

## Highlights

**Determination** the iron concentration in silicon solar cells using photovoltaic parameters and machine learning

Oleg Olikh, Oleksii Zavhorodnii

- 80 ML models for iron quantification in silicon solar were tested
- XGB and DNN models outperform RF, GB, and SVR models
- MSE, MAPE, and  $R^2$  reach up to 0.003, 3%, and 0.997 for synthetic data and 0.004, 9%, and 0.987 for experimental data
- Photovoltaic parameter variations due to FeB dissociation used as ML model inputs.

# Determination of the iron concentration in silicon solar cells using photovoltaic parameters and machine learning

Oleg Olikh\*, Oleksii Zavhorodnii

Taras Shevchenko National University of Kyiv, 64/13, Volodymyrska Street, Kyiv, 01601, Ukraine

## ARTICLE INFO

### Keywords:

photovoltaic defect quantification  
silicon  
solar cells  
iron contamination  
machine learning  
current-voltage curves  
feature selection

## ABSTRACT

This study introduces a pioneering machine learning (ML)-based methodology for quantifying iron impurities in silicon solar cells. A comprehensive analysis was done on 80 models, utilizing algorithms such as Random Forest (RF), Gradient Boosting (GB), eXtreme Gradient Boosting (XGB), Support Vector Regression (SVR), and Deep Neural Networks (DNN) to predict iron concentration based on variations in photovoltaic parameters caused by FeB pair dissociation. The conditions the training dataset must meet to minimize forecasting errors were identified, along with the feature combinations that yield the most accurate predictions. Furthermore, the effectiveness of using Principal Component Analysis for data pre-processing was assessed. The results demonstrate that XGB and DNN outperform other models, achieving MSE, MAPE, and  $R^2$  values of up to 0.003, 3%, and 0.997 for synthetic data and 0.004, 9%, and 0.987 for experimental data.

## 1. Introduction

The urgent challenges of global warming and fossil fuel depletion have spurred the search for renewable energy sources, accelerating the rapid advancement of photovoltaic (PV) energy and the widespread deployment of solar panels [1, 2]. However, PV energy faces several significant challenges: the need to optimize solar cell manufacturing processes, improve the efficiency and stability of PV devices (either by discovering new materials or by enhancing existing structures), develop energy management systems designed to regulate the energy flow in real-time and establish methods for monitoring the condition of photoelectric converters, specifically through defect characterization. Simultaneously, the application of machine learning (ML) methods, which can identify patterns and correlations that are not readily apparent through traditional analysis [3], opens new avenues for addressing the issues above in photovoltaics. For instance, ML methods optimize individual production stages — such as crystal growth [4] and plasma-enhanced vapor deposition [5] — and streamline entire solar cell production lines, which may involve up to ten processing steps and forty-seven process parameter inputs [6]. Moreover, ML techniques enhance PV devices by identifying key photovoltaic materials, analyzing references to various structural configurations in the literature [7] or optimizing perovskite solar cells [8], designing renewable energy-based demand-side management systems [9], and extracting additional information from internal quantum efficiency measurements [10]. Numerous reviews provide further insights into ML applications in PV systems [11, 12, 13, 14, 15].

We will now explore the application of artificial intelligence in greater depth for defect analysis. Most relevant studies focus on image analysis of solar cells, including electroluminescence, photoluminescence, and infrared thermography [16, 17, 18, 19, 20, 21, 22, 23]. These methods facilitate detecting and classifying defects such as cracks, finger failures, hot spots, scratches, and horizontal dislocations and are predominantly implemented using convolutional neural networks. Another widely adopted approach involves applying ML models to current-voltage ( $I$ - $V$ ) curves, enabling the identification of permanent and temporal faults in PV arrays [24, 25, 26, 27]. A key advantage of this defect characterization method is its reliance on  $I$ - $V$  measurements, a standard procedure for PV device assessment. Additionally, ML techniques specifically designed for analyzing point defects warrant particular attention. For instance, researchers have developed methods for determining the electronic structure of intrinsic defects [28, 29], detecting radiation-induced defects via Raman spectroscopy [30, 3], and extracting recombination-active center parameters from temperature- and injection-dependent lifetime measurements [31, 32, 33]. Although such defects represent one of the main limitations of PV devices [34, 35], studies in this area remain scarce. Moreover, existing research primarily

\*Corresponding author. E-mail address: nevermor464@gmail.com (O. Zavhorodnii).  
ORCID(s): 0000-0003-0633-5429 (O. Olikh); 0000-0001-8080-7661 (O. Zavhorodnii)

focuses on characterizing defects in PV materials rather than solar cells and relies on data obtained through specialized equipment.

This study proposes an ML-based approach to extract impurity concentrations from  $I$ - $V$  curves. Specifically, we focus on quantifying iron in boron-doped crystalline silicon solar cells. This constraint is not overly restrictive, given that (i) Si-based solar cells dominate the current PV market [36, 37], with most being manufactured from boron-doped crystalline silicon (c-Si:B); and (ii) iron is one of the most prevalent, ubiquitous, and efficiency-limiting metallic impurities [38, 39]. It is well established that in  $p$ -type material, iron tends to bind with dopant atoms such as boron, forming iron–boron pairs under equilibrium conditions or existing as interstitial species only in the presence of sufficiently high free electron densities [40, 41]. The deliberate transition between these states can be readily induced through intense illumination, electron injection, or heating up to 200 °C and is commonly employed in various methods for assessing iron concentration [42, 43, 44, 45, 46, 47, 48, 49]. The approach proposed in this study leverages changes in photovoltaic parameters (PVPs) (short-circuit current, open-circuit voltage, efficiency, and fill factor) resulting from FeB dissociation as input features for ML algorithms. The specified PVPs can be easily extracted from  $I$ - $V$  characteristics, making this method advantageous compared to existing approaches:

- (1) Unlike glow discharge mass spectrometry or secondary ion mass spectrometry [50], it is non-destructive;
- (2) It does not require specialized equipment or specially prepared samples, in contrast to Fourier-transform infrared spectroscopy, electron paramagnetic resonance, deep-level transient spectroscopy (DLTS), Laplace DLTS, carrier lifetime measurements, photoluminescence, or photoconductance [51, 52, 53, 43, 44, 47, 45];
- (3) It is relatively simple and fast compared to other methods that also rely on measuring PV parameters, such as the kinetics of short-circuit current [46] and open-circuit voltage [48], but require lengthy experimental procedures or multiple illumination levels.

In our previous work [54], we employed a deep learning approach to estimate iron concentration based on the ideality factor, which was also derived from  $I$ - $V$  characteristics. However, the method proposed in this study imposes fewer constraints on the accuracy of  $I$ - $V$  measurements across the entire voltage range and the model used to describe the  $I$ - $V$  characteristics of actual structures. Thus, our approach allows for the simultaneous determination of iron impurity concentration along with key electrical parameters. These inline characterization techniques are crucial for ensuring efficient production lines and optimizing processes to produce reliable solar cells. Notably, luminescence imaging is increasingly used for solar cell characterization. However, ML methods have been proposed to extract  $I$ - $V$  characteristic data from such images [55, 56]. Integrating these methods with our approach into a unified pipeline would enable the extraction of iron concentration data from luminescence measurements.

## 2. Methodology

The workflow of the ML pipeline for extracting iron contamination from IV data is shown in Fig. 1. A more detailed description follows.

### 2.1. Data Collection

Our research aimed to develop ML models for estimating iron concentration in silicon solar cells. Selecting relevant descriptors is a critical first step in building robust and effective predictive models. It is well established that the presence of iron induces the formation of recombination centers, which, in turn, affect the photoelectric conversion process and key PV parameters, including short-circuit current ( $I_{SC}$ ), open-circuit voltage ( $V_{OC}$ ), efficiency ( $\eta$ ), and fill factor ( $FF$ ). Both theoretical and experimental studies [57, 39, 58] have demonstrated that the restructuring of iron-related defects (the dissociation of FeB pairs) leads to variations in PVPs that depend on iron concentration ( $N_{Fe}$ ). As a result, this investigation focused on using iron presence as a descriptor of the relative changes in PVPs caused by the  $Fe_iB_{Si} \rightleftharpoons Fe_i + B_{Si}$  reconstruction. Specifically, the relative changes in short-circuit current,  $\epsilon I_{SC}$ , were calculated as

$$\epsilon I_{SC} = \frac{I_{SC}^{FeB} - I_{SC}^{Fe}}{I_{SC}^{FeB}} \times 100\%, \quad (1)$$

where  $I_{SC}^{FeB}$  and  $I_{SC}^{Fe}$  represent the short-circuit current values before and after pair dissociation, respectively. The relative changes in the other parameters ( $\epsilon V_{OC}$ ,  $\epsilon \eta$  and  $\epsilon FF$ ) were determined similarly. Using relative rather than absolute changes helped, to some extent, isolate iron-related defect contributions from those of other recombination centers and mitigate potential fluctuations in external conditions, such as illumination intensity.

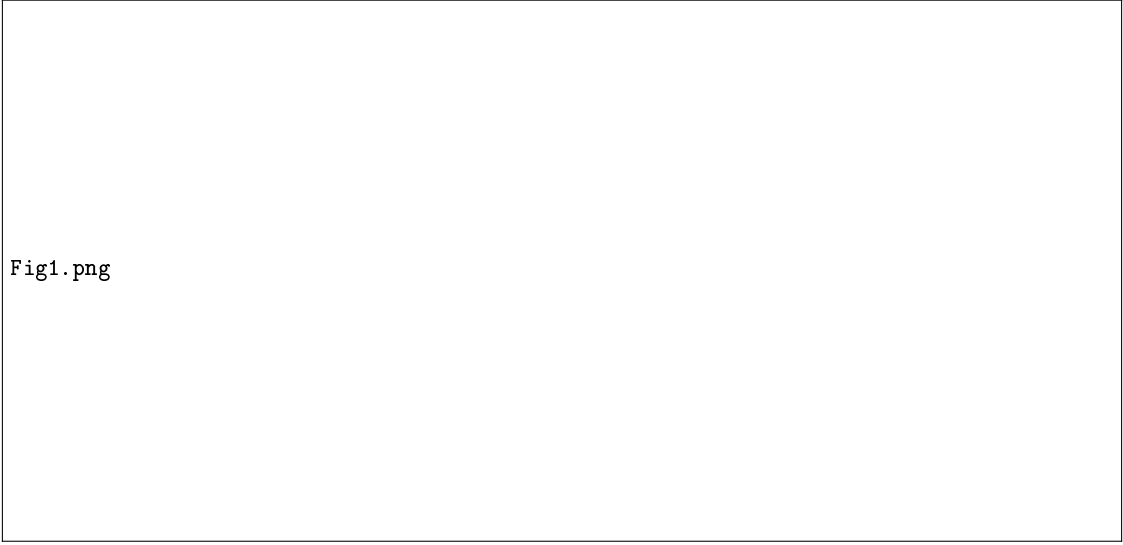


Fig1.png

**Figure 1:** The workflow of the ML pipeline.

At the same time, changes in PVPs are shown [57, 58] to be influenced not only by iron concentration but also by specific characteristics of the solar cell structure, such as doping level and base thickness, as well as external factors, including temperature and illumination spectrum. Therefore, these effects must be considered when selecting an appropriate set of descriptors. In our study, we incorporated features such as boron concentration in the base ( $N_B$ ), base thickness ( $d_p$ ), and temperature ( $T$ ) and developed separate models for different illumination conditions.

