

Dear Editor and Reviewers,

We sincerely thank you for taking the time to review our manuscript “Computer vision-based method for quantifying iron-related defects in silicon solar cells” (Ref. No.: SST-111269). Your insightful comments and constructive suggestions have greatly helped us improve the quality of our work. We particularly appreciate your careful reading and thoughtful feedback, which have led to significant improvements in both the technical content and presentation clarity of our manuscript. We have carefully addressed all the comments and made corresponding revisions to the manuscript. The location of revisions is pointed by red color and highlighted in yellow in “CompleteDocumentForReview.pdf”. Below we provide our detailed point-by-point responses to each comment. We hope the revised manuscript better meets your expectations and standards for publication in Semiconductor Science and Technology.

#### *Response to Reviewer #1*

**Major Comment 1.** *Novelty and Contribution: The idea of applying pre-trained CV models to wavelet-transformed kinetic data is interesting and potentially generalizable. However, similar signal-to-image ML transformations exist in other domains. The manuscript should emphasize what new physical or methodological insight this work provides beyond prior Fourier/wavelet-based ML approaches..*

**Reply:** We thank the Reviewer for this insightful comment.

To summarize the novelty and contribution of our work, as well as its distinction from previous studies on semiconductor defects, we note the following. The use of CNNs for analysing defect-related electrophysical dependencies has been explored previously [1]. In that study, however, image construction required a set of curves measured under different conditions, specifically at various temperatures, and the model was trained from scratch, which demanded a very large training dataset. In contrast, our approach relies on a single kinetic dependency and leverages the capabilities of pre-trained computer vision (CV) models. Standard CV models for defect detection in solar cells have also been reported [2, 3, 4, 5, 6, 7, 8, 9], although prior work focused on macro-defects and processed naturally acquired images from conventional cameras. In our case, the emphasis is on point-defect characteristics, and the images used as input are generated from electrophysical measurements. In the analysis of solar cells, wavelet transforms have been applied to one-dimensional dependencies [10] in addition to their use in improving defect detection in photographic images [11]. In those studies, however, the resulting wavelet coefficients were used as features in regression algorithms rather than for constructing images, which is the approach adopted in the present work. More broadly, to the best of our knowledge, one-dimensional signal-to-image conversion for CNN input preparation has typically been achieved either by employing a set of curves [1] or by digitizing standard graphs produced in software such as Origin [12]. The use of the wavelet transform as a preprocessing step for CNNs is therefore novel. Finally, our methodology is designed to function effectively with extremely small datasets, which facilitates the practical application of the proposed approach.

This information has been added to the revised manuscript in the final paragraph of the Introduction.

**Major Comment 2.** *Dataset Size and Overfitting Risk: The study relies on*

*extremely small datasets (25 simulated and 28 experimental samples). Although data augmentation is performed, flipping or rotating spectrograms likely introduces redundant samples rather than independent data points. The near-perfect  $R^2$  values (0.996–0.999) strongly suggest overfitting. A robust cross-validation (e.g., k-fold or leave-one-out) with uncertainty quantification is needed.*

**Reply:**

We thank the reviewer for this important comment and fully agree that, when working with small datasets, careful validation is essential to avoid overfitting. We apologize for the insufficient detail in the description of the network configuration procedure and would like to clarify that robust cross-validation was an integral part of the training and model selection process. Specifically, 5-fold cross-validation was employed during hyperparameter optimization using the Optuna framework. For each trial, the model was trained and evaluated across multiple folds, and the objective function was defined as the mean performance metric over all folds rather than being based on a single train-validation split. In most cases, the standard deviation of the  $R^2$  metric across cross-validation folds did not exceed 0.02, while for other metrics, the variability remained within 15% of the mean value. Collectively, these results indicate a low risk of overfitting. This strategy was applied consistently to all models considered in the study.

After hyperparameter tuning, each model was retrained on the entire training dataset using the selected hyperparameters and evaluated exclusively on a fully independent hold-out test set, which was not involved in either tuning or calibration. This separation ensures an unbiased assessment of generalization performance and corresponds to a practical form of nested cross-validation.

Regarding data augmentation, we acknowledge that geometric transformations of spectrograms (e.g., flipping and rotation) do not create fully independent samples. In this work, augmentation was used solely as a regularization technique to improve model robustness, not as a means to increase the effective dataset size for validation or testing. Importantly, cross-validation and final testing were always performed at the level of the original samples, ensuring that augmented versions of a given sample never appeared simultaneously in training and validation/test folds.

Concerning the near-perfect  $R^2$  values (0.996–0.999), we emphasize that a high  $R^2$  alone does not indicate overfitting when it is: (i) obtained on an independent test set (for both synthetic and experimental data); (ii) accompanied by low and stable values of other error metrics; (iii) consistent across cross-validation folds (standard deviation approximately 0.02); and (iv) associated with a small gap between training and test performance ( $\pm 0.05$ ). Together, these observations indicate that the models did not rely on memorization but instead captured meaningful patterns in the data.

We have updated the manuscript to provide a more precise description of the models tuning procedure, to include information on the standard deviation of the metrics during cross-validation (page 8, the second paragraph in left column), and to highlight the small difference in  $R^2$  values between the training and test datasets (page 11, fist paragraph in left column).

