

Solar Energy

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

--Manuscript Draft--

Manuscript Number:	SEJ-D-25-01089R1
Article Type:	Research paper
Section/Category:	Photovoltaic Cells and Cell Physics
Keywords:	photovoltaic defect quantification; silicon; Solar Cells; iron contamination; Machine learning; current-voltage curves
Corresponding Author:	Oleg Olikh, Dr. Hab. Taras Shevchenko National University of Kyiv Kyiv, UKRAINE
First Author:	Oleg Olikh, Dr. Hab.
Order of Authors:	Oleg Olikh, Dr. Hab. Oleksii Zavhorodnii
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 R2 values of up to 0.003, 3%, and 0.997 for synthetic data and 0.004, 9%, and 0.987 for experimental data.

To: Solar Energy Editorial Board
Subject: Manuscript Submission

Dear Editors,

Enclosed with this letter you will find the electronic submission of manuscript entitled “Extracting the iron concentration in silicon solar cells using photovoltaic parameters and machine learning” by Oleg Olikh and Oleksii Zavhorodnii.

The integration of machine learning (ML) technology is becoming increasingly widespread across various domains of human activity, and photovoltaics is no exception. Thus, researchers have extensively applied ML techniques to analyze electroluminescence and photoluminescence images of solar cells, facilitating the detection and classification of defects such as cracks, finger failures, hot spots, scratches, and dislocations. However, impurity quantification has received relatively little attention.

In this study, we propose a ML-based approach for extracting information about iron contamination in silicon solar cells from current-voltage measurements. This manuscript presents the results of a comparative analysis of 80 different models, which vary in terms of the ML algorithm used, feature extraction conditions, and their combinations, as well as the application of data preprocessing techniques. Testing was conducted on both simulated data and experimentally obtained data. Unlike traditional non-ML techniques, the proposed method does not rely on complex equipment, enabling faster, non-destructive, and inline-compatible analysis of silicon solar cells. This approach has the potential to offer a more efficient and cost-effective solution for impurity quantification in photovoltaic devices. We believe this work will be highly relevant to researchers focused on photovoltaic device characterization methods and the optimization of solar cell manufacturing processes. Given the growing interest in renewable energy technologies and machine learning, we are confident that our findings will resonate with *Solar Energy* readership and stimulate further exploration in this area.

This manuscript is original and has not been submitted elsewhere, in whole or in part. No aspects of this work have been published in any form. There are no conflicts of interest related to this submission.

We would very much appreciate if you would consider the manuscript for publication in the *Solar Energy*.

Sincerely yours,
Oleg Olikh and Oleksii Zavhorodnii
Taras Shevchenko National University of Kyiv
Kyiv 01601, Ukraine
E-mail: olegolikh@knu.ua

Dear Editor and Reviewers,

We sincerely thank you for taking the time to review our manuscript “Extracting the iron concentration in silicon solar cells using photovoltaic parameters and machine learning” (Ms. Ref. No.: SEJ-D-25-01089). 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 “MarkedManuscript.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 Solar Energy.

Response to Reviewer #1

Comment 1. *The introduction lacks the role of iron valency (ferrous or ferric) in silicon solar cells.*

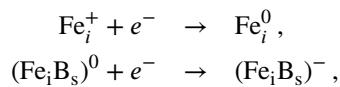
Reply: Reviewer is correct that ferrous (Fe^{2+}) and ferric (Fe^{3+}) states deserve consideration, as they are the most common and stable charge states of iron. These ionic forms typically occur in compounds where iron forms chemical bonds (either ionic or covalent) with other elements, as well as in cases where iron is present as an impurity in solid materials. In silicon, iron can also exist in a trivalent (ferric) state when it substitutes for a silicon atom at a lattice site. However, under normal conditions, the concentration of substitutional iron is extremely low, less than 1% of the total iron impurity atoms [1]. Increasing the concentration of substitutional iron requires special sample processing, such as high-temperature annealing or irradiation. Moreover, substitutional iron acts as a weak recombination center, so its influence on the properties of silicon solar cells can be neglected. The ferrous form is virtually absent in silicon. The majority of iron impurity atoms in silicon occupy positions, where they can exist in either a neutral (Fe_i^0) or positively charged (Fe_i^+) state, depending on the position of the Fermi level [2]. In *n*-type silicon, Fe_i is more likely to exist in a neutral state, whereas in *p*-type silicon, it is more likely to be positively charged. Even when interstitial iron forms complex point defects, such as iron–boron pairs (Fe_iB_s), no valence bonds are formed. In interstitial configuration, iron acts as an active recombination center.

Consequently, in silicon solar cells, the role of iron valence is negligible, in contrast to other types of solar cells, such as perovskite solar cells [3]. Specifically, in MAPbI_3 -based devices, both Fe^{3+} and Fe^{2+} point defects are observed, with Fe^{3+} being electronically inactive in terms of recombination.

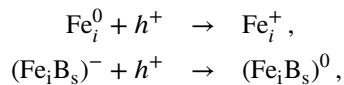
We have added the relevant information to the Introduction (page 2, third paragraph from the top).

Comment 2. *The work didn't deal with the redox reaction of iron in silicon solar cells.*

Reply: Initially, it is important to note that redox reactions do not play a central role in silicon solar cells, unlike dye-sensitized solar cells or photoelectrochemical cells. However, changes in the charge states of individual atoms — whether intrinsic or impurity-related — occur during the photoelectric conversion process. For example, during the recombination of charge carriers at iron-related defects (such as interstitial iron atoms or iron–boron pairs), electron capture occurs in the initial stage:



which typically corresponds to a reduction reaction. In the subsequent stage, hole capture alters the defect's charge state again



formally representing an oxidation reaction. When FeB pairs dissociate due to intense illumination, electron injection, or heating up to 200 °C, the iron atom becomes neutral by capturing an electron (reduction) during the first step of this two-stage process. After dissociation and the cessation of the external stimulus, the interstitial iron atom loses an electron (oxidation), becoming singly positively charged, which allows it to re-form a pair with a negatively charged dopant (boron). However, transitions between different charge states of iron-related defects in silicon are governed by electron and hole capture and emission processes, involving the exchange of carriers between the defect levels and the

conduction or valence band. These processes are described within Shockley–Read–Hall theory and do not constitute redox reactions in the classical chemical sense.

Consequently, the original manuscript did not emphasize redox reactions of iron in silicon solar cells, and the Reviewer's observation is valid. In the revised manuscript, we have included a note on the distinct nature of charge-state transitions in iron-related defects, highlighting how they differ from classical redox reactions (page 2, fourth paragraph from the top).

Comment 3. *In the title, "Extracting the iron concentration" should be changed to "Determination of iron concentration."*

Reply: We concur that the phrase "Determination of iron concentration" better emphasizes the quantitative aspect of the work and aligns well with the standard terminology used in similar studies. Accordingly, we have revised the manuscript title to: "Determination of iron concentration in silicon solar cells using photovoltaic parameters and machine learning."

Response to Reviewer #2

Comment 1. *Why temperature range (290–340) K? what was the dissociation efficiency of FeB pairs at this temperature? How much is translated to cell degradation? Since at this temperature range, other factors also go side on that impact device characteristic.*

Reply: We hope that the proposed approach will be applied to evaluate the concentration of iron in already installed solar modules and during the certification of photovoltaic converters. On the one hand, the selected temperature range of 290–340 K reflects the typical operating conditions of most silicon solar cells. On the other hand, the IEC 61215-2:2021 [4] standard requires that temperature coefficients of short-circuit current, open-circuit voltage, and maximum power be determined based on current–voltage measurements conducted within a temperature range of approximately +15 °C to +75 °C (288–348 K). In other words, the standard solar cell testing results can serve as input data for machine learning models trained within selected temperature range.

Additional information justifying the choice of temperature range has been added in the revised manuscript (page 4, second paragraph).

Complete dissociation of the FeB pairs was achieved during experiments conducted within the specified temperature range (further clarified below in response to the following Reviewer's comment).

The changes in photovoltaic parameter values resulting from FeB pair dissociation depend on both the concentration of iron impurities and the solar cell parameters (such as base thickness and dopant concentration), as well as the measurement conditions (illumination intensity and spectral composition, temperature). This issue is discussed in more detail in [5], which presents the results of silicon solar cell modeling used to create the training dataset. For instance, under AM1.5 illumination, the relative changes in short-circuit current range from –40.88% (for $N_B = 10^{17} \text{ cm}^{-3}$, $T = 290 \text{ K}$, $d_p = 380 \mu\text{m}$) to 6.11% (for $N_B = 10^{15} \text{ cm}^{-3}$, $T = 315 \text{ K}$, $d_p = 380 \mu\text{m}$). Similarly, the relative changes in open-circuit voltage vary from –12.95% (for $N_B = 10^{15} \text{ cm}^{-3}$, $T = 340 \text{ K}$, $d_p = 180 \mu\text{m}$) to 1.92% (for $N_B = 1.778 \times 10^{15} \text{ cm}^{-3}$, $T = 340 \text{ K}$, $d_p = 230 \mu\text{m}$). Notably, the most pronounced changes in solar cell characteristics do not necessarily occur at the highest iron concentrations. For example, the maximum negative change in open-circuit voltage (–12.95%) is observed at an iron concentration of $3.162 \times 10^{12} \text{ cm}^{-3}$, rather than at the highest tested level.

This suggests that FeB pair dissociation, depending on the specific parameters, can not only degrade solar cell performance but also improve certain characteristics [5, 6]. However, it is important to emphasize that light-induced dissociation of FeB pairs does not lead to irreversible changes in photovoltaic parameters. Once the illumination ceases, the FeB pairs re-form, and the solar cell characteristics return to their original values [7, 8, 9, 10, 7, 8].

Information regarding the ranges of changes in photovoltaic parameters and their reversibility has been added to the revised manuscript (page 4, fourth paragraph from the top; page 3, first paragraph in Section 2.1).

We fully agree with the Reviewer that, within the chosen temperature range, factors beyond iron concentration and the selected descriptors may also influence solar cell characteristics. This is especially relevant for real solar cells, which, in addition to iron, may contain other recombination-active impurities. To better isolate iron-related effects from the influence of other factors, we used relative changes in PV parameters as model inputs. Since models trained

in this way provided sufficiently accurate predictions of iron concentration from both simulated and experimentally measured current–voltage characteristics, we consider this approach well justified.

Comment 2. Complete details about the dissociation of FeB pairs using halogen lamp illumination? What was the time duration? How much was the decay efficiency.

Reply: It is known [11] that strong optical illumination ($>100 \text{ W/m}^2$) results in almost complete (>99%) dissociation of the FeB pairs. The dissociation rate R_d is influenced by the pair concentration N_{FeB} and overall carrier generation rate G , which is proportional to the illumination intensity [7, 8, 9, 10]:

$$R_d = K \left(\frac{G}{N_{\text{FeB}}} \right)^2, \quad (1)$$

where K is the constant of the material. Additionally, it is well established that R_d depends on temperature [12, 13], the spectral composition of illumination [14], and the presence of recombination channels other than those associated with iron-related defects [7, 8].

For experimental validation of the proposed models, samples with an iron concentration of 2×10^{11} to $4 \times 10^{13} \text{ cm}^{-3}$ were used. Dissociation of FeB pairs was induced by intense illumination from a halogen lamp (7000 W/m^2). The illumination duration required to achieve near-complete dissociation depended on the iron concentration and reached up to 400 s for the solar cell with the highest N_{FeB} values. When selecting the illumination intervals, we considered the results of a previous study [14], which showed that in samples with a similar defect composition and exposed to a comparable light source, complete dissociation at $N_{\text{Fe}} = 9 \times 10^{12} \text{ cm}^{-3}$ occurred within 20 s.

Details regarding the selection of the illumination duration can be found on page 4 (first paragraph).

Comment 3. Most of the models significantly fail for the AM1.5 illumination? What could be the reason for that?

Reply: Indeed, models trained on data obtained under AM1.5 illumination perform worse in almost all cases. We believe this can be attributed to the following reasons. It is well known that a regression model may achieve high accuracy by having the target variable be an unambiguous function of the descriptor. Moreover, the absolute value of the function's derivative should be as significant as possible. As shown earlier [5], the condition of uniqueness in the relationship between N_{Fe} and the changes in photovoltaic parameters under AM1.5 illumination is often not fulfilled in contrast to the case of monochromatic illumination. In addition, the ranges of changes in ϵI_{SC} , ϵV_{OC} , and $\epsilon \eta$ under monochromatic illumination exceed those observed under AM1.5 illumination. For illustration, Fig. 1 shows some examples from [5]. The physical origin of the differences in the relationship between iron concentration and parameter changes under AM1.5 and monochromatic illumination lies in the fact that, in the former case, the generation of non-equilibrium charge carriers occurs throughout the entire volume of the solar cell, rather than being limited to the base region. As a result, the processes involved in photovoltaic conversion are more complex. In our view, these differences explain why the performance of ML models trained on AM1.5 illumination data is generally poorer. One way to address the ambiguity in the relationship between features and the target variable is to incorporate additional features into the model. As shown in the study (e.g., see Fig. 7 in manuscript), for AM1.5 illumination, increasing the number of descriptors often improves prediction accuracy, supporting our hypothesis.

This answer is incorporated in the text on page 17, second paragraph from the bottom.

Comment 4. Authors are suggested to provide more explanation behind failure of the particular model, infact, for the results obtained.

Reply: We thank the Reviewer for the valuable comment. We fully agree that discussing potential reasons for the discrepancies in model performance is essential. In fact, our response to the previous comment already addresses part of this issue. Upon further examination of other cases, we note the following:

Increasing the number of descriptors while keeping the sample size constant leads to higher sparsity in the data space due to increased dimensionality, which in turn makes it more difficult for the model to generalize. In such cases, the model may also learn spurious correlations that do not generalize to test data — a risk that is particularly pronounced

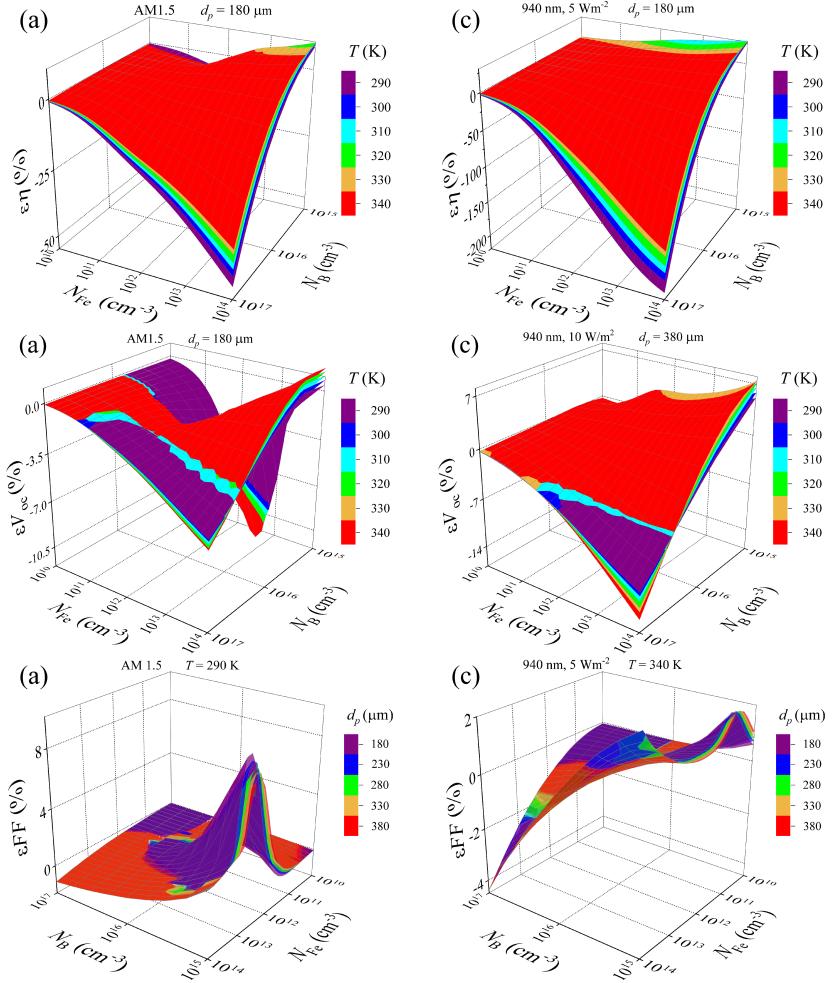


Figure 1: Relative changes in solar cell efficiency (first row), open-circuit voltage (second row), fill factor (third row) caused by a complete dissociation of Fe_3B_s . The dependencies are shown for two cases: AM1.5 illumination (left column) and 940 nm monochromatic light (right column). The figures are taken from reference [5].

in deep neural networks and deep tree-based ensemble methods. Moreover, a larger number of features results in more model parameters, increasing the risk of overfitting. These factors can collectively lead to a reduction in prediction accuracy. In our opinion, for $A^{AM1.5}$ models, both the positive effects achieved by increasing the number of descriptors (enhancing overall information variance and reducing ambiguity between the target variable and the input features) were dominant. However, for A^{940} models where this ambiguity was already low, the negative effects described above outweighed the benefits, which explains why prediction accuracy did not improve when using an expanded feature set.

