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

Oleg Olikh‡, Oleksii Zavhorodnii, Yulia Perets
Taras Shevchenko National University of Kyiv, Kyiv 01601, Ukraine
E-mail: olegolikh@knu.ua

Abstract. abstract abstract abstract abstract abstract abstract

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

Submitted to: *Semicond. Sci. Technol.*

‡ Author to whom any correspondence should be addressed.

1. Introduction

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

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

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

One of the main challenges in applying ML methods effectively is that training the models requires a large amount of labelled data [12]. In practice, researchers often cannot obtain such large volumes of experimental data; therefore, they commonly employ approaches such as simulations, in which hundreds of thousands of dependencies are computed [17, 18, 19, 24, 23]; Physics-Informed Neural Networks (PINNs), which incorporate physical laws into the loss function to generate synthetic data [25, 26]; or Transfer Learning, in which a model trained on one task is adapted to another related task [27, 28]. However, simulations can be highly demanding in terms of time and computational resources; PINNs are primarily suitable for phenomena described by partial differential equations, and pre-trained models are not available for all types of physical problems. At the same time, one of the most extensively studied tasks in machine learning is computer vision (CV), for which many pre-trained models have been publicly released. Moreover, these models are typically trained on extremely large standard datasets. For example, EfficientNetB7 was trained on approximately 1.2 million images from the ImageNet dataset.

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

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

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

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

2. Methodology

2.1. General outline of the method

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

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

The obtained feature sets served as inputs to regression models based on one of the standard algorithms described in Subsection 2.5, which aimed to predict the iron concentration ($N_{\mathtt{Fe}}$) in the solar cell.

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

2.2. Simulation details

2.3. Experiment details

2.4. Computer vision models

2.5. Regression algorithms

2.6. Model evaluation

3. Results and discussion

4. Conclusion

References

[1] International Energy Agency 2024 Renewables 2024 Tech. rep. IEA Paris licence: CC BY 4.0 URL `https://www.iea.org/reports/renewables-2024`

[2] Osama A, Tina G M, Gagliano A, Jimenez-Castillo G and Munoz-Rodríguez F J 2025 *Sol. Energ. Mat. Sol.* **287** 113625

[3] Institute P 2025 Solar generation grew by 30% in 2024, says IEA `https://prometheus.org/2025/02/23/solar-generation-grew-by-30-in-2024-says-iea/` accessed: 2025-10-10

[4] Fischer M, Woodhouse M, Brammer T and Puzant B 2025 International technology roadmap for photovoltaic (itrpv) Tech. rep. VDMA e. V. Frankfurt am Main, Germany

[5] Thome F T, Garashli E, Kwapil W, Schindler F and Schubert M C 2025 *Sol. Energ. Mat. Sol.* **293** 113854

[6] Jia Y, Chen G and Zhao L 2024 *Sci. Rep.* **14** 15170

[7] Hijjawi U, Lakshminarayana S, Xu T, Piero Malfense Fierro G and Rahman M 2023 *Sol. Energy* **266** 112186 ISSN 0038-092X

[8] Liu Y, Wu Y, Yuan Y and Zhao L 2024 *Opt. Express* **32** 17295–17317

[9] Doll B, Hepp J, Hoffmann M, Schüler R, Buerhop-Lutz C, Peters I M, Hauch J A, Maier A and Brabec C J 2021 *IEEE J. Photovolt.* **11** 1419–1429

[10] Datta S, Baul A, Sarker G C, Sadhu P K and Hodges D R 2023 *IEEE Access* **11** 77750–77778

[11] Jaiswal R, Martínez-Ramón M and Busani T 2023 *IEEE J. Photovolt.* **13** 2–15

[12] Buratti Y, Javier G M, Abdullah-Vetter Z, Dwivedi P and Hameiri Z 2024 *Renewable Sustainable Energy Rev.* **202** 114617 ISSN 1364-0321