**Major Comment 3.** *Simulation–Experiment Gap: A major discrepancy is observed between simulated and experimental predictions, requiring a post-hoc quadratic correction. This implies that the CNN–regressor models primarily learn the synthetic*

*data distribution rather than physical correlations. The authors should explore physics-based domain adaptation or partial fine-tuning using experimental data instead of empirical correction.*

**Reply:**

We thank the reviewer for this insightful comment. We agree that CNN regressor models trained on simulated data primarily learn the synthetic data distribution. However, it is important to emphasize that the synthetic data faithfully reflect real physical principles. These data were generated on the basis of explicit and transparent physical laws, as described in Section 2.2, and the calculations employed realistic parameter values reported in the literature, as summarized in Table 1 of the revised manuscript. The close correspondence between the synthetic and experimental domains is demonstrated by the high  $R^2$  values (0.95–0.98) and non-catastrophic error magnitudes, for example a MedAPE of 16–23% in the best-case scenarios, observed when models trained exclusively on simulated data were applied directly to experimental measurements without any post-hoc correction, as shown in Figure S6. We therefore maintain that the observed simulation–experiment gap represents a systematic domain shift rather than a lack of learned physical correlations. This shift may arise from limitations of the simulation framework, such as deviations from the assumed uniform distribution of iron atoms across the solar cell thickness, or from minor imprecisions in the parameter values, as further discussed in our response to Minor Comment 6. Accordingly, the post-hoc quadratic correction was introduced as a transparent and controlled means of compensating for this systematic domain shift.

Importantly, as an alternative to the partial fine-tuning suggested by the reviewer, the second part of our study explicitly examines models trained exclusively on experimental data. This strategy completely avoids reliance on synthetic distributions, employs no correction procedure, and directly evaluates model performance under purely experimental conditions.

An alternative approach involves training models on a combined dataset comprising both synthetic and experimental data. The corresponding calculations were performed using this hybrid strategy, in which CNN+regression models were trained on the combined synthetic and experimental datasets and evaluated on an experimental test set. However, the resulting predictive accuracy were low. For example, the best performance was achieved for the YL:FE2:P+GB combination, yielding a MAPE of  $34 \pm 15\%$ , whereas five other model combinations exhibited median absolute errors of approximately 40% accompanied by substantial dispersion. These results indicate that the effective integration of simulated and measured data presents additional challenges and requires further investigation. Moreover, this hybrid strategy deviates from the primary objective of the present work, which is to demonstrate the feasibility of learning from extremely small training datasets. Consequently, these results were not included in the revised manuscript.

The revised manuscript incorporates additional evidence supporting the conclusion that the discrepancy observed for models trained exclusively on simulated data arises from a systematic prediction bias, rather than from a fundamental loss of correlation with real-world physical phenomena.

**Major Comment 4.** *Physical Model Validation: The SCAPS-1D simulations use fixed FeB parameters (binding energy, migration energy, and pre-exponential factors).*

*Yet, these parameters vary widely in literature (0.55–0.69 eV for migration energy). Without sensitivity analysis or error quantification, the generated synthetic dataset may not reflect realistic kinetics. Validation against first-principles or experimental benchmarks would strengthen the study.*

**Reply:**

We thank the reviewer for highlighting this important point regarding the physical assumptions underlying the SCAPS-1D simulations. We acknowledge that FeB-related parameters, including binding energy, migration energy, and pre-exponential factors, show a range of reported values in the literature. The manuscript also provides a concise literature review of the values adopted for these and several other parameters used in the calculations (page 13, second paragraph of the left column). In this study, these parameters were fixed to representative values corresponding to the most commonly cited data in the defect-physics literature and have been consistently employed in numerous previous investigations of FeB kinetics in silicon. We recognize that these parameter values may deviate from the exact experimental conditions, which is reflected in the systematic shift observed between the predicted and experimental results. Consequently, the generated synthetic dataset may not reproduce the short-circuit current kinetics in every quantitative detail. However, the simulated data reliably capture the fundamental trends of the  $I_{\text{SC}}(t)$  dependence as a function of iron concentration. The primary purpose of using the artificial dataset was to demonstrate the feasibility of employing CV models to extract physically meaningful features from wavelet-transformed representations, thereby enabling the estimation of uncontrolled metallic impurity concentrations. This objective was successfully achieved using the simulated data. Importantly, the key conclusions derived from these synthetic datasets were subsequently validated through experiments using measured data later in the study.

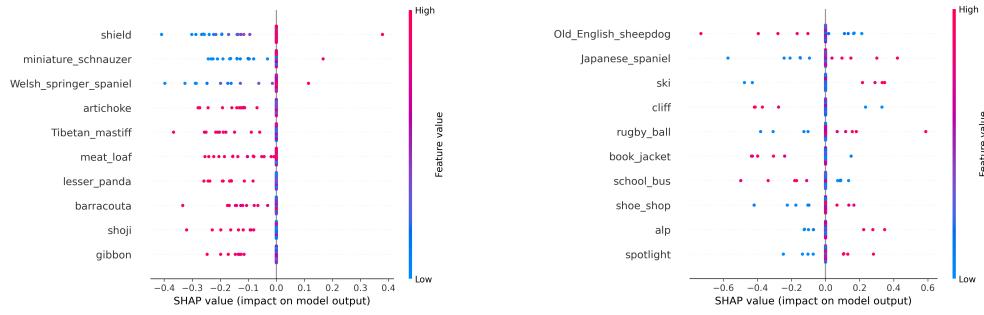
The revised manuscript now includes a discussion of the limitations of the SCAPS-1D simulation results while emphasizing the main outcomes derived from the synthetic data (pages 15–16, last paragraph in Section 3.1).

**Major Comment 5. Regression and Feature Interpretation:** *Although multiple regressors (SVR, XGB, DNN, RF, GB) are compared, no insight is given into the learned features or their physical correlation with iron concentration. Incorporating explainable AI techniques (e.g., SHAP, PCA loading analysis) would add interpretability to what the CNN features represent.*

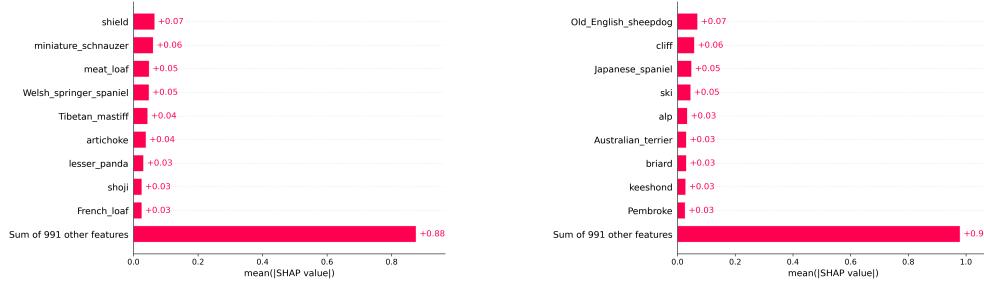
**Reply:** We thank the Reviewer for the comment on model interpretability. In this study, CV models were employed solely as feature extractors. The extracted features correspond either to class probability outputs or to high-level latent embeddings from the penultimate layer, representing abstract, non-linear image descriptors rather than directly interpretable variables.