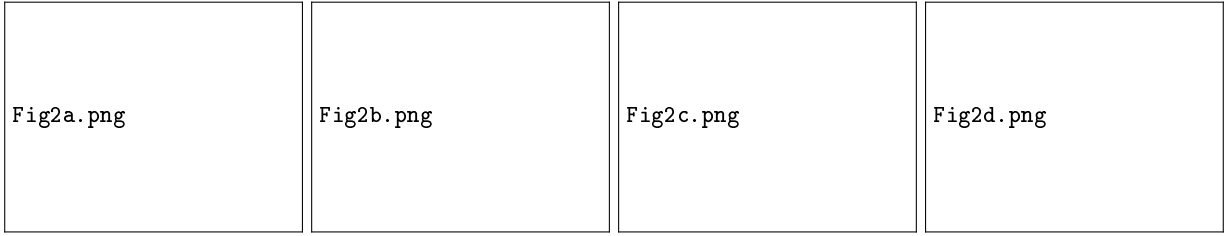
For this research, datasets were obtained from both simulations and experiments. Specifically,  $I$ - $V$  curve simulations were performed for a silicon  $n^+$ - $p$ - $p^+$  structure using SCAPS 3.3.11. The SCAPS-1D software [59] is a widely used tool [60, 61, 62, 63, 64, 65] that enables modeling of barrier structure  $I$ - $V$  characteristics while accounting for defect states. PVPs were extracted from pairs of  $I$ - $V$  curves simulated for a solar cell with a known structure and a specified iron concentration under defined external conditions, both before and after FeB pair dissociation. This approach allowed us to determine the values of  $\epsilon I_{SC}$ ,  $\epsilon V_{OC}$ ,  $\epsilon \eta$ , and  $\epsilon F/F$ . A detailed description of the modeling approach, including the used temperature dependencies of silicon and defect parameters, is provided elsewhere [66, 58].

To generate the training dataset, we performed simulations over a temperature range of (290 – 340) K for solar cells with a base thickness of (180 – 380)  $\mu\text{m}$ , a boron concentration of ( $10^{15}$  –  $10^{17}$ )  $\text{cm}^{-3}$ , and an iron concentration of ( $10^{10}$  –  $10^{14}$ )  $\text{cm}^{-3}$ . The simulations covered all parameter combinations within a four-dimensional grid, consisting of 11 values along the  $T$  axis, 5 along the  $d_p$  axis, 9 along the  $N_B$  axis, and 25 along the  $N_{Fe}$  axis. The values were evenly distributed within the specified ranges, with  $N_B$  and  $N_{Fe}$  spaced logarithmically. Each train dataset consisted of 12,375 samples obtained from simulations of 24,750  $I$ - $V$  characteristics.

We also generated four test datasets through simulation, each representing different scenarios that may arise in the practical application of ML models. In one test dataset, the iron concentration values used in the simulation were absent from the training set, while the other parameters ( $T$ ,  $d_p$ ,  $N_B$ ) matched the grid node values from the training simulations. This dataset was designated as “ $N_{Fe}$ -altered” and contained 1,034 samples. Similarly, we created the “ $N_B$ -altered” (1,100 samples) and “ $T$ -altered” (1,200 samples) test datasets using boron concentration and temperature values that were not present in the training dataset. In the “All-altered” dataset (1,190 samples), all four simulation parameters differed from those used in the training dataset.

The specified simulations for both the training and test datasets were performed under two lighting scenarios: AM1.5, which corresponds to standard characterization conditions for PV devices, and monochromatic light with a wavelength of 940 nm and an intensity of 5  $\text{W/m}^2$ , reflecting the remarkable sensitivity of PVPs to iron [58]. A total of 67,596  $I$ - $V$  curves were simulated.

For experimental validation of the proposed models, we used data obtained from measurements of real solar cells. We used a set of  $n^+$ - $p$ - $p^+$ -Si samples in the experiment. The structure was fabricated from a 380  $\mu\text{m}$  thick  $p$ -type boron-doped Czochralski silicon (100) wafer with a doping level of  $N_B = 1.36 \times 10^{15} \text{ cm}^{-3}$ . Iron concentration, determined



**Figure 2:** Cumulative explained variance versus number of PC components. Feature dimension: 4 (a), 5 (b), 6 (c), 7 (d).

using the methodology described in [67, 46], ranged from  $2 \times 10^{11}$  to  $4 \times 10^{13} \text{ cm}^{-3}$  across different samples. To create the experimental test dataset, we measured the  $I$ - $V$  characteristics under monochromatic illumination from a light-emitting diode (SN–HPIR940nm–1W) with a wavelength of 940 nm and an intensity of approximately  $5 \text{ W/m}^2$ , within a temperature range of (305 – 340) K. The decay of FeB pairs was induced using intensive halogen lamp illumination ( $7000 \text{ W/m}^2$ ). The number of samples in experimental dataset varied from 34 to 12, depending on the feature dimensionality.

## 2.2. Data Pre-Processing

Feature selection is a crucial pre-processing step in developing a forecasting model. Selecting the appropriate number of features requires balancing two competing factors. On one hand, fewer features simplify data collection and reduce computational costs. On the other hand, incorporating more features may enhance prediction accuracy by providing additional information. In this study, various feature combinations were explored. As shown in [58],  $\epsilon I_{SC}$  is a key metric for quantifying iron impurities due to its monotonic dependence on iron concentration and relatively large absolute values during the  $\text{Fe}_i\text{B}_{Si} \rightleftharpoons \text{Fe}_i + \text{B}_{Si}$  transformation. The following most suitable parameters for estimating  $N_{Fe}$  are  $\epsilon\eta$ ,  $\epsilon V_{OC}$ , and  $\epsilon FF$  [58].

Moreover,  $N_B$  (one of the selected descriptors) and  $N_{Fe}$  (target output) span several orders of magnitude. To achieve high prediction accuracy, we transformed the doping level and iron concentration into  $\log N_B$  and  $\log N_{Fe}$ . This approach is standard for quantities that vary over a wide range [68, 69].

Thus, we used the following sets to predict  $y = \log N_{Fe}$ :  $(T, d_p, N_B, \epsilon I_{SC})$ ;  $(T, d_p, N_B, \epsilon I_{SC}, \epsilon\eta)$ ;  $(T, d_p, N_B, \epsilon I_{SC}, \epsilon\eta, \epsilon V_{OC})$ ;  $(T, d_p, N_B, \epsilon I_{SC}, \epsilon\eta, \epsilon V_{OC}, \epsilon FF)$ . For simplicity, we will refer to the number of features as “dimension” from here on. That is, features with dimensions 4, 5, 6, and 7 were used in various models.

Data normalization transforms a dataset’s values to a standard scale, which can improve model accuracy. To standardize our features, we normalized them to have a mean of zero and a standard deviation of one.

Increasing the number of features does not always enhance informativeness if the descriptors are not independent. This study used the Pearson correlation coefficient to assess the relationships between input features. We found that changes in PV parameters are not entirely independent, particularly under monochromatic illumination (see Fig. S1 in the Supplementary Material). This is not surprising, since each of the PV parameters is linked to the diffusion and recombination of photo-induced charge carriers. Besides,  $\epsilon I_{SC}$ ,  $\epsilon V_{OC}$ ,  $\epsilon\eta$ , and  $\epsilon FF$  exhibit a strong correlation with boron concentration.  $N_B$  determines the position of the Fermi level, which, in turn, significantly influences the intensity of recombination processes within the Shockley–Read–Hall approximation. This provides the physical basis for the observed correlation.

To ensure that ML algorithms are trained on independent input features while reducing data redundancy and memory usage, we applied Principal Component Analysis (PCA). PCA constructs new, uncorrelated features (principal components, PCs) and allows us to evaluate each PC’s contribution to the total variance. Fig.2 shows that increasing the feature dimension does not always enhance overall information variance. To assess the impact of redundant data, we trained ML models using features selected as follows: we computed PCs for the original dataset and then reduced the total dimension by discarding PCs whose contribution to the total variance was below 1.5%. The final dimensions of the different feature sets are provided in Table 1.

## 2.3. Machine Learning Algorithms

We used five ML algorithms to develop regression models for predicting iron concentration.

**Table 1**

Feature dimension after PCA applying

Initial dimension	Final dimension	
	AM1.5 illumination	940 nm illumination
4	4	4
5	4	4
6	5	4
7	6	5

*Random Forest* (RF) improves predictive accuracy by training multiple decision trees on different subsets of the dataset. RF aggregates predictions from all trees using majority voting for classification or averaging for regression, reducing overfitting and increasing robustness to noise [70].

*Gradient Boosting* (GB) combines multiple weak learners, typically decision trees, to improve predictive performance. It sequentially adds predictors, with each new model correcting the errors of its predecessor, thereby enhancing overall accuracy. The final prediction is obtained by aggregating forecasts, usually through a weighted sum [71].

*eXtreme Gradient Boosting* (XGB) is an advanced ensemble method for gradient boosting machines [72]. Unlike standard GB, XGB utilizes the Newton–Raphson method with second-order derivatives of the loss function, improving accuracy, computational speed, and efficiency for large datasets.

*Support Vector Regression* (SVR) finds a hyperplane that maximizes the margin while minimizing errors within a specified tolerance. It maps input data into a high-dimensional feature space using kernel functions, effectively capturing nonlinear relationships [73].

*Deep Neural Network* (DNN) comprises multiple layers of interconnected neurons that process input data through successive nonlinear transformations [8].

It is worth noting that all the algorithms used have demonstrated effectiveness in previous defect studies [33, 32, 54, 14].

Considering different machine learning algorithms, data collected under various lighting conditions, features with varying dimensions, and the inclusion or exclusion of PCA, we evaluated 80 distinct models. From this point forward, the models will be referred to as

$$A_{\text{feat}}^{\text{illum}},$$

where “A” represents the ML algorithm used, and it can take one of the following values: (RF, GB, XGB, SVR, DNN); “illum” indicates the solar cell illumination type:  $\text{illum} \in (\text{AM1.5}, 940)$ ; “feat” represents the feature dimension and the application of PCA, where  $\text{feat} \in (4, 5, 6, 7, 4:\text{PC}, 5:\text{PC}, 6:\text{PC}, 7:\text{PC})$ .

