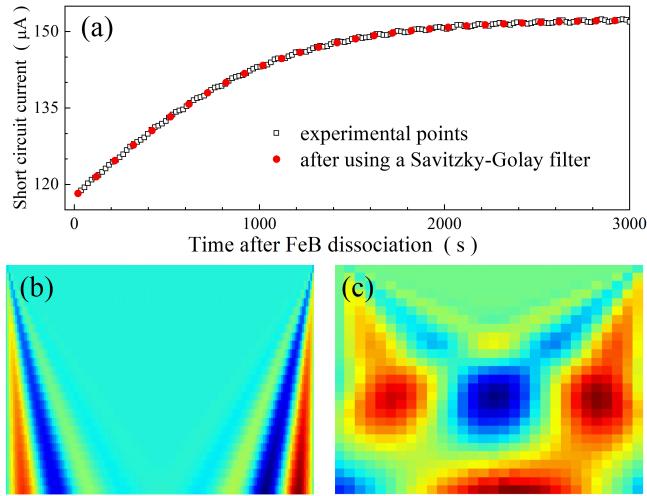


Figure 3. (a) Experimentally measured time dependence of the short-circuit current for a sample with $N_{\text{Fe}} = 2.8 \cdot 10^{13} \text{ cm}^{-3}$ (a, open squares) and the same dependence after applying the Savitzky–Golay filter (filled circles). Panels (b) and (c) show the wavelet spectrograms corresponding to the curves with filled squares and open circles, respectively.

served as a monochromatic light source, with its output stabilized by a W1209 thermostat and a feedback-controlled power supply. The cell temperature was regulated by a thermoelectric cooler with an STS-21 sensor and maintained constant by a PID algorithm in the control software. FeB-pair dissociation was induced by intense (about 7000 W/m^2) halogen-lamp illumination. The illumination intervals were selected according to a previous study [59]. The kinetics of the short-circuit current were measured in the dark at 340 K for 3000 s. According to Eq. (3), this interval is sufficient for the complete restoration of the iron–boron pairs to their equilibrium concentration.

Figure 3a presents an example of a measured $I_{\text{sc}}(t)$ dependence. The signal contains some noise because, despite using a thermostat, the LED temperature fluctuated by approximately 0.4 K. A Savitzky–Golay filter was applied for smoothing, with the window lengths and filter order selected adaptively according to Krishnan and Seelamantula [43]. Only current values corresponding to the time points used in the simulations were retained for the wavelet transformation. The smoothed curve is shown in Figure 3a, while the remaining panels of the figure display the spectrograms obtained from the raw experimental curve and the processed dependence.

The iron concentration N_{Fe} was determined using a methodology described in [60, 33], which is based on fitting the kinetics of the short circuit current following FeB pairs dissociation. A total of 28 samples with iron concentrations ranging from 10^{11} cm^{-3} to $2 \times 10^{13} \text{ cm}^{-3}$ were examined. To evaluate models trained

on simulated data, the entire experimental dataset was used as the test set. In cases where the models were trained using experimental data, 20 randomly selected samples were included in the training set, while the remaining eight samples were used for testing.

2.4. Computer vision models

To extract graphical features from the wavelet spectrograms, several computer vision models available in Keras were employed, namely EfficientNetB7, ResNet152V2, MobileNetV2, Xception, and NASNet-Large. Although these models have different architectures, they all belong to the CNN class, are designed for object classification, and have previously been applied successfully to processing EL images of solar cells [6, 35, 36, 38, 39]. Two feature extraction strategies were evaluated for all models: in the first, the class-specific probability distributions (soft labels) were passed to the subsequent stage of the pipeline, while in the second, the raw feature vectors directly extracted by the computer vision model were utilized.

Furthermore, the CSPDarknet53 model was employed, which serves as the CNN backbone for YOLOv4. Models of this family feature a more sophisticated CNN architecture optimized not for single-object classification but for multi-object detection in images. They are widely used in imaging-based techniques [8, 34, 36]. The employed model produces three feature maps, and for subsequent processing, either only the highest-level layer or the two deepest layers were selected.

It is well known that increasing the feature dimensionality does not necessarily enhance the total information variance. To mitigate the impact of redundant data, Principal Component Analysis (PCA) was applied, which constructs new, uncorrelated features (principal components). PCA is a widely used and effective technique in machine learning, particularly for improving performance in Electrical Testing Techniques [61, 62]. In this study, PCA was applied to the training datasets with an explained variance threshold of 99.9%. In other words, the principal components explaining no less than 99.9% of the total variance in the original features were selected, thus achieving a substantial reduction in feature dimensionality. This pre-processing procedure was selectively applied to a subset of the computer vision models — specifically, those demonstrating good performance on the test sets without PCA — with the aim of assessing the feasibility and effectiveness of this approach.

Given the remarkably high dimensionality of the features produced by YOLOv4, the feasibility of applying an alternative dimensionality reduction technique was examined. Specifically, global average

pooling was applied to each convolutional feature map, replacing the spatial map with its mean value and thereby yielding a single scalar value per channel.

The configurations of the computer vision models used in this study are summarized in table 1. The table also lists the notations that are subsequently used to refer to these configurations.

2.5. Regression algorithms

Five ML algorithms were employed to develop regression models for predicting iron concentration: Random Forest (RF), Gradient Boosting (GB), eXtreme Gradient Boosting (XGB), Support Vector Regression (SVR), and Deep Neural Network (DNN). The models were implemented using Python libraries: Keras for DNN, Scikit-learn for RF, GB, and SVR, and XGBoost for XGB.

Each regression model was trained using features obtained from all configurations listed in table 1 and subsequently used to make predictions. The only exception involved the uncompressed features extracted by YOLOv4, for which the available computational resources (2.9 GHz AMD Ryzen 7 4800H CPU, 8 GB RAM, GeForce GTX 1650 4 GB) permitted the use of SVR only. The target variable of all models was $\log N_{\text{Fe}}$. Such logarithmic transformation is a standard approach for achieving higher prediction accuracy when the target quantity spans several orders of magnitude [63, 64]. Both input features and target values were normalized to have zero mean and unit standard deviation within the training set.

For each scenario, regression models were optimized to enhance predictive performance. Hyperparameter tuning was performed using the Optuna toolkit, employing the TPE sampler and the Hyperband pruner to ensure efficient selection. The complete list of tuned hyperparameters and their respective search ranges is provided in Tables S1–S5 (Supplementary Material). Five-fold cross-validation was implemented during model tuning, with 20% of the training data used as a validation set to evaluate models trained on the remaining 80%. The resulting optimal hyperparameter combinations are presented in Tables S6–S10.

Consequently, 87 distinct combinations of computer vision and regression models were investigated. Each combination was subsequently trained and evaluated using both simulated and experimental data. To identify the results for each case, a composite label was employed, derived from the last column of table 1 and the abbreviated name of the regression algorithm.

2.6. Model evaluation

A rigorous assessment of model performance across diverse metrics is essential for constructing a robust regression model. The evaluation metrics for iron quantification were the mean squared error (MSE), mean absolute percentage error (MAPE), median absolute percentage error (MedAPE) and coefficient of determination (R^2), as defined in Eqs. (5)–(9).

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2, \quad (5)$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \text{MAPE}_i, \quad (6)$$

$$\text{MAPE}_i = \frac{|N_{\text{Fe},\text{PRED},i} - N_{\text{Fe},\text{TRUE},i}|}{N_{\text{Fe},\text{TRUE},i}} \times 100\%, \quad (7)$$

$$\text{MedAPE} = \frac{1}{2} [\text{MAPE}_{\lceil \frac{N}{2} \rceil} + \text{MAPE}_{\lfloor \frac{N}{2} + 1 \rfloor}], \quad (8)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (N_{\text{Fe},\text{TRUE},i} - N_{\text{Fe},\text{PRED},i})^2}{\sum_{i=1}^N (N_{\text{Fe},\text{TRUE},i} - \bar{N}_{\text{Fe},\text{TRUE}})^2}, \quad (9)$$

where where \hat{y}_i is the predicted value of the target variable for the i -th data point, y_i is the corresponding known value (obtained by logarithmic transformation and normalization of the iron concentration); N denotes the number of samples in the dataset ($N = 25$ for the simulated training set, 20 for the experimental training set, 10 for the simulated test set, and 28 and 8 for the experimental datasets used to test models trained on the simulated and experimental sets, respectively); MAPE_i is the absolute percentage error for i -th data point; $N_{\text{Fe},\text{PRED},i}$ is the predicted iron concentration, $N_{\text{Fe},\text{TRUE},i}$ is the known value (either the parameter used in the simulation or obtained from experimental iron determination); $\bar{N}_{\text{Fe},\text{TRUE}}$ is the mean of the true values in the dataset.

MSE is one of the most widely used metrics for evaluating model accuracy, and the training objective was specifically defined to minimize this quantity. However, since the computation of y_i involves both normalization and a logarithmic transformation of N_{Fe} , the MSE metric alone does not fully reflect the accuracy of the iron concentration estimation. Therefore, the MAPE, which quantifies the mean relative deviation, was also employed. In addition, the MedAPE, representing the error value below which 50% of the predictions fall, provides a more robust measure against the influence of individual outliers, which can have a particularly large effect on mean-based metrics in smaller datasets. Finally, the R^2

Table 1. Summary of used pretrained CV models and feature extraction variants

Base model	Model type	Feature processing	Output dimension	Model Label
EfficientNetB7	Classifier	None	1000	ENB7:CL
	Feature extractor	None	2560	ENB7:FE
MobileNetV2	Classifier	PCA	39	ENB7:FE:P
	Feature extractor	None	1000	MNV2:CL
NASNetLarge	Classifier	None	1280	MNV2:FE
	Feature extractor	PCA	124	MNV2:FE:P
ResNet152V2	Classifier	None	1000	NAS:CL
	Feature extractor	PCA	30	NAS:CL:P
Xception	Classifier	None	4032	NAS:FE
	Feature extractor	None	1000	R152:CL
YOLOv4 (CSPDarknet53)	Classifier	None	2048	R152:FE
	Feature extractor (raw, top layer)	None	1000	XCP:CL
YOLOv4 (CSPDarknet53)	Feature extractor (raw, top & penultimate layers)	PCA	2048	XCP:FE
	Feature extractor (pooled, top layer)	None	86528	YL:FE1
YOLOv4 (CSPDarknet53)	Feature extractor (pooled, top & penultimate layers)	PCA	137	YL:FE1:P
	Feature extractor (raw, top & penultimate layers)	None	433640	YL:FE2
YOLOv4 (CSPDarknet53)	Feature extractor (pooled, top layer)	PCA	142	YL:FE2:P
	Feature extractor (pooled, top & penultimate layers)	None	512	YL:FP1
YOLOv4 (CSPDarknet53)	Feature extractor (pooled, top & penultimate layers)	None	1024	YL:FP2

was used to quantify the fraction of the variance in the target variable explained by the model, thereby indicating how well the predicted values reproduce the observed data; a value of 1 corresponds to perfect agreement.

3. Results and discussion

3.1. Simulated data

Figure 4 illustrates representative prediction results obtained from models trained on the simulated training dataset. The complete set of results, covering all 87 investigated configurations, is provided in Figure S1 of the Supplementary Material.

Figure 5 presents the MAPE and R^2 scores obtained when applying the models to the training dataset. Although these metrics are sufficiently representative, the corresponding MedAPE and MSE values are provided in Figure S2 (Supplementary Material). The most notable observation from Figure 4 and Figure 5 is that, despite the very limited size of the training dataset (25 samples), most models demonstrate high training performance: in many cases, the mean relative error is approximately 1% or lower, and it rarely exceeds 10%. At the same time, the R^2 score drops below 0.980 in only 8 out of 87 cases. It is also worth noting that the MedAPE values generally do not exceed the MAPE and are, in fact, smaller in most instances (see Figure S2). These consistently high performance metrics indicate that (i) the wavelet transformation produced highly informative images

that effectively encoded information about the Fe concentration, and (ii) the CV models successfully extracted features correlated with the concentration.

Among the regression models, GB and SVR exhibited the best performance, whereas the DNN produced the least favorable results. This outcome is entirely consistent with expectations, since both Gradient Boosting and Support Vector Regression are well known to perform effectively when the number of samples is limited and the feature space exhibits a low noise level, as is characteristic of synthetic datasets. In contrast, neural networks contain a large number of parameters and therefore do not tend to exhibit perfect generalization under such conditions.

Among the CV models, EfficientNetB7 and NASNetLarge demonstrated the best performance, whereas ResNet152V2 and YOLOv4 showed the weakest results. This discrepancy can be attributed to the fact that the former two are relatively modern architectures specifically designed to extract generalizable features and are therefore well suited to wavelet spectrograms, which are characterized by a multi-scale structure. In contrast, ResNet152V2 is primarily optimized for object classification, whereas YOLO is less effective for regression tasks involving global image patterns, as it focuses on localized object detection.

It is also evident that using class probabilities as descriptors degrades prediction quality compared with the cases where image features are employed directly. This indicates that the internal feature maps of CNN