Examples of feature attributions using SHAP are shown in Figure 1 and Figure 2. However, owing to the abstract and highly non-linear nature of CV-derived features, such as wavelet spectrogram embeddings, these attributions do not correspond to directly interpretable physical quantities.

Consequently, applying SHAP or similar explainable AI techniques to these CNN-derived features primarily reflects sensitivities in an abstract latent space and does not provide direct physical insight into iron concentration. As the CNN features



**Figure 1.** Global SHAP importance of the top ten deep features used by the XGBoost model for iron concentration prediction. The features were extracted using EfficientNetB7 from simulated data (left panel) and NASNetLarge from experimental data (right panel).



**Figure 2.** Bar plots of mean absolute SHAP values showing the ten most influential deep features in the XGBoost-based prediction of iron concentration. Features were extracted using EfficientNetB7 from simulated data (left panel) and NASNetLarge from experimental data (right panel).

lack a clear physical correspondence, their interpretability is inherently limited, and therefore, explainability analysis was not included.

This clarification has been added to the revised manuscript (page 8, last paragraph in Section 2.5).

**Major Comment 6.** *Post-Hoc Correction:* The quadratic correction (Eq. 10) is an empirical adjustment that artificially improves metrics but lacks theoretical justification. The authors should either (i) replace it with a physics-informed calibration (e.g., temperature or diffusion based scaling) or (ii) clearly acknowledge its heuristic nature and limitations.

#### Reply:

We fully acknowledge that the post-hoc calibration procedure employed is purely heuristic. As a result, it can improve predictive accuracy primarily within the training range while providing no physical insight into the underlying system. Its applicability beyond the calibrated domain is therefore limited, and it does not correct structural model errors.

A clear acknowledgment of the heuristic nature of this correction and its associated limitations has been added to the revised manuscript (page 16, last

paragraph of Section 3.1).

**Major Comment 7.** *Statistical Reporting: The reported MSE, MAPE, and  $R^2$  values are given without variance or confidence intervals. Given the small datasets, reporting mean  $\pm$  standard deviation across multiple random splits would be essential to establish statistical robustness.*

**Reply:**

In the revised manuscript, confidence intervals were estimated using a hierarchical resampling procedure. Regression models were trained in five independent runs with different random initializations and stochastic optimization trajectories while maintaining a fixed set of hyperparameters determined during the tuning phase. For each trained model, performance was evaluated on 100 bootstrap resamples of the test dataset generated with replacement. Confidence intervals were then derived from the empirical distribution of the resulting performance metrics.

The selected numbers of stochastic training runs (5) and bootstrap resamples (100) represent a balance between statistical robustness and computational cost and are sufficient to capture the dominant sources of variability in the reported metrics.

The revised manuscript now includes a detailed description of the confidence interval estimation procedure (page 9, last paragraph in Section 2.6). In addition, confidence intervals have been incorporated into all heatmaps presenting performance metrics in both the main text and the Supplementary Materials.

**Minor Comment 1.** *The introduction is overly broad; it should focus more on ML for microscopic defects rather than general PV or macro-defect analysis.*

**Reply:**

The body of work specifically focused on applying ML to microscopic defect analysis remains relatively limited. However, we conducted a detailed review of the publications cited in the initial manuscript draft and carried out an extensive supplementary literature search. As a result, it has been established that existing applications of ML in microscopic defect characterization can be broadly categorized into several distinct approaches. One such approach focuses on enhancing conventional defect-analysis techniques through the integration of Artificial Intelligence methods for processing and interpreting the resulting experimental signals.

For example, Buratti *et al.* [13] employed regression algorithms, including Random Forest (RF), Gradient Boosting (GB), and Deep Neural Networks (DNN), to analyze dependencies derived from temperature- and injection-dependent lifetime spectroscopy (TIDLS). They trained the models on more than one million simulated curves, which enabled accurate estimation of silicon defect energy levels and carrier capture cross-sections. In addition, unlike the conventional fitting of signals with the Shockley–Read–Hall equation, their approach can also predict the energy level position at half of the bandgap. An extension of this approach was presented in [1], where the methodology incorporated a Convolutional Neural Network (CNN) to analyse images derived from a family of lifetime curves measured at different temperatures, in addition to applying a RF model to the standard TIDLS signal. In that study, the CNN was used both to perform the classification of the half-bandgap position of the energy level and to extract features, which were subsequently used by the RF. As in

the earlier work, the models were trained on a dataset consisting of several hundred thousand synthetic samples. An alternative TIDLS signal processing strategy was also investigated by Wang *et al.* [14], who used CNNs to analyse one-dimensional signals and thereby extract parameters associated with two-energy-level defects in silicon. Machine-learning methods are also employed for the analysis of Raman spectra [15]. In this study, the spectra of electron-irradiated GaAs were examined using linear discriminant analysis models. These models were trained on 6,000 experimentally acquired spectra, enabling the identification of radiation-induced defects.