The models are implemented using Python toolkits: Keras for DNN, Scikit-learn for RF, GB, and SVR, and XGBoost for XGB. Hyperparameter tuning is known [74] to be essential for optimizing model performance. We used the Optuna toolkit to optimize model parameters, employing the TPE sampler and Hyperband pruner for efficient hyperparameter selection. Tables S1–S5 (Supplementary Material) provide the tuned hyperparameters and their search ranges. It is worth noting that 5-fold cross-validation was employed during the model tuning process, with 20% of the train data used as the validation set to evaluate models trained on the remaining 80%. The chosen hyperparameter combinations are presented in Tables S6–S10.

## 2.4. Model evaluation

To build a regression model, it is crucial to evaluate its performance using various metrics. These metrics assess how well the model has learned and predicted outcomes. The evaluation metrics for iron quantification were the mean squared error (MSE), mean absolute percentage error (MAPE), and coefficient of determination ( $R^2$ ), as defined in Eqs. (2)–(4).

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

where  $\hat{y}_i$  represents the predicted value for the  $i$ -th data point,  $y_i$  is the known value for the  $i$ -th data point, and  $N$  is the number of samples in the dataset. MSE is one of the most commonly used metrics for evaluating model accuracy.

However, since the computation of  $\hat{y}_i$  involves the normalization and logarithm transformation of  $N_{\text{Fe}}$ , this metric does not fully reflect the accuracy of iron contamination estimation. Therefore, we used MAPE, which determines the mean relative error:

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \frac{|N_{\text{Fe,PRED},i} - N_{\text{Fe,TRUE},i}|}{N_{\text{Fe,TRUE},i}} \times 100\%, \quad (3)$$

where  $N_{\text{Fe,PRED},i}$  is the predicted value of iron concentration,  $N_{\text{Fe,TRUE},i}$  is the known value (used in simulation or obtained from experimental iron determination).

The  $R^2$  score is often interpreted as the percentage of explained variance and measures how well the predicted and true values align; a value of unity indicates a perfect prediction:

$$R^2 = 1 - \frac{\sum_{i=1}^N (N_{\text{Fe,TRUE},i} - N_{\text{Fe,PRED},i})^2}{\sum_{i=1}^N (N_{\text{Fe,TRUE},i} - \overline{N_{\text{Fe,TRUE}}})^2}, \quad (4)$$

where  $\overline{N_{\text{Fe,TRUE}}}$  is the mean of the true values.

In particular, Table S11 (Supplementary Materials) presents the performance metrics obtained using 5-fold cross-validation on the train dataset with the selected hyperparameter combinations.

The MSE and MAPE are highly sensitive to even a few low-accuracy predictions. To better assess the models and account for the impact of individual outliers, we also used the median absolute percentage error (MdAPE), which indicates the error value below which 50% of the predictions in the dataset fall. Furthermore, we evaluated the metric  $p$ , representing the percentage of samples (feature vectors) in a dataset with an absolute error below a specified threshold. Specifically, we computed the values of  $p01$  and  $p10$ , which represent the proportion of predictions in the dataset with an accuracy of 1% or better and 10% or better, respectively. The decrease in MdAPE and the increase in  $p01$  and  $p10$  indicate improved model performance due to a higher proportion of accurate predictions.

### 3. Results and Discussion

#### 3.1. Train dataset

Fig. 3 presents the models' prediction results on the training set. The Supplementary Materials (Figs.S4–S7) provide more detailed results, while Tables S12 and S13 list the performance metrics. The RF and GB models yield the most accurate predictions, whereas SVR performs the worst. Applying PCA generally reduces prediction accuracy on the training set.

As expected, increasing the number of descriptors enhances model performance (Fig. 4 and Fig.S7). The only exception occurs under AM1.5 illumination with PCA, where adding a fifth descriptor may degrade predictions rather than improve them. Overall, models trained on AM1.5 data exhibit significantly lower accuracy than those using monochromatic illumination, with errors approximately an order of magnitude higher. Notably, models with four descriptors (or five with PCA) under AM1.5 appear unpromising, even for RF and GB. Meanwhile, increasing the number of descriptors from six to seven offers little to no additional benefit in high-accuracy models such as RF, GB, XGB, and, to some extent, DNN.

Fig. 5 illustrates typical dependencies of the proportion of predictions with a given accuracy on  $T$ ,  $d_p$ ,  $N_B$ , and  $N_{\text{Fe}}$ , enabling an assessment of the models' predictive capabilities based on measurement conditions, solar cell structure, and iron concentration. The data indicate that the most challenging cases for prediction involve  $N_{\text{Fe}} < 10^{11} \text{ cm}^{-3}$  and solar cells with a base doping level around  $10^{16} \text{ cm}^{-3}$ . The last difficulty arises because, as shown by simulation [58], variations in PV parameters during FeB pair dissociation remain minimal. Meanwhile, measurement temperature and base thickness (when feature dimensionality exceeds 4) have virtually no impact on prediction accuracy.

#### 3.2. $N_{\text{Fe}}$ -altered test dataset

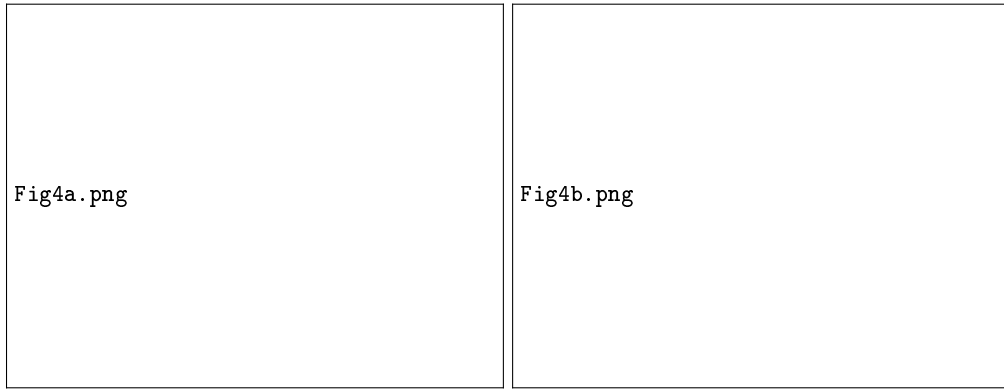
The  $N_{\text{Fe}}$ -altered test dataset is the most realistic, as it mimics real-world conditions where the goal is to predict iron concentration based on PVP variations measured under standardized conditions for predefined structures, using the same temperature and solar cell parameters as in model training.

Fig3a.png	Fig3b.png	Fig3c.png	Fig3d.png
Fig3e.png	Fig3f.png	Fig3g.png	Fig3h.png
Fig3i.png	Fig3j.png	Fig3k.png	Fig3l.png
Fig3m.png	Fig3n.png	Fig3o.png	Fig3p.png
Fig3q.png	Fig3r.png	Fig3s.png	Fig3t.png

**Figure 3:** Scatter plots compare reference iron concentrations with ML-predicted values during the training phase. The ML algorithms include RF (a–d), GB (e–h), XGB (i–l), SVR (m–p), and DNN (q–t). The data come from simulation under monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 illumination (c, d, g, h, k, l, o, p, s, t). Panels b, d, f, h, j, l, n, p, r, and t include PCA. The input feature dimensions are 4 (a, e, i, m, q), 5 (b, f, j, n, r), 6 (c, g, k, o, s), and 7 (d, h, l, p, t). The black lines are the identified lines serving as the references.

Fig. 6 and Figs.S8–S11 present the prediction results, while Fig. 7 and Table2 summarize the performance metrics. SVR consistently performs poorly, a trend observed across all test datasets. Since this pattern remains unchanged in subsequent analyses, we do not discuss it further. For the remaining models, it is crucial to differentiate between results obtained under monochromatic and solar illumination. Under 940 nm illumination without PCA, nearly all algorithms achieve reasonably accurate predictions, with a MAPE of approximately 10. DNN models yield slightly worse results, whereas XGB demonstrates marginally better performance. Interestingly, increasing feature dimensionality does not improve prediction accuracy but instead reduces it. Notably, incorporating the relative changes in the fill factor into the descriptor set significantly degrades DNN predictions. As shown in Table 2, RF, GB, and XGB, which are the most accurate models, exhibit a noticeable increase in MdAPE and a rise in p01 and p10 as the number of descriptors grows.





**Figure 4:** MSE dependence on input feature dimension for training data obtained under monochromatic (a) and AM1.5 illumination (b). ML algorithms: RF (circles), GB (rectangles), XGB (triangles), and DNN (stars). Closed markers indicate results with PCA, while open markers represent results without PCA.



**Figure 5:** Typical dependencies of the proportion of predictions with an accuracy of at least 10% (a, b, c) or 1% (d) on iron concentration, alloying level, substrate thickness, and temperature.

These trends indicate that expanding feature dimensionality leads to a higher proportion of predictions with significant errors, as illustrated in Fig. 8,a for GB. Fig. 8,a and Table 2 also show that PCA-based model simplification increases the proportion of high-error predictions in XGB while reducing it in RF and GB. In summary,  $GB_{4:PC}^{940}$ ,  $GB_{5:PC}^{940}$ , and  $XGB_5^{940}$  yield the most precise results for this test dataset under monochromatic illumination, with the proportion of predictions achieving at least 10% accuracy reaching 80.3%, 82.4%, and 78.6%, respectively.

