Reviewer Report on Manuscript SST-111269

Title: Computer vision-based method for quantifying iron-related defects in silicon solar cells

Overall Evaluation

The manuscript presents a hybrid transfer-learning framework that leverages pre-trained computer vision (CV) models to quantify iron contamination in boron-doped crystalline silicon solar cells. The authors transform short-circuit current ($I_{SC}(t)$) relaxation curves following FeB pair dissociation into wavelet spectrogram images and subsequently use convolutional neural networks (CNNs) to extract high-dimensional features for regression-based prediction of iron concentration. The proposed method achieves high accuracy (R^2 up to 0.999) using both simulated and experimental datasets.

While the topic is timely and the concept of applying transfer learning to point-defect quantification is innovative, the study suffers from limitations related to dataset size, potential overfitting, and insufficient physical validation. The manuscript would benefit from substantial revision before being suitable for publication.

Major Comments

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.

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.

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.

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.

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.

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.

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.

Minor Comments

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

2. Some typographical errors exist (e.g., "where where" in Eq. 1). Please proofread carefully.

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

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

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

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

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

Recommendation

The paper presents an innovative methodology that bridges computer vision and defect physics; however, its claims are currently overstated. Key issues—especially overfitting risk, lack of independent validation, and absence of physical justification for corrections—must be addressed. Expanding the dataset, improving cross-validation, and providing uncertainty analysis would substantially enhance credibility.

Recommendation: Major Revision