An alternative approach is based on determining defect parameters by analysing the characteristics of devices, primarily solar cells, that are directly affected by such defects. For silicon solar cells, for example, a methodology has been proposed to estimate the concentration of contaminant impurities from the magnitude of the ideality factor obtained from current–voltage ( $I$ – $V$ ) characteristics [16] or from variations in photovoltaic conversion parameters [17]. In both studies, classical regression algorithms (DNN, RF, Support Vector Regression (SVR), and GB) were employed. The numerical values of parameters extracted from the  $I$ – $V$  characteristics served as input features, and the models were trained on tens of thousands of current–voltage curves simulated under different defect parameters. A closely related approach was presented by Haidari *et al.* [18], who used thirteen parameters extracted from the  $I$ – $V$  curves of CIGS solar cells as inputs to a DNN to predict the spatial distribution and concentration of six bulk and surface defects. The inverse problem, namely the determination of photovoltaic conversion parameters based on predefined defect concentrations, was examined by Kim *et al.* [19] for perovskite solar cells. In that study, RF, XGBoost, Linear Regression, and Multilayer Perceptron algorithms were evaluated, and the RF model delivered the highest performance. In all four studies mentioned above, the SCAPS-1D simulation tool was consistently used to generate the synthetic  $I$ – $V$  characteristics that formed the training datasets.

Beyond the detection and characterization of defects in actual devices, a distinct research direction focuses on accelerating and improving density functional theory (DFT) and molecular dynamics (MD) calculations of defect parameters. For example, several studies have demonstrated the use of Graph Neural Networks, trained on DFT-calculated data, to estimate vacancy formation energies [20, 21] and to evaluate the electronic structure of charged defects in GaAs [22]. Graph Convolutional Networks have also been applied to MD-generated datasets for predicting vacancy diffusion paths in high-entropy alloys [23] and intrinsic defects in perovskites [24]. In addition, DFT datasets have been integrated with ML methods to identify formation enthalpies and ionization energies of impurity defects [25] and to determine the equilibrium configurations of defects in emerging materials [26].

The relevant information has been incorporated into the Introduction section (pages 2–3, paragraphs three through six).

**Minor Comment 2.** Some typographical errors exist (e.g., “where where” in Eq. 1). Please proofread carefully.

**Reply:** We carefully reviewed the manuscript and corrected multiple typographical errors. Namely:

- changed phrase “Standard approaches to solving such problems involve the use of Fourier or wavelet transforms, and last were applied in this study” to “Standard

approaches to solving such problems involve the use of Fourier or wavelet transforms, and **the latter** were applied in this study" on page 2;

- changed "where where" to "where" after Eq. (1);
- changed "A is the constant" to "A is the pre-exponential constant" after Eq. (3);
- changed phase "Panels (b) and (c) show the wavelet spectrograms corresponding to the curves with filled squares and open circles, respectively" to "Panels (b) and (c) show the wavelet spectrograms corresponding to the curves with **open** squares and **filled** circles, respectively" in the caption of Fig 3;
- a sentence was added to clarify the applicability of the formula for the median absolute percentage error "Eq. (8) implies that **MAPE<sub>i</sub>** must be arranged in order of magnitude;" after Eq. (8)
- etc.

We hope that these revisions have resolved the issue as thoroughly as possible.

**Minor Comment 3.** *Figures should include axis units, consistent color scales, and indicate whether values are in linear or logarithmic scale.*

**Reply:**

We thank the Reviewer for identifying the shortcomings in the figures. Axes, units of measurement, and color scales have been added to Figures 2b, 2c, 3b, and 3c. The units of measurement (%) corresponding to the MAPE and MedAPE values have also been included next to the color scale labels, as shown in Figures 5, 6, 7, 8, 10, and 11. The mutual arrangement of the ticks and tick labels in all figures enables unambiguous identification of the scale type.

**Minor Comment 4.** *The Supplementary Figures (S1–S10) are repeatedly referenced but insufficiently summarized in the main text. A concise overview table would be helpful.*

**Reply:**

We agree with the Reviewer and have improved the presentation of the Supplementary Figures. A brief summary of Supplementary Figures S1–S10 has been added to the main text at their first reference (page 9, second paragraph in Section 3.1), clarifying their scope and role in supporting the reported results. In addition, the Supplementary Materials were updated with a short introductory paragraph describing their overall content. Besides, we have added a concise overview table summarizing the content and main conclusions of the Supplementary Figures.

**Minor Comment 5.** *The data availability statement ("upon reasonable request") should be replaced with a public repository link for transparency.*

**Reply:**

We have made the data, including simulated and experimentally measured short-circuit current kinetic dependencies as well as wavelet spectrogram images, together with the trained models, available in a public repository ([https://github.com/olegolikh/CV\\_Fe\\_SiSC.git](https://github.com/olegolikh/CV_Fe_SiSC.git)).

**Table 1.** Summary of 1D-CNN model performance metrics

Training Data Type	Evaluation Dataset	MSE ( $10^{-3}$ )	$R^2$	MAPE (%)	MedAPE (%)	Reference (CV-based Models)*
Simulated	Train	$31 \pm 5$	$0.89 \pm 0.02$	$35 \pm 5$	$32 \pm 2$	Fig.5, Fig.S2
	Test	$30 \pm 5$	$0.90 \pm 0.04$	$35 \pm 5$	$32 \pm 3$	Fig.6, Fig.S3
	Experimental, without post-hoc calibration	$530 \pm 70$	$< 0.1$	$480 \pm 60$	$400 \pm 20$	Fig.8(a,b), Fig.S6
	Experimental, with post-hoc calibration	$130 \pm 30$	$0.6 \pm 0.1$	$50 \pm 10$	$44 \pm 8$	Fig.8(c,d), Fig.S7
Experimental	Train	$27 \pm 7$	$0.2 \pm 0.1$	$33 \pm 6$	$26 \pm 8$	Fig.10, Fig.S9
	Test	$30 \pm 10$	$< 0.1$	$40 \pm 10$	$30 \pm 15$	Fig.11, Fig.S10

\* The column indicates the figures in which performance metrics obtained for CV-based models under comparable conditions are reported.

**Minor Comment 6.** A comparison with simpler ML baselines (e.g., direct regression on  $ISC(t)$  data without wavelet transformation) would contextualize the improvement due to CV-based transfer learning.