Another effect that may arise when increasing the number of features is multicollinearity. It occurs when new features are highly correlated with existing ones and can lead to instability in model estimates. However, in our case, it is unlikely that multicollinearity among the photovoltaic parameters (incidentally, their correlation is significantly higher in the 940 nm case than in the AM1.5 case — see Fig. S1) affects prediction accuracy. First, all algorithms used are sufficiently robust to multicollinearity. Second, a standard method for mitigating multicollinearity is PCA. Despite previous successes in applying principal component analysis to various photovoltaics-related machine learning tasks [15, 16, 17, 18, 19] — and the presence of favorable conditions for its use in our case, such as the high correlation between changes in photoelectric parameters caused by FeB pair dissociation — applying PCA as a data pre-processing step for evaluating iron impurity concentration did not improve the maximum prediction accuracy in most cases. One likely reason is that PCA preserves the overall variance in the data but does not account for the specific relationships between individual features and the target variable. For instance, some photovoltaic parameters may be

more informative for predicting low iron concentrations, while others may be more relevant for high concentrations. For instance, ϵ_{FF}) are shown [5] to be most pronounced at low iron concentrations, unlike other PVPs. When these components are merged through PCA transformation, such distinctions may be lost, leading to reduced accuracy. Another possible reason is that principal components — being linear combinations of original features — do not support localized decision splits (i.e., those based on individual feature thresholds), which are particularly effective in tree-based models.

When comparing the effectiveness of different algorithms, several points are worth noting. The poorest performance observed with SVR is likely due to its limited ability to model complex nonlinear relationships — such as those that apparently exist between iron concentration and variations in photoelectric parameters. In contrast, the superior performance of the XGB and DNN models is attributable to their ability to effectively capture nonlinear patterns, which in our case become more prominent with increasing descriptor dimensionality. Specifically, XGB performs best in relatively simpler scenarios, such as those involving 940 nm illumination (see Figs. 7i-7l, 9i-9l). This is because XGB requires less training data than DNN (12,000 examples are sufficient, though this is not a particularly large amount for a neural network), handles high-quality structured data well (simulated data are noise-free), and is more resilient to the influence of correlated features. At the same time, DNNs are more powerful tools for nonlinear modeling, capable of uncovering hidden structures in the data. This enables them to achieve better results on more complex tasks — such as the transition to A^{AM1.5}-models with high-dimensional features (see Figs. 7m-p, 9m-9p). As shown by the results in Subsection 3.4, successful prediction within the N_B -altered also requires accounting for higher-order relationships, which explains why DNN models achieved the best performance (Figs. 11e-11l). Regarding the RF and GB algorithms, as indicated in Tables 1 and 2, they tend to perform best only on less complex data (i.e., small sets of descriptors), where more powerful models are at higher risk of overfitting.

Reply has been incorporated into revised manuscript as Subsection 3.5.

Comment 5. What can be done to further increase the accuracy of these models?

Reply: When exploring potential strategies for improving model accuracy, it is essential to consider the efforts already undertaken towards this goal. Specifically, data normalization was applied, multiple descriptor sets were tested, and the impact of reducing feature correlation using PCA was evaluated. In addition, the main model parameters were optimized (see Tables S1–S5 and Figure S2) through bayesian optimization using the Optuna package, and cross-validation was employed.

Among the remaining justified options, the following approaches can be highlighted:

1. Expanding the training dataset, primarily by incorporating a greater number of boron concentration values into the database. This strategy aligns with a fundamental principle of machine learning—the need for as much data as possible.
2. Deriving new descriptors from existing ones that are more directly and simply related to iron concentration. These descriptors should not be limited to linear combinations, such as those used in PCA; instead, their development should be guided by underlying physical processes. Alternatively, generative models, such as variational autoencoders, can be employed to discover latent variables. Notably, this approach may also facilitate easier dataset expansion.
3. Incorporating additional features, especially those obtainable without extra measurements, is desirable because the requirements of additional experiments would substantially reduce the practicality of the proposed method. For example, the minority carrier diffusion length could serve as a valuable descriptor; however, to our knowledge, no methods exist for its determination based solely on single or dual I - V characteristic measurements. The most apparent step in this direction involves the use of relative changes in voltage (ϵV_{MP}) and current (ϵI_{MP}), corresponding to the maximum power point, as supplementary features. However, it should be noted that these quantities exhibit correlation with other photovoltaic parameters (see Fig. 2).

Separately, a strategy for improving the quality of the training dataset, which involves increasing the realism of the simulation model, can be distinguished. For example, instead of employing SCAPS, three-dimensional simulators such as Synopsys Sentaurus TCAD, Silvaco ATLAS TCAD, or Solcore can be used to generate labeled datasets that account for the presence of additional impurities commonly found in Cz-Si, such as oxygen and carbon. Furthermore, incorporating experimental uncertainty into the training data by introducing noise can enhance the model's robustness and predictive accuracy.

This answer is incorporated in the revised text in Subsection 3.7.



Figure 2: Correlation plot of features in hypothetical expanded training set. Data above and below the main diagonal correspond to AM1.5 and 940 nm illumination, respectively.

Comment 6. There are some typing mistakes as well. for e.g. Page 14, line 44, t-ltered...

Reply: We thank the Reviewer for pointing out the typographical errors and apologize for their presence. We have carefully reviewed the entire manuscript and corrected typographical issues to improve its clarity and professionalism.

Response to Reviewer #4

Comment 1. Clearly describe the main objectives as to why and how using this ML approach can be better than traditional ways in the Introduction section.

Reply: The authors fully agree with the reviewer's assessment concerning the importance and, indeed, the necessity of including a synthesis in the Introduction that outlines the primary objective of this work and highlights the advantages of the proposed approach.

The primary objective of this study is to introduce a novel methodology for quantifying iron impurities in silicon solar cells and to examine the specific aspects of its implementation. The proposed approach utilizes variations in photovoltaic parameters resulting from FeB pair dissociation as input features for machine learning models. Due to the ability of machine learning algorithms to capture latent correlations and the simplicity of applying trained models, the method is characterized by notable ease of use. The required measurements for generating input data, specifically the acquisition of two current–voltage characteristics or potentially only two short-circuit current values, are fast, non-destructive, do not require prior sample preparation, and can be performed on fully fabricated solar cells without specialized equipment. These features clearly distinguish the proposed method from traditional techniques. An additional advantage is that the necessary equipment is also used for solar cell characterization, which enables straightforward integration of the proposed approach into routine procedures employed in the quality control or certification of photovoltaic devices.

The text has been revised according to the Reviewer's advice — see page 2-3, last paragraph in Section 1.

Comment 2. Explain the Fe concentration measurement procedure and methods followed here for the experimental validation. In section 2.1, further clarification of experimental validation with details of number of cells per data samples, data collection methods and tools used will be helpful.

Reply: We used a set of n^+ - p - p^+ -Si samples for the experimental validation. The structure was fabricated from a 380 μm thick p -type boron-doped Czochralski silicon (100) wafer with a doping level of $N_{\text{B}} = 1.36 \times 10^{15} \text{ cm}^{-3}$. The n^+ -layer, with a sheet resistance of 20 – 30 Ω/\square and a thickness of 0.7 μm , and the p^+ -layer, with a sheet resistance of 10 – 20 Ω/\square and a thickness of 0.6 μm , were formed by phosphorus and boron diffusion, respectively.

The iron concentration was determined using a methodology described in [20, 21], which is based on fitting the kinetics of the short-circuit current following FeB pairs dissociation, measured under low-intensity monochromatic illumination. The kinetics of I_{SC} were measured at 340 K, which allowed tracking the complete recovery of FeB pairs within 1 hour only (the characteristic recovery time being approximately 600 s at 340 K, compared to 13,000 s at 300 K [8, 9]). The determined iron concentration ranged from 2×10^{11} to $4 \times 10^{13} \text{ cm}^{-3}$ across the various solar cells.

To create the experimental test dataset, we measured the I - V characteristics under monochromatic illumination, both before and immediately after FeB pair dissociation. Photovoltaic parameters were extracted from the I - V curves using a standard procedure [22], followed by calculations of their relative changes. For each solar cell, measurements were performed at 1-3 temperatures within the range of 305–340 K.

The similar experimental setup was used both for N_{Fe} determination and experimental test dataset creation. The current-voltage characteristics and I_{sc} kinetics were measured using a Keithley 2450 source meter. The light-emitting diode SN-HPIR940nm-1W with light wavelength 940 nm and intensity of about 5 W/m^2 was used as low-intensity monochromatic light source. The LED radiation intensity was stabilized by a W1209 thermostat and a power supply regulated by a circuit incorporating positive feedback and digital control. The solar cell temperature was driven by a thermoelectric cooler controlled by an STS-21 sensor and maintained constant by a PID algorithm embedded in the software that serves the experimental setup.

A total of 20 solar cells were used for experimental validation, and 30 I - V curves were analyzed. However, at low temperatures and low iron concentrations, the expected relative changes in the fill factor were below 0.5%, making it difficult to determine the ϵFF with sufficient accuracy. As a result, ϵFF could be reliably determined in only 8 cases (corresponding to 5 samples). Consequently, the number of samples in the experimental datasets used to test models trained on different descriptor sets (described in detail in Subsection 2.2) ranged from 30 to 8.

The detailed information can be found in revised manuscript (last five paragraphs in Subsection 2.1).

Comment 3. Table 1 and PCA does not add much value.

Reply: In general, principal component analysis (PCA) is a widely used and effective tool in machine learning, particularly for addressing problems in photovoltaics. For example, it has been successfully applied to fault detection and classification based on I-V curves [16, 15], performance prediction of organic [17] and perovskite [18] solar cells, as well as measurement of the internal quantum efficiency in GaAs solar cells [19].

Despite previous successes in applying principal component analysis to various photovoltaics-related machine learning tasks — and the presence of favorable conditions for its use in our case, such as the high correlation between changes in photoelectric parameters caused by FeB pair dissociation — applying PCA as a data pre-processing step for evaluating iron impurity concentration did not improve the maximum prediction accuracy. One likely reason is that PCA preserves the overall variance in the data but does not account for the specific relationships between individual features and the target variable. For instance, some photovoltaic parameters may be more informative for predicting low iron concentrations, while others may be more relevant for high concentrations. For instance, ϵFF are shown [5] to be most pronounced at low iron concentrations, unlike other PVPs. When these components are merged through PCA transformation, such distinctions may be lost, leading to reduced accuracy. Another possible reason is that principal components — being linear combinations of original features — do not support localized decision splits (i.e., those based on individual feature thresholds), which are particularly effective in tree-based models.

At the same time, we consider it important to document that this commonly used technique was systematically evaluated and found to be less effective than using the original features for iron concentration determination. Conversely, we fully agree that Table 1 does not offer substantial additional value and have therefore excluded it from the revised manuscript.

Comment 4. In section 3.1" 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." And similar discussions about the impact of changing number of features and dimensions is included with respect to different algorithms. It will be good to have some insight into the reasons for this. Also, the physical significance or meaning of it.

Reply: We thank the Reviewer for the valuable comment. We fully agree that discussing potential reasons for the discrepancies in model performance is essential.

We note that this comment is similar in content to Comments 3 and 4 from Reviewer #2. We trust that Reviewer #4 will be understanding that, in response to this remark, we do not repeat the content of previous replies, but instead note that the relevant information has been incorporated into the revised manuscript in Subsection 3.5.

Comment 5. *Include reasons for improvement in prediction with certain combinations of features in the discussion section.*

Reply: First of all, it is important to note that the selection of feature combinations was guided by the degree of correlation between variations in individual photovoltaic parameters and the iron concentration (see Fig. S1 in the Supplementary Material). Specifically, the 4-dimensional feature set included the measurement conditions (T), solar cell characteristics (d_p and N_B), and the PVP most strongly correlated with $\log N_{Fe}$, i.e. ϵI_{SC} . In extending the set to five dimensions, we added the parameter most strongly correlated with N_{Fe} among the remaining unused ones — namely, $\epsilon\eta$ — to the existing features. Subsequent steps followed a similar procedure.

The presence of iron impurities in silicon increases the recombination rate of charge carriers. However, each photovoltaic parameter captures a different aspect of how iron-related recombination influences solar cell performance. Specifically, the short-circuit current reflects the photogenerated current (I_{ph}), which primarily depends on the diffusion length and recombination in the quasi-neutral region. In contrast, the open-circuit voltage is influenced not only by I_{ph} but also by the saturation current (I_0) and the ideality factor (n), both of which are determined by recombination processes in the space-charge region [23]: $V_{OC} = nkT [\ln(I_{ph}/I_0) + 1]$. The fill factor, in turn, is less sensitive to variations in n [24]:

$$FF \approx \frac{v_{OC} - \ln(0.72 + v_{OC})}{v_{OC} + 1}, \quad (2)$$

where v_{OC} being the normalized open-circuit voltage $v_{OC} = qV_{OC}/nkT$. Ultimately, the efficiency integrates all of these effects: $\eta \propto I_{SC} \cdot V_{OC} \cdot FF$.

Incorporating multiple descriptors associated with variations in photovoltaic parameters provides a more comprehensive representation of the effects of iron impurities. This expanded descriptor set increases the informational diversity available for model training, which can enhance predictive accuracy. However, adding more features also raises the risk of overfitting and increases data sparsity, which may hinder the model's ability to generalize. Therefore, simply increasing the number of features does not guarantee improved performance. Instead, optimal predictive accuracy is typically achieved with specific, well-chosen combinations of descriptors.

The rationale for selecting specific feature combinations was added in the third paragraph of Subsection 2.2, while the physical basis for the observed improvement in prediction accuracy with an increased number of descriptors was included in the second paragraph of Subsection 3.5. A more detailed discussion of the adverse effects associated with expanding the feature set is provided in the response to Comment 4 from Reviewer #2 and in the third and fourth paragraphs of Subsection 3.5 of the revised manuscript.

Comment 6. *Would this prediction be applicable beyond the range of N_B , d_p , T that was in the study? why or why not?*

Reply: We want to thank the Referee for the insightful question. To tackle this problem, it is first necessary to consider each parameter separately. The ability of the models to make predictions beyond the range of the training dataset depends on the relationship between the target variable and the descriptors. As previously shown [5], the relationship between the relative changes in the photoelectric parameters and iron concentration can vary significantly depending on the temperature and the concentration of the doping impurity in the base material. This is because a number of parameters, including those of silicon such as bandgap, carrier mobilities, intrinsic recombination coefficients, and Fermi level position, as well as iron-related defects such as carrier capture cross-sections, depend on temperature and

carrier concentration. These relationships, which were taken into account when creating the training set, are sufficiently complex, particularly with respect to N_B , as demonstrated by the results obtained for the N_B -altered dataset. As previously mentioned, even the task of interpolation within the range of N_B used during training was only satisfactorily addressed by the DNN models due to their high ability to identify hidden patterns. Predicting beyond the training range of N_B is even more challenging. Therefore, in our view, the capabilities of the developed models are quite limited.

The temperature-related relationships are simpler, as demonstrated by the reasonably good results obtained using tree-based models for the T -altered dataset (see Fig. 9, Table 2). However, for the training set that included T values beyond the original range, the performance of these models degraded significantly (see Fig. 10). Once again, the best average results were achieved by the deep neural network models (see Fig. 9), and only with their use can relatively accurate predictions beyond the training range of temperature be expected.

The situation is more straightforward concerning the base width, as its influence on the dependence of N_{Fe} on variations in PVPs value is much weaker and often approximately linear [5]. Therefore, in our opinion, the developed models can be expected to remain applicable for predictions beyond the range selected in this study. Naturally, the extension of the base thickness range is only possible within certain limits, until the ratio between d_p and the diffusion length of minority carriers changes significantly.

In summary, we note that if predictions are required for cases outside the range of parameter values used during training, it is advisable to rely exclusively on DNN models. Among the considered parameters, base thickness allows for the largest permissible deviations, whereas for temperature, deviations are unlikely to exceed 10 K. The use of the developed models with doping impurity concentrations different from those in the training set is not recommended. This requirement effectively serves as an additional condition that the training set must satisfy to ensure satisfactory prediction accuracy. It is worth noting that in our study, the parameter ranges were selected to closely correspond to those of real solar cells and their operating or testing conditions.

A detailed discussion of the applicability of the models beyond the selected parameter ranges is provided in Subsection 3.7 (last four paragraphs).

Comment 7. Discuss clearly the advantages and limitations of this model in conclusion. Eg practical significance, is it only for B doped cells only this thickness etc.

Reply: From a practical standpoint, the developed models are applicable for iron quantification in monocrystalline silicon solar cells with an $n^+ - p - p^+$ structure, featuring a boron-doped base and a wide range of base thicknesses (180 – 380 μm) and doping concentrations ($10^{15} – 10^{17} \text{ cm}^{-3}$). The proposed ML-based approach provides a non-destructive, fast, and simple method for determining iron concentrations that does not require specialized equipment. The methodology used to develop models can be adapted for other types of solar cells and/or for quantifying other impurities.

Thanks to the Reviewer's comment, we supplemented Conclusion section (second to last paragraph).