[13] Mahdavipour Z 2024 *Sol. Energ. Mat. Sol.* **278** 113210

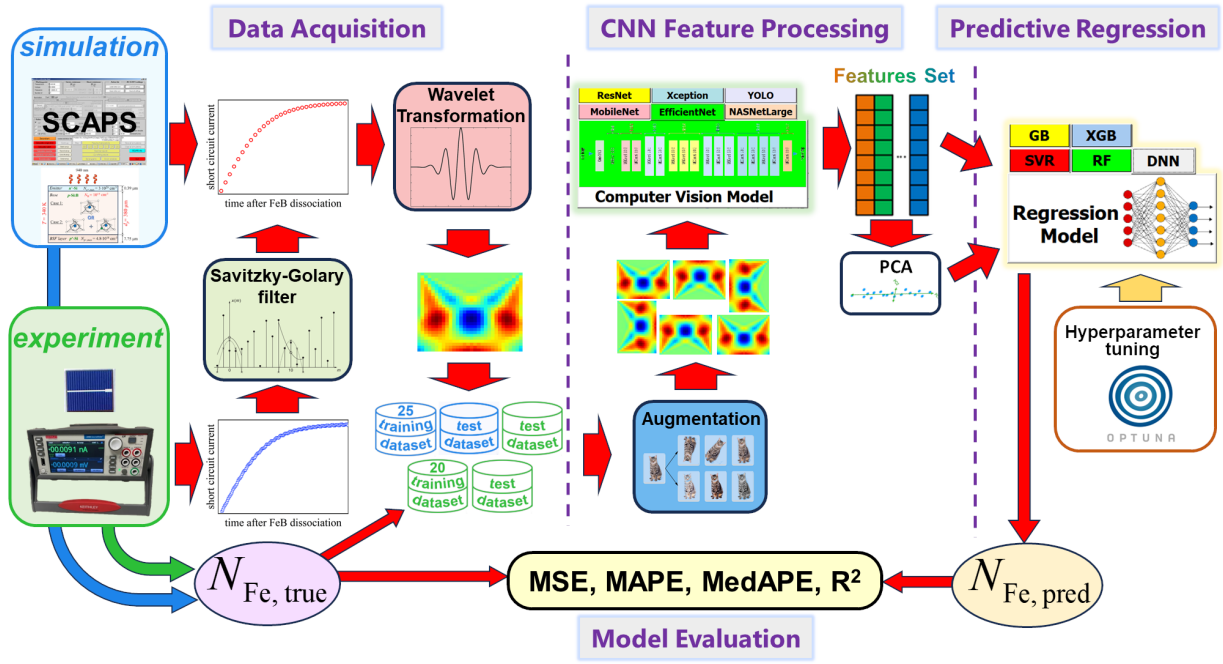

Figure 1. The workflow of the ML pipeline

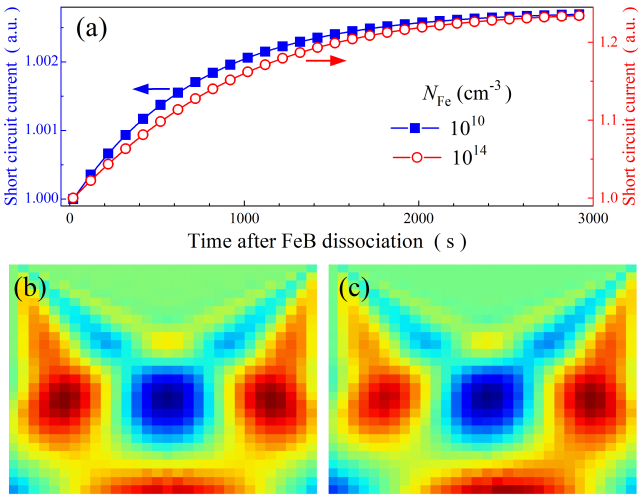

Figure 2. Simulated time dependencies of the short-circuit current (a) and the corresponding wavelet spectrograms for iron concentrations of 10^{10} cm^{-3} (b) and 10^{14} cm^{-3} (c). The data in panel a are shown with filled squares for the concentration corresponding to panel b and with open circles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

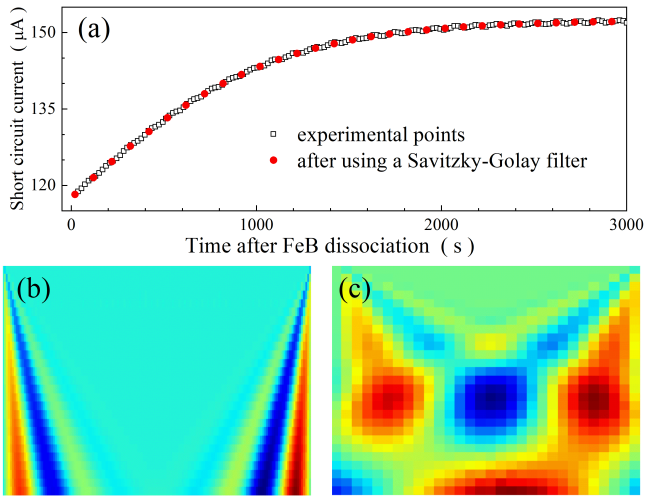

Figure 3. (a) Experimentally measured time dependence of the short-circuit current for a sample with $N_{Fe} = 2.8 \cdot 10^{13}$ cm^{-3} (a, filled squares) and the same dependence after applying the Savitzky–Golay filter (open circles). Panels (b) and (c) show the wavelet spectrograms corresponding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