**Reply:**

Following the Reviewer's suggestion, CNN models were designed that directly process the kinetic dependencies of the short-circuit current to predict iron concentration. The structure of the model was identical to that previously used to successfully analyse one-dimensional signal associated with defects [14]. Specifically, the 1D-CNN model used in this study comprises two one-dimensional convolutional layers followed by global average pooling and three fully connected layers. The first convolutional layer has a kernel size of three, a padding of one, and eight output channels. The second convolutional layer also uses a kernel size of three and a padding of one, but with sixteen output channels. The resulting feature maps are aggregated using a global average pooling layer. The aggregated feature vector is then passed through a fully connected layer with 64 neurons, followed by a second fully connected layer with 32 neurons. The network output is generated by a final fully connected layer with a single neuron. During network tuning, the configuration of batch normalization and regularization in the convolutional layers, dropout between fully connected layers, activation functions, and weight initialization was performed. The models were trained and tested on the same dataset, comprising both artificial and experimental data, from which the wavelet spectrograms were generated. The resulting performance metrics are presented in the table 1.

First, a small difference between the metrics obtained for the training and test sets is observed, as indicated by the pairwise comparison of rows 1 and 2 and rows 3 and 4. This behaviour suggests an appropriate model complexity of the 1D-CNN, similar distributions of the training and test data, and a correct train–test split. At the same time, the absolute values of the metrics are low, indicating underfitting, which is expected given the extremely small size of the training dataset. Moreover, these values are substantially lower than those achieved by the best computer vision based models. This result suggests, on the one hand, that computer vision based transfer learning is highly effective and, on the other hand, that not every computer vision model is

suitable for identifying features relevant to the restructuring of iron-containing defects.

The corresponding information has been added to the revised manuscript.

**Minor Comment 7.** *References [6], [35], [38] should be verified for year and page accuracy. Some reference formatting inconsistencies (journal abbreviations, italics) should be corrected.*

**Reply:** We have thoroughly verified all references, with particular attention given to the citations specifically requested by the Reviewer. Several images containing the references included in the manuscript, as well as screenshots of the corresponding publisher pages, are attached — see Figure 3.

In general, the reference list is generated using .bib files downloaded directly from the publishers' websites. The final formatting of the references is applied automatically according to the IOPscience style guidelines for manuscript preparation. Journal abbreviations follow the options provided by JabRef 5.9. We therefore trust that the final list of literature sources is formatted correctly.

#### *Response to Reviewer #2*

**Comment 1.** *It is additionally essential to examine temperatures between 270 to 350 kelvin.*

**Reply:**

We concur with the reviewer that solar cells typically operate over a broad temperature range, which necessitates the evaluation of various temperature-dependent effects. In the present case, temperature governs the thermodynamic equilibrium between  $\text{Fe}_i$  and FeB defect concentrations, the carrier capture cross-sections of these defects, which together determine the recombination activity of iron-related centers, as well as the silicon parameters that control the photogeneration and transport of excess carriers. In addition, the characteristic time of FeB complex association is temperature-dependent and, in the present system, is described by the relation given below [27, 28, 29]:

$$\tau_{\text{ass}} = 4.2 \times 10^{-10} \frac{\text{s}}{\text{K}} \times T \exp\left(\frac{0.66 \text{ eV}}{kT}\right). \quad (1)$$

Collectively, these factors influence the kinetics of the short-circuit current and its dependence on the iron impurity concentration.

Calculations were performed for both the  $I_{\text{sc}}(t)$  dependencies and the corresponding wavelet spectrograms at 270 K. Representative results are presented in Figure 4. For comparison, the corresponding dependencies obtained at 340 K are shown alongside. As can be seen, at the lower temperature, variations in iron concentration exert a more pronounced effect on the  $I_{\text{sc}}$  kinetics and, consequently, on the appearance of the spectrograms. Accordingly, model predictions are expected to be more accurate under these conditions.

Figure 5 presents the performance metrics obtained from testing models trained on data simulated at different temperatures. As shown, the model performance is, on average, superior for the 270 K case.

However, it should be emphasized that the proposed methodology for iron concentration assessment is primarily intended for practical application. Calculations based on Eq. (1) indicate that  $\tau_{\text{ass}}$  is approximately 850 s at 340 K, whereas it

Schubert M C 2025 *Sol. Energ. Mat. Sol.* **293** 113854  
[6] Jia Y, Chen G and Zhao L 2024 *Sci. Rep.* **14** 15170  
nature > scientific reports > articles > article

Article | [Open access](#) | Published: 02 July 2024  
**Defect detection of photovoltaic modules based on improved VarifocalNet**  
Yanfei Jia, Guangda Chen & Liquan Zhao  
*Scientific Reports* **14**, Article number: 15170 (2024) | [Cite this article](#)  
6152 Accesses | 68 Citations | 7 Altmetric | [Metrics](#)

[35] Otamendi U, Martinez I, Quartulli M, Olaizola I G, Viles E and Cambarau W 2021 *Sol. Energy* **220** 914–926  
  
Volume 220, 15 May 2021, Pages 914-926

Segmentation of cell-level anomalies in electroluminescence images of photovoltaic modules  
Urtzi Otamendi <sup>a</sup> , Iñigo Martínez <sup>b</sup> , Marco Quartulli <sup>a</sup> , Igor G. Olaizola <sup>a</sup> , Elisabeth Viles <sup>b</sup>   
<sup>a</sup>  <sup>b</sup>   
[Show more](#)

[38] Abdelsattar M, Abdelmoety A, Ismeil M A and Emad-Eldeen A 2025 *IEEE Access* **13** 4136–4157

**Automated Defect Detection in Solar Cell Images Using Deep Learning Algorithms**  
Publisher: IEEE | [Cite This](#) | [PDF](#)

Montaser Abdelsattar  ; Ahmed Abdelmoety  ; Mohamed A. Ismeil  ; Ahmed Emad-Eldeen [All Authors](#)

46 Cites in Papers | 2124 Full Text Views

 Open Access  Comment(s)