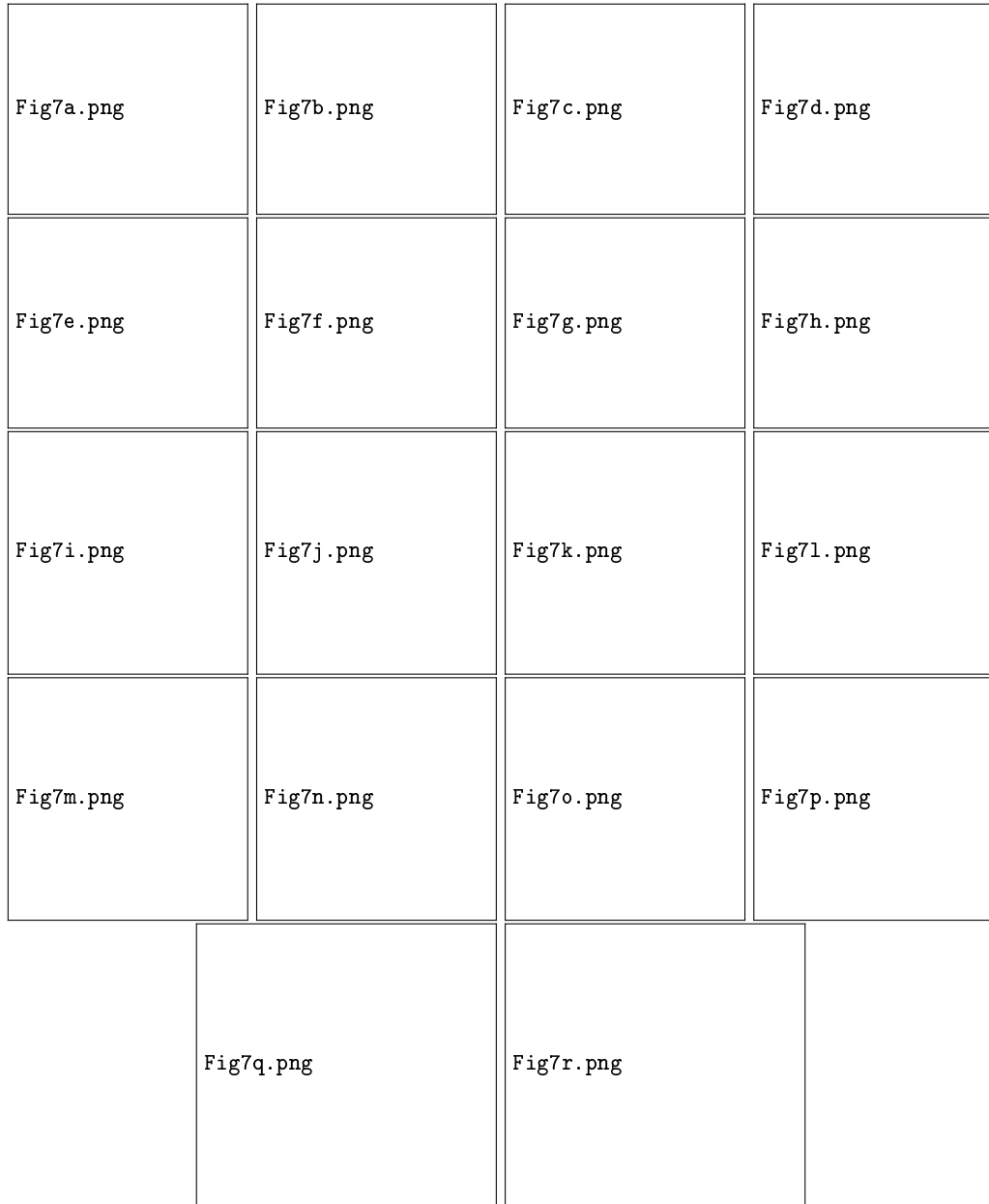
When estimating iron concentration based on PVP variations obtained under AM1.5 illumination, using only four descriptors does not yield sufficiently accurate predictions with any model. However, adding efficiency variations (increasing the feature dimension from 4 to 5) enables DNN and XGB to predict  $NFe$  with MAPE slightly above 10%,

Fig6a.png	Fig6b.png	Fig6c.png	Fig6d.png
Fig6e.png	Fig6f.png	Fig6g.png	Fig6h.png
Fig6i.png	Fig6j.png	Fig6k.png	Fig6l.png
Fig6m.png	Fig6n.png	Fig6o.png	Fig6p.png
Fig6q.png	Fig6r.png	Fig6s.png	Fig6t.png

**Figure 6:** Scatter plots compare reference iron concentrations with ML-predicted values during the  $N_{Fe}$ -altered test phase. The ML algorithms include RF (a-d), GB (e-h), XGB (i-l), SVR (m-p), and DNN (q-t). The data come from simulation under monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 illumination (c, d, g, h, k, l, o, p, s, t). Panels b, d, f, h, j, l, n, p, r, and t include PCA. The input feature dimensions are 7 (a, e, i, m, q), 6 (b, f, j, n, r), 5 (c, g, k, o, s), and 4 (d, h, l, p, t). The black lines are the identified lines serving as the references.

though the results remain inferior to those under monochromatic illumination. The use of PCA significantly worsens the results (see Fig. 7,f, Fig. 7,n, Table 2). A further increase in feature dimensionality improves prediction accuracy and reduces the performance gap between models with and without PCA. The best results are observed when using seven descriptors. For RF and GB, applying PCA decreases both the mean errors (Fig. 7,p) and the proportion of predictions with large deviations from the true values (Table 2). For XGB and DNN — the most efficient models across the entire  $N_{Fe}$ -altered test dataset, including data under 940nm illumination — PCA provides no benefit (Fig. 7,p, Fig. 8,b). Notably, in these cases, the proportion of predictions with accuracy below 10% does not exceed 1.3%.

These results underscore the importance of selecting an optimal model and applying appropriate data preprocessing to achieve accurate iron concentration predictions in silicon solar cells.



**Figure 7:** MSE (a–h), MAPE (i–p), and  $R^2$  (q, r) scores for various models, feature combinations, and illumination conditions on the  $N_{\text{Fe}}$ -altered test dataset. Illumination: 940 nm (a–d, i–l, q), AM1.5 (e–h, m–p, r). Feature dimensions: 4 (a, e, i, m), 5 (b, f, j, n), 6 (c, g, k, o), and 7 (d, h, l, p). Results with PCA are shown as circles (a–h) and shaded areas (i–r), while results without PCA are represented by squares (a–h) and filled regions (i–r). The numbers in panels (a–h) represent MSE values multiplied by 1000, while the numbers in panels (i–p) indicate MAPE values in percentage.

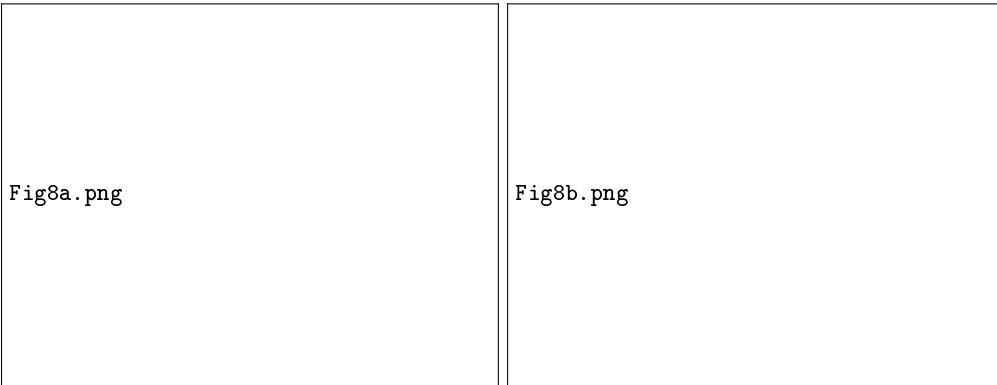
### 3.3. $T$ -altered test dataset

The  $T$ -altered test dataset simulates a scenario in which impurity concentration estimates are based on measurements taken under conditions different from those used for model training. Additionally, this dataset includes temperature values (287 K, 341 K, and 342 K) that extend beyond the training range (290 – 340) K, enabling an assessment of the models' extrapolation capability.

Figs. S12–S15 present the prediction results, while Fig. 9 and Table 3 summarize the performance metrics. The data show that the average metrics for all models worsen slightly compared to the  $N_{\text{Fe}}$ -altered dataset (Fig. 7, Fig. 9).

**Table 2**Performance metrics of the ML models for  $N_{\text{Fe}}$ -altered dataset

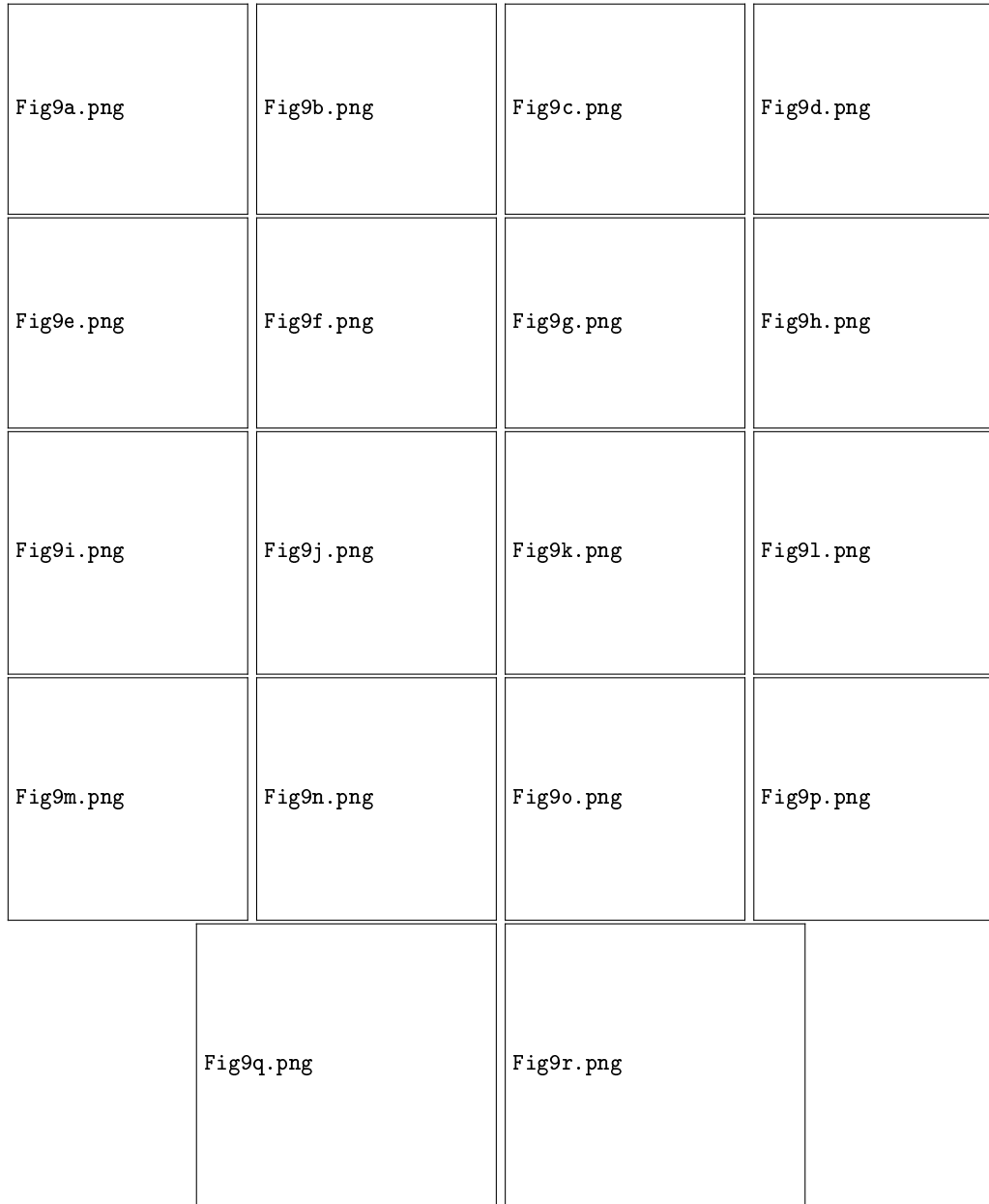
Algorithm	Feature dimension	MdAPE (%)				$p01$ (%)				$p10$ (%)			
		940 nm		AM1.5		940 nm		AM1.5		940 nm		AM1.5	
		Init	PCA	Init	PCA	Init	PCA	Init	PCA	Init	PCA	Init	PCA
RF	4	6.47	6.38	8.29	8.00	10.1	9.48	7.54	8.12	68.0	70.2	57.0	57.9
	5	7.67	<b>6.00</b>	6.96	9.46	7.06	8.51	10.1	6.96	60.5	72.2	64.2	52.1
	6	8.01	7.34	5.46	5.66	6.58	7.74	11.0	10.9	59.5	60.4	72.4	72.9
	7	8.89	8.74	6.27	4.62	5.71	5.42	10.4	13.3	54.2	56.4	68.8	82.2
GB	4	4.63	4.35	6.95	6.27	<b>13.6</b>	<b>14.1</b>	9.87	9.38	76.5	80.3	62.4	67.2
	5	4.96	4.63	6.55	8.05	<b>13.6</b>	12.5	11.2	7.25	72.7	<b>82.4</b>	66.6	57.4
	6	6.78	6.73	5.31	4.41	8.03	9.09	11.9	15.4	65.5	64.0	74.4	79.1
	7	7.87	7.41	4.97	3.73	7.16	7.64	12.2	17.4	60.7	61.8	75.3	84.3
XGB	4	4.82	6.91	6.43	7.75	11.0	9.77	10.1	8.32	75.6	63.5	62.7	60.2
	5	<b>4.46</b>	6.20	4.26	9.04	11.7	10.2	13.8	6.58	<b>78.6</b>	68.6	74.7	52.3
	6	6.49	7.57	4.13	4.79	9.87	8.51	14.8	11.8	68.7	59.9	83.2	77.2
	7	6.16	7.96	3.71	3.94	8.80	7.54	15.2	15.6	66.0	60.9	83.3	83.0
SVR	4	31.1	26.7	41.2	40.5	1.45	1.35	1.06	0.77	14.9	18.2	11.4	11.8
	5	36.3	42.7	39.7	51.9	1.64	0.87	1.74	1.35	14.4	10.8	10.9	9.77
	6	37.0	44.8	37.3	44.2	1.16	0.48	1.64	0.77	13.4	8.99	14.4	10.1
	7	39.1	38.1	36.5	32.5	1.16	1.06	0.68	1.26	9.48	13.7	13.4	15.2
DNN	4	8.06	7.48	11.4	13.0	7.74	7.64	3.97	4.35	58.8	60.5	44.8	41.6
	5	7.42	9.02	5.97	6.84	6.48	6.67	10.1	6.77	66.0	54.7	72.3	61.4
	6	6.98	8.15	6.30	3.18	7.06	5.80	8.80	17.8	64.8	59.3	69.9	90.2
	7	17.3	6.93	<b>1.63</b>	<b>2.40</b>	4.06	8.90	<b>32.6</b>	<b>23.5</b>	33.1	64.7	<b>98.7</b>	<b>96.7</b>