References

- [1] E. Wright, J. Coutinho, S. Öberg, V. J. B. Torres, Mössbauer parameters of fe-related defects in group-iv semiconductors: First principles calculations, *J. Appl. Phys.* 119 (2016) 181509.
- [2] A. A. Istratov, H. Hieslmair, E. Weber, Iron and its complexes in silicon, *Appl. Phys. A: Mater. Sci. Process.* 69 (1999) 13–44.
- [3] J. R. Poindexter, R. L. Z. Hoye, L. Nienhaus, R. C. Kurchin, A. E. Morishige, E. E. Looney, A. Osherov, J.-P. Correa-Baena, B. Lai, V. Bulović, V. Stevanović, M. G. Bawendi, T. Buonassisi, High tolerance to iron contamination in lead halide perovskite solar cells, *ACS Nano* 11 (2017) 7101–7109.
- [4] IEC 61215-2:2021 - Terrestrial photovoltaic (PV) modules – Design qualification and type approval – Part 2: Test procedures, International Standard IEC 61215-2:2021, Geneva, Switzerland, 2021. <https://webstore.iec.ch/publication/61174>.
- [5] 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.
- [6] 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.
- [7] L. J. Geerligs, D. Macdonald, Dynamics of light-induced feb pair dissociation in crystalline silicon, *Appl. Phys. Lett.* 85 (2004) 5227–5229.
- [8] 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.
- [9] N. Khelifati, H. S. Laine, V. Vähäniemi, H. Savin, F. Z. Bouamama, D. Bouhafs, Dissociation and formation kinetics of iron–boron pairs in silicon after phosphorus implantation gettering, *Phys Status Solidi A* 216 (2019) 1900253.

- [10] 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.
- [11] 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.
- [12] J. Lagowski, P. Edelman, A. M. Kontkiewicz, O. Milic, W. Henley, M. Dexter, L. Jastrzebski, A. M. Hoff, Iron detection in the part per quadrillion range in silicon using surface photovoltage and photodissociation of iron–boron pairs, *Appl. Phys. Lett.* 63 (1993) 3043–3045.
- [13] K. Lauer, C. Möller, D. Debbih, M. Auge, D. Schulze, Determination of activation energy of the iron acceptor pair association and dissociation reaction, in: *Gettering and Defect Engineering in Semiconductor Technology XVI*, volume 242 of *Solid State Phenomena*, Trans Tech Publications Ltd, 2016, pp. 230–235.
- [14] O. Olikh, O. Datsenko, S. Kondratenko, Influence of illumination spectrum on dissociation kinetics of iron–boron pairs in silicon, *Phys. Status Solidi A* 221 (2024) 2400351.
- [15] W. Gao, R.-J. Wai, A novel fault identification method for photovoltaic array via convolutional neural network and residual gated recurrent unit, *IEEE Access* 8 (2020) 159493–159510.
- [16] S. Fadhel, C. Delpha, D. Diallo, I. Bahri, A. Migan, M. Trabelsi, M. Mimouni, PV shading fault detection and classification based on I-V curve using principal component analysis: Application to isolated PV system, *Sol. Energy* 179 (2019) 1–10.
- [17] T. W. David, G. A. Soares, N. Bristow, D. Bagnis, J. Kettle, Predicting diurnal outdoor performance and degradation of organic photovoltaics via machine learning: relating degradation to outdoor stress conditions, *Prog. Photovoltaics Res. Appl.* 29 (2021) 1274–1284.
- [18] Y. Liu, W. Yan, S. Han, H. Zhu, Y. Tu, L. Guan, X. Tan, How machine learning predicts and explains the performance of perovskite solar cells, *Sol. RRL* 6 (2022) 2101100.
- [19] 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.
- [20] O. Olikh, V. Kostylyov, V. Vlasiuk, 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.
- [21] O. Olikh, V. Kostylyov, V. Vlasiuk, 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.
- [22] B. Paviet-Salomon, J. Levrat, V. Fakhfouri, Y. Pelet, N. Rebeaud, M. Despeisse, C. Ballif, New guidelines for a more accurate extraction of solar cells and modules key data from their current–voltage curves, *Prog. Photovoltaics Res. Appl.* 25 (2017) 623–635.
- [23] D. Yang (Ed.), *Handbook of Photovoltaic Silicon*, Springer Berlin, Heidelberg, first edition, 2019.
- [24] M. A. Green, Accuracy of analytical expressions for solar cell fill factors, *Sol. Cells* 7 (1982) 337–340.

Rev. 1
Com. 3

Determination 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 R² 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 [3, 30], 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

 olegoliuh@knu.ua (O. Olikh); 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].

Rev. 1

Com. 1

It is well known that ferrous (Fe^{2+}) and ferric (Fe^{3+}) are the most common and stable charge states of iron in most materials. But the majority of iron impurity atoms in silicon occupy interstitial positions, where they can exist in either a neutral (Fe_i^0) or positively charged (Fe_i^+) state, depending on the position of the Fermi level [40, 41]. In *n*-type silicon, Fe_i is more likely to exist in a neutral state, whereas in *p*-type silicon, it is more likely to be positively charged. Even when interstitial iron forms complex point defects no valence bonds are formed. In silicon, iron can also exist in a trivalent (ferric) state when it substitutes for a silicon atom at a lattice site. However, under normal conditions, the concentration of substitutional iron is extremely low, less than 1% of the total iron impurity atoms [42]. The ferrous form is virtually absent in silicon. Consequently, in silicon solar cells, the role of iron valence is negligible, in contrast to other photovoltaic technologies, such as perovskite-based devices [43].

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 [44, 45]. 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 [40, 46, 47, 48, 49, 50, 51, 52].

It is worth noting that the charge-state transitions of iron-related defects — whether during recombination activity or FeB pair dissociation — formally resemble redox reactions.

Rev. 1
Com. 2 However, these processes involve the exchange of charge carriers between defect levels and the conduction or valence band, are governed by Shockley–Read–Hall theory, and are not described by classical redox chemistry.

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 [53], 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 [54, 55, 56, 47, 48, 51, 49];
- (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 [50] and open-circuit voltage [52], but require lengthy experimental procedures or multiple illumination levels.

In our previous work [57], 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 [58, 59]. Integrating these methods with our approach into a unified pipeline would enable the extraction of iron concentration data from luminescence measurements.

Rev. 4

Com. 1 Thus, the primary objective of this study is to introduce a novel methodology for quantifying iron impurities in silicon solar cells and to examine the specific aspects of its implementation. The proposed approach utilizes variations in photovoltaic parameters resulting from FeB pair dissociation as input features for machine learning models. Due to the ability of machine learning algorithms to capture latent correlations and the simplicity of applying trained models, the method is characterized by notable ease of use. The required measurements for generating input data, specifically the acquisition of two current–voltage characteristics or potentially only two short-circuit current values, are fast, non-destructive, do not require prior sample preparation, and can be performed on fully fabricated solar cells.

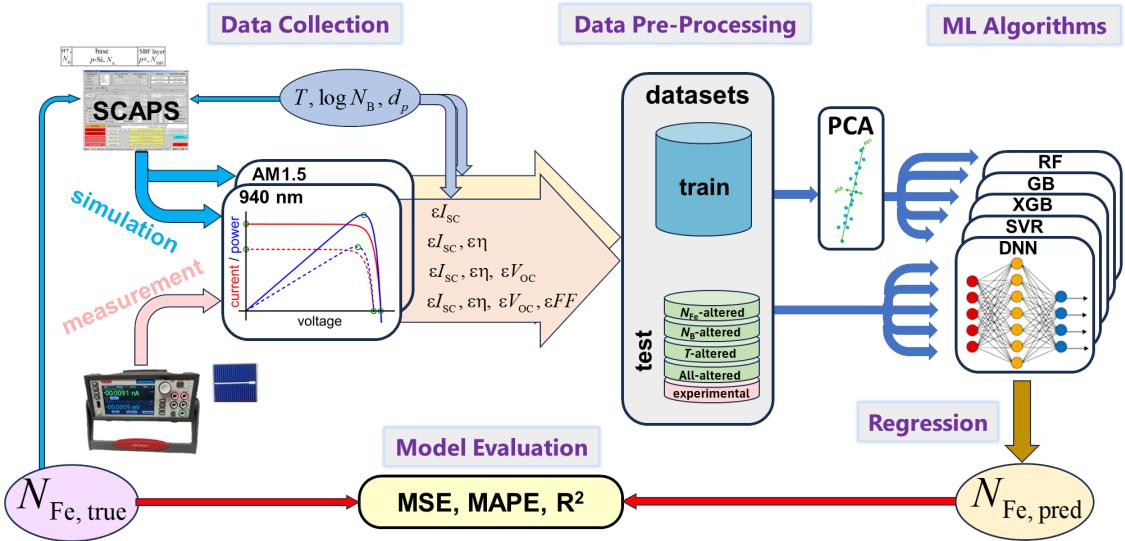


Figure 1: The workflow of the ML pipeline.

without specialized equipment. These features clearly distinguish the proposed method from traditional techniques. An additional advantage is that the necessary equipment is also used for solar cell characterization, which enables straightforward integration of the proposed approach into routine procedures employed in the quality control or certification of photovoltaic devices.

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 (η), and fill factor (FF). Both theoretical and experimental studies [39, 60, 61] have demonstrated that the restructuring of iron-related defects (the dissociation of FeB pairs) leads to reversible 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, ϵI_{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 (ϵV_{OC} , $\epsilon \eta$ and ϵFF) 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.

At the same time, changes in PVPs are shown [60, 61] 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 [62] is a widely used tool [63, 64, 65, 66, 67, 68] 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 ϵI_{SC} , ϵV_{OC} , $\epsilon \eta$, and ϵFF . A detailed description of the modeling approach, including the used temperature dependencies of silicon and defect parameters, is provided elsewhere [61, 69].

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) μm , a boron concentration of (10^{15} – 10^{17}) cm^{-3} , and an iron concentration of (10^{10} – 10^{14}) 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 emphasize that the parameter values used in our study closely reflect those of real solar cells and their operating (testing) conditions. For instance, the selected temperature range corresponds both to the typical operating conditions of most silicon solar cells and to the (15–75) °C range recommended by the IEC 61215-2:2021 standard for determining the temperature coefficients of PVPs.

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 W/m², reflecting the remarkable sensitivity of PVPs to iron [61]. The obtained parameter limits are as follows: ϵI_{SC} ranges from –40.88% to 6.11% under AM1.5 illumination and from –170.58% to 24.66% under monochromatic light; $\epsilon \eta$ ranges from –50.37% to 9.90% (AM1.5) and from –210.32% to 31.53% (monochromatic); ϵV_{OC} ranges from –12.95% to 1.92% (AM1.5) and from –16.33% to 8.15% (monochromatic); and ϵFF ranges from –1.42% to 11.88% (AM1.5) and from –4.39% to 2.98% (monochromatic). 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 μm thick p -type boron-doped Czochralski silicon (100) wafer with a doping level of $N_B = 1.36 \times 10^{15}$ cm^{–3}. The n^+ -layer, with a sheet resistance of 20 – 30 Ω/\square and a thickness of 0.7 μm , and the p^+ -layer, with a sheet resistance of 10 – 20 Ω/\square and a thickness of 0.6 μm , were formed by phosphorus and boron diffusion, respectively. A more detailed description of the sample preparation procedure is available elsewhere [50].

The iron concentration was determined using a methodology described in [70, 50], which is based on fitting the kinetics of the short-circuit current following FeB pairs dissociation, measured under low-intensity monochromatic illumination. The kinetics of I_{SC} were measured at 340 K, which allowed tracking the complete recovery of FeB pairs within 1 hour only (the characteristic recovery time being approximately 600 s at 340 K, compared to 13,000 s at 300 K [45, 71]). The determined iron concentration ranged from 2×10^{11} to 4×10^{13} cm^{–3} across the various solar cells.

To create the experimental test dataset, we measured the I - V characteristics under monochromatic illumination, both before and immediately after FeB pair dissociation. Photovoltaic parameters were extracted from the I - V curves using a standard procedure [72], followed by calculations of their relative changes. For each solar cell, measurements were performed at 1–3 temperatures within the range of 305–340 K.

The similar experimental setup was used both for N_{Fe} determination and experimental test dataset creation. The current-voltage characteristics and I_{SC} kinetics were measured using a Keithley 2450 source meter. The light-emitting diode SN-HPIR940nm-1W with light wavelength 940 nm and intensity of about 5 W/m² was used as low-intensity monochromatic light source. The LED radiation intensity was stabilized by a W1209 thermostat and a power supply regulated by a circuit incorporating positive feedback and digital control. The solar cell temperature was driven by

a thermoelectric cooler controlled by an STS-21 sensor and maintained constant by a PID algorithm embedded in the software that serves the experimental setup. The decay of FeB pairs was induced using intensive halogen lamp illumination (7000 W/m^2). In general, the dissociation rate of FeB pairs depends on the illumination intensity, N_{Fe}

Rev. 2 [45, 51, 73], temperature [74], the spectral composition of illumination [75], and the presence of recombination channels other than those associated with iron-related defects [45, 73]. The selected illumination duration required to achieve near-complete dissociation depended on the iron concentration and reached up to 400 s for the solar cells with the highest N_{Fe} values. The illumination intervals were determined based on the results of a previous study [75], which showed that under similar conditions, complete dissociation at $N_{\text{Fe}} = 9 \times 10^{12} \text{ cm}^{-3}$ occurred within 20 s.

A total of 20 solar cells were used for experimental validation, and 30 $I-V$ curves were analyzed. However, at low temperatures and low iron concentrations, the expected relative changes in the fill factor were below 0.5%, making it difficult to determine the ϵFF with sufficient accuracy. As a result, ϵFF could be reliably determined in only 8 cases (corresponding to 5 samples). Consequently, the number of samples in the experimental datasets used to test models trained on different descriptor sets (described in detail in Subsection 2.2) ranged from 30 to 8.

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 [61], ϵI_{SC} is a key metric for quantifying iron impurities due to its monotonic dependence on iron concentration and relatively large absolute values during the $\text{Fe}_i\text{B}_{\text{Si}} \rightleftharpoons \text{Fe}_i + \text{B}_{\text{Si}}$ transformation. The following most suitable parameters for estimating N_{Fe} are $\epsilon\eta$, ϵV_{OC} , and ϵFF [61].

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 $\log N_{\text{B}}$ and $\log N_{\text{Fe}}$. This approach is standard for quantities that vary over a wide range [76, 77].

Thus, we used the following sets to predict $y = \log N_{\text{Fe}}$: $(T, d_p, N_{\text{B}}, \epsilon I_{\text{SC}})$; $(T, d_p, N_{\text{B}}, \epsilon I_{\text{SC}}, \epsilon\eta)$; $(T, d_p, N_{\text{B}}, \epsilon I_{\text{SC}}, \epsilon\eta, \epsilon V_{\text{OC}})$; $(T, d_p, N_{\text{B}}, \epsilon I_{\text{SC}}, \epsilon\eta, \epsilon V_{\text{OC}}, \epsilon FF)$. 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. The selection of feature combinations was guided by the degree of correlation between variations in individual PVPs and the iron concentration (see Fig. S1 in the Supplementary Material). (see Fig. S1 in the Supplementary Material). Rev. 4 Specifically, the 4-dimensional feature set included the measurement conditions (T), solar cell characteristics (d_p and N_{B}), and the PVP most strongly correlated with $\log N_{\text{Fe}}$, i.e. ϵI_{SC} . In extending the set to five dimensions, we added the parameter most strongly correlated with N_{Fe} among the remaining unused ones — namely, $\epsilon\eta$ — to the existing features. Subsequent steps followed a similar procedure. Com. 5

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). This is not surprising, since each of the PV parameters is linked to the diffusion and recombination of photo-induced charge carriers. Besides, ϵI_{SC} , ϵV_{OC} , $\epsilon\eta$, and ϵFF 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 also trained ML models using features selected via the following procedure: PCs were computed for the original dataset, and the dimensionality was then reduced by discarding PCs whose contribution to the total variance was below 1.5%. As a result, after applying PCA, the feature dimensionality for the different models was reduced from the original sizes of 4, 5, 6, and 7 to 4, 4, 5, and 6 in the AM1.5 case, and to 4, 4, 4, and 5 in the 940 nm case. In general, PCA is a widely used and effective tool in machine learning, particularly for addressing problems in photovoltaics. For example,

Rev. 4
Com. 3 O.Olikh, O.Zavhorodnii: Preprint submitted to Elsevier

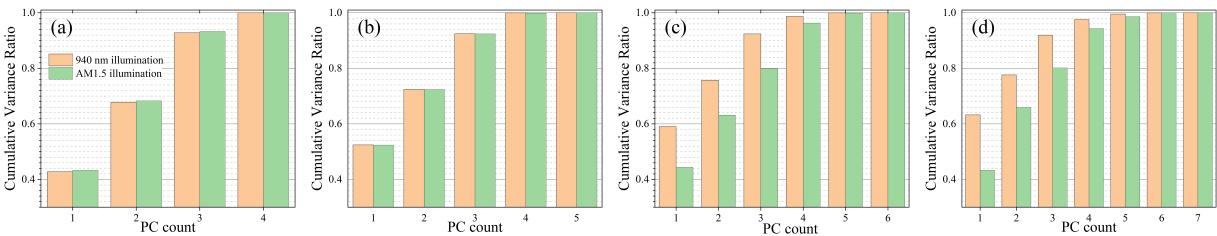


Figure 2: Cumulative explained variance versus number of PC components. Feature dimension: 4 (a), 5 (b), 6 (c), 7 (d).

it has been successfully applied to fault detection and classification based on *I-V* curves [78, 79], performance Rev. 4 prediction of organic [80] and perovskite [81] solar cells, as well as measurement of the internal quantum efficiency Com. 3 in GaAs solar cells [10].