Under a Creative Commons License

**Abstract**  
This research study introduces a unique method that makes use of a wide range of deep learning (DL) techniques for automated flaw identification in solar cell images. The research paper investigates how well 24 distinct convolutional neural network (CNN) architectures—Residual network (ResNet), densely connected convolutional networks (DenseNet), visual geometry group (VGG), Inception, mobile network (MobileNet), Xception, SqueezeNet, and AlexNet—classify solar cells into defected and non-defective categories. This study is interesting since it does a thorough assessment of a wide variety of models and concentrates on high-performance architectures and lightweight models that may be used in contexts with limited resources. The research paper performed our studies using a balanced and well-curated dataset of 3,102 images of solar cells with a range of common faults. MobileNetV2 and Xception demonstrated excellent performance in defect identification, with accuracy rates of 99.95% and 99.29% respectively, with minimal validation losses. This study demonstrates the potential of efficient models such as MobileNetV2 for real-world use in solar energy generation. It also provides a detailed comparison of several DL models. The results suggest that the inclusion of these models might significantly enhance quality control systems, offering a reliable and efficient method for detecting flaws in solar cells.

**Document Sections**  
I. Introduction  
II. Literature Review  
III. Methodology  
IV. Results and Discussion  
V. Conclusion  
Show Full Outline ▾

**Authors**

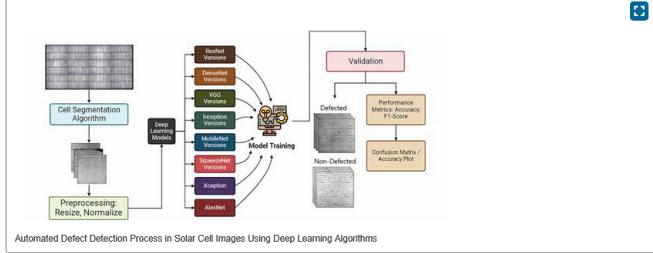
**Figures**

**References**

**Citations**

**Keywords**

**Metrics**

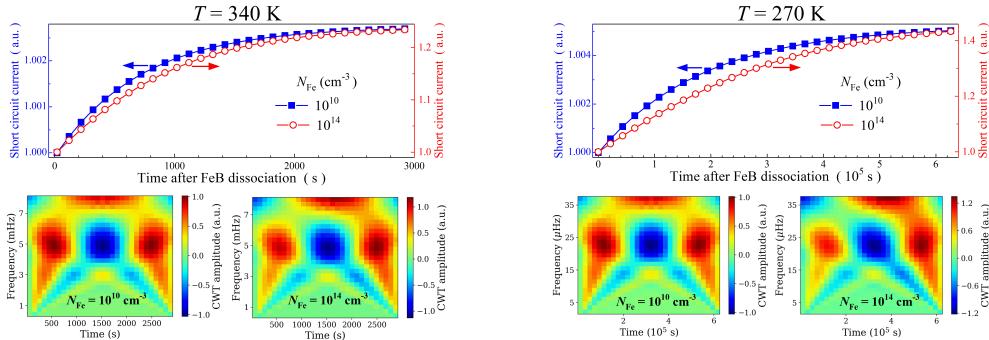


```

graph LR
    Input[Input Image] --> Segmentation[Cell Segmentation Algorithm]
    Segmentation --> Preprocess[Preprocessing: Resize, Normalize]
    Preprocess --> DL[Deep Learning Model]
    DL --> ModelTraining[Model Training]
    ModelTraining --> Validation[Validation]
    Validation --> Performance[Performance Metrics: F1-score]
    Validation --> Confusion[Confusion Matrix / Accuracy Plot]
    ModelTraining --> Defected[Defected]
    ModelTraining --> NonDefected[Non-Defected]
    
```

Published in: *IEEE Access* (Volume: 13)  
Page(s): 4136 - 4157  
Date of Publication: 03 January 2025   
Electronic ISSN: 2169-3536  
Funding Agency:

**Figure 3.** Three images are provided to illustrate the correctness of the references. The upper part of each image contains an excerpt from the bibliography, and the lower part shows a screenshot of the corresponding publisher page.



**Figure 4.** Simulated time dependencies of the short-circuit current (upper curves) and corresponding wavelet spectrograms (lower plots) for iron concentrations of  $10^{10}$  and  $10^{14} \text{ cm}^{-3}$ . The left panel corresponds to a temperature of 340 K, and the right panel corresponds to 270 K.

increases to about 230,000 s at 270 K. The simulated  $I_{\text{SC}}(t)$  dependencies at 340 K were computed assuming a measurement interval of 100 s, while at 270 K the interval increases to 21,600 s, corresponding to six measurements per day and requiring nearly one working week to acquire a complete curve. Experimental validation at 270 K is therefore impractical and lacks clear practical relevance due to the infeasibility of such measurement protocols. Accordingly, we consider it appropriate to focus this study on the upper limit of the temperature range suggested by the Reviewer.

A brief discussion of the advantages and limitations of analyzing the  $I_{\text{SC}}(t)$  kinetic at 270 K, has been added to the revised manuscript (Figure 7, last two paragraphs on page 11, first two paragraphs on page 13). In addition, the performance metrics for models trained on data obtained at 270 K have been included in the Supplementary Materials for both the training and testing sets.