[14] Hopwood M W, Gunda T, Seigneur H and Walters J 2020 *IEEE Access* **8** 161480–161487
[15] Li B, Delpha C, Diallo D and Migan-Dubois A 2021 *Renewable Sustainable Energy Rev.* **138** 110512 ISSN 1364-0321
[16] Liu Y, Ding K, Zhang J, Li Y, Yang Z, Zheng W and Chen X 2021 *Energy Convers. Manage.* **245** 114603
[17] Wang S, Wright B, Zhu Y, Buratti Y and Hameiri Z 2024 *Sol. Energ. Mat. Sol.* **277** 113123
[18] Buratti Y, Dick J, Le Gia Q and Hameiri Z 2022 *ACS Appl. Mater. Interfaces* **14** 48647–48657
[19] Buratti Y, Le Gia Q T, Dick J, Zhu Y and Hameiri Z 2020 *npj Computational Materials* **6** 142
[20] Park S, Lee J, Khan S, Wahab A and Kim M 2022 *Sensors* **22** 596
[21] Chia J Y, Thamrongsiripak N, Thongphanit S and Nuntawong N 2024 *J. Appl. Phys.* **135** 025701
[22] Olikh O, Lozitsky O and Zavhorodnii O 2022 *Prog. Photovoltaics Res. Appl.* **30** 648–660
[23] Olikh O and Zavhorodnii O 2025 *Sol. Energy* **300** 113754

[24] Olikh O and Zavhorodnii O 2025 *Materials Science and Engineering: B* **317** 118192
[25] Wang S, Sankaran S and Perdikaris P 2024 *Comput. Methods Appl. Mech. Eng.* **421** 116813
[26] Li W K and Zhang Y T 2025 *J. Appl. Phys.* **137** 203304
[27] Kaya M and Hajimirza S 2019 *Sci. Rep.* **9** 5034
[28] Kim Q, Lee S, Ma A, Kim J, Noh H K, Chang K B, Cheon W, Yi S, Jeong J, Kim B, Kim Y S and Kim D S 2023 *Solid-State Electron.* **201** 108568
[29] Buonassisi T, Istratov A A, Pickett M D, Heuer M, Kalejs J P, Hahn G, Marcus M A, Lai B, Cai Z and Heald S M 2006 *Prog. Photovolt.: Res. Appl.* **14** 513–531 ISSN 1062-7995
[30] Schubert M, Padilla M, Michl B, Mundt L, Giesecke J, Hohl-Ebinger J, Benick J, Warta W, Tajima M and Ogura A 2015 *Sol. Energy Mater. Sol. Cells* **138** 96–101
[31] Kimerling L and Benton J 1983 *Physica B+C* **116** 297–300
[32] Möller C, Bartel T, Gibaja F and Lauer K 2014 *J. Appl. Phys.* **116** 024503
[33] Olikh O, Kostylyov V, Vlasiuk V, Korkishko R, Olikh Y and Chupryna R 2021 *J. Appl. Phys.* **130** 235703
[34] Li Z, Zhang S, Qu C, Zhang Z and Sun F 2024 *PLOS ONE* **19** 1–16
[35] Otamendi U, Martinez I, Quartulli M, Olaizola I G, Viles E and Cambarau W 2021 *Sol. Energy* **220** 914–926
[36] Chen X, Karin T and Jain A 2022 *Sol. Energy* **242** 20–29
[37] Munawer Al-Otum H 2024 *Sol. Energy* **278** 112803
[38] Abdelsattar M, Abdelmoety A, Ismeil M A and Emad-Eldeen A 2025 *IEEE Access* **13** 4136–4157
[39] Tella H, Hussein A, Rehman S, Liu B, Balghonaim A and Mohandes M 2025 *Case Studies in Thermal Engineering* **66** 105749
[40] Li W C and Tsai D M 2012 *Pattern Recogn.* **45** 742–756
[41] dela Rosa M E C, Mateo-Romero H F, Alonso-Gómez V, Ngungu V N, Nava R, Aragonés J I M, Plaza A R, Ángel González-Rebollo M, Isaza J R F and Cardeñoso-Payo V 2024 *Renewable Energies* **2** 27533735241304090
[42] Khanna M, Srinath N K and Mendiratta J K 2020 *Application of Neural Networks and Lifting Wavelet Transform for Long Term Solar Radiation Power Prediction* (Singapore: Springer Singapore) pp 95–105 ISBN 978-981-15-3125-5
[43] Krishnan S R and Seelamantula C S 2013 *IEEE Trans. Signal Process.* **61** 380–391
[44] Torrence C and Compo G P 1998 *Bull. Am. Meteorol. Soc.* **79** 61–78
[45] Ahmad A, Jin Y, Zhu C, Javed I, Maqsood A and Akram M W 2020 *IET Renew. Power Gener.* **14** 2693–2702