**Figure 8:** The proportion of samples in the  $N_{\text{Fe}}$ -altered dataset with an absolute error below a given threshold, plotted as a function of the threshold value.

In most cases, PCA reduces model performance, except for AM1.5 at a feature dimensionality of 7. For the 940 nm case, using six descriptors proves more effective than using four, as observed in the previous test dataset.

A comparison of Table 2 and Table 3 reveals that, unlike the mean characteristics, MdAPE,  $p01$ , and  $p10$  are often better than those in the  $N_{\text{Fe}}$ -altered dataset. For example, for  $\text{XGB}_6^{940}$ , MdAPE and  $p01$  reach 2.47% and 33.6%, whereas in the previous case, the best values were 4.46% and 13.6%. These differences in the behavior of MSE, MAPE, and  $R^2$ , compared to other performance metrics, suggest a higher proportion of accurate predictions while increasing error values in less successful forecasts. In other words, prediction disparity increases — the rich get richer, and the poor get poorer.

Fig. 10 illustrates the dependence of the proportion of predictions with an error below 10% on temperature for the two training datasets. Within the (290–340) K range, where the networks were trained, both datasets exhibit similar  $p10$  values and overall trends, despite the temperature values in the  $T$ -altered test dataset differing from those used during



**Figure 9:** MSE (a–h), MAPE (i–p), and  $R^2$  (q, r) scores for various models, feature combinations, and illumination conditions on the  $T$ -altered test dataset. Illumination: 940 nm (a–d, i–l, q), AM1.5 (e–h, m–p, r). Feature dimensions: 4 (a, e, i, m), 5 (b, f, j, n), 6 (c, g, k, o), and 7 (d, h, l, p). Results with PCA are shown as circles (a–h) and shaded areas (i–r), while results without PCA are represented by squares (a–h) and filled regions (i–r). The numbers in panels (a–h) represent MSE values multiplied by 1000, while the numbers in panels (i–p) indicate MAPE values in percentage.

training. Specifically,  $p10$  remains nearly independent of temperature, consistent with the trend observed in Fig. 5d. However, a slight deviation of just a few kelvins from the training range significantly reduces prediction accuracy. This behavior highlights the models' limited extrapolation capability compared to their interpolation performance. The degradation of the average metrics (MSE, MAPE,  $R^2$ ) primarily results from lower prediction accuracy at temperatures beyond the training range.

Thus, applying the models to measurement data obtained at temperatures that slightly deviate from the training conditions but remain within the training range is justified.

**Table 3**Performance metrics of the ML models for  $T$ -altered dataset

Algorithm	Feature dimension	MdAPE (%)				$p01$ (%)				$p10$ (%)			
		940 nm		AM1.5		940 nm		AM1.5		940 nm		AM1.5	
		Init	PCA	Init	PCA	Init	PCA	Init	PCA	Init	PCA	Init	PCA
RF	4	6.50	13.1	<b>1.88</b>	11.6	18.2	8.92	<b>28.1</b>	11.8	60.4	40.8	<b>95.6</b>	45.2
	5	5.59	13.2	9.58	5.90	22.4	6.75	12.1	16.6	62.9	40.7	51.3	62.7
	6	4.61	13.3	14.2	2.13	20.8	8.58	9.50	35.3	68.8	43.5	42.1	77.3
GB	7	5.07	9.82	5.05	<b>1.94</b>	21.5	12	17.7	38.7	67.4	50.4	68.3	79.8
	4	5.29	12.5	3.47	10.1	20.7	6.67	21.5	5.08	62.0	42.9	80.6	49.8
	5	4.91	12.6	9.34	4.89	19.1	7.00	6.58	22.5	66.2	44.7	51.8	65.0
XGB	6	3.41	9.12	12.8	1.81	22.3	11.2	4.67	<b>40.9</b>	73.1	52.4	43.2	78.3
	7	5.03	9.72	5.14	2.44	15.6	10.2	13.9	31.3	68.3	51.0	70.7	79.3
	4	5.03	10.9	3.40	8.72	22.8	8.92	20.8	8.42	64.6	46.9	80.8	53.8
SVR	5	3.64	9.01	9.88	5.14	28.4	8.75	7.00	23.7	67.3	53.0	50.3	66.3
	6	<b>2.47</b>	10.4	10.5	2.08	<b>33.6</b>	6.75	6.58	34.4	<b>73.7</b>	49.2	48.6	78.0
	7	9.93	<b>5.37</b>	5.78	3.38	9.92	<b>14.9</b>	12.3	19.3	50.3	<b>67.9</b>	69.2	78.3
DNN	4	30.9	27.8	5.11	36.4	1.50	1.58	13.6	0.67	15.3	18.3	72.9	12.8
	5	33.7	38.0	35.7	37.5	1.75	1.00	1.17	0.92	15.3	12.9	13.5	12.8
	6	38.3	40.9	47.6	37.1	1.50	1.50	0.67	1.25	14.3	15.0	9.83	12.8
	7	18.6	6.85	46.2	36.3	3.83	7.50	0.75	1.83	29.8	66.3	9.58	15.4
	4	6.92	7.54	10.5	14.2	7.83	8.33	4.58	4.67	65.3	59.0	47.3	37.2
	5	6.53	8.90	5.46	7.44	9.17	5.50	10.9	7.42	68.8	53.8	71.6	58.4
	6	7.09	7.18	4.92	2.99	7.50	9.25	10.7	18.3	66.8	61.1	79.2	90.9
	7	39.3	35.9	32.2	2.53	1.25	1.50	1.00	23.2	10.7	13.3	14.7	<b>92.5</b>

Fig10.png

**Figure 10:** Typical dependencies of the fraction of predictions with an accuracy within 10% as a function of temperature for the  $T$ -altered and  $N_{Fe}$ -altered datasets.

### 3.4. $N_B$ -altered and All-altered test datasets

The  $N_B$ -altered dataset represents a case where a model trained on a specific type of solar cell is used to evaluate iron impurity concentration in photovoltaic converters with slightly different parameters. The boron concentration in the base is one of the parameters that can vary during the production process of PV devices. Therefore, the model's ability to adapt to these variations enhances its applicability to real-world manufacturing conditions.

The optimal performance of 16 models utilizing RF, GB, XGB, and DNN algorithms is shown in Figs. 11a–11d, respectively. The remaining panels of Fig. 11 illustrate the MAPE values. The Supplementary Materials (Figs. S16–S20, Table S14) provide more detailed prediction results.

Fig11a.png	Fig11b.png	Fig11c.png	Fig11d.png
Fig11e.png	Fig11f.png	Fig11g.png	Fig11h.png
Fig11i.png	Fig11j.png	Fig11k.png	Fig11l.png

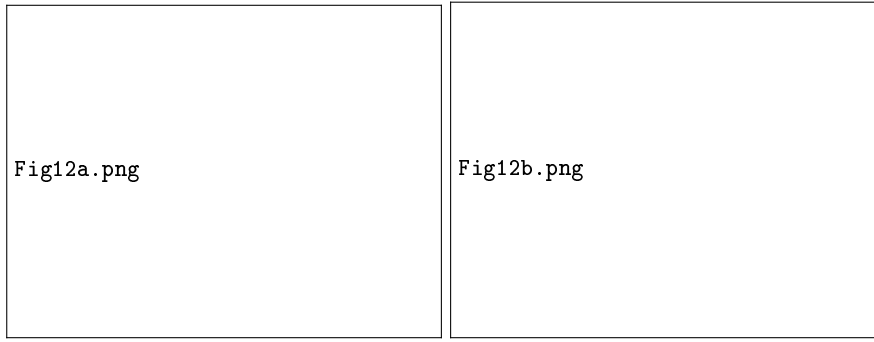
**Figure 11:** (a)–(d) Scatter plots comparing reference iron concentrations with ML–predicted values for the  $N_B$ –altered dataset for the best models based on RF (a), GB (b), XGB (c), and DNN (d). (e)–(l) MAPE scores for various models, feature combinations, and illumination conditions. Illumination: 940 nm (e–h), AM1.5 (i–l). Feature dimension: 4 (e, i), 5 (f, j), 6 (g, k), and 7 (h, l). The numbers represent error values as percentages.