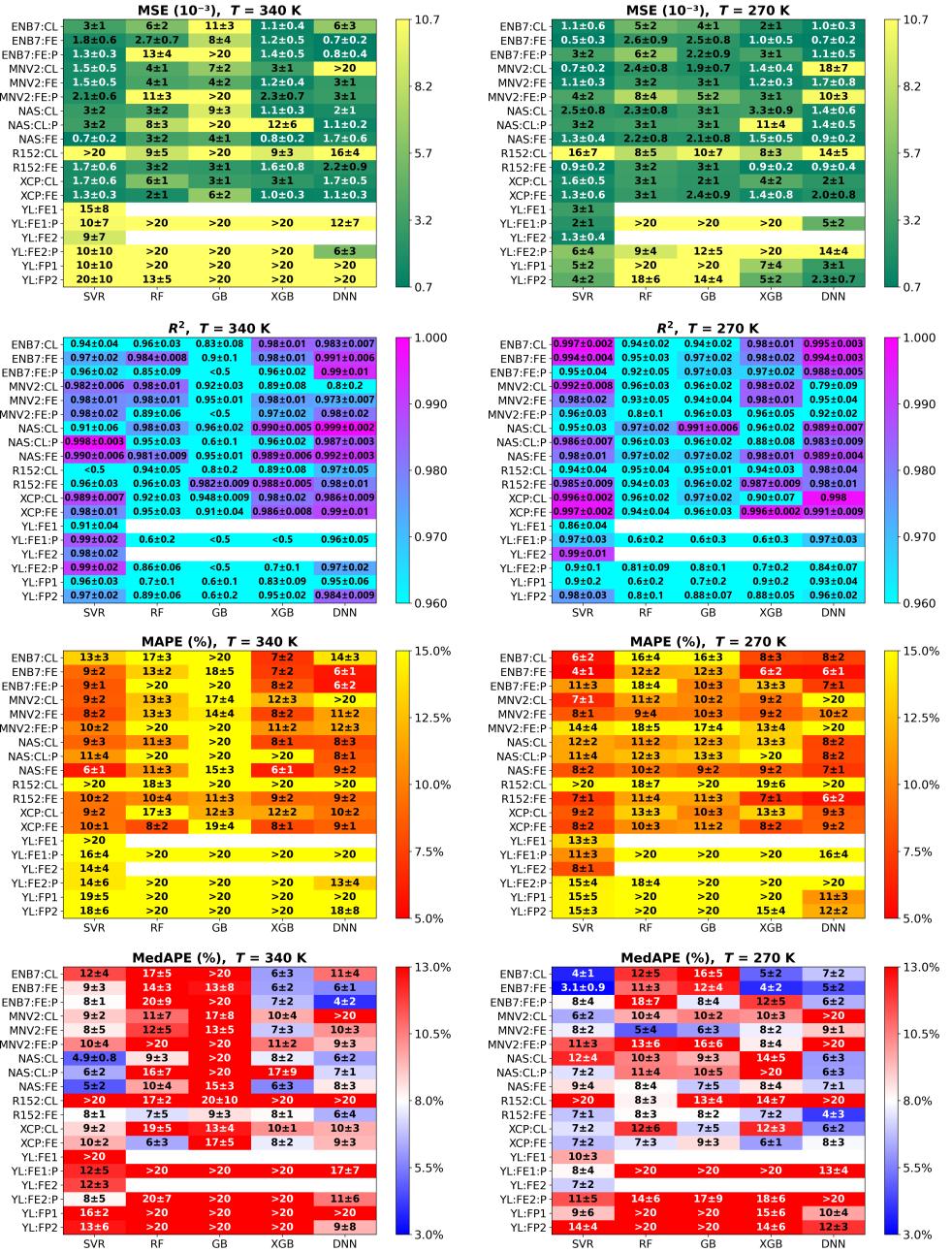
**Comment 2.** You should compare and review your manuscript with other new articles such as “Novel Design of Multi-Layer Cubic Nanoparticles for Achieving Efficient Thin-Film Perovskite Solar Cells”

**Reply:** We have expanded the Introduction section, as detailed in our responses to Minor Comment 1 and Major Comment 1 from Reviewer 1. In particular, the article mentioned in Reviewer 2’s comment is now cited as Reference [38] in the revised manuscript.

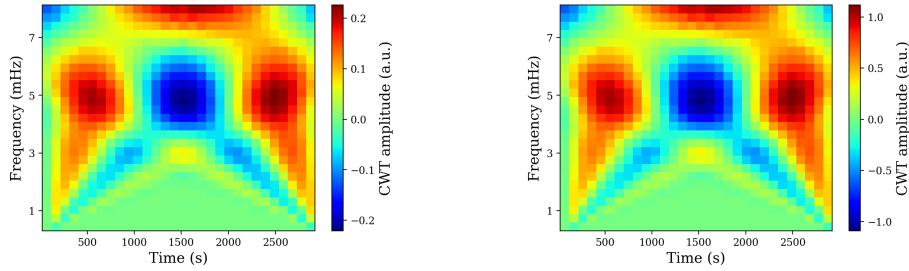
**Comment 3.** Put the solar cell parameters in a table with references.

**Reply:** We added Table 1 to the revised manuscript (page 7). This table summarizes the parameters of the solar cell, silicon, and defect states at the temperature used in the main calculations.

**Comment 4.** Actually, all solar cells have  $R_s$  and  $R_{\text{sh}}$  values. By investigating parasitic losses on cell performance, the article could be made more interesting.



**Figure 5.** Metrics for different combinations of CV models (vertical axis) and regression models (horizontal axis) during test phase with the simulated dataset. The models were trained using the simulated dataset. The left column corresponds to calculations performed at 340 K, and the right column corresponds to 270 K.



**Figure 6.** Wavelet spectrograms corresponding to the simulated short-circuit curves for a solar cell with  $N_{\text{Fe}} = 10^{13} \text{ cm}^{-3}$ , obtained before (left) and after (right) normalization.

**Reply:** We fully agree with the Reviewer that, for real solar cells, it is essential to consider the influence of series and shunt resistances on the efficiency of photovoltaic conversion. In particular, the presence of these resistances reduces the short-circuit current. In an ideal case, the short-circuit current  $I_{\text{SC}}$  is equal to the photogenerated current  $I_{\text{ph}}$ :

$$I_{\text{SC}} = I_{\text{ph}}.$$

In the presence of resistances, within the single-diode model approximation

$$I_{\text{SC}} \approx I_{\text{ph}} \frac{R_{\text{sh}}}{R_{\text{sh}} + R_s}.$$

However, when FeB pairs are restored, the values of  $R_{\text{sh}}$  and  $R_s$  remain unchanged. Therefore, the presence of parasitic resistances will effectively scale the dependence  $I_{\text{SC}}(t)$  by a constant factor only. In this case, the resulting wavelet spectrogram remains unaffected. As an example, Figure 6 presents the spectrograms obtained for the kinetic dependencies of the short-circuit current before and after normalization to the initial current value. It can be seen that normalization affects the amplitude of the continuous wavelet transform, but this change is proportional across all frequency and time values. As a result, the graphical appearance of the spectrogram remains unchanged, and computer vision models extract the same features regardless of normalization (and, consequently, regardless of the presence of parasitic resistance).