2.3. Machine Learning Algorithms

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

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 [82].

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 [83].

Extreme Gradient Boosting (XGB) is an advanced ensemble method for gradient boosting machines [84]. 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 [85].

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, 57, 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: illum ∈ (AM1.5, 940); “feat” represents the feature dimension and the application of PCA, where feat ∈ (4, 5, 6, 7, 4:PC, 5:PC, 6:PC, 7: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 [86] 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_{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 p_{01} and p_{10} , 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 p_{01} and p_{10} 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_{Fe} , enabling an assessment of the models' predictive capabilities based on measurement conditions, solar cell structure,

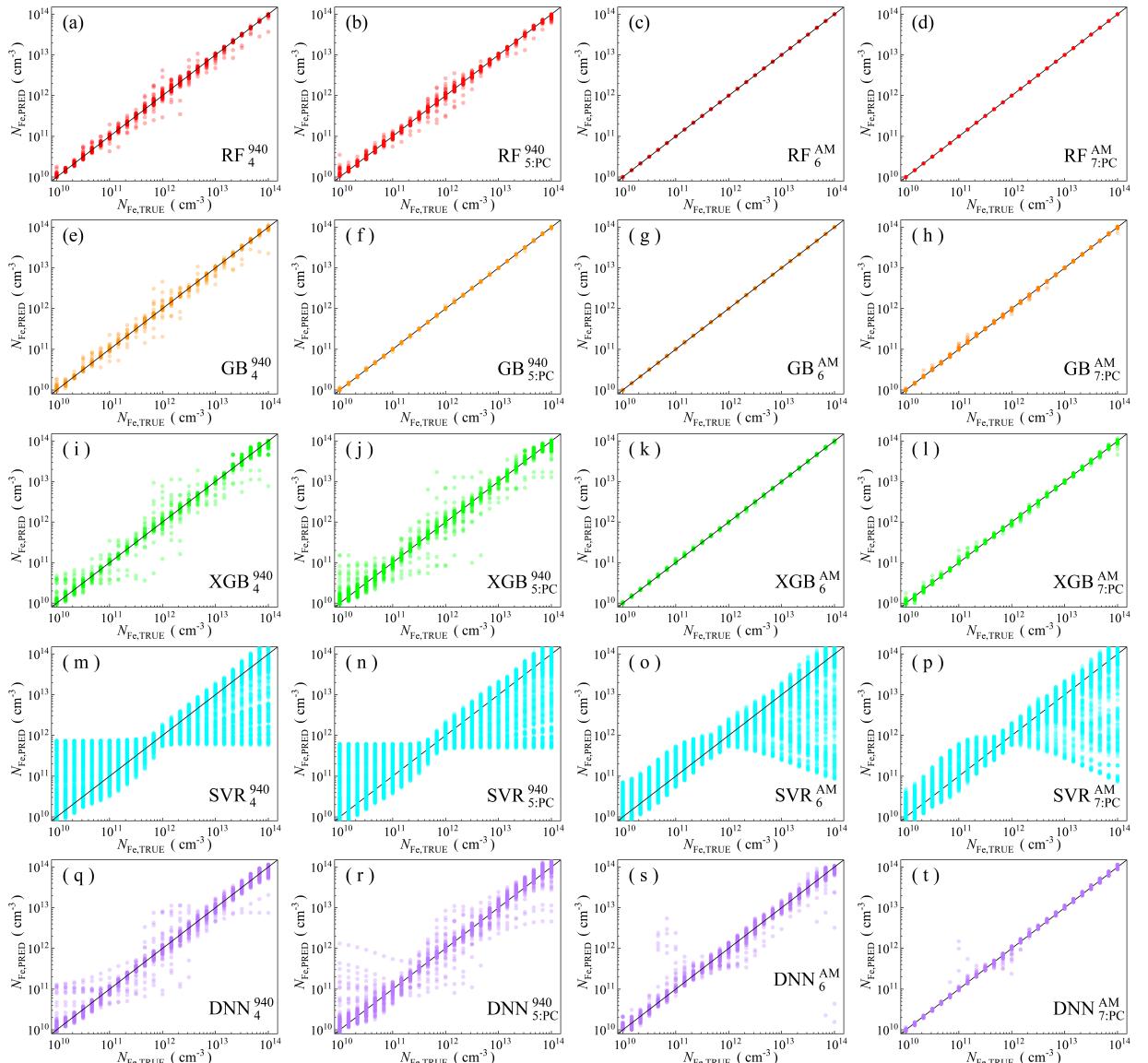


Figure 3: Scatter plots compare reference iron concentrations with ML-predicted values during the training phase. The ML algorithms include RF (a–d), GB (e–h), XGB (i–l), SVR (m–p), and DNN (q–t). The data come from simulation under monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 illumination (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.

and iron concentration. The data indicate that the most challenging cases for prediction involve $N_{Fe} < 10^{11} \text{ cm}^{-3}$ and solar cells with a base doping level around 10^{16} cm^{-3} . The last difficulty arises because, as shown by simulation [61], 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_{Fe} -altered test dataset

The N_{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.

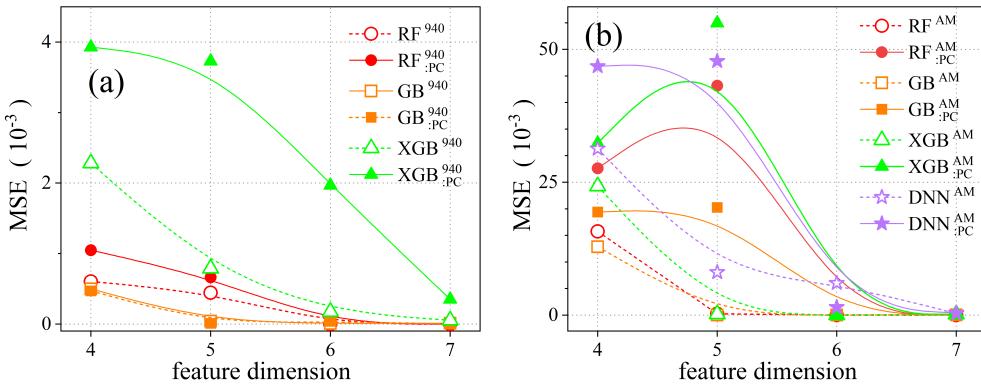


Figure 4: MSE dependence on input feature dimension for training data obtained under monochromatic (a) and AM1.5 illumination (b). ML algorithms: RF (circles), GB (rectangles), XGB (triangles), and DNN (stars). Closed markers indicate results with PCA, while open markers represent results without PCA.

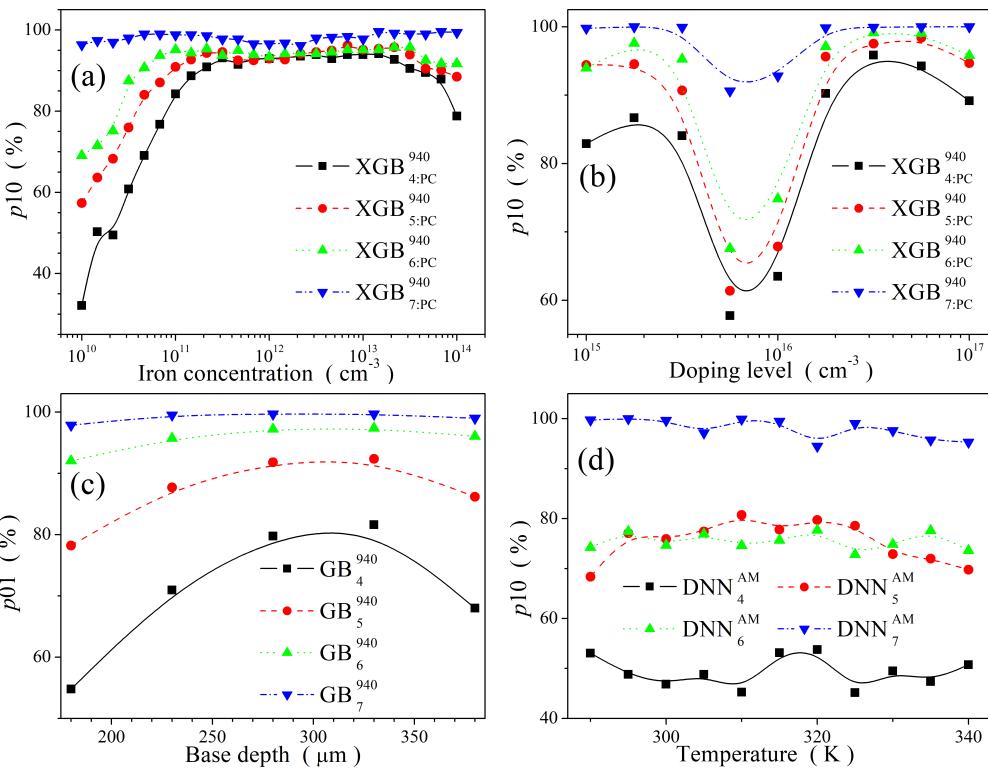


Figure 5: Typical dependencies of the proportion of predictions with an accuracy of at least 10% (a, b, c) or 1% (d) on iron concentration, alloying level, substrate thickness, and temperature.

Rev. 2

Com. 6 Fig. 6 and Figs. S8–S11 in Supplementary Materials present the prediction results, while Fig. 7 and Table 1 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

Rev. 2 940 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

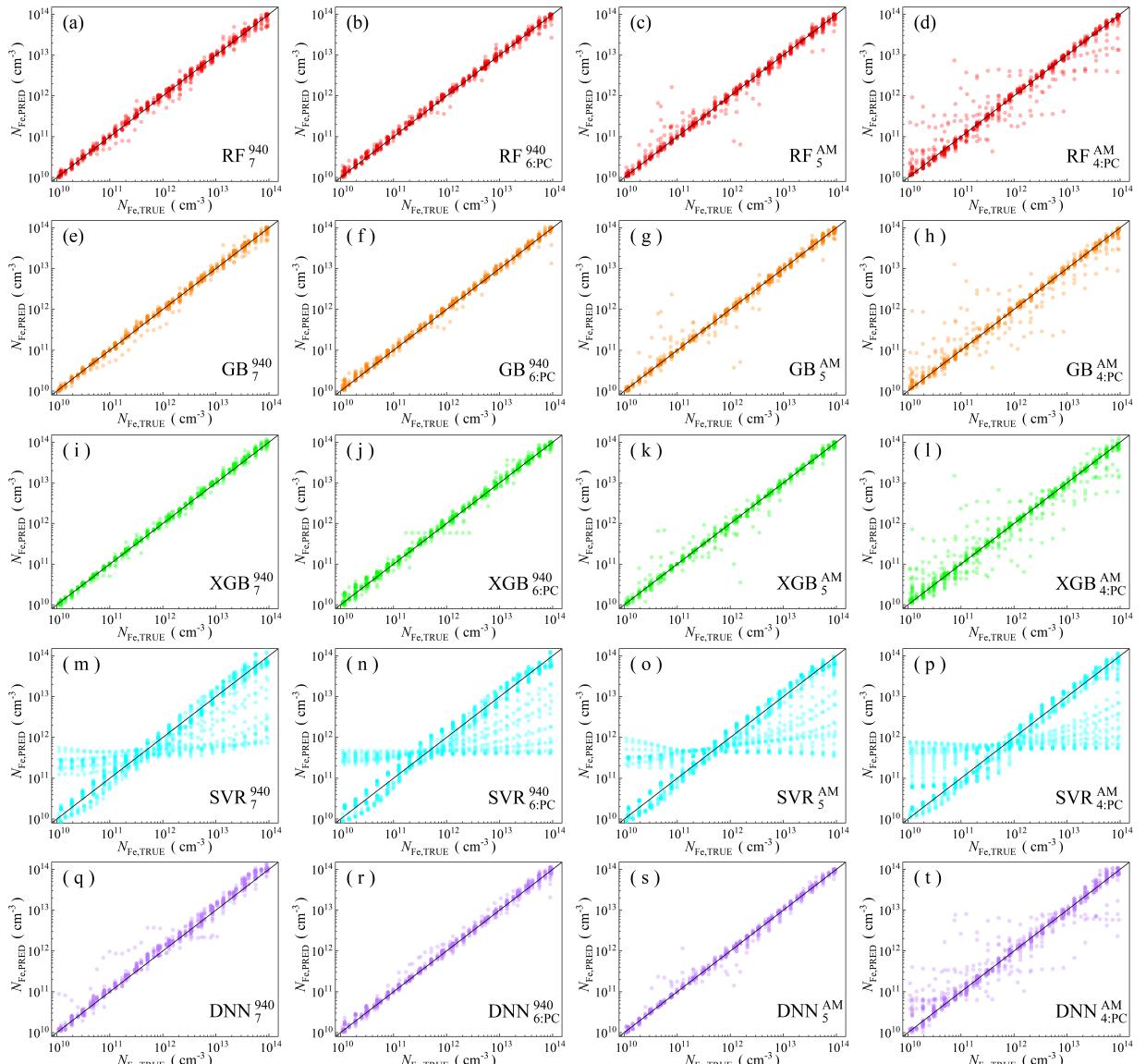


Figure 6: Scatter plots compare reference iron concentrations with ML-predicted values during the N_{Fe} -altered test phase. The ML algorithms include RF (a-d), GB (e-h), XGB (i-l), SVR (m-p), and DNN (q-t). The data come from simulation under monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 illumination (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.

predictions. As shown in Table 1, 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. These trends indicate that expanding feature dimensionality leads to a higher proportion of predictions with significant errors, as illustrated in Fig. 8a for GB. Fig. 8a and Table 1 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, $GB^{940}_{4:PC}$, $GB^{940}_{5:PC}$, and XGB^{940}_5 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

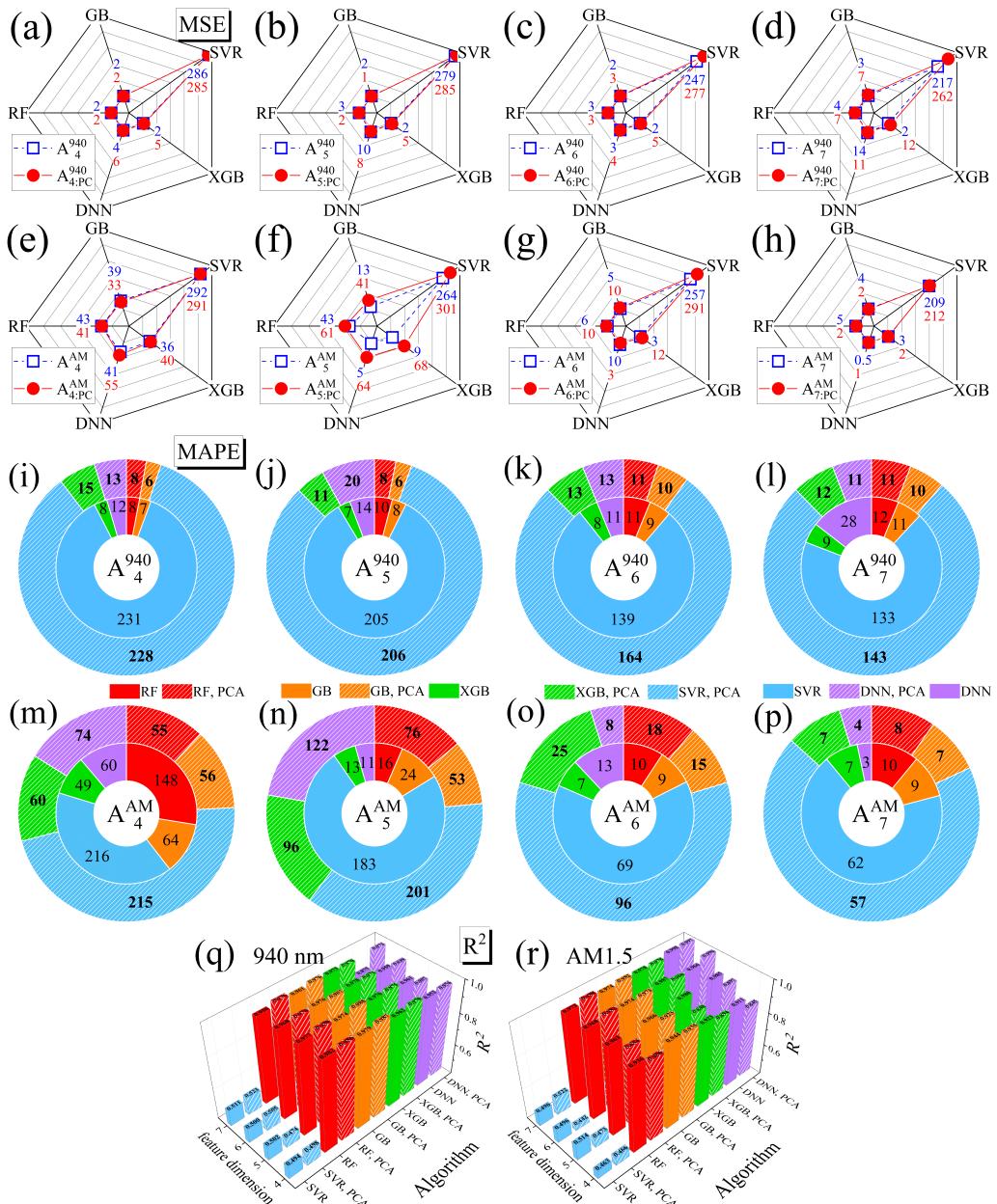
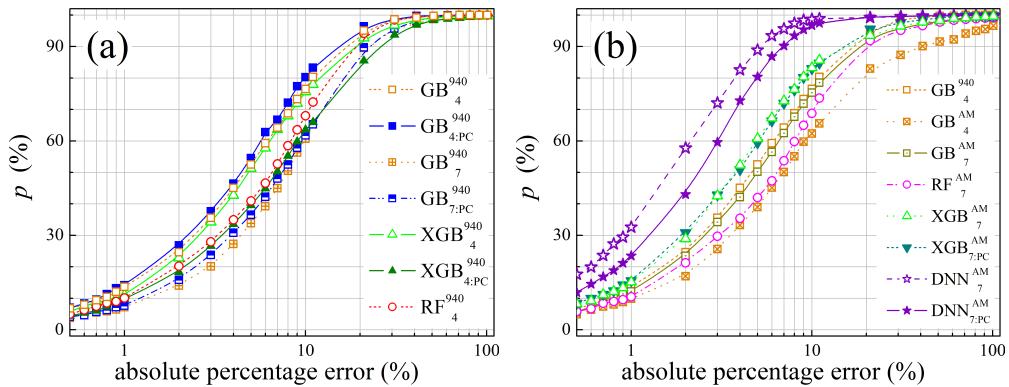


Figure 7: MSE (a–h), MAPE (i–p), and R^2 (q, r) scores for various models, feature combinations, and illumination conditions on the N_{Fe} -altered test dataset. Illumination: 940 nm (a–d, i–l, q), AM1.5 (e–h, m–p, r). Feature dimensions: 4 (a, e, i, m), 5 (b, f, j, n), 6 (c, g, k, o), and 7 (d, h, l, 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.