The most striking aspect is the subpar performance of the RF, GB, and XGB models, which had previously demonstrated stable and satisfactory accuracy, at least for certain combinations of descriptor sets and illumination conditions. Notably, even SVR models do not always produce the worst results. The application of PCA slightly improves prediction accuracy, except under monochromatic illumination with seven descriptors. However, the overall error metrics remain excessively bad. Another important finding is that the best predictive performance is achieved with six features, even though incorporating the ideality factor does not enhance model accuracy.

DNN models exhibit the highest performance for the  $N_B$ –altered dataset, achieving results comparable to the best outcomes obtained for the  $N_{Fe}$ –altered and  $T$ –altered datasets. Specifically, the most accurate predictions occur under 940 nm illumination with four or five descriptors and under AM1.5 illumination with six or seven descriptors. The application of PCA improves results only when the feature dimension is six or seven.

At the same time, it is important to note that in DNN model predictions, approximately 10% of forecasts exhibit significant errors (greater than 50%), as shown in Fig. 11d. Fig. 12a illustrates the distribution of prediction errors for the  $N_{Fe}$ –altered,  $N_B$ –altered, and  $T$ –altered datasets. While the dependencies remain similar for APE less than 10%, for higher errors, the dependence  $p$  for the  $N_{Fe}$ –altered dataset increases more slowly, indicating a higher proportion of less accurate predictions in this case. Fig. 12b presents the dependence of  $p_{10}$  on boron concentration across different test datasets. The observed dependencies are similar to those in the training set (Fig. 5b) and highlight the most significant challenges for the developed models in predicting solar cells with a base doping level around  $10^{16} \text{ cm}^{-3}$ . However, for the  $N_{Fe}$ –altered set, prediction accuracy in this problematic range is significantly lower, resulting in a higher occurrence of poor-quality forecasts. In contrast, for  $N_B < 3 \times 10^{15} \text{ cm}^{-3}$ , DNN models demonstrate adaptability to previously unknown doping levels.

The results of this section highlight the importance of training models with  $N_B$  values expected in solar cells requiring iron quantification, particularly within the range  $N_B = (5 \times 10^{15} - 3 \times 10^{16}) \text{ cm}^{-3}$ . If it is necessary to use models trained on boron concentrations in the base that differ from those in the analyzed sample, the DNN algorithm should be preferred.



**Figure 12:** The proportion of samples in the test dataset with an absolute error below a given threshold as a function of the threshold value (a) and the dependencies of the fraction of predictions with an accuracy not exceeding 10% on the doping level for various test datasets (b). The vertical lines in panel (b) indicate the boron concentration values used in the training dataset.



**Figure 13:** (a)–(d) Scatter plots comparing reference iron concentrations with ML–predicted values for the All–altered test dataset for the best models based on RF (a), GB (b), XGB (c), and DNN (d). (e)–(l) MSE scores for various models, feature combinations, and illumination conditions. Illumination: 940 nm (e–h), AM1.5 (i–l). Feature dimension: 4 (e, i), 5 (f, j), 6 (g, k), and 7 (h, l). The numbers represent MSE values multiplied by 1000.

Given the results obtained for the  $N_B$ –altered dataset, testing the All–altered dataset presented a notable challenge for the developed models, as it aimed to assess their interpolation capabilities. Notably, this dataset does not test extrapolation capabilities: all four features (including temperature), which are unrelated to PVP modifications, have different magnitudes from those in the train dataset but still fall within the corresponding range. The All–altered dataset encapsulates most of the challenges that may arise under practical conditions when characterizing solar cells.

As in the previous case, Figs. 13a–13d presents the best results from the four algorithmic families, while Figs. 13e–13l displays the MSE values. Figs. S21–S25 and Table S15 provide additional statistical details.

The presented data reveal several similarities between the results for the All–altered and  $N_B$ –altered datasets: i) RF, GB, XGB, and SVR models exhibit low prediction accuracy; although PCA enhances accuracy, the improvement remains insufficient; ii) DNN models demonstrate relatively high average predictive performance, particularly with four and five features under 940 nm illumination and six features under AM1.5 illumination; iii) PCA reduces DNN model



error when applied to datasets with six and seven features; iv) approximately 5% of DNN predictions have APE > 50%. These similarities suggest that the primary source of increased errors in the All–altered dataset is the deviation of boron concentration values from those used in train phase.

### 3.5. Experimental validation

To demonstrate the effectiveness of the proposed method, real solar cells were used. The base thickness (380  $\mu\text{m}$ ) and the measurement temperatures (300, 305, 310, 320, 330, and 340 K) match those used in  $I$ – $V$  curve modeling for the train dataset. However, the base doping level,  $N_B = 1.36 \times 10^{15} \text{ cm}^{-3}$ , was not among the  $N_B$  values used for training. Therefore, the experimental test dataset is, to some extent, analogous to the  $N_B$ –altered test dataset or even the All–altered dataset, considering the spread of iron concentration values.

That is, even before testing began, some difficulties with prediction based on measurement data could be foreseen (not only ML models can make predictions). Meanwhile, the  $N_B$ –altered test dataset included some data simulated for  $N_B = 1.36 \times 10^{15} \text{ cm}^{-3}$ . Therefore, we tested the experimental data using following models: i) models trained on the training dataset for 940 nm illumination; these models had been previously validated on artificial test datasets; ii) models trained on all simulated data in monochromatic illumination case (full dataset), incorporating train,  $N_{\text{Fe}}$ –altered,  $T$ –altered,  $N_B$ –altered, and All–altered test datasets. In the latter case, we will use the following notation

$$A_{\text{feat}}^{\text{full}},$$

where the upper index specifies the completeness of the data for training and accounts for the fact that applying models trained for the AM1.5 case to the experimental dataset is unjustified due to differing measurement conditions.

Fig. 14 and Figs. S26–S29 present the prediction results of different models. The experimental test datasets contained between 8 and 30 samples, with the exact number depending on the number of features. The referenced figures display predictions for these datasets. To ensure a more precise quantitative comparison, we analyzed a case where all models used an input dataset of the same size, derived from  $I$ – $V$  measurements of the same solar cells. The resulting performance metrics are shown in Fig. 15. The Supplementary Materials (Fig. S30) also include averaged quantitative characteristics of model performance across datasets of different sample sizes, which align with the trends observed in Fig. 15.

Notably, the results for the experimental test dataset closely resemble those observed for synthetic test datasets simulated under the assumption of monochromatic illumination. Specifically, adding descriptors to the  $(T, d_p, N_B, \varepsilon I_{\text{SC}})$  set does not enhance prediction accuracy, at least. Likewise, reducing dimensionality and simplifying the model using PCA proves ineffective: error reduction occurs only in a few cases (all SVR models, GB<sup>940</sup>) and only when the initial (non-PCA) error is already substantial. However, even in such cases, PCA does not reduce the error to acceptable levels. Similar to the results in Fig. 7, the XGB algorithm consistently delivers the best performance. When training excluded  $N_B$  values corresponding to real solar cells, the highest prediction accuracy was achieved by DNN<sup>940</sup>, aligning with the findings from the  $N_B$ –altered test dataset.

Including samples in the training dataset that correspond to the base doping level of the real solar cell improves the performance of RF, GB, and XGB models, similar to the transition from the  $N_B$ –altered to the  $N_{\text{Fe}}$ –altered test dataset. It is also noteworthy to highlight the high quantitative prediction metrics: for a feature dimension of 4, the MSE is approximately  $3 \times 10^{-3}$ , and the MAPE falls within the 10–13% range. These values are comparable to the best results obtained on simulated test datasets, despite the latter representing idealized cases that do not account for real-world factors affecting PVP values, such as series and shunt resistances. A key distinction in predictions on the experimental dataset is a narrower error range: while predictions with accuracy better than 1% are nearly absent, the proportion of predictions with an APE exceeding 50% is also lower compared to synthetic data.

## 4. Conclusion

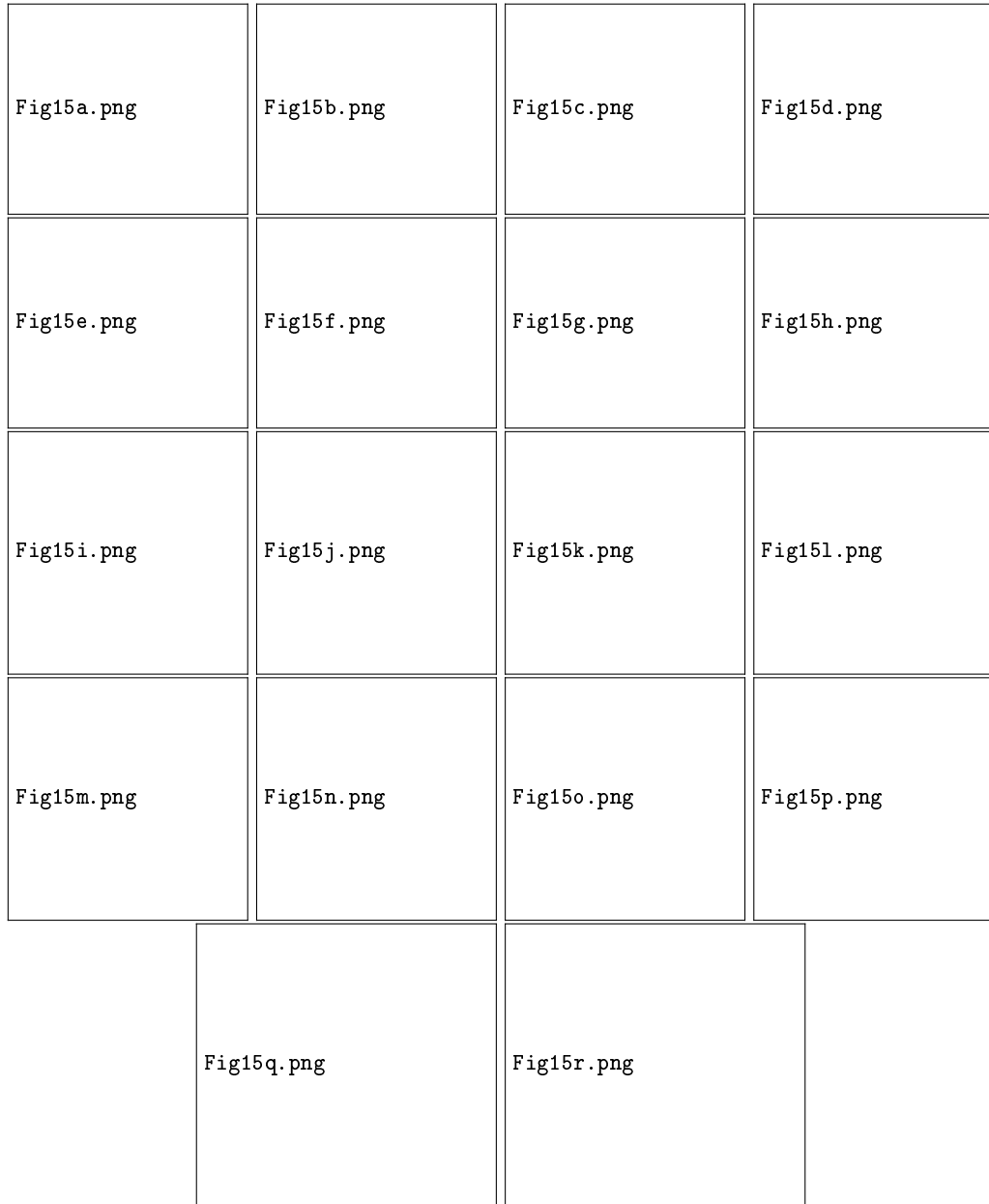
This study presents a novel ML–based approach for addressing the challenging problem of iron quantification in silicon solar cells. The proposed models predict iron concentration based on variations in photovoltaic parameters caused by FeB pair dissociation, solar cell characteristics, and measurement conditions. The datasets were obtained under AM1.5 illumination and monochromatic illumination at 940 nm, where photon absorption primarily occurs in the base of the photovoltaic converter. A total of 80 models were tested using data from both simulated and experimentally measured  $I$ – $V$  characteristics.