Thus, the method proposed in this work for estimating the concentration of impurity iron is inherently resistant to the presence of series and shunt resistances. The corresponding information has been added to the manuscript in the final paragraph preceding the Conclusion section. We are grateful to the Reviewer for highlighting this additional advantage of the proposed approach.

A direct investigation of the influence of parasitic losses on cell performance was not the objective of this study.

#### *Response to EDITOR REPORT*

**REPORT.** *We have found that your manuscript contains text which appears to have been replicated from the following published articles:*

[www.sciencedirect.com/science/article/abs/pii/S0038092X25005171?via%3Dihub](http://www.sciencedirect.com/science/article/abs/pii/S0038092X25005171?via%3Dihub)

*Please reduce the level of overlap in your revised manuscript by rewriting the appropriate sections.*

### Reply:

First, we apologize for the observed similarities. The cited article is our own and also addresses the determination of iron concentration in silicon solar cells. However, the approaches used in the cited work and in the present study are fundamentally different. In the former, regression models are employed that utilize changes in photoelectric parameters during the decay of FeB pairs, whereas in the present manuscript, the primary approach involves converting the kinetic dependencies of the short-circuit current into images and extracting features using computer vision models. Nevertheless, both studies use similar solar cell models, apply standard regression algorithms and metrics, and perform testing on experimental samples from the same batch. This explains certain similarities in wording.

We have revised the relevant sections and reduced the extent of overlap.

### References

- [1] Buratti Y, Dick J, Le Gia Q and Hameiri Z 2022 *ACS Appl. Mater. Interfaces* **14** 48647–48657
- [2] Liu Y, Wu Y, Yuan Y and Zhao L 2024 *Opt. Express* **32** 17295–17317
- [3] Li Z, Zhang S, Qu C, Zhang Z and Sun F 2024 *PLOS ONE* **19** 1–16
- [4] Jia Y, Chen G and Zhao L 2024 *Sci. Rep.* **14** 15170
- [5] Otamendi U, Martinez I, Quartulli M, Olaizola I G, Viles E and Cambarau W 2021 *Sol. Energy* **220** 914–926
- [6] Chen X, Karin T and Jain A 2022 *Sol. Energy* **242** 20–29
- [7] Munawer Al-Otum H 2024 *Sol. Energy* **278** 112803
- [8] Abdelsattar M, Abdelmoety A, Ismeil M A and Emad-Eldeen A 2025 *IEEE Access* **13** 4136–4157
- [9] Tellal H, Hussein A, Rehman S, Liu B, Balghonaim A and Mohandes M 2025 *Case Studies in Thermal Engineering* **66** 105749
- [10] 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
- [11] Li W C and Tsai D M 2012 *Pattern Recogn.* **45** 742–756
- [12] Held M, Bulling J, Lugovtsova Y and Prager J 2024 *Ultrasonics* **143** 107403
- [13] Buratti Y, Le Gia Q T, Dick J, Zhu Y and Hameiri Z 2020 *npj Computational Materials* **6** 142
- [14] Wang S, Wright B, Zhu Y, Buratti Y and Hameiri Z 2024 *Sol. Energ. Mat. Sol.* **277** 113123
- [15] Chia J Y, Thamrongtsiripak N, Thongphanit S and Nuntawong N 2024 *J. Appl. Phys.* **135** 025701
- [16] Olikh O, Lozitsky O and Zavhorodnii O 2022 *Prog. Photovoltaics Res. Appl.* **30** 648–660
- [17] Olikh O and Zavhorodnii O 2025 *Sol. Energy* **300** 113754
- [18] Haidari G 2025 *Appl. Phys. A* **131** 850
- [19] Kim S, Jeong Y, Han D W and Mo C B 2023 *J. Electron. Mater.* **52** 5861–5871
- [20] Choudhary K and Sumpter B G 2023 *AIP Adv.* **13** 095109
- [21] Kumagai Y 2025 *Materia Japan* **64** 518–523
- [22] Ma Y, Yu H, Zhong Y, Chen S, Gong X and Xiang H 2025 *Appl. Phys. Lett.* **126** 044103
- [23] Reimer C, Saidi P, Casert C, Beeler C, Tetsassi Feugmo C G, Whitelam S, Mansouri E, Martinez A, Beland L and Tamblyn I 2025 *J. Appl. Phys.* **138** 074306
- [24] Tyagi V, Pols M, Brocks G and Tao S 2025 *The Journal of Physical Chemistry Letters* **16** 5153–5159 pMID: 40371867 (*Preprint* <https://doi.org/10.1021/acs.jpclett.5c01139>) URL <https://doi.org/10.1021/acs.jpclett.5c01139>
- [25] Mannodi-Kanakkithodi A, Xiang X, Jacoby L, Biegaj R, Dunham S T, Gamelin D R and Chan M K 2022 *Patterns* **3** 100450 ISSN 2666-3899 URL <https://www.sciencedirect.com/science/article/pii/S266638992200023X>
- [26] Mosquera-Lois I, Kavanagh S R, Ganose A M and Walsh A 2024 *npj Comput. Mater.* **10** 121
- [27] Khelifati N, Laine H S, Vähäniemi V, Savin H, Bouamama F Z and Bouhafs D 2019 *Phys Status Solidi A* **216** 1900253
- [28] Möller C, Bartel T, Gibaja F and Lauer K 2014 *J. Appl. Phys.* **116** 024503
- [29] Tan J, Macdonald D, Rougieux F and Cuevas A 2011 *Semicond Sci. Technol.* **26** 055019