(increasing the feature dimension from 4 to 5) enables DNN and XGB to predict N_{Fe} with MAPE slightly above 10%, though the results remain inferior to those under monochromatic illumination. The use of PCA significantly worsens the results (see Fig. 7f, Fig. 7n, Table 1). 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. 7p) and the proportion of predictions with large deviations from the true values (Table 1). For XGB and DNN — the most efficient models across the entire

Table 1Performance metrics of the ML models for N_{Fe} -altered dataset

Algorithm	Feature dimension	MdAPE (%)				p01 (%)				p10 (%)			
		940 nm		AM1.5		940 nm		AM1.5		940 nm		AM1.5	
		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	6.00	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	13.6	14.1	9.87	9.38	76.5	80.3	62.4	67.2
	5	4.96	4.63	6.55	8.05	13.6	12.5	11.2	7.25	72.7	82.4	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	4.46	6.20	4.26	9.04	11.7	10.2	13.8	6.58	78.6	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	1.63	2.40	4.06	8.90	32.6	23.5	33.1	64.7	98.7	96.7

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

N_{Fe} -altered test dataset, including data under 940 nm illumination — PCA provides no benefit (Fig. 7p, Fig. 8b). 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.

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.

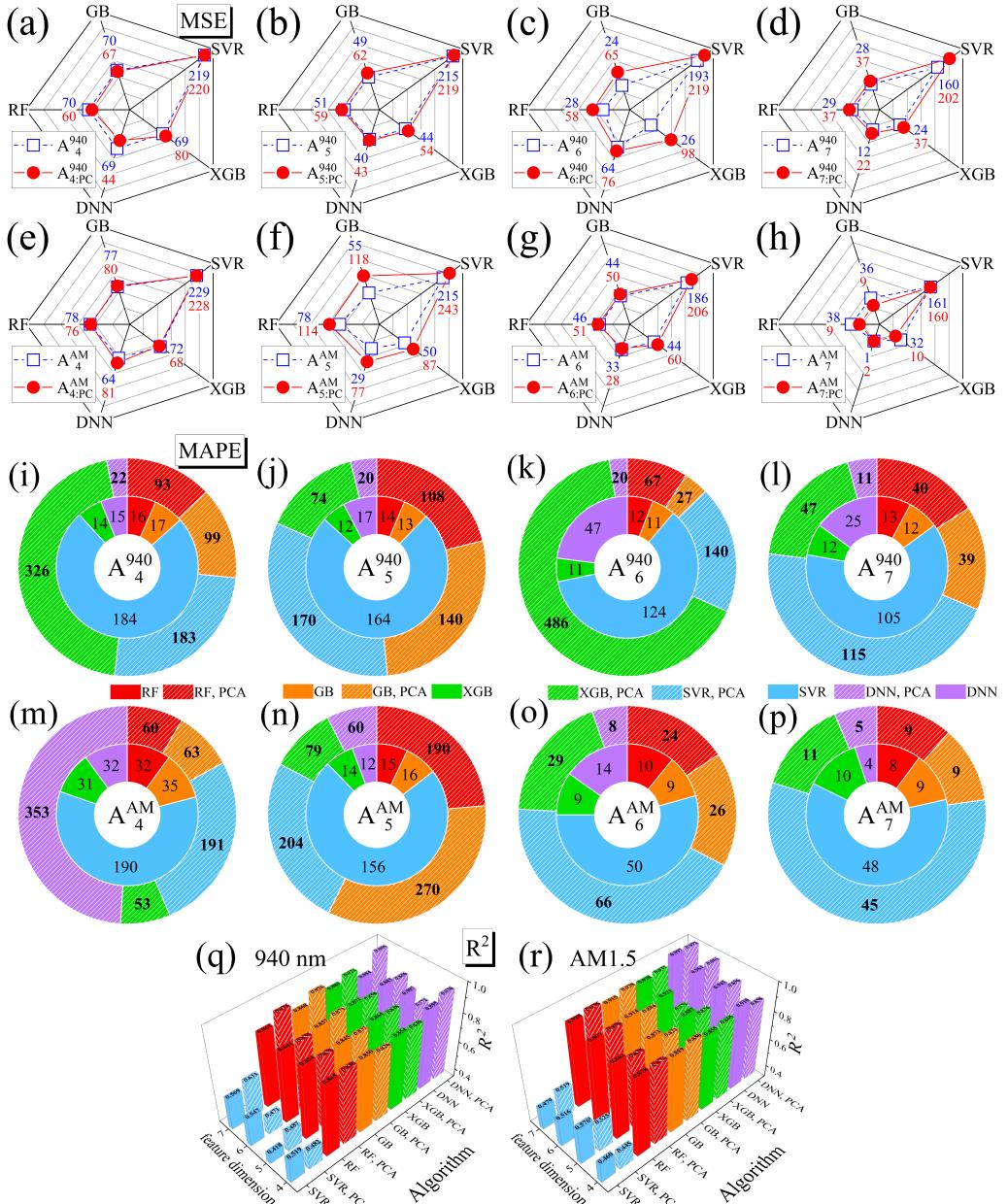


Figure 9: MSE (a–h), MAPE (i–p), and R^2 (q, r) scores for various models, feature combinations, and illumination conditions on the T -altered test dataset. Illumination: 940 nm (a–d, i–l, q), AM1.5 (e–h, m–p, r). Feature dimensions: 4 (a, e, i, m), 5 (b, f, j, n), 6 (c, g, k, o), and 7 (d, h, l, p). 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.

Figs. S12–S15 present the prediction results, while Fig. 9 and Table 2 summarize the performance metrics. The data show that the average metrics for all models worsen slightly compared to the N_{Fe} -altered dataset (Fig. 7, Fig. 9). In most cases, PCA reduces model performance. 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 1 and Table 2 reveals that, unlike the mean characteristics, MdAPE, $p01$, and $p10$ are often better than those in the N_{Fe} -altered dataset. For example, for XGB₆⁹⁴⁰, 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,

Table 2Performance 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	11.6	9.58	18.2	8.92	11.8	12.1	60.4	40.8	45.2	51.3
	5	5.59	13.2	5.90	14.2	22.4	6.75	16.6	9.5	62.9	40.7	62.7	42.1
	6	4.61	13.3	2.13	5.05	20.8	8.58	35.3	17.7	68.8	43.5	77.3	68.3
	7	5.07	9.82	1.94	3.47	21.5	12.0	38.7	21.5	67.4	50.4	79.8	80.6
GB	4	5.29	12.5	10.07	9.34	20.7	6.67	5.08	6.58	62.0	42.9	49.8	51.8
	5	4.91	12.6	4.89	12.7	19.1	7.00	22.2	4.67	66.2	44.7	65.0	70.7
	6	3.41	9.12	1.81	5.14	22.3	11.2	40.6	13.92	73.1	52.4	78.3	70.7
	7	5.03	9.93	2.44	3.40	15.6	9.92	31.3	20.8	68.3	50.3	79.3	80.8
XGB	4	5.03	10.9	8.72	9.88	22.8	8.92	8.42	7.00	64.6	46.9	53.8	50.3
	5	3.65	9.00	5.14	10.5	28.4	8.75	23.7	6.58	67.3	53.0	66.3	48.6
	6	2.47	10.4	2.08	5.78	33.6	6.75	34.4	12.3	73.7	49.2	78.0	69.2
	7	5.37	9.72	3.38	5.11	14.9	10.2	19.3	13.6	67.9	51.0	78.3	72.9
SVR	4	30.9	27.8	36.4	35.7	1.50	1.58	0.67	1.17	15.3	18.3	12.8	13.5
	5	33.7	38.0	37.5	47.6	1.75	1.00	0.92	0.67	15.3	12.9	12.8	9.8
	6	38.3	40.9	37.1	46.2	1.50	1.50	1.25	0.75	14.3	15.0	12.8	9.58
	7	39.3	35.91	36.3	32.2	1.25	1.50	1.83	1.00	10.7	13.3	15.4	14.7
DNN	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	18.6	6.85	1.88	2.53	3.83	7.5	28.1	23.2	29.8	66.3	95.58	92.5

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 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.

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.

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 descriptors 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

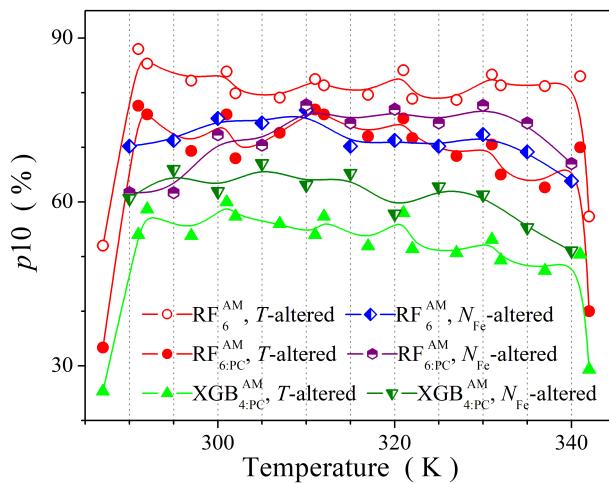


Figure 10: Typical dependencies of the fraction of predictions with an accuracy within 10% as a function of temperature for the T -altered and N_{Fe} -altered datasets. [The vertical lines indicate the temperature values used in the training dataset.]

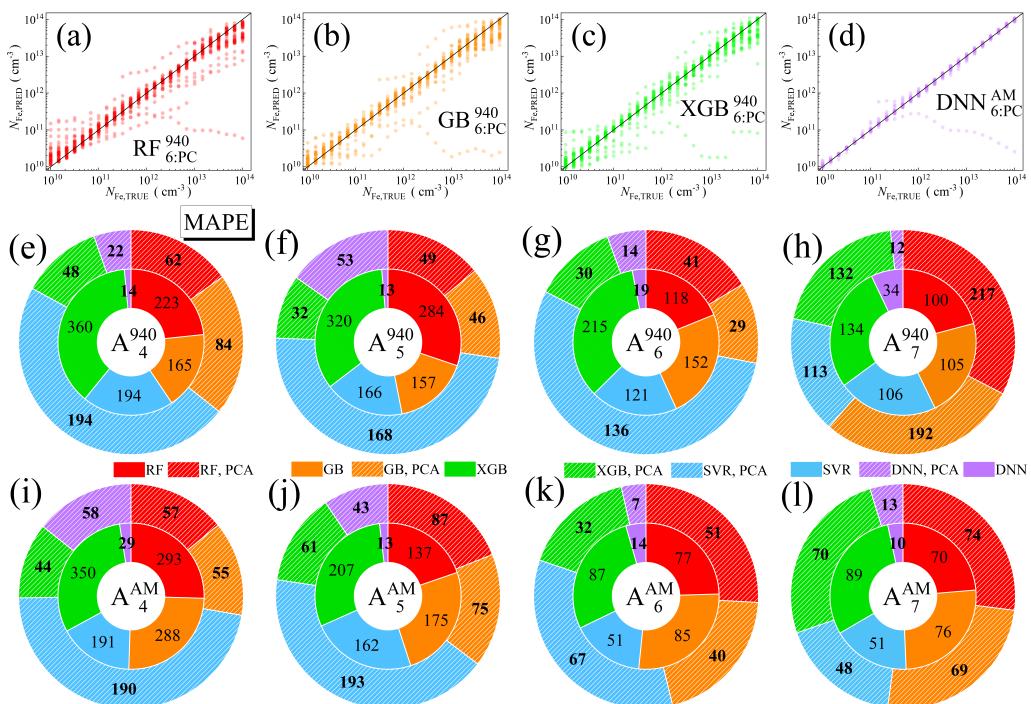


Figure 11: (a)–(d) Scatter plots comparing reference iron concentrations with ML-predicted values for the N_{B} -altered dataset for the best models based on RF (a), GB (b), XGB (c), and DNN (d). (e)–(l) MAPE scores for various models, feature combinations, and illumination conditions. Illumination: 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.

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.

Rev. 2 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.

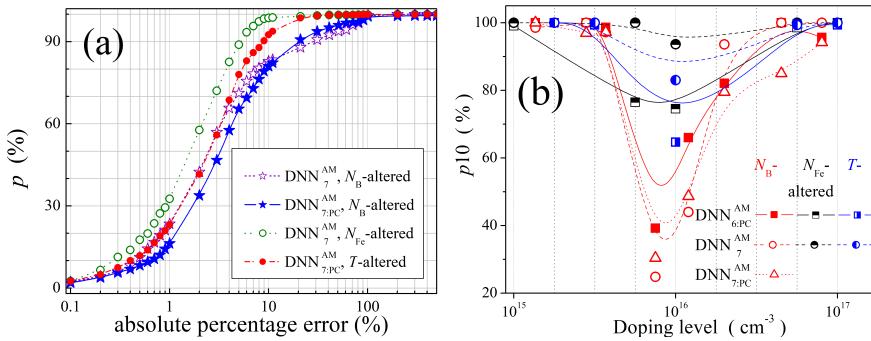


Figure 12: The proportion of samples in the test dataset with an absolute error below a given threshold as a function of the threshold value (a) and the dependencies of the fraction of predictions with an accuracy not exceeding 10% on the doping level for various test datasets (b). The vertical lines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

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 $p10$ 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} 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.

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. Comparative Analysis

Before presenting the results based on experimental data, a few summarizing remarks are in order.

Rev. 4 The presence of iron impurities in silicon increases the recombination rate of charge carriers. However, each
Com. 5 photovoltaic parameter captures a different aspect of how iron-related recombination influences solar cell performance. Specifically, the short-circuit current reflects the photogenerated current (I_{ph}), which primarily depends on the diffusion

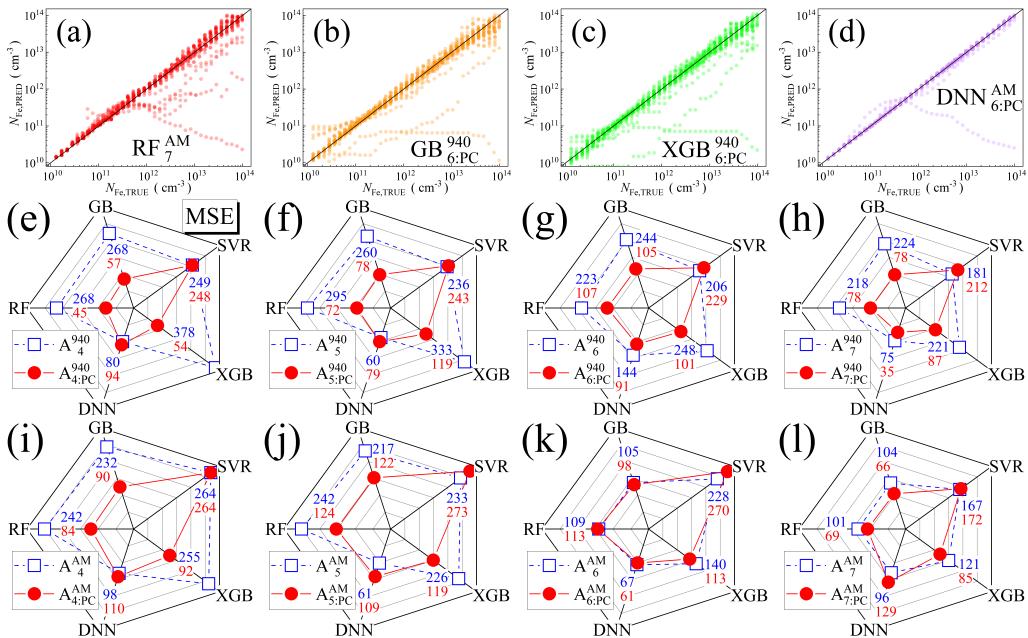


Figure 13: (a)–(d) Scatter plots comparing reference iron concentrations with ML-predicted values for the All-altered test dataset for the best models based on RF (a), GB (b), XGB (c), and DNN (d). (e)–(l) MSE scores for various models, feature combinations, and illumination conditions. Illumination: 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.