Fig14a.png	Fig14b.png	Fig14c.png	Fig14d.png
Fig14e.png	Fig14f.png	Fig14g.png	Fig14h.png
Fig14i.png	Fig14j.png	Fig14k.png	Fig14l.png
Fig14m.png	Fig14n.png	Fig14o.png	Fig14p.png
Fig14q.png	Fig14r.png	Fig14s.png	Fig14t.png

**Figure 14:** Scatter plots compare reference iron concentrations with ML-predicted values during the experimental test phase. The ML algorithms include RF (a–d), GB (e–h), XGB (i–l), SVR (m–p), and DNN (q–t). Panels (b, d, f, h, j, l, n, p, r, t) incorporate PCA pre-processing. The input feature dimensions are 4 (a, b, i, j, q, r), 5 (e, f, m, n), 6 (c, d, k, l), and 7 (g, h, o, p, s, t). The black lines are the identified lines serving as the references. Marks with a black point inside correspond to models trained on the complete set of synthetic data.

The results emphasize the critical role of feature selection in achieving accurate predictions. Under 940 nm illumination, simple features such as relative changes in short-circuit current, base doping level, base thickness, and temperature yielded the most precise predictions. For AM1.5 illumination, incorporating changes in open-circuit voltage, conversion efficiency, and fill factor further improved accuracy. It was found that pre-processing the input data using Principal Component Analysis, despite simplifying the models with minimal loss of overall variance, led to a decrease in prediction accuracy.

This study establishes key criteria for datasets used in training ML models for iron quantification. Due to the models' limited extrapolation ability, the training data range must be carefully selected to encompass temperature and solar cell parameter values relevant to the model's intended application. The constraints on base doping levels are even



**Figure 15:** MSE (a–h), MAPE (i–p), and  $R^2$  (q, r) scores for various models and feature sets on the experimental test set. The models used the training dataset (a–d, i–l, q) or the full synthetic dataset (e–h, m–p, r) for training. The feature dimensions are as follows: 4 (a, e, i, m), 5 (b, f, j, n), 6 (c, g, k, o), and 7 (d, h, l, p). Results with PCA are shown as circles (a–h) and shaded areas (i–r), while results without PCA appear as squares (a–h) and filled regions (i–r). Numbers in (a–h) represent MSE values multiplied by 1000, while (i–p) show MAPE as percentages. The samples in the experimental test set were kept identical for all models.

more stringent: training the model with boron concentration values that align with real-world conditions is preferable, particularly for  $N_B$  values around  $10^{16} \text{ cm}^{-3}$ .

The results indicate that eXtreme Gradient Boosting and Deep Neural Network models outperform other approaches, achieving MSE, MAPE, and  $R^2$  values of up to 0.003, 3%, and 0.997 for synthetic data, and 0.004, 9%, and 0.987 for experimental data. When doping levels in the test data deviate from those in the training set, the DNN model remains the most reliable choice. In contrast, Support Vector Regression proves ineffective in all cases.

The obtained results lay the foundation for improving the standard method of silicon solar cell characterization, which relies on measuring  $I$ - $V$  curves, by enabling the determination of iron impurity concentration.

## Acknowledgments

O.O. would like to acknowledge the financial support by National Research Foundation of Ukraine (Project No. 2023.03/0252 “Development of principles for the creation and machine-oriented characterization of porous silicon nanostructures with optimal heat transport properties”)

## Supplementary data

Supplementary data to this article can be found online at <https://tinyurl.com/yzn9xkdm>

## Data availability

Data will be made available on request.

## References

- [1] S. Karlilar Pata, M. Balcilar, Decarbonizing energy: Evaluating fossil fuel displacement by renewables in oecd countries, *Environ. Sci. Pollut. Res.* 31 (2024) 31304–31313.
- [2] J. L. Holechek, H. M. E. Geli, M. N. Sawalhah, R. Valdez, A global assessment: Can renewable energy replace fossil fuels by 2050?, *Sustainability* 14 (2022) 4792.
- [3] S. Park, J. Lee, S. Khan, A. Wahab, M. Kim, Machine learning-based heavy metal ion detection using surface-enhanced raman spectroscopy, *Sensors* 22 (2022) 596.
- [4] X. Qi, W. Ma, Y. Dang, W. Su, L. Liu, Optimization of the melt/crystal interface shape and oxygen concentration during the czochralski silicon crystal growth process using an artificial neural network and a genetic algorithm, *J. Cryst. Growth* 548 (2020) 125828.
- [5] L. Rachdi, M. Hofmann, Use of optical emission spectroscopy to predict silicon nitride layer properties, *Vacuum* 191 (2021) 110322.
- [6] Y. Buratti, C. Eijkens, Z. Hameiri, Optimization of solar cell production lines using neural networks and genetic algorithms, *ACS Appl. Energy Mater.* 3 (2020) 10317–10322.
- [7] L. Zhang, M. He, Unsupervised machine learning for solar cell materials from the literature, *J. Appl. Phys.* 131 (2022) 064902.
- [8] T. Liu, S. Wang, Y. Shi, L. Wu, R. Zhu, Y. Wang, J. Zhou, W. C. H. Choy, Machine-learning accelerating the development of perovskite photovoltaics, *Sol. RRL* 7 (2023) 2300650.
- [9] Z. Asghar, K. Hafeez, D. Sabir, B. Ijaz, S. S. H. Bukhari, J. Ro, Reclaim: Renewable energy based demand-side management using machine learning models, *IEEE Access* 11 (2023) 3846–3857.
- [10] Z. Abdullah-Vetter, B. Wright, T.-C. Wu, A. Shakiba, Z. Hameiri, Automatic quantitative analysis of internal quantum efficiency measurements of gaas solar cells using deep learning, *Adv. Sci.* 12 (2025) 2407048.
- [11] M. Di Sabatino, R. Hendawi, A. S. Garcia, Silicon solar cells: Trends, manufacturing challenges, and ai perspectives, *Crystals* 14 (2024) 167.
- [12] S. Datta, A. Baul, G. C. Sarker, P. K. Sadhu, D. R. Hodges, A comprehensive review of the application of machine learning in fabrication and implementation of photovoltaic systems, *IEEE Access* 11 (2023) 77750–77778.
- [13] R. Jaiswal, M. Martínez-Ramón, T. Busani, Recent advances in silicon solar cell research using data science-based learning, *IEEE J. Photovolt.* 13 (2023) 2–15.
- [14] Y. Buratti, G. M. Javier, Z. Abdullah-Vetter, P. Dwivedi, Z. Hameiri, Machine learning for advanced characterisation of silicon photovoltaics: A comprehensive review of techniques and applications, *Renewable Sustainable Energy Rev.* 202 (2024) 114617.
- [15] S. Bhatti, H. U. Manzoor, B. Michel, R. S. Bonilla, R. Abrams, A. Zoha, S. Hussain, R. Ghannam, Revolutionizing low-cost solar cells with machine learning: A systematic review of optimization techniques, *Advanced Energy and Sustainability Research* 4 (2023) 2300004.
- [16] H. Munawer Al-Otun, Classification of anomalies in electroluminescence images of solar pv modules using cnn-based deep learning, *Sol. Energy* 278 (2024) 112803.
- [17] L. Pratt, D. Govender, R. Klein, Defect detection and quantification in electroluminescence images of solar pv modules using u-net semantic segmentation, *Renew. Energ.* 178 (2021) 1211–1222.
- [18] Z. Li, S. Zhang, C. Qu, Z. Zhang, F. Sun, Research on multi-defects classification detection method for solar cells based on deep learning, *PLOS ONE* 19 (2024) 1–16.
- [19] H.-H. Lin, H. K. Dandage, K.-M. Lin, Y.-T. Lin, Y.-J. Chen, Efficient cell segmentation from electroluminescent images of single-crystalline silicon photovoltaic modules and cell-based defect identification using deep learning with pseudo-colorization, *Sensors* 21 (2021) 4292.
- [20] W. Tang, Q. Yang, K. Xiong, W. Yan, Deep learning based automatic defect identification of photovoltaic module using electroluminescence images, *Sol. Energy* 201 (2020) 453–460.
- [21] C. Bu, T. Liu, R. Li, R. Shen, B. Zhao, Q. Tang, Electrical pulsed infrared thermography and supervised learning for pv cells defects detection, *Sol. Energ. Mat. Sol.* 237 (2022) 111561.
- [22] M. Turek, M. Meusel, Automated classification of electroluminescence images using artificial neural networks in correlation to solar cell performance parameters, *Sol. Energ. Mat. Sol.* 260 (2023) 112483.