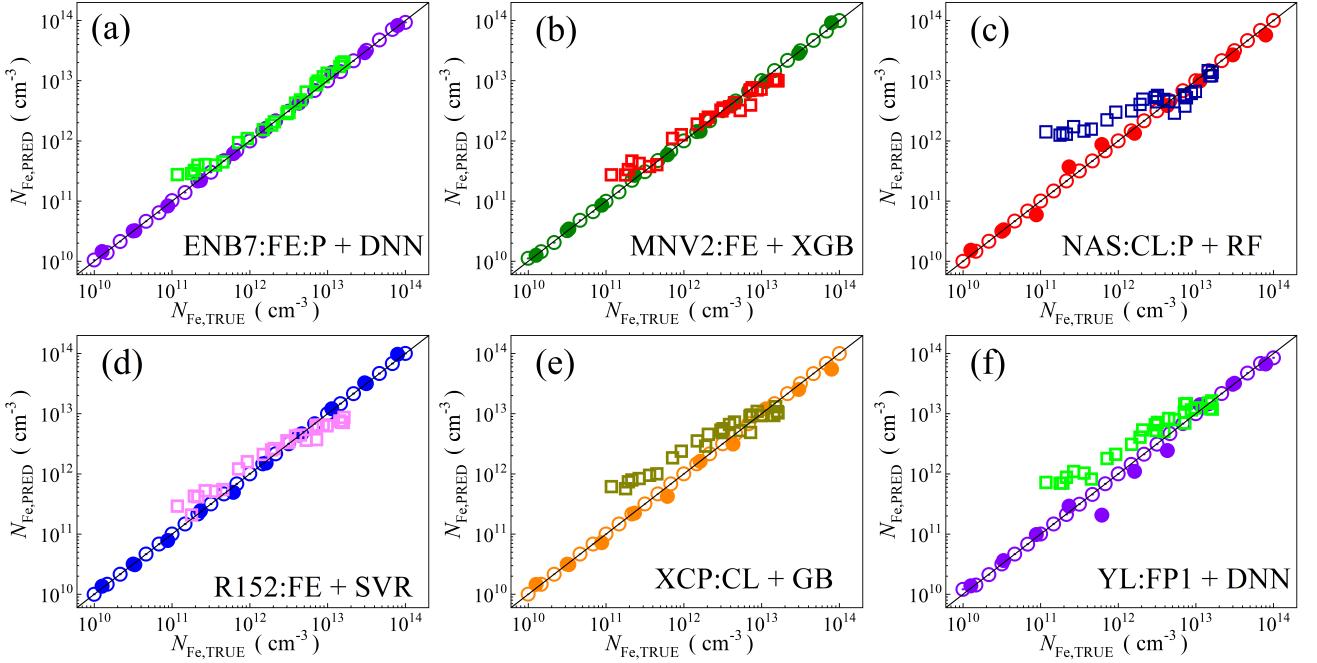


Figure 4. Scatter plots compare the reference iron $N_{\text{Fe},\text{TRUE}}$ with ML-predicted values $N_{\text{Fe},\text{PRED}}$, obtained using feature vectors extracted from various CV models combined with different regression algorithms (specific models are indicated in the figures). The ML models were trained on a simulated dataset. The open circles correspond to the training phase, while the filled circles and open squares correspond to the test phase, representing the simulated and experimental datasets, respectively. The black lines are the identified lines serving as the references.

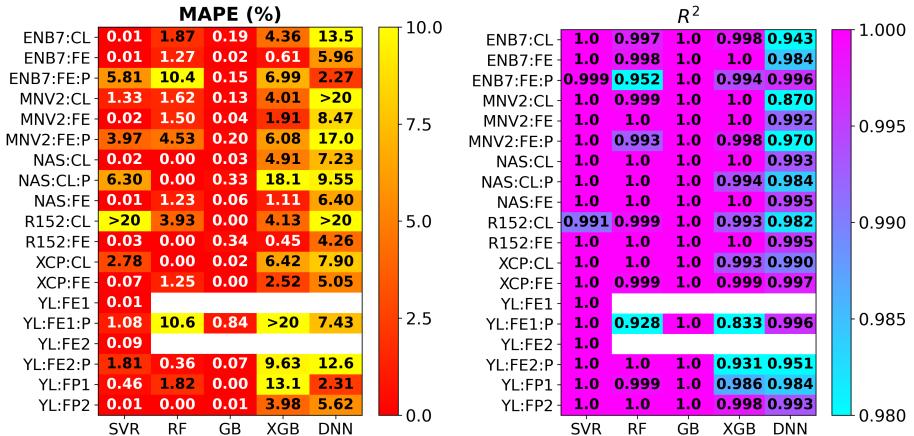


Figure 5. Mean absolute percentage error (left panel) and coefficient of determination (right panel) for different combinations of CV models (vertical axis) and regression models (horizontal axis) during the training phase. The models were trained on a simulated dataset.

models provide a more informative representation of visual patterns, enabling the regressor to establish a stronger relationship with concentration. Moreover, the application of PCA, even though it retains 99.9% of the variance, results in reduced predictive accuracy. This observation suggests that the image patterns associated with variations in iron concentration may constitute only a minor portion of the overall data variance.

The ability of models to achieve high accuracy on

the training set is a necessary prerequisite for effective performance on unseen data; however, it does not guarantee reliable prediction outcomes. Consequently, evaluation on an independent test set is essential. This step becomes particularly critical when the training set is small, as the test results provide the primary evidence of model generalizability. Figure 4 and Figure 6 present the prediction results obtained for the test set generated from synthetic data (more detailed versions are available in Figures S1 and S3 of the

Supplementary Material). As expected, prediction performance declined; however, the reduction in accuracy varied across the regression algorithms. Specifically, this degradation was least pronounced for the DNN, which achieved the best overall performance. The poorest metrics were obtained for RF and GB, whereas SVR and XGB performed slightly worse than the DNN but with a relatively small margin. This behavior can be attributed to the DNN’s ability to approximate continuous dependencies smoothly. In contrast to GB and RF, which rely on the formation of local decision rules, neural networks construct a continuous surface in the feature space. This property enhances interpolation for iron concentration values not represented in the training data. From this perspective, XGB and SVR occupy an intermediate position, explaining their slightly lower performance relative to the DNN.

As shown, the relative performance of the computer vision models on the simulated test dataset remained consistent, with no change in the top- and bottom-performing architectures. Specifically, EfficientNetB7 and NASNetLarge yielded the most favorable results, confirming their effectiveness in extracting wavelet image features relevant for concentration prediction. In contrast, ResNet152V2 and YOLOv4 produced the least satisfactory outcomes. A notable exception arises in configurations where YOLOv4 features from two layers are combined with either a DNN or SVR. In these cases, performance metrics were considerably improved. This enhancement suggests that the use of a larger number of features enabled the capture of more diverse patterns, which, when coupled with flexible regressors, partially mitigated the limitations observed in the standalone YOLO model.

Interestingly, the PCA application to the test set did not substantially degrade performance; in certain cases, it even produced slight improvements. This indicates that PCA is not universally detrimental, although the resulting gains are generally marginal and unpredictable. Moreover, the differences in metrics between models using class probabilities and those employing raw image features were less pronounced than observed for the training dataset. Nevertheless, uncompressed feature representations consistently maintained a performance advantage.

The top-performing model combinations were identified as follows: ENB7:FE+DNN (MAPE = 5.90%, MedAPE = 4.83%, R^2 = 0.996), ENB7:FE:P+DNN (MAPE = 5.96%, MedAPE = 3.68%, R^2 = 0.997), NAS:FE+SVR (MAPE = 5.82%, MedAPE = 5.42%, R^2 = 0.987), NAS:FE+XGB (MAPE = 6.19%, MedAPE = 5.12%, R^2 = 0.987), and NAS:CL:P+SVR (MAPE = 10.3%, MedAPE = 4.52%, R^2 = 0.999). These results correspond to exceptionally high abso-

lute performance metrics, providing clear evidence that the models effectively capture the underlying relationship between the wavelet-based images and the iron concentration.

Having established the effectiveness of the models on the simulated test dataset, we next evaluated their performance on experimental measurements. This step is critical for assessing the generalizability of the models to real-world data, where additional sources of noise and variability may be present. By comparing the results obtained from experimental data with those from the simulated dataset, it becomes possible to identify potential limitations of the models and to confirm whether the features extracted from wavelet spectrograms remain informative under practical conditions. Figure 7a and Figure 7b show the performance metrics of models trained on synthetic data when applied to experimental measurements. The relationship between predicted and actual concentrations is illustrated in Figure 4, with additional results provided in Figures S1 and S4. As observed, the mean and median prediction errors fall within the (15–25)% range for only a limited number of configurations, specifically certain combinations of EfficientNetB7 or NASNetLarge with DNN or SVR. Although this outcome is not catastrophic, considering the approximately 10% inherent experimental error in N_{Fe} determination, it falls short of ideal expectations. At the same time, the R^2 metric remains acceptably high. Further analysis (Figure 4) indicates that the prediction error depends on the iron concentration level: the relationship between $N_{\text{Fe},\text{PRED}}$ and $N_{\text{Fe},\text{TRUE}}$ is linear on a logarithmic scale, but its slope deviates from the line of unity. This observation suggests a systematic prediction bias rather than a complete loss of correlation, implying that the models are capable of capturing relative differences in concentration but do not accurately predict absolute values.

While the presence of residual noise patterns in the experimental curves — insufficiently suppressed by filtering — could contribute to the observed discrepancies, a more plausible explanation lies in the incomplete correspondence between the physical model used for data synthesis and the actual behavior of solar cells. This mismatch is likely associated with the numerical parameters employed in the model’s fundamental equations (Eqs. (1)-(3)). For instance, the calculation of the characteristic FeB association time (Eq. (3)) adopted values of $A = 5.7 \times 10^5 \frac{\text{s}}{\text{K cm}^3}$ and $E_m = 0.66 \text{ eV}$, which are among the most frequently reported in the literature. However, a considerable scatter exists in these parameters. Specifically, E_m values ranging from 0.55 eV [65] to 0.69 eV [66] have been reported, including intermediate estimates of 0.64

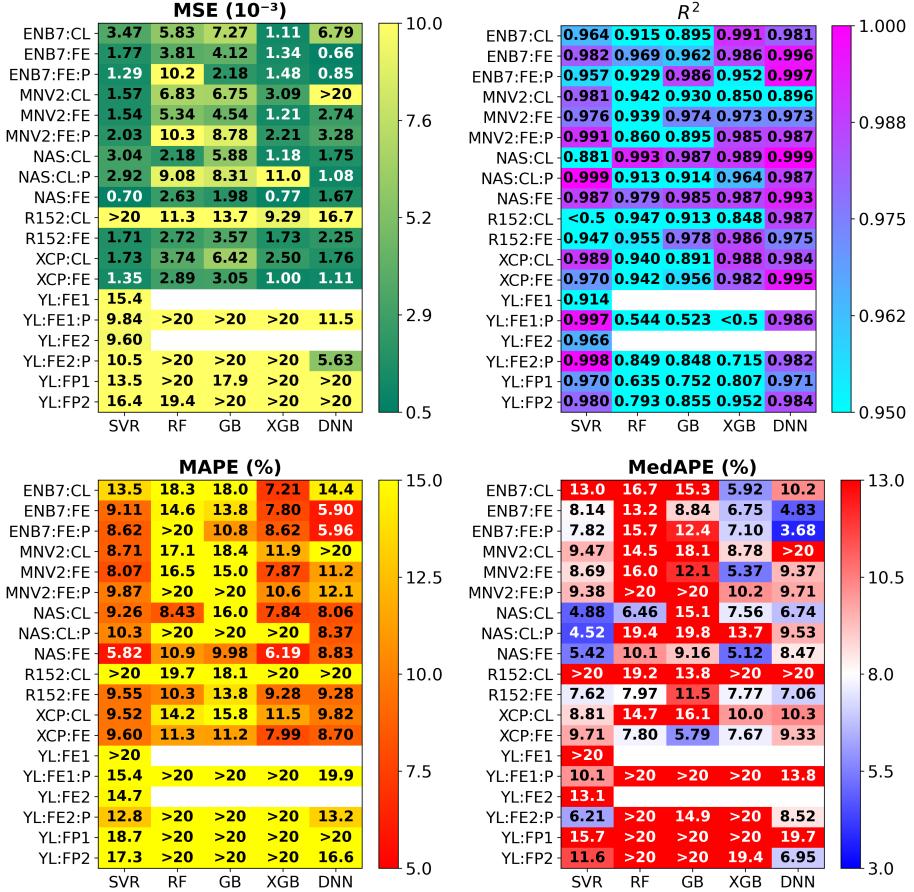


Figure 6. Mean squared error, coefficient of determination, mean absolute percentage error, and median absolute percentage error for different combinations of CV models (vertical axis) and regression models (horizontal axis) during the test phase with simulated dataset. The models were trained on a simulated training dataset.

eV [67], 0.65 eV [68], 0.66 eV [57, 69, 56, 70], 0.67 eV [71], and 0.68 eV [55, 72, 73]. Similarly, the pre-exponential factor A has been cited as 4.3×10^5 [74] or 5×10^5 [70, 75]. Moreover, the calculations assumed a spatially invariant E_m across the device. In practice, the reported diffusion barriers are typically derived for bulk p -Si, whereas the energy value can vary in regions with different Fermi level positions [76], such as the space-charge region in the present structures. Similar variability exists for other key parameters. The FeB pair binding energy has been reported to range from 0.45 to 0.67 eV [68, 71, 55, 77], the Fe_i donor level position varies between 0.38 and 0.394 eV above the valence band maximum [32, 72, 78, 79], and the pre-exponential factor in denominator of Eq. (2) A_z has been cited as either 10^{-23} cm^3 or $2.7 \times 10^{-22} \text{ cm}^{-3} \text{ cm}^3$ [71]. A deviation of any of these parameter values from their actual physical magnitudes could account for the observed prediction errors. Furthermore, earlier studies have demonstrated that both the non-uniform distribution of iron across the base thickness and the variation in the base thickness itself can significantly

affect iron concentration estimation [68]. Neither of these effects was incorporated into the present modeling framework.

A common strategy to improve prediction accuracy involves post-hoc calibration, where a corrective function is applied to model outputs using parameters derived from a limited subset of experimental data. In the present case, the analysis indicated that a quadratic correction of the target variable provides the most suitable adjustment, expressed as follows:

$$\log N_{\text{Fe,PRED}} = 9.51 - 1.71 \cdot \log N_{\text{Fe,PRED}}^* + 0.079 \cdot (\log N_{\text{Fe,PRED}}^*)^2, \quad (10)$$

where $N_{\text{Fe,PRED}}^*$ denotes the direct model prediction. The performance metrics after post-hoc calibration are presented in Figure 7c and Figure 7d. As shown by the data, applying this correction substantially reduced the prediction errors. Specifically, the mean relative error across the 28-sample experimental dataset now lies within the (13–17)% range for the best-performing models (approximately 20 out of 87 configurations) and remains below 25% for most of the others. The median error is even lower, reaching (7–10)%

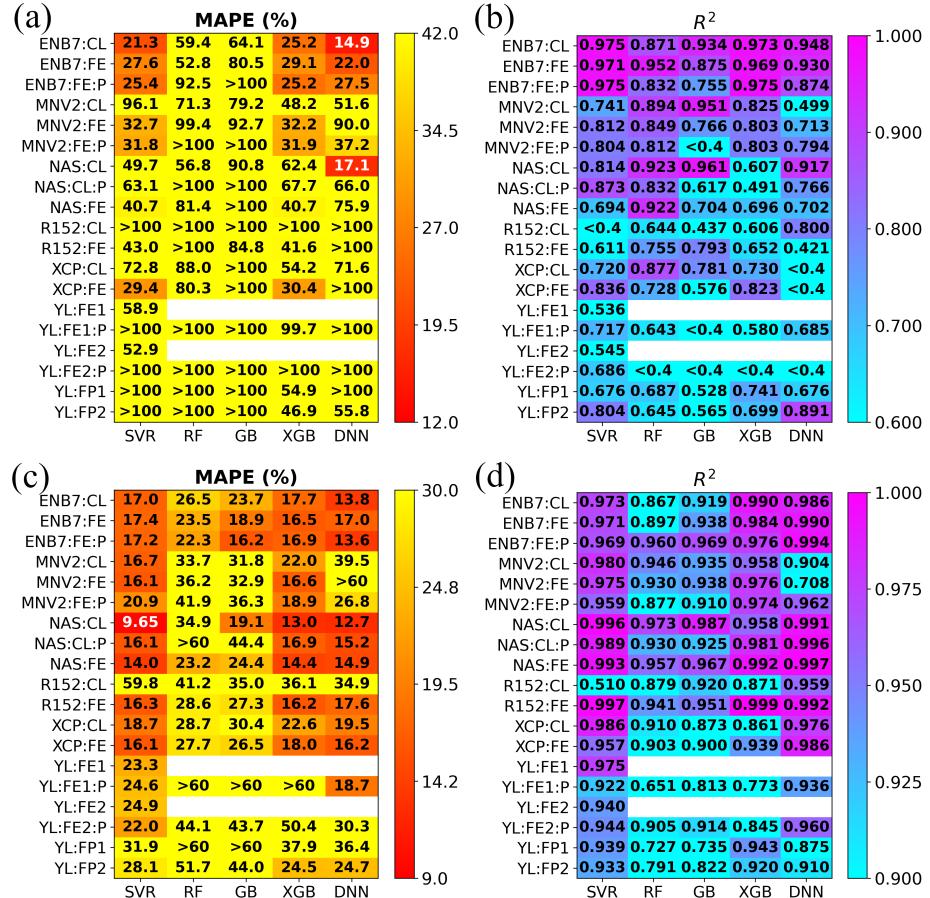


Figure 7. Mean absolute percentage error (a,c) and coefficient of determination (b, d) for different combinations of CV models (vertical axis) and regression models (horizontal axis) during the test phase with experimental dataset without (a, b) and with (c, d) post-hoc calibration according to Eq. (10). The models were trained on a simulated dataset.

in the most favorable cases. In line with the results obtained for the simulated test dataset, the most accurate configurations involve EfficientNetB7, NASNetLarge, DNN, SVR, and XGB. Interestingly, MobileNetV2 also appeared among the top-performing combinations, yielding the lowest MedAPE value (7.64%) for the MNV2:CL configuration. This outcome may suggest that the features extracted by MobileNetV2, though less informative for the training and simulated test datasets, capture specific image patterns more relevant to the experimental measurements and applied correction could have further enhanced the effectiveness of these features.

It is worth noting that applying the correction increased the gap between the mean and median absolute percentage errors. This observation suggests that, although the correction improved the overall agreement between the predicted and true values, a few samples still exhibit relatively large residual errors.

The correction also led to a decrease in the R^2 value. This outcome is expected, as post-processing can weaken the linear correspondence between the

initial predictions and the experimental values, even when the overall prediction errors are simultaneously reduced. Therefore, the reduction in R^2 should be regarded as a side effect of enhancing the model's accuracy rather than as evidence of its degradation.

Although the quadratic correction proved effective, it does not represent a universal solution for transferring models from synthetic to experimental data. A more robust approach involves integrating experimental data into the training process, as implemented in this study below.

3.2. Experimental data

Figure 8 shows the correlation between the iron concentrations predicted by models trained on experimental data and reference values derived according to [60, 33]. Representative outcomes for selected CV-regression model combinations are shown for both training and testing, with extended data provided in Figure S6 of the Supplementary Material.

It should be noted that, compared to training

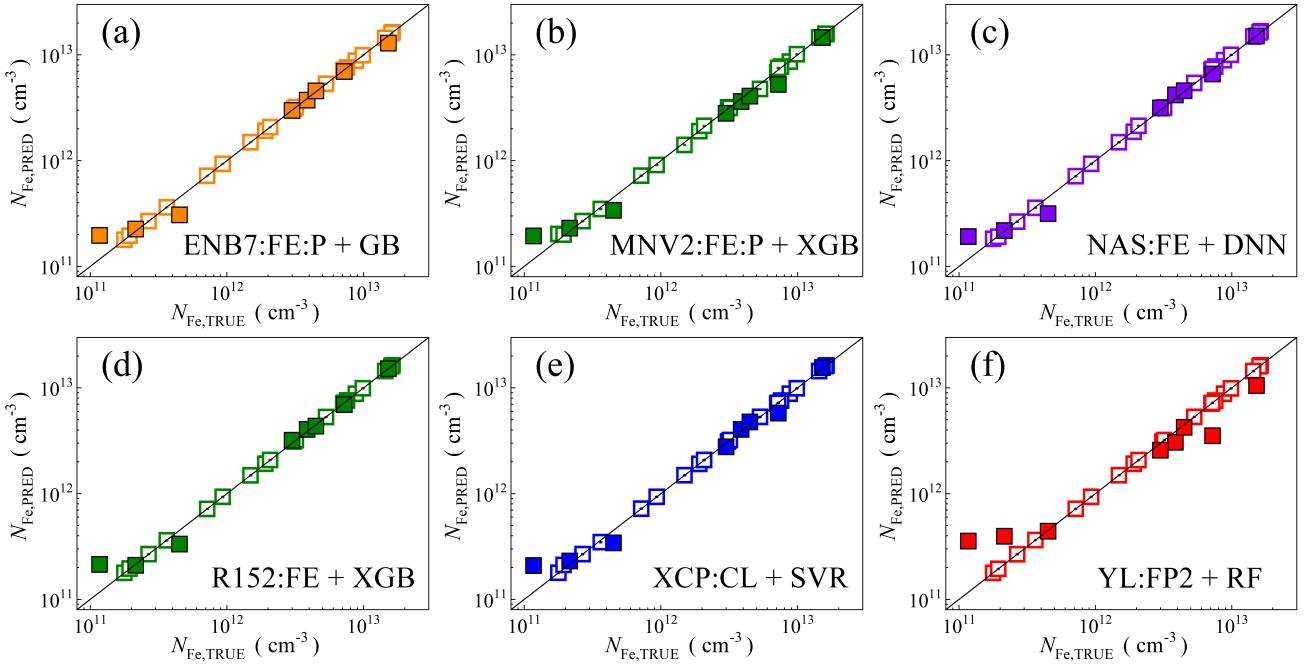


Figure 8. Scatter plots compare the reference iron concentrations $N_{\text{Fe},\text{TRUE}}$ with ML-predicted values $N_{\text{Fe},\text{PRED}}$, obtained using feature vectors extracted from various CV models combined with different regression algorithms (specific models are indicated in the figures). The ML models were trained on a dataset derived from experimental measurements. The open and filled squares correspond to the training and test phases, respectively. The black lines are the identified lines serving as the references.

on synthetic data, this case employed an even smaller sample set (20 samples versus 25). However, these samples covered a narrower range of iron concentrations (10^{11} - $2 \times 10^{13} \text{ cm}^{-3}$ versus 10^{10} - 10^{14} cm^{-3}). Figure 9 presents a subset of the performance metrics obtained during the training phase (a more comprehensive overview is provided in Figure S7). Overall, the results are similar to those shown in Figure 5. In many cases, extremely low errors (below 0.5%) and high coefficients of determination (approaching unity) were observed. This behavior is particularly characteristic of GB, RF, and SVR. As in the previous case, the DNN exhibited weaker performance relative to the other algorithms, although its results slightly improved compared to those for the simulated training dataset. This improvement may stem from the greater homogeneity of the experimental data. The leading performance of EfficientNetB7 and NASNetLarge among the CV models was again confirmed, indicating that these architectures produce the most relevant features for this task. However, the performance gap relative to other CV models has become less pronounced.

Interestingly, the application of PCA often enhances the performance of the DNN. For example, the MAPE values for ENB7:FE and ENB7:FE:P are 7.3% and 0.97%, respectively. This observation suggests that PCA effectively mitigates the adverse effects of high feature dimensionality when the sample

size is limited. In contrast, a similar improvement is not observed for other regression algorithms. Moreover, for the experimental training dataset, the difference between MAPE and MedAPE becomes smaller. This reduction implies a more symmetric error distribution and a decreased occurrence of extreme deviations.

Figure 10 and Figure S8 show the heatmaps of prediction metrics for the test experimental dataset using models trained on a separate subset of experimental data. Compared to models trained on simulated data, the predictive performance for experimental test dataset improved substantially. In particular, for several of the best-performing CV-regressor combinations, MAPE and MedAPE values fall within 7–17%, markedly better than the 20–30% observed previously. Moreover, the R^2 values predominantly exceed 0.97, indicating a strong agreement with the actual dependencies. The accuracy achieved through direct training on experimental data is comparable to that obtained with post-hoc correction; however, the higher correlation coefficients suggest a more faithful representation of both the scale and the variations of the underlying dependency. SVR, DNN, and XGBoost remain the top-performing regressors, whereas EfficientNetB7, NASNetLarge, and — surprisingly — ResNet152V2 (when using image features) are the strongest CV models. YOLOv4 and MobileNetV2 consistently exhibit the weakest

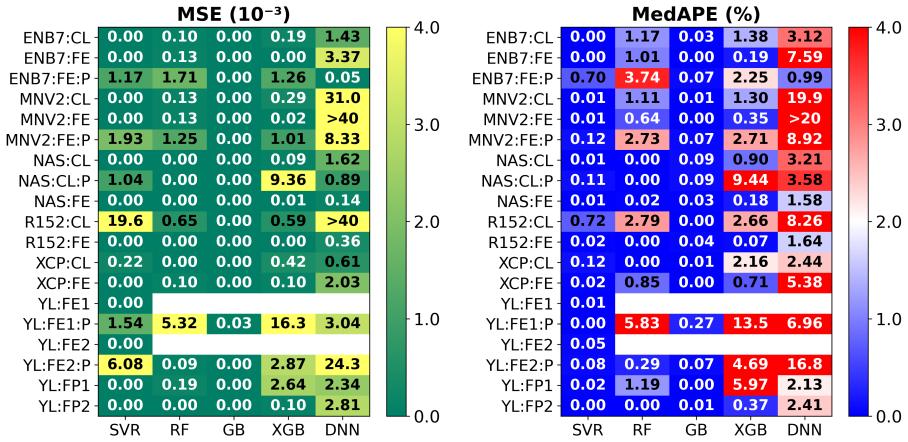


Figure 9. Mean squared error (left panel) and median absolute percentage error (right model) for different combinations of CV models (vertical axis) and regression models (horizontal axis) during the training phase. The models were trained on an experimental dataset.

performance. Interestingly, for experimental data, using class probabilities as descriptors (:CL) yields results comparable to, and in some cases slightly better than, direct image features (:FE), particularly when strong CV architectures are combined with flexible regressors. This likely reflects the more compact and aggregated nature of class features, which makes them less sensitive to experimental noise. Conversely, for weaker CV models, the :FE configurations retain their advantage, consistent with previous findings. Finally, applying PCA to :FE models improved prediction accuracy, whereas it did not benefit models based on class features. This suggests that reducing the dimensionality of highly noisy features contributes to improving prediction accuracy on experimental data.

In summary, three approaches for predicting iron concentrations from experimental data were evaluated: (i) models trained on simulated data, (ii) models trained on simulated data with post-hoc correction, and (iii) models trained directly on experimental data. Training on simulated data alone yielded moderate agreement with experimental measurements, but systematic biases were observed due to differences between the synthetic and real systems. Post-hoc correction effectively reduced these biases, lowering the mean and median errors, yet the correlation with the actual variation remained limited. Direct training on experimental data provided the most balanced outcome, achieving both low prediction errors and high correlation coefficients, thereby demonstrating the importance of incorporating real measurements during model development. Thus, training directly on experimental data improves prediction accuracy and eliminates the systematic biases observed in models transferred from synthetic data. At the same time, achieving optimal performance requires

both the careful selection of the CV model-regressor combination and the inclusion of real experimental data.

4. Conclusion

This study demonstrates that Transfer Learning from pretrained Computer Vision (CV) models enables accurate regression modeling even with the extremely small training datasets typical of experimental materials research. The proposed workflow involves measuring a characteristic kinetic dependency, transforming it into an image via wavelet analysis, extracting features using a pretrained CV model, and training a regression model on these features to predict material properties. The feasibility of this approach is illustrated by predicting the iron impurity concentration in silicon solar cell from short-circuit current kinetics following FeB pair dissociation using a training dataset of only 20–25 samples.

The performance of models trained on both synthetic and experimental datasets was evaluated. In both cases, EfficientNetB7 and NASNetLarge provided the most informative features, while Deep Neural Networks and Support Vector Regression yielded the highest prediction accuracy. The best-performing models achieved MSE, MAPE, MedAPE, and R^2 values of 0.001, 6%, 4%, and 0.999 for synthetic data, and 0.008, 10%, 5%, and 0.996 for experimental data.

When training models on synthetic data, image feature vectors serve as the most suitable descriptors for the regression model. In contrast, when experimental data are used, prediction accuracy can be improved and the influence of noise reduced by utilizing class probability distributions or applying Principal Component Analysis.

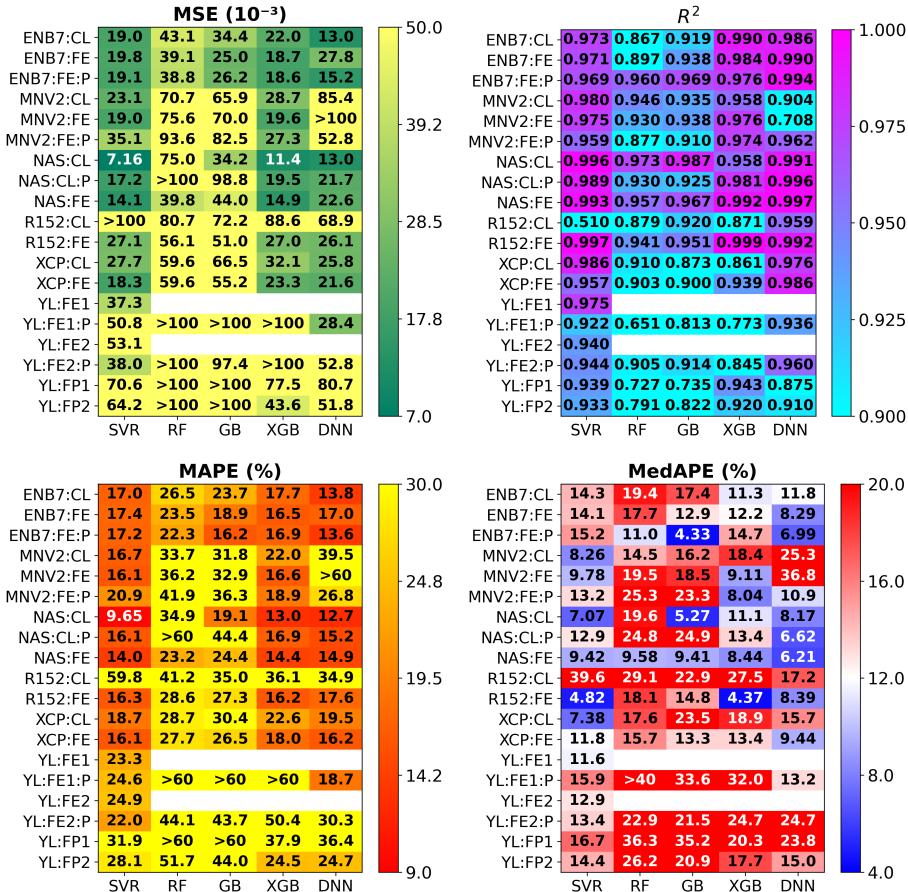


Figure 10. Mean squared error, coefficient of determination, mean absolute percentage error, and median absolute percentage error for different combinations of CV models (vertical axis) and regression models (horizontal axis) during the test phase. The models were trained on an experimental dataset.

The combination of Transfer Learning from CNNs with an appropriate choice of descriptor type and regression algorithm represents a promising strategy for materials research in general and for defect characterization in particular, especially when the acquisition of large datasets is difficult or impractical.

References

- [1] International Energy Agency 2024 Renewables 2024 Tech. rep. IEA Paris licence: CC BY 4.0 URL <https://www.iea.org/reports/renewables-2024>
- [2] Osama A, Tina G M, Gagliano A, Jimenez-Castillo G and Munoz-Rodríguez F J 2025 *Sol. Energ. Mat. Sol.* **287** 113625
- [3] Institute P 2025 Solar generation grew by 30% in 2024, says IEA <https://prometheus.org/2025/02/23/solar-generation-grew-by-30-in-2024-says-iea/> accessed: 2025-10-10
- [4] Fischer M, Woodhouse M, Brammer T and Puzant B 2025 International technology roadmap for photovoltaic (itrv) Tech. rep. VDMA e. V. Frankfurt am Main, Germany
- [5] Thome F T, Garashli E, Kwapił W, Schindler F and Schubert M C 2025 *Sol. Energ. Mat. Sol.* **293** 113854
- [6] Jia Y, Chen G and Zhao L 2024 *Sci. Rep.* **14** 15170
- [7] Hijjawi U, Lakshminarayana S, Xu T, Piero Malfense Fierro G and Rahman M 2023 *Sol. Energy* **266** 112186 ISSN 0038-092X
- [8] Liu Y, Wu Y, Yuan Y and Zhao L 2024 *Opt. Express* **32** 17295–17317
- [9] Doll B, Hepp J, Hoffmann M, Schüler R, Buerhop-Lutz C, Peters I M, Hauch J A, Maier A and Brabec C J 2021 *IEEE J. Photovolt.* **11** 1419–1429
- [10] Datta S, Baul A, Sarker G C, Sadhu P K and Hodges D R 2023 *IEEE Access* **11** 77750–77778
- [11] Jaiswal R, Martínez-Ramón M and Busani T 2023 *IEEE J. Photovolt.* **13** 2–15
- [12] Buratti Y, Javier G M, Abdullah-Vetter Z, Dwivedi P and Hameiri Z 2024 *Renewable Sustainable Energy Rev.* **202** 114617 ISSN 1364-0321
- [13] Mahdavipour Z 2024 *Sol. Energ. Mat. Sol.* **278** 113210
- [14] Hopwood M W, Gunda T, Seigneur H and Walters J 2020 *IEEE Access* **8** 161480–161487
- [15] Li B, Delpha C, Diallo D and Migan-Dubois A 2021 *Renewable Sustainable Energy Rev.* **138** 110512 ISSN 1364-0321
- [16] Liu Y, Ding K, Zhang J, Li Y, Yang Z, Zheng W and Chen X 2021 *Energy Convers. Manage.* **245** 114603
- [17] Wang S, Wright B, Zhu Y, Buratti Y and Hameiri Z 2024 *Sol. Energ. Mat. Sol.* **277** 113123
- [18] Buratti Y, Dick J, Le Gia Q and Hameiri Z 2022 *ACS Appl.*

- Mater. Interfaces* **14** 48647–48657
- [19] Buratti Y, Le Gia Q T, Dick J, Zhu Y and Hameiri Z 2020 *npj Computational Materials* **6** 142
- [20] Park S, Lee J, Khan S, Wahab A and Kim M 2022 *Sensors* **22** 596
- [21] Chia J Y, Thamrong siripak N, Thongphanit S and Nuntawong N 2024 *J. Appl. Phys.* **135** 025701
- [22] Olikh O, Lozitsky O and Zavhorodnii O 2022 *Prog. Photovoltaics Res. Appl.* **30** 648–660
- [23] Olikh O and Zavhorodnii O 2025 *Sol. Energy* **300** 113754
- [24] Olikh O and Zavhorodnii O 2025 *Materials Science and Engineering: B* **317** 118192
- [25] Wang S, Sankaran S and Perdikaris P 2024 *Comput. Methods Appl. Mech. Eng.* **421** 116813
- [26] Li W K and Zhang Y T 2025 *J. Appl. Phys.* **137** 203304
- [27] Kaya M and Hajimirza S 2019 *Sci. Rep.* **9** 5034
- [28] Kim Q, Lee S, Ma A, Kim J, Noh H K, Chang K B, Cheon W, Yi S, Jeong J, Kim B, Kim Y S and Kim D S 2023 *Solid-State Electron.* **201** 108568
- [29] Buonassisi T, Istratov A A, Pickett M D, Heuer M, Kalejs J P, Hahn G, Marcus M A, Lai B, Cai Z and Heald S M 2006 *Prog. Photovolt.: Res. Appl.* **14** 513–531 ISSN 1062-7995
- [30] Schubert M, Padilla M, Michl B, Mundt L, Giesecke J, Hohl-Ebinger J, Benick J, Warta W, Tajima M and Ogura A 2015 *Sol. Energy Mater. Sol. Cells* **138** 96–101
- [31] Kimerling L and Benton J 1983 *Physica B+C* **116** 297–300
- [32] Möller C, Bartel T, Gibaja F and Lauer K 2014 *J. Appl. Phys.* **116** 024503
- [33] Olikh O, Kostylyov V, Vlasiuk V, Korkishko R, Olikh Y and Chupryna R 2021 *J. Appl. Phys.* **130** 235703
- [34] Li Z, Zhang S, Qu C, Zhang Z and Sun F 2024 *PLOS ONE* **19** 1–16
- [35] Otamendi U, Martinez I, Quartulli M, Olaizola I G, Viles E and Cambarau W 2021 *Sol. Energy* **220** 914–926
- [36] Chen X, Karin T and Jain A 2022 *Sol. Energy* **242** 20–29
- [37] Munawer Al-Otum H 2024 *Sol. Energy* **278** 112803
- [38] Abdelsattar M, Abdelmoety A, Ismeil M A and Emad-Eldeen A 2025 *IEEE Access* **13** 4136–4157
- [39] Tella H, Hussein A, Rehman S, Liu B, Balghonaim A and Mohandes M 2025 *Case Studies in Thermal Engineering* **66** 105749
- [40] Li W C and Tsai D M 2012 *Pattern Recogn.* **45** 742–756
- [41] dela Rosa M E C, Mateo-Romero H F, Alonso-Gómez V, Ngungu V N, Nava R, Aragónés J I M, Plaza A R, Gonzalez-Rebollo M A, Isaza J R F and Cardeñoso-Payo V 2024 *Renewable Energies* **2** 27533735241304090
- [42] Khanna M, Srinath N K and Mendiratta J K 2020 *Application of Neural Networks and Lifting Wavelet Transform for Long Term Solar Radiation Power Prediction* (Singapore: Springer Singapore) pp 95–105 ISBN 978-981-15-3125-5
- [43] Krishnan S R and Seelamantula C S 2013 *IEEE Trans. Signal Process.* **61** 380–391
- [44] Torrence C and Compo G P 1998 *Bull. Am. Meteorol. Soc.* **79** 61–78
- [45] Ahmad A, Jin Y, Zhu C, Javed I, Maqsood A and Akram M W 2020 *IET Renew. Power Gener.* **14** 2693–2702
- [46] Burgelman M, Nollet P and Degrave S 2000 *Thin Solid Films* **361–362** 527–532 ISSN 0040-6090
- [47] Masum Mia M, Faruk Hossain M, Rahman M, Badi N, Irfan A and Ferdous Rahman M 2025 *Materials Science and Engineering: B* **311** 117817
- [48] Joshi T K, Sharma G, Sharma Y R and Verma A S 2024 *Phys. B Condens. Matter* **682** 415793
- [49] Ravidas B K, Das A, Agnihotri S K, Pandey R, Madan J, Hossain M K, Roy M K and Samajdar D 2024 *Sol. Energ. Mat. Sol.* **267** 112688 ISSN 0927-0248
- [50] Liu H, Xiang L, Liu Q, Gao P, Zhang Y, Li S and Gao F 2024 *IEEE J. Photovoltaic* **14** 59–64
- [51] You L, Zhang X, Ma Q, Zhu W and Wu J 2023 *Phys. Status Solidi A* **220** 2300071
- [52] Abdulmalik M and Danladi E 2023 *Semiconductor Physics, Quantum Electronics & Optoelectronics* **26** 321–331
- [53] Paviet-Salomon B, Levrat J, Fakhfouri V, Pelet Y, Rebeaud N, Despeisse M and Ballif C 2017 *Prog. Photovoltaics Res. Appl.* **25** 623–635
- [54] Murphy J D, Bothe K, Olmo M, Voronkov V V and Falster R J 2011 *J. Appl. Phys.* **110** 053713
- [55] Wijaranakula W 1993 *J. Electrochem. Soc.* **140** 275–281
- [56] Khelifati N, Laine H S, Vähänissi V, Savin H, Bouamama F Z and Bouhafs D 2019 *Phys. Status Solidi A* **216** 1900253
- [57] Tan J, Macdonald D, Rougieux F and Cuevas A 2011 *Semicond Sci. Technol.* **26** 055019
- [58] Olikh O 2019 *Superlattices Microstruct.* **136** 106309
- [59] Olikh O, Datsenko O and Kondratenko S 2024 *Phys. Status Solidi A* **221** 2400351
- [60] Olikh O, Kostylyov V, Vlasiuk V, Korkishko R and Chupryna R 2022 *J. Mater. Sci.: Mater. Electron.* **33** 13133–13142
- [61] Fadhel S, Delpha C, Diallo D, Bahri I, Migan A, Trabelsi M and Mimouni M 2019 *Sol. Energy* **179** 1–10
- [62] Gao W and Wai R J 2020 *IEEE Access* **8** 159493–159510
- [63] Srivastava Y and Jain A 2023 *J. Appl. Phys.* **134** 225101
- [64] Minagawa H, Tezuka T and Tsuchida H 2024 *Nucl. Instrum. Methods Phys. Res. Sect. B Beam Interact. Mater. At.* **553** 165383
- [65] Lauer K, Möller C, Debbih D, Auge M and Schulze D 2016 Determination of activation energy of the iron acceptor pair association and dissociation reaction *Gettinger and Defect Engineering in Semiconductor Technology XVI (Solid State Phenomena vol 242)* (Trans Tech Publications Ltd) pp 230–235
- [66] Zhu X, Yang D, Yu X, He J, Wu Y, Vanhellemont J and Que D 2013 *AIP Adv.* **3** 082124
- [67] Zhu X, Yu X, Li X, Wang P and Yang D 2011 *Scr. Mater.* **64** 217–220
- [68] Kimerling L and Benton J 1983 *Physica B+C* **116** 297–300
- [69] Le T T, Zhou Z, Chen A, Yang Z, Rougieux F, Macdonald D and Liu A 2024 *J. Appl. Phys.* **135** 133107
- [70] Macdonald D, Roth T, Deenapanray P N K, Bothe K, Pohl P and Schmidt J 2005 *J. Appl. Phys.* **98** 083509
- [71] Zhu X, Yu X, Chen P, Liu Y, Vanhellemont J and Yang D 2015 *Int. J. Photoenergy* **2015** 154574
- [72] Macdonald D H, Geerligs L J and Azzizi A 2004 *J. Appl. Phys.* **95** 1021–1028
- [73] Zoth G and Bergholz W 1990 *J. Appl. Phys.* **67** 6764–6771 ISSN 0021-8979 (*Preprint* https://pubs.aip.org/aip/jap/article-pdf/67/11/6764/18634381/6764_1_online.pdf) URL <https://doi.org/10.1063/1.345063>
- [74] Geerligs L J and Macdonald D 2004 *Appl. Phys. Lett.* **85** 5227–5229
- [75] Macdonald D, Cuevas A and Geerligs L J 2008 *Appl. Phys. Lett.* **92** 202119
- [76] Murphy J D, McGuire R E, Bothe K, Voronkov V V and Falster R J 2014 *J. Appl. Phys.* **116** 053514 ISSN 0021-8979
- [77] Hayamizu Y, Hamaguchi T, Ushio S, Abe T and Shimura F 1991 *J. Appl. Phys.* **69** 3077–3081
- [78] Schmidt J 2005 *Progress in Photovoltaics: Research and Applications* **13** 325–331
- [79] Næland T, Bernardini S, Stoddard N, Good E, Augusto A and Bertoni M 2017 *Energy Procedia* **124** 138–145