Rev. 4 length and recombination in the quasi-neutral region. In contrast, the open-circuit voltage is influenced not only by I_{ph} Com. 5 but also by the saturation current (I_0) and the ideality factor (n), both of which are determined by recombination processes in the space-charge region [87]: $V_{OC} = nkT [\ln (I_{ph}/I_0) + 1]$. The fill factor, in turn, is less sensitive to variations in n [88]:

$$FF \approx \frac{v_{OC} - \ln (0.72 + v_{OC})}{v_{OC} + 1}, \quad (5)$$

where v_{OC} being the normalized open-circuit voltage $v_{OC} = qV_{OC}/nkT$. Ultimately, the efficiency integrates all of these effects: $\eta \propto I_{SC} \cdot V_{OC} \cdot FF$. Incorporating multiple descriptors associated with variations in photovoltaic parameters provides a more comprehensive representation of the effects of iron impurities. This expanded descriptor set increases the informational diversity available for model training, which can enhance predictive accuracy.

One can see that models trained on data obtained under AM1.5 illumination perform worse in almost all cases. Rev. 2 We believe this can be attributed to the following reasons. It is well known that a regression model may achieve high Com. 3 accuracy by having the target variable be an unambiguous function of the descriptor. Moreover, the absolute value of the function's derivative should be as significant as possible. As shown earlier [61], the condition of uniqueness in the relationship between N_{Fe} and the changes in PVPs under AM1.5 illumination is often not fulfilled in contrast to the case of monochromatic illumination. In addition, the ranges of changes in ϵI_{SC} , ϵV_{OC} , and $\epsilon \eta$ under monochromatic illumination exceed those observed under AM1.5 illumination (see Subsection 2.1). The physical origin of the differences in the relationship between iron concentration and parameter changes under AM1.5 and monochromatic illumination lies in the fact that, in the former case, the generation of non-equilibrium charge carriers occurs throughout the entire volume of the solar cell, rather than being limited to the base region. As a result, the processes involved in photovoltaic conversion are more complex. In our view, these differences explain why the performance of ML models trained on AM1.5 illumination data is generally poorer. One way to address the ambiguity in the relationship

between features and the target variable is to incorporate additional features into the model. As it is shown, for AM1.5 illumination, increasing the number of descriptors often improves prediction accuracy, supporting our hypothesis.

Increasing the number of descriptors while keeping the sample size constant leads to higher sparsity in the data space due to increased dimensionality, which in turn makes it more difficult for the model to generalize. In such cases, the model may also learn spurious correlations that do not generalize to test data — a risk that is particularly pronounced in deep neural networks and deep tree-based ensemble methods. Moreover, a larger number of features results in more model parameters, increasing the risk of overfitting. These factors can collectively lead to a reduction in prediction accuracy. In our opinion, for A^{AM1.5} models, both the positive effects achieved by increasing the number of descriptors (enhancing overall information variance and reducing ambiguity between the target variable and the input features) were dominant. However, for A⁹⁴⁰ models where this ambiguity was already low, the negative effects described above outweighed the benefits, which explains why prediction accuracy did not improve when using an expanded feature set. Therefore, simply increasing the number of features does not guarantee improved performance. Instead, optimal predictive accuracy is typically achieved with specific, well-chosen combinations of descriptors.

Another effect that may arise when increasing the number of features is multicollinearity. It occurs when new features are highly correlated with existing ones and can lead to instability in model estimates. However, in our case, it is unlikely that multicollinearity among the photovoltaic parameters (incidentally, their correlation is significantly higher in the 940 nm case than in the AM1.5 case — see Fig. S1) affects prediction accuracy. First, all algorithms used are sufficiently robust to multicollinearity. Second, a standard method for mitigating multicollinearity is PCA. Despite previous successes in applying principal component analysis to various photovoltaics-related machine learning tasks [10, 78, 79, 80, 81] — and the presence of favorable conditions for its use in our case, such as the high correlation between changes in photoelectric parameters caused by FeB pair dissociation — applying PCA as a data pre-processing step for evaluating iron impurity concentration did not improve the maximum prediction accuracy in most cases. One likely reason is that PCA preserves the overall variance in the data but does not account for the specific relationships between individual features and the target variable. For instance, some photovoltaic parameters may be more informative for predicting low iron concentrations, while others may be more relevant for high concentrations. So ϵ_{FF} are shown [61] to be most pronounced at low iron concentrations, unlike other PVPs. When these components are merged through PCA transformation, such distinctions may be lost, leading to reduced accuracy. Another possible reason is that principal components — being linear combinations of original features — do not support localized decision splits (i.e., those based on individual feature thresholds), which are particularly effective in tree-based models.

Rare cases where PCA pre-processing leads to improved prediction accuracy typically occur when the original model (without PCA) performs poorly. In this context, one might consider using PCA as a diagnostic indicator of model quality when evaluating iron concentration in silicon solar cells: if PCA improves prediction accuracy, it suggests that the baseline model is weak; if not, the model is likely to be strong. However, this approach relies on access to the true N_{Fe} values, significantly restricting its practical applicability.

When comparing the effectiveness of different algorithms, several points are worth noting. The poorest performance observed with SVR is likely due to its limited ability to model complex nonlinear relationships — such as those that apparently exist between iron concentration and variations in photoelectric parameters. In contrast, the superior performance of the XGB and DNN models is attributable to their ability to effectively capture nonlinear patterns, which in our case become more prominent with increasing descriptor dimensionality. Specifically, XGB performs best in relatively simpler scenarios, such as those involving 940 nm illumination (see Figs. 7i-7l, 9i-9l). This is because XGB requires less training data than DNN (12,000 examples are sufficient, though this is not a particularly large amount for a neural network), handles high-quality structured data well (simulated data are noise-free), and is more resilient to the influence of correlated features. At the same time, DNNs are more powerful tools for nonlinear modeling, capable of uncovering hidden structures in the data. This enables them to achieve better results on more complex tasks — such as the transition to A^{AM1.5}-models with high-dimensional features (see Figs. 7m-7p, 9m-9p). As shown by the results in Subsection 3.4, successful prediction within the N_B -altered also requires accounting for higher-order relationships, which explains why DNN models achieved the best performance (Figs. 11e-11l). Regarding the RF and GB algorithms, as indicated in Tables 1 and 2, they tend to perform best only on less complex data (i.e., small sets of descriptors), where more powerful models are at higher risk of overfitting.

3.6. 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, $N_{\text{B}} = 1.36 \times 10^{15}\ \text{cm}^{-3}$, 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 $N_{\text{B}} = 1.36 \times 10^{15}\ \text{cm}^{-3}$. 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_{Fe} -altered, T -altered, N_{B} -altered, and All-altered test datasets. In the latter case, we will use the following notation

$$\mathbf{A}_{\text{feat}}^{\text{full}},$$

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 $(T, d_p, N_{\text{B}}, \varepsilon I_{\text{SC}})$ set does not enhance prediction accuracy, at least. Likewise, reducing dimensionality and simplifying the model Rev. 2 using PCA proves ineffective: error reduction occurs only in a few cases (all SVR models, GB⁹⁴⁰₆) and only when the Com. 6 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 DNN⁹⁴⁰₄, aligning Rev. 2 Com. 6 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_{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 3×10^{-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.

3.7. Potential directions for enhancing predictive accuracy

When exploring potential strategies for improving model accuracy, it is essential to consider the efforts already undertaken towards this goal. Specifically, data normalization was applied, multiple descriptor sets were tested, and the impact of reducing feature correlation using PCA was evaluated. In addition, the main model parameters were optimized (see Tables S1–S5 and Figure S2) through bayesian optimization using the Optuna package, and cross-validation was employed.

Among the remaining justified options, the following approaches can be highlighted:

1. Expanding the training dataset, primarily by incorporating a greater number of boron concentration values into the database. This strategy aligns with a fundamental principle of machine learning—the need for as much data as possible.

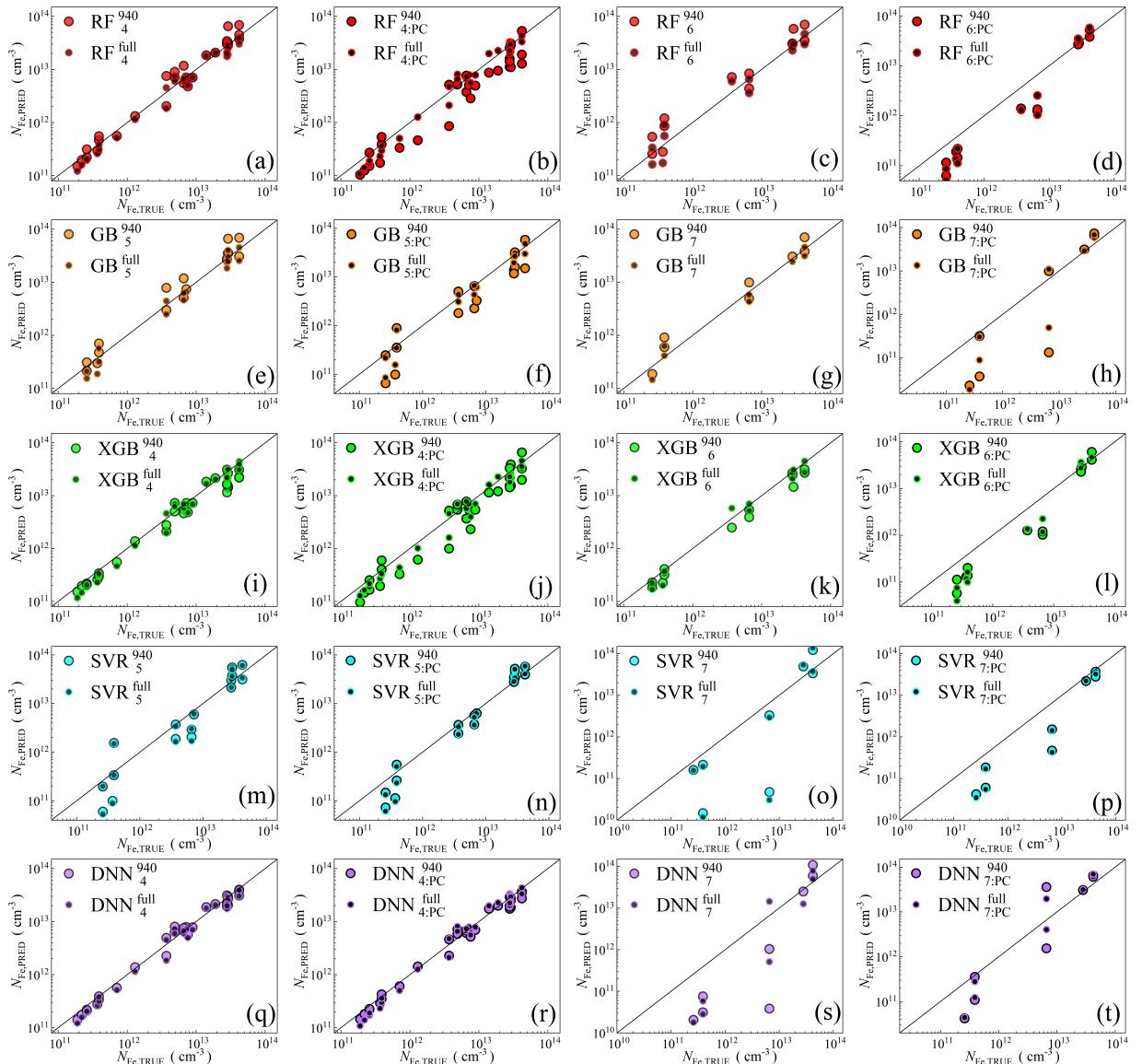


Figure 14: Scatter plots compare reference iron concentrations with ML-predicted values during the experimental test phase. The ML algorithms include RF (a-d), GB (e-h), XGB (i-l), SVR (m-p), and DNN (q-t). Panels (b, d, f, h, j, l, n, p, r, t) incorporate PCA pre-processing. The input feature dimensions 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.

Rev. 2
Com. 5

2. Deriving new descriptors from existing ones that are more directly and simply related to iron concentration. These descriptors should not be limited to linear combinations, such as those used in PCA; instead, their development should be guided by underlying physical processes. Alternatively, generative models, such as variational autoencoders, can be employed to discover latent variables. Notably, this approach may also facilitate easier dataset expansion.

3. Incorporating additional features, especially those obtainable without extra measurements, is desirable because the requirements of additional experiments would substantially reduce the practicality of the proposed method. For example, the minority carrier diffusion length could serve as a valuable descriptor; however, to our knowledge, no methods exist for its determination based solely on single or dual $I-V$ characteristic measurements. The most apparent step in this direction involves the use of relative changes in voltage (εV_{MP}) and current (εI_{MP}), corresponding to the

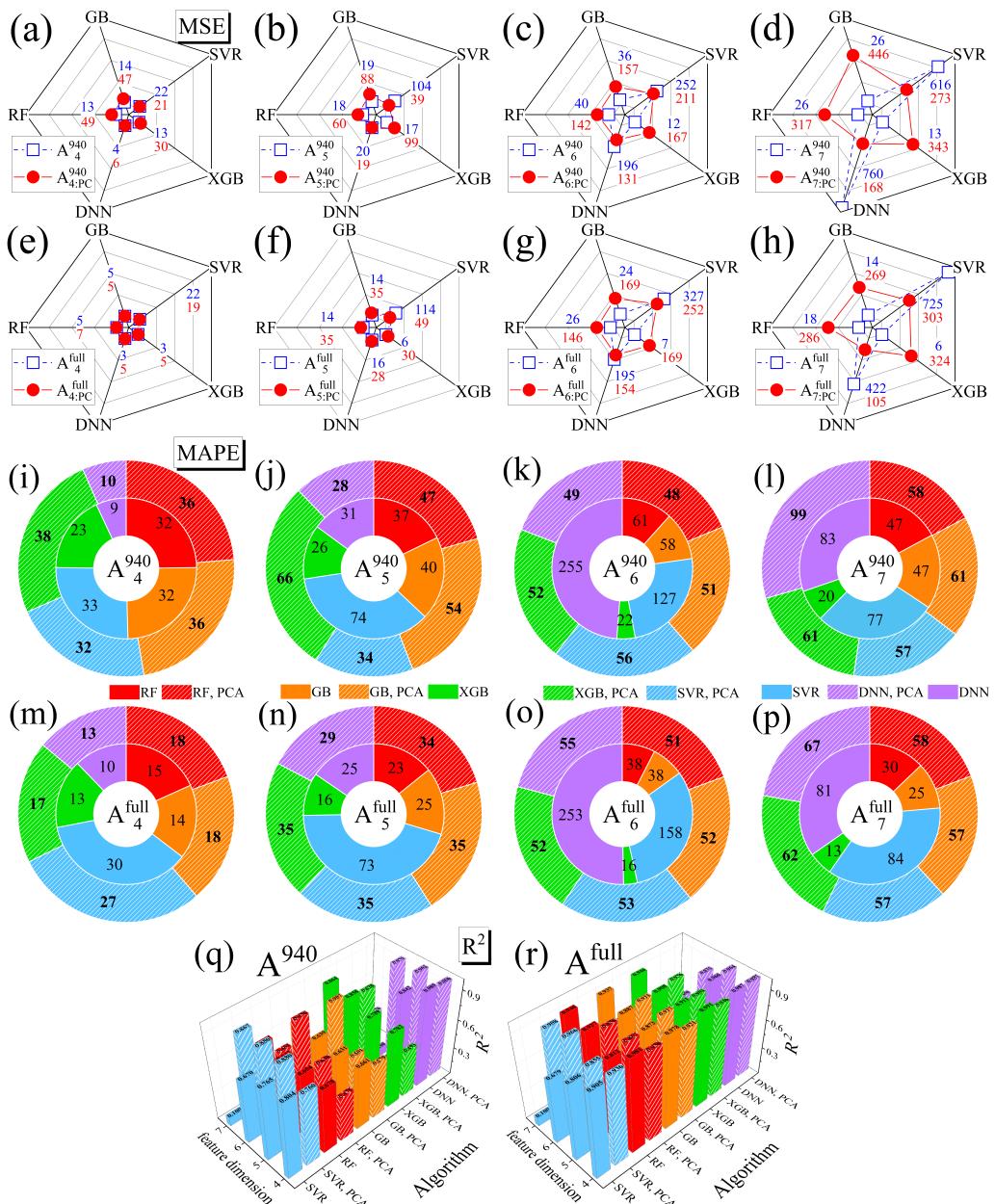


Figure 15: MSE (a–h), MAPE (i–p), and R^2 (q, r) scores for various models and feature sets on the experimental test set. The models used the training dataset (a–d, i–l, q) or the full synthetic dataset (e–h, m–p, r) for training. The feature dimensions are as follows: 4 (a, e, i, m), 5 (b, f, j, n), 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.

maximum power point, as supplementary features. However, it should be noted that these quantities exhibit correlation with other PVPs (see Fig. S31).