- [23] C. Huang, Z. Zhang, L. Wang, Psoprner: Pso-based deep convolutional neural network pruning method for pv module defects classification, *IEEE J. Photovolt.* 12 (2022) 1550–1558.
- [24] Y. Liu, K. Ding, J. Zhang, Y. Li, Z. Yang, W. Zheng, X. Chen, Fault diagnosis approach for photovoltaic array based on the stacked auto-encoder and clustering with i-v curves, *Energy Convers. Manage.* 245 (2021) 114603.
- [25] Z. Chen, Y. Chen, L. Wu, S. Cheng, P. Lin, Deep residual network based fault detection and diagnosis of photovoltaic arrays using current-voltage curves and ambient conditions, *Energy Convers. Manage.* 198 (2019) 111793.
- [26] M. W. Hopwood, T. Gunda, H. Seigneur, J. Walters, Neural network-based classification of string-level iv curves from physically-induced failures of photovoltaic modules, *IEEE Access* 8 (2020) 161480–161487.
- [27] A. Mellit, S. Kalogirou, Artificial intelligence and internet of things to improve efficacy of diagnosis and remote sensing of solar photovoltaic systems: Challenges, recommendations and future directions, *Renewable Sustainable Energy Rev.* 143 (2021) 110889.
- [28] Y. Ma, H. Yu, Y. Zhong, S. Chen, X. Gong, H. Xiang, Transferable machine learning approach for predicting electronic structures of charged defects, *Appl. Phys. Lett.* 126 (2025) 044103.
- [29] K. Choudhary, B. G. Sumpter, Can a deep-learning model make fast predictions of vacancy formation in diverse materials?, *AIP Adv.* 13 (2023) 095109.
- [30] J. Y. Chia, N. Thamongsiripak, S. Thongphanit, N. Nuntawong, Machine learning-enhanced detection of minor radiation-induced defects in semiconductor materials using raman spectroscopy, *J. Appl. Phys.* 135 (2024) 025701.
- [31] S. Wang, B. Wright, Y. Zhu, Y. Buratti, Z. Hameiri, Extracting the parameters of two-energy-level defects in silicon wafers using machine learning models, *Sol. Energ. Mat. Sol.* 277 (2024) 113123.
- [32] Y. Buratti, J. Dick, Q. Le Gia, Z. Hameiri, Deep learning extraction of the temperature-dependent parameters of bulk defects, *ACS Appl. Mater. Interfaces* 14 (2022) 48647–48657.
- [33] Y. Buratti, Q. T. Le Gia, J. Dick, Y. Zhu, Z. Hameiri, Extracting bulk defect parameters in silicon wafers using machine learning models, *npj Computational Materials* 6 (2020) 142.
- [34] T. T. Le, Z. Zhou, A. Chen, Z. Yang, F. Rougieux, D. Macdonald, A. Liu, Reassessing iron–gallium recombination activity in silicon, *J. Appl. Phys.* 135 (2024) 133107.
- [35] M. Yamaguchi, K.-H. Lee, K. Araki, N. Kojima, Y. Ohshita, Analysis for efficiency potential of crystalline si solar cells, *J. Mater. Res.* 33 (2018) 2621–2626.
- [36] M. Bošnjaković, Advance of sustainable energy materials: Technology trends for silicon-based photovoltaic cells, *Sustainability* 16 (2024) 7962.
- [37] J. Zhang, Solar pv market research and industry competition report, *IOP Conf. Ser.: Earth Environ. Sci.* 632 (2021) 032047.
- [38] T. Buonassisi, A. A. Istratov, M. D. Pickett, M. Heuer, J. P. Kalejs, G. Hahn, M. A. Marcus, B. Lai, Z. Cai, S. M. Heald, Chemical natures and distributions of metal impurities in multicrystalline silicon materials, *Prog. Photovolt.: Res. Appl.* 14 (2006) 513–531.
- [39] M. Schubert, M. Padilla, B. Michl, L. Mundt, J. Giesecke, J. Hohl-Ebinger, J. Benick, W. Warta, M. Tajima, A. Ogura, Iron related solar cell instability: Imaging analysis and impact on cell performance, *Sol. Energy Mater. Sol. Cells* 138 (2015) 96–101.
- [40] L. Kimerling, J. Benton, Electronically controlled reactions of interstitial iron in silicon, *Physica B+C* 116 (1983) 297–300.
- [41] C. Möller, T. Bartel, F. Gibaja, K. Lauer, Iron-boron pairing kinetics in illuminated p-type and in boron/phosphorus co-doped n-type silicon, *J. Appl. Phys.* 116 (2014) 024503.
- [42] G. Zoth, W. Bergholz, A fast, preparation-free method to detect iron in silicon, *J. Appl. Phys.* 67 (1990) 6764–6771.
- [43] S. Rein, S. W. Glunz, Electronic properties of interstitial iron and iron-boron pairs determined by means of advanced lifetime spectroscopy, *J. Appl. Phys.* 98 (2005) 113711.
- [44] J. Schmidt, D. Macdonald, Recombination activity of iron-gallium and iron-indium pairs in silicon, *J. Appl. Phys.* 97 (2005) 113712.
- [45] M. Goodarzi, R. A. Sinton, H. Jin, P. Zheng, W. Chen, Q. Wang, D. Macdonald, Accuracy of interstitial iron measurements on p-type multicrystalline silicon blocks by quasi-steady-state photoconductance, *IEEE J. Photovolt.* 7 (2017) 1216–1223.
- [46] O. Olikh, V. Kostylov, V. Vlasuk, R. Korkishko, Y. Olikh, R. Chupryna, Features of FeB pair light-induced dissociation and repair in silicon n+-p-p+ structures under ultrasound loading, *J. Appl. Phys.* 130 (2021) 235703.
- [47] S. Herlufsen, D. Macdonald, K. Bothe, J. Schmidt, Imaging of the interstitial iron concentration in crystalline silicon by measuring the dissociation rate of iron–boron pairs, *Phys. Status Solidi RRL* 6 (2012) 1–3.
- [48] A. Herguth, Quantification of iron in boron-doped silicon solar cells from open circuit voltage measurements, *IEEE J. Photovolt.* 12 (2022) 937–947.
- [49] D. H. Macdonald, L. J. Geerligs, A. Azzizi, Iron detection in crystalline silicon by carrier lifetime measurements for arbitrary injection and doping, *J. Appl. Phys.* 95 (2004) 1021–1028.
- [50] M. Di Sabatino, Detection limits for glow discharge mass spectrometry (gdms) analyses of impurities in solar cell silicon, *Measurement* 50 (2014) 135–140.
- [51] D. K. Schroder, Semiconductor Material and Device Characterization, John Wiley & Sons, New Jersey, third edition, 2006.
- [52] R. C. Kurchin, J. R. Poindexter, V. Vähänissi, H. Savin, C. del Cañizo, T. Buonassisi, How much physics is in a current-voltage curve? inferring defect properties from photovoltaic device measurements, *IEEE J. Photovolt.* 10 (2020) 1532–1537.
- [53] A. R. Peaker, V. P. Markevich, J. Coutinho, Tutorial: Junction spectroscopy techniques and deep-level defects in semiconductors, *J. Appl. Phys.* 123 (2018) 161559.
- [54] O. Olikh, O. Lozitsky, O. Zavhorodnii, Estimation for iron contamination in si solar cell by ideality factor: Deep neural network approach, *Prog. Photovoltaics Res. Appl.* 30 (2022) 648–660.
- [55] P. Kunze, J. M. Greulich, A. Tummali, W. Wirtz, H. Hoeffler, N. Woehle, S. Glunz, S. Rein, M. Demant, Contactless inline iv measurement of solar cells using an empirical model, *Sol. RRL* 7 (2023) 2200599.
- [56] M. Battaglia, E. Comi, T. Stadelmann, R. Hiestand, B. Ruhstaller, E. Knapp, Deep ensemble inverse model for image-based estimation of solar cell parameters, *APL Machine Learning* 1 (2023) 036108.

- [57] J. Schmidt, Effect of dissociation of iron–boron pairs in crystalline silicon on solar cell properties, *Progress in Photovoltaics: Research and Applications* 13 (2005) 325–331.
- [58] O. Olikh, O. Zavhorodnii, Iron's impact on silicon solar cell execution: Comprehensive modeling across diverse scenarios, *Materials Science and Engineering: B* 317 (2025) 118192.
- [59] M. Burgelman, P. Nollet, S. Degraeve, Modelling polycrystalline semiconductor solar cells, *Thin Solid Films* 361–362 (2000) 527–532.
- [60] M. Masum Mia, M. Faruk Hossain, M. Rahman, N. Badi, A. Irfan, M. Ferdous Rahman, Unveiling the impact of Se based HTM on BaZrSe<sub>3</sub> perovskites solar cell and improving the theoretical efficiency above 32%, *Materials Science and Engineering: B* 311 (2025) 117817.
- [61] T. K. Joshi, G. Sharma, Y. R. Sharma, A. S. Verma, Spectroscopic screening and performance parameters of hybrid perovskite (CH<sub>3</sub>CH<sub>2</sub>PH<sub>3</sub>PbI<sub>3</sub>) using WIEN2k and SCAPS-1d, *Phys. B Condens. Matter* 682 (2024) 415793.
- [62] B. K. Ravidas, A. Das, S. K. Agnihotri, R. Pandey, J. Madan, M. K. Hossain, M. K. Roy, D. Samajdar, Design principles of crystalline silicon/CsGeI<sub>3</sub> perovskite tandem solar cells using a combination of density functional theory and SCAPS-1D frameworks, *Sol. Energ. Mat. Sol.* 267 (2024) 112688.
- [63] H. Liu, L. Xiang, Q. Liu, P. Gao, Y. Zhang, S. Li, F. Gao, Toward high-performance HTL-free all-perovskite tandem solar cells: SCAPS simulation, *IEEE J. Photovolt.* 14 (2024) 59–64.
- [64] L. You, X. Zhang, Q. Ma, W. Zhu, J. Wu, Optimization of electron transport layer-free Cs<sub>2</sub>TiBr<sub>6</sub>/MASnBr<sub>3</sub> laminated structure perovskite solar cells by SCAPS-1D simulation, *Phys. Status Solidi A* 220 (2023) 2300071.
- [65] M. Abdulmalik, E. Danladi, Influence of perovskite thickness on the performance of silver-doped NaZnBr<sub>3</sub> perovskite solar cells using SCAPS software, *Semiconductor Physics, Quantum Electronics & Optoelectronics* 26 (2023) 321–331.
- [66] O. Olikh, Relationship between the ideality factor and the iron concentration in silicon solar cells, *Superlattices Microstruct.* 136 (2019) 106309.
- [67] O. Olikh, V. Kostylyov, V. Vlasniuk, R. Korkishko, R. Chupryna, Intensification of iron–boron complex association in silicon solar cells under acoustic wave action, *J. Mater. Sci.: Mater. Electron.* 33 (2022) 13133–13142.
- [68] Y. Srivastava, A. Jain, End-to-end material thermal conductivity prediction through machine learning, *J. Appl. Phys.* 134 (2023) 225101.
- [69] H. Minagawa, T. Tezuka, H. Tsuchida, Effective combinations of features in predicting the range of incident ions using machine learning, *Nucl. Instrum. Methods Phys. Res. Sect. B Beam Interact. Mater. At.* 553 (2024) 165383.
- [70] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32.
- [71] A. Natekin, A. Knoll, Gradient boosting machines, a tutorial, *Front. Neurobot.* 7 (2013).
- [72] S. Akinpelu, S. Abolade, E. Okafor, D. Obada, A. Ukpog, S. Kumar R., J. Healy, A. Akande, Interpretable machine learning methods to predict the mechanical properties of abx<sub>3</sub> perovskites, *Results Phys.* 65 (2024) 107978.
- [73] W. Cao, X. Liu, J. Ni, Parameter optimization of support vector regression using henry gas solubility optimization algorithm, *IEEE Access* 8 (2020) 88633–88642.
- [74] M. F. Hanif, M. U. Siddique, J. Si, M. S. Naveed, X. Liu, J. Mi, Enhancing solar forecasting accuracy with sequential deep artificial neural network and hybrid random forest and gradient boosting models across varied terrains, *Advanced Theory and Simulations* 7 (2024) 2301289.