Separately, a strategy for improving the quality of the training dataset, which involves increasing the realism of the simulation model, can be distinguished. For example, instead of employing SCAPS, three-dimensional simulators such as Synopsys Sentaurus TCAD, Silvaco ATLAS TCAD, or Solcore can be used to generate labeled datasets that account for the presence of additional impurities commonly found in Cz-Si, such as oxygen and carbon. Furthermore,

incorporating experimental uncertainty into the training data by introducing noise can enhance the model's robustness and predictive accuracy.

To some extent, the issue of accuracy is related to the applicability of the developed models beyond the selected range of N_B , d_p , and T . To tackle this problem, it is first necessary to consider each parameter separately. The ability of the models to make predictions beyond the range of the training dataset depends on the relationship between the target variable and the descriptors. As previously shown [61], the relationship between the relative changes in the photoelectric parameters and iron concentration can vary significantly depending on the temperature and the concentration of the doping impurity in the base material. This is because a number of parameters, including those of silicon such as bandgap, carrier mobilities, intrinsic recombination coefficients, and Fermi level position, as well as iron-related defects such as carrier capture cross-sections, depend on temperature and carrier concentration. These relationships, which were taken into account when creating the training set, are sufficiently complex, particularly with respect to N_B , as demonstrated by the results obtained for the N_B -altered dataset. As previously mentioned, even the task of interpolation within the range of N_B used during training was only satisfactorily addressed by the DNN models due to their high ability to identify hidden patterns. Predicting beyond the training range of N_B is even more challenging. Therefore, in our view, the capabilities of the developed models are quite limited.

The temperature-related relationships are simpler, as demonstrated by the reasonably good results obtained using tree-based models for the T -altered dataset (see Fig. 9, Table 2). However, for the training set that included T values beyond the original range, the performance of these models degraded significantly (see Fig. 10). Once again, the best average results were achieved by the deep neural network models (see Fig. 9), and only with their use can relatively accurate predictions beyond the training range of temperature be expected.

The situation is more straightforward concerning the base width, as its influence on the dependence of N_{Fe} on variations in PVPs value is much weaker and often approximately linear [61]. Therefore, in our opinion, the developed models can be expected to remain applicable for predictions beyond the range selected in this study. Naturally, the extension of the base thickness range is only possible within certain limits, until the ratio between d_p and the diffusion length of minority carriers changes significantly.

In summary, we note that if predictions are required for cases outside the range of parameter values used during training, it is advisable to rely exclusively on DNN models. Among the considered parameters, base thickness allows for the largest permissible deviations, whereas for temperature, deviations are unlikely to exceed 10 K. The use of the developed models with doping impurity concentrations different from those in the training set is not recommended. This requirement effectively serves as an additional condition that the training set must satisfy to ensure satisfactory prediction accuracy.

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.

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 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} cm^{-3} .

The results indicate that eXtreme Gradient Boosting and Deep Neural Network models outperform other approaches, achieving MSE, MAPE, and R² 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 factors contributing to differences in the predictive accuracy of various models are analyzed, and potential measures for its further improvement are proposed.**

Rev. 4 Com. 7 From a practical standpoint, the developed models are applicable for iron quantification in monocrystalline silicon solar cells with an $n^+ - p-p^+$ structure, featuring a boron-doped base and a wide range of base thicknesses (180 – 380 μm) and doping concentrations ($10^{15} - 10^{17} \text{ cm}^{-3}$). The proposed ML-based approach provides a non-destructive, fast, and simple method for determining iron concentrations that does not require specialized equipment. The methodology used to develop models can be adapted for other types of solar cells and/or for quantifying other impurities.

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

Oleg Olikh 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-Otum, 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, Psopruner: 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. Thamrongsrirapak, 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] 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.
- [41] A. A. Istratov, H. Hieslmair, E. Weber, Iron and its complexes in silicon, *Appl. Phys. A: Mater. Sci. Process.* 69 (1999) 13–44.
- [42] E. Wright, J. Coutinho, S. Öberg, V. J. B. Torres, Mössbauer parameters of fe-related defects in group-iv semiconductors: First principles calculations, *J. Appl. Phys.* 119 (2016) 181509.
- [43] J. R. Poindexter, R. L. Z. Hoye, L. Nienhaus, R. C. Kurchin, A. E. Morishige, E. E. Looney, A. Osherov, J.-P. Correa-Baena, B. Lai, V. Bulović, V. Stevanović, M. G. Bawendi, T. Buonassisi, High tolerance to iron contamination in lead halide perovskite solar cells, *ACS Nano* 11 (2017) 7101–7109.
- [44] L. Kimerling, J. Benton, Electronically controlled reactions of interstitial iron in silicon, *Physica B+C* 116 (1983) 297–300.
- [45] 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.
- [46] G. Zoth, W. Bergholz, A fast, preparation-free method to detect iron in silicon, *J. Appl. Phys.* 67 (1990) 6764–6771.
- [47] 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.
- [48] J. Schmidt, D. Macdonald, Recombination activity of iron-gallium and iron-indium pairs in silicon, *J. Appl. Phys.* 97 (2005) 113712.

- [49] 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.
- [50] O. Olikh, V. Kostylyov, V. Vlasiuk, 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.
- [51] 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.
- [52] A. Herguth, Quantification of iron in boron-doped silicon solar cells from open circuit voltage measurements, *IEEE J. Photovolt.* 12 (2022) 937–947.
- [53] M. Di Sabatino, Detection limits for glow discharge mass spectrometry (gdms) analyses of impurities in solar cell silicon, *Measurement* 50 (2014) 135–140.
- [54] D. K. Schroder, *Semiconductor Material and Device Characterization*, John Wiley & Sons, New Jersey, third edition, 2006.
- [55] R. C. Kurchin, J. R. Poindexter, V. Vähäniemi, 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.
- [56] A. R. Peaker, V. P. Markevich, J. Coutinho, Tutorial: Junction spectroscopy techniques and deep-level defects in semiconductors, *J. Appl. Phys.* 123 (2018) 161559.
- [57] 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.
- [58] P. Kunze, J. M. Greulich, A. Tummalieh, W. Wirtz, H. Hoeffler, N. Woehrle, S. Glunz, S. Rein, M. Demant, Contactless inline iv measurement of solar cells using an empirical model, *Sol. RRL* 7 (2023) 2200599.
- [59] 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.
- [60] 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.
- [61] 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.
- [62] M. Burgelman, P. Nollet, S. Degrave, Modelling polycrystalline semiconductor solar cells, *Thin Solid Films* 361–362 (2000) 527–532.
- [63] M. Masum Mia, M. Faruk Hossain, M. Rahman, N. Badi, A. Irfan, M. Ferdous Rahman, Unveiling the impact of Se based HTM on BaZrSe₃ perovskites solar cell and improving the theoretical efficiency above 32%, *Materials Science and Engineering: B* 311 (2025) 117817.
- [64] T. K. Joshi, G. Sharma, Y. R. Sharma, A. S. Verma, Spectroscopic screening and performance parameters of hybrid perovskite (CH₃CH₂PH₃PbI₃) using WIEN2k and SCAPS-1d, *Phys. B Condens. Matter* 682 (2024) 415793.
- [65] 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/CsGeI₃ perovskite tandem solar cells using a combination of density functional theory and SCAPS-1D frameworks, *Sol. Energ. Mat. Sol.* 267 (2024) 112688.
- [66] 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.
- [67] L. You, X. Zhang, Q. Ma, W. Zhu, J. Wu, Optimization of electron transport layer-free Cs₂TiBr₆/MASnBr₃ laminated structure perovskite solar cells by SCAPS-1D simulation, *Phys. Status Solidi A* 220 (2023) 2300071.
- [68] M. Abdulmalik, E. Danladji, Influence of perovskite thickness on the performance of silver-doped NaZnBr₃ perovskite solar cells using SCAPS software, *Semiconductor Physics, Quantum Electronics & Optoelectronics* 26 (2023) 321–331.
- [69] O. Olikh, Relationship between the ideality factor and the iron concentration in silicon solar cells, *Superlattices Microstruct.* 136 (2019) 106309.
- [70] O. Olikh, V. Kostylyov, V. Vlasiuk, 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.
- [71] N. Khelifati, H. S. Laine, V. Vähäniemi, H. Savin, F. Z. Bouamama, D. Bouhafs, Dissociation and formation kinetics of iron–boron pairs in silicon after phosphorus implantation gettering, *Phys Status Solidi A* 216 (2019) 1900253.
- [72] B. Paviet-Salomon, J. Levrat, V. Fakhfouri, Y. Pelet, N. Rebeaud, M. Despeisse, C. Ballif, New guidelines for a more accurate extraction of solar cells and modules key data from their current–voltage curves, *Prog. Photovoltaics Res. Appl.* 25 (2017) 623–635.
- [73] L. J. Geerligs, D. Macdonald, Dynamics of light-induced feb pair dissociation in crystalline silicon, *Appl. Phys. Lett.* 85 (2004) 5227–5229.
- [74] K. Lauer, C. Möller, D. Debbih, M. Auge, D. Schulze, Determination of activation energy of the iron acceptor pair association and dissociation reaction, in: *Gettering and Defect Engineering in Semiconductor Technology XVI*, volume 242 of *Solid State Phenomena*, Trans Tech Publications Ltd, 2016, pp. 230–235.
- [75] O. Olikh, O. Datsenko, S. Kondratenko, Influence of illumination spectrum on dissociation kinetics of iron–boron pairs in silicon, *Phys. Status Solidi A* 221 (2024) 2400351.
- [76] Y. Srivastava, A. Jain, End-to-end material thermal conductivity prediction through machine learning, *J. Appl. Phys.* 134 (2023) 225101.
- [77] 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.
- [78] S. Fadhel, C. Delpha, D. Diallo, I. Bahri, A. Migan, M. Trabelsi, M. Mimouni, PV shading fault detection and classification based on I-V curve using principal component analysis: Application to isolated PV system, *Sol. Energy* 179 (2019) 1–10.
- [79] W. Gao, R.-J. Wai, A novel fault identification method for photovoltaic array via convolutional neural network and residual gated recurrent unit, *IEEE Access* 8 (2020) 159493–159510.
- [80] T. W. David, G. A. Soares, N. Bristow, D. Bagnis, J. Kettle, Predicting diurnal outdoor performance and degradation of organic photovoltaics via machine learning: relating degradation to outdoor stress conditions, *Prog. Photovoltaics Res. Appl.* 29 (2021) 1274–1284.

- [81] Y. Liu, W. Yan, S. Han, H. Zhu, Y. Tu, L. Guan, X. Tan, How machine learning predicts and explains the performance of perovskite solar cells, Sol. RRL 6 (2022) 2101100.
- [82] L. Breiman, Random forests, Mach. Learn. 45 (2001) 5–32.
- [83] A. Natekin, A. Knoll, Gradient boosting machines, a tutorial, Front. Neurorob. 7 (2013).
- [84] S. Akinpelu, S. Abolade, E. Okafor, D. Obada, A. Ukpong, S. Kumar R., J. Healy, A. Akande, Interpretable machine learning methods to predict the mechanical properties of abx₃ perovskites, Results Phys. 65 (2024) 107978.
- [85] W. Cao, X. Liu, J. Ni, Parameter optimization of support vector regression using henry gas solubility optimization algorithm, IEEE Access 8 (2020) 88633–88642.
- [86] 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.
- [87] D. Yang (Ed.), Handbook of Photovoltaic Silicon, Springer Berlin, Heidelberg, first edition, 2019.
- [88] M. A. Green, Accuracy of analytical expressions for solar cell fill factors, Sol. Cells 7 (1982) 337–340.

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 R² 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 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), cXtreme 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 R² 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 [3, 30], 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

*olegolikh@knu.ua (O. Olikh); nevermore464@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 known that ferrous (Fe^{2+}) and ferric (Fe^{3+}) are the most common and stable charge states of iron in most materials. But the majority of iron impurity atoms in silicon occupy interstitial positions, where they can exist in either a neutral (Fe_i^0) or positively charged (Fe_i^+) state, depending on the position of the Fermi level [40, 41]. In *n*-type silicon, Fe_i is more likely to exist in a neutral state, whereas in *p*-type silicon, it is more likely to be positively charged. Even when interstitial iron forms complex point defects no valence bonds are formed. In silicon, iron can also exist in a trivalent (ferric) state when it substitutes for a silicon atom at a lattice site. However, under normal conditions, the concentration of substitutional iron is extremely low, less than 1% of the total iron impurity atoms [42]. The ferrous form is virtually absent in silicon. Consequently, in silicon solar cells, the role of iron valence is negligible, in contrast to other photovoltaic technologies, such as perovskite-based devices [43].

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 [44, 45]. 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 [40, 46, 47, 48, 49, 50, 51, 52]. It is worth noting that the charge-state transitions of iron-related defects — whether during recombination activity or FeB pair dissociation — formally resemble redox reactions. However, these processes involve the exchange of charge carriers between defect levels and the conduction or valence band, are governed by Shockley–Read–Hall theory, and are not described by classical redox chemistry.

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 [53], 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 [54, 55, 56, 47, 48, 51, 49];
- (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 [50] and open-circuit voltage [52], but require lengthy experimental procedures or multiple illumination levels.

In our previous work [57], 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 [58, 59]. Integrating these methods with our approach into a unified pipeline would enable the extraction of iron concentration data from luminescence measurements.

Thus, the primary objective of this study is to introduce a novel methodology for quantifying iron impurities in silicon solar cells and to examine the specific aspects of its implementation. The proposed approach utilizes variations in photovoltaic parameters resulting from FeB pair dissociation as input features for machine learning models. Due to the ability of machine learning algorithms to capture latent correlations and the simplicity of applying trained models, the method is characterized by notable ease of use. The required measurements for generating input data, specifically the acquisition of two current–voltage characteristics or potentially only two short-circuit current values, are fast, non-destructive, do not require prior sample preparation, and can be performed on fully fabricated solar cells without specialized equipment. These features clearly distinguish the proposed method from traditional techniques. An additional advantage is that the necessary equipment is also used for solar cell characterization, which enables

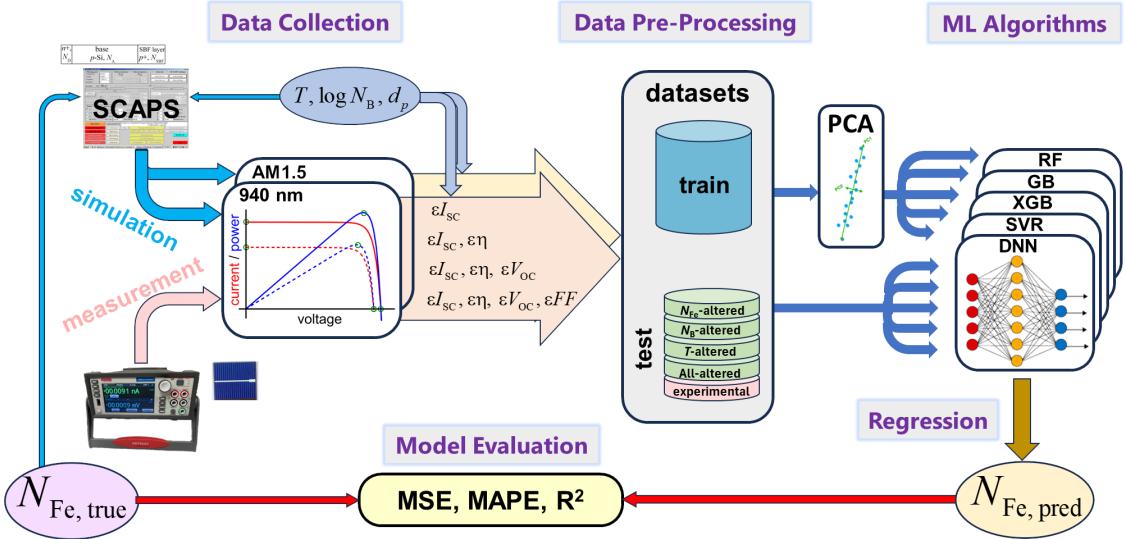


Figure 1: The workflow of the ML pipeline.

straightforward integration of the proposed approach into routine procedures employed in the quality control or certification of photovoltaic devices.

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 (η), and fill factor (FF). Both theoretical and experimental studies [39, 60, 61] have demonstrated that the restructuring of iron-related defects (the dissociation of FeB pairs) leads to reversible 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, εI_{SC} , were calculated as

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

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 (εV_{OC} , $\varepsilon \eta$ and εFF) 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.

At the same time, changes in PVPs are shown [60, 61] 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 [62] is a

widely used tool [63, 64, 65, 66, 67, 68] 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 εI_{SC} , εV_{OC} , $\varepsilon\eta$, and εFF . A detailed description of the modeling approach, including the used temperature dependencies of silicon and defect parameters, is provided elsewhere [61, 69].

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) μm , a boron concentration of ($10^{15} – 10^{17}$) cm^{-3} , and an iron concentration of ($10^{10} – 10^{14}$) 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 emphasize that the parameter values used in our study closely reflect those of real solar cells and their operating (testing) conditions. For instance, the selected temperature range corresponds both to the typical operating conditions of most silicon solar cells and to the (15–75) °C range recommended by the IEC 61215-2:2021 standard for determining the temperature coefficients of PVPs.

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 W/m^2 , reflecting the remarkable sensitivity of PVPs to iron [61]. The obtained parameter limits are as follows: εI_{SC} ranges from –40.88% to 6.11% under AM1.5 illumination and from –170.58% to 24.66% under monochromatic light; $\varepsilon\eta$ ranges from –50.37% to 9.90% (AM1.5) and from –210.32% to 31.53% (monochromatic); εV_{OC} ranges from –12.95% to 1.92% (AM1.5) and from –16.33% to 8.15% (monochromatic); and εFF ranges from –1.42% to 11.88% (AM1.5) and from –4.39% to 2.98% (monochromatic). 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 μm thick p -type boron-doped Czochralski silicon (100) wafer with a doping level of $N_B = 1.36 \times 10^{15} \text{ cm}^{-3}$. The n^+ -layer, with a sheet resistance of 20 – 30 Ω/\square and a thickness of 0.7 μm , and the p^+ -layer, with a sheet resistance of 10 – 20 Ω/\square and a thickness of 0.6 μm , were formed by phosphorus and boron diffusion, respectively. A more detailed description of the sample preparation procedure is available elsewhere [50].

The iron concentration was determined using a methodology described in [70, 50], which is based on fitting the kinetics of the short-circuit current following FeB pairs dissociation, measured under low-intensity monochromatic illumination. The kinetics of I_{SC} were measured at 340 K, which allowed tracking the complete recovery of FeB pairs within 1 hour only (the characteristic recovery time being approximately 600 s at 340 K, compared to 13,000 s at 300 K [45, 71]). The determined iron concentration ranged from 2×10^{11} to $4 \times 10^{13} \text{ cm}^{-3}$ across the various solar cells.

To create the experimental test dataset, we measured the I - V characteristics under monochromatic illumination, both before and immediately after FeB pair dissociation. Photovoltaic parameters were extracted from the I - V curves using a standard procedure [72], followed by calculations of their relative changes. For each solar cell, measurements were performed at 1-3 temperatures within the range of 305–340 K.

The similar experimental setup was used both for N_{Fe} determination and experimental test dataset creation. The current-voltage characteristics and I_{SC} kinetics were measured using a Keithley 2450 source meter. The light-emitting diode SN-HPIR940nm-1W with light wavelength 940 nm and intensity of about 5 W/m^2 was used as low-intensity monochromatic light source. The LED radiation intensity was stabilized by a W1209 thermostat and a power supply regulated by a circuit incorporating positive feedback and digital control. The solar cell temperature was driven by a thermoelectric cooler controlled by an STS-21 sensor and maintained constant by a PID algorithm embedded in the software that serves the experimental setup. The decay of FeB pairs was induced using intensive halogen lamp illumination (7000 W/m^2). In general, the dissociation rate of FeB pairs depends on the illumination intensity, N_{Fe} [45, 51, 73], temperature [74], the spectral composition of illumination [75], and the presence of recombination

channels other than those associated with iron-related defects [45, 73]. The selected illumination duration required to achieve near-complete dissociation depended on the iron concentration and reached up to 400 s for the solar cells with the highest N_{Fe} values. The illumination intervals were determined based on the results of a previous study [75], which showed that under similar conditions, complete dissociation at $N_{\text{Fe}} = 9 \times 10^{12} \text{ cm}^{-3}$ occurred within 20 s.

A total of 20 solar cells were used for experimental validation, and 30 I - V curves were analyzed. However, at low temperatures and low iron concentrations, the expected relative changes in the fill factor were below 0.5%, making it difficult to determine the ϵFF with sufficient accuracy. As a result, ϵFF could be reliably determined in only 8 cases (corresponding to 5 samples). Consequently, the number of samples in the experimental datasets used to test models trained on different descriptor sets (described in detail in Subsection 2.2) ranged from 30 to 8.

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 [61], ϵI_{SC} is a key metric for quantifying iron impurities due to its monotonic dependence on iron concentration and relatively large absolute values during the $\text{Fe}_i\text{B}_{\text{Si}} \rightleftharpoons \text{Fe}_i + \text{B}_{\text{Si}}$ transformation. The following most suitable parameters for estimating N_{Fe} are $\epsilon\eta$, ϵV_{OC} , and ϵFF [61].

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 $\log N_{\text{Fe}}$. This approach is standard for quantities that vary over a wide range [76, 77].

Thus, we used the following sets to predict $y = \log N_{\text{Fe}}$: $(T, d_p, N_B, \epsilon I_{\text{SC}})$; $(T, d_p, N_B, \epsilon I_{\text{SC}}, \epsilon\eta)$; $(T, d_p, N_B, \epsilon I_{\text{SC}}, \epsilon\eta, \epsilon V_{\text{OC}})$; $(T, d_p, N_B, \epsilon I_{\text{SC}}, \epsilon\eta, \epsilon V_{\text{OC}}, \epsilon FF)$. 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. The selection of feature combinations was guided by the degree of correlation between variations in individual PVPs and the iron concentration (see Fig. S1 in the Supplementary Material). (see Fig. S1 in the Supplementary Material). Specifically, the 4-dimensional feature set included the measurement conditions (T), solar cell characteristics (d_p and N_B), and the PVP most strongly correlated with $\log N_{\text{Fe}}$, i.e. ϵI_{SC} . In extending the set to five dimensions, we added the parameter most strongly correlated with N_{Fe} among the remaining unused ones — namely, $\epsilon\eta$ — to the existing features. Subsequent steps followed a similar procedure.

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). This is not surprising, since each of the PV parameters is linked to the diffusion and recombination of photo-induced charge carriers. Besides, ϵI_{SC} , ϵV_{OC} , $\epsilon\eta$, and ϵFF 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 also trained ML models using features selected via the following procedure: PCs were computed for the original dataset, and the dimensionality was then reduced by discarding PCs whose contribution to the total variance was below 1.5%. As a result, after applying PCA, the feature dimensionality for the different models was reduced from the original sizes of 4, 5, 6, and 7 to 4, 4, 5, and 6 in the AM1.5 case, and to 4, 4, 4, and 5 in the 940 nm case. In general, PCA is a widely used and effective tool in machine learning, particularly for addressing problems in photovoltaics. For example, it has been successfully applied to fault detection and classification based on I - V curves [78, 79], performance prediction of organic [80] and perovskite [81] solar cells, as well as measurement of the internal quantum efficiency in GaAs solar cells [10].

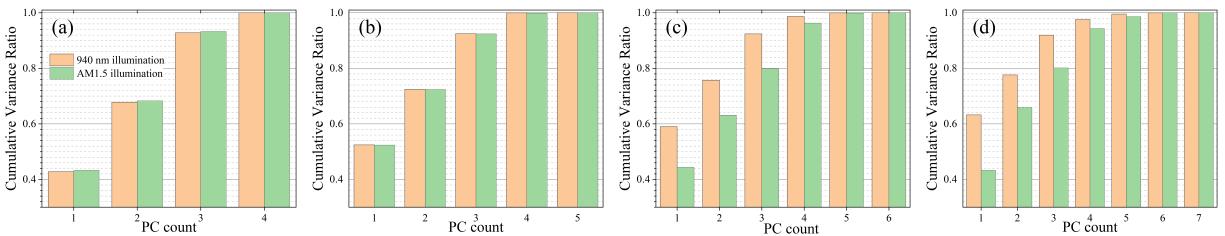


Figure 2: Cumulative explained variance versus number of PC components. Feature dimension: 4 (a), 5 (b), 6 (c), 7 (d).

2.3. Machine Learning Algorithms

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

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 [82].

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 [83].

eXtreme Gradient Boosting (XGB) is an advanced ensemble method for gradient boosting machines [84]. 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 [85].

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, 57, 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: illum ∈ (AM1.5, 940); “feat” represents the feature dimension and the application of PCA, where feat ∈ (4, 5, 6, 7; 4:PC, 5:PC, 6:PC, 7: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 [86] 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_{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 p_{01} and p_{10} , 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 p_{01} and p_{10} 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_{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} cm^{-3} . The last difficulty arises because, as shown by simulation [61], 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.

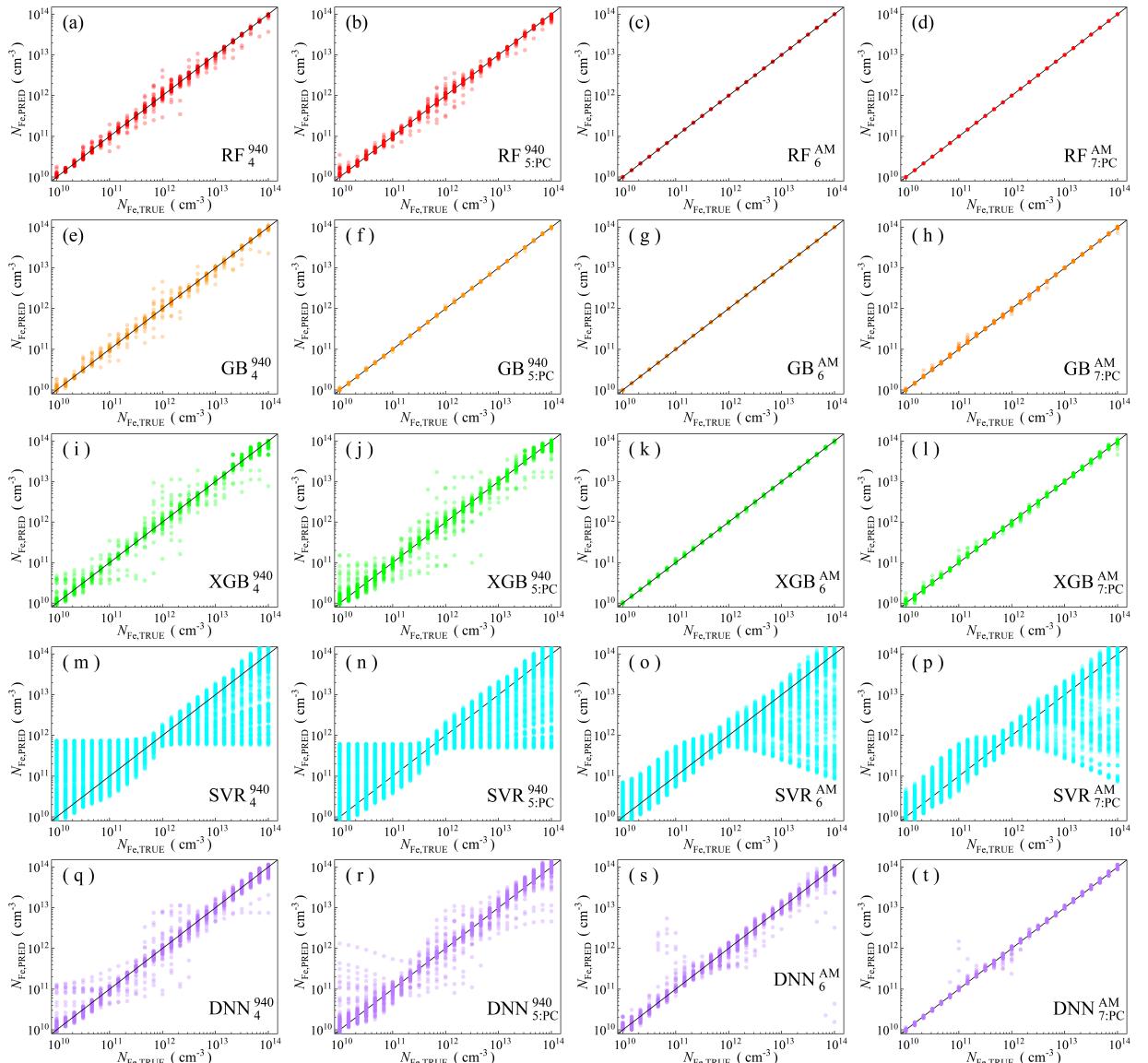


Figure 3: Scatter plots compare reference iron concentrations with ML-predicted values during the training phase. The ML algorithms include RF (a–d), GB (e–h), XGB (i–l), SVR (m–p), and DNN (q–t). The data come from simulation under monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 illumination (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.

3.2. N_{Fe} -altered test dataset

The N_{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.

Fig. 6 and Figs. S8–S11 in Supplementary Materials present the prediction results, while Fig. 7 and Table 1 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

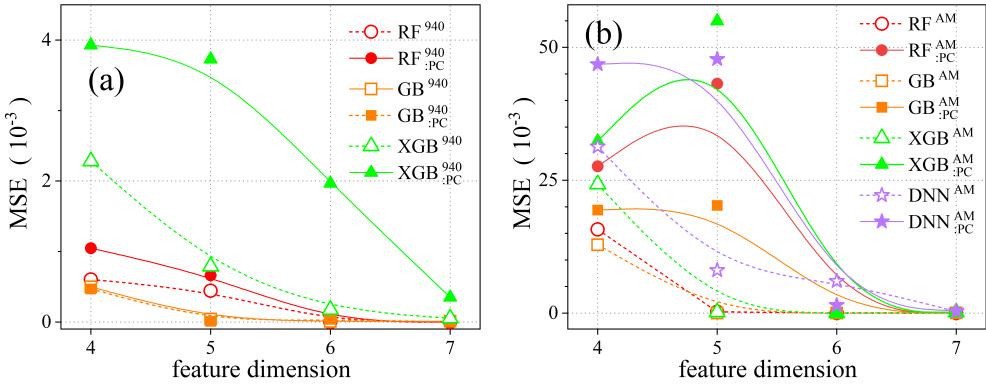


Figure 4: MSE dependence on input feature dimension for training data obtained under monochromatic (a) and AM1.5 illumination (b). ML algorithms: RF (circles), GB (rectangles), XGB (triangles), and DNN (stars). Closed markers indicate results with PCA, while open markers represent results without PCA.

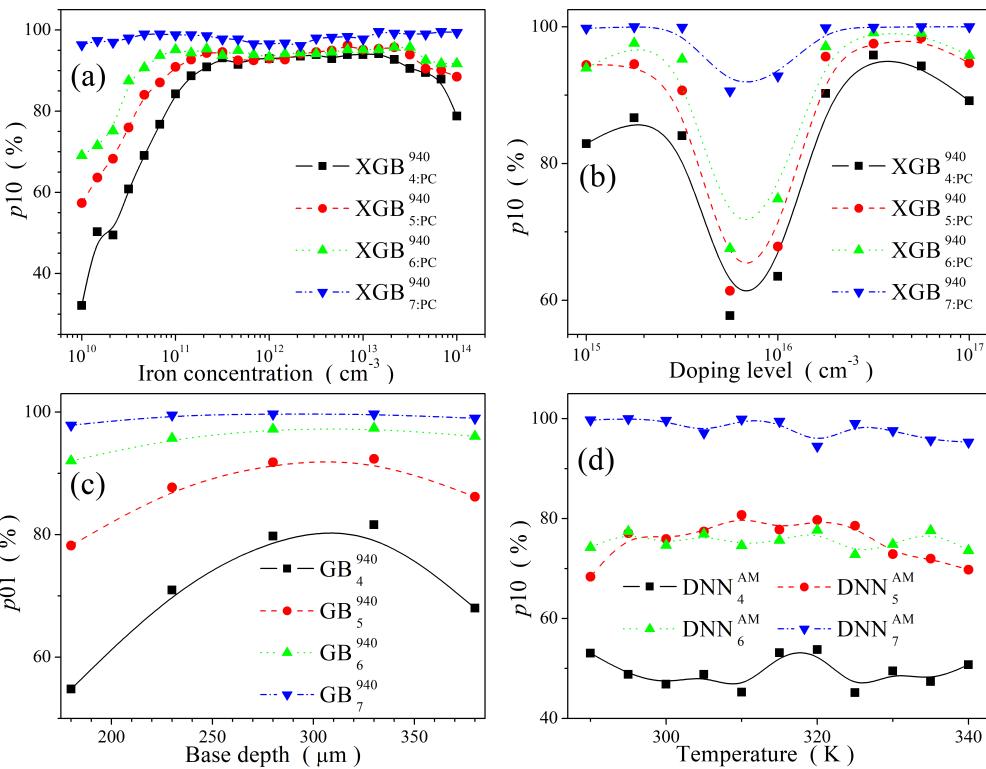


Figure 5: Typical dependencies of the proportion of predictions with an accuracy of at least 10% (a, b, c) or 1% (d) on iron concentration, alloying level, substrate thickness, and temperature.

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 1, RF, GB, and XGB, which are the most accurate models, exhibit a noticeable increase in MdAPE and a rise in p_{01} and p_{10} as the number of descriptors grows. These trends indicate that expanding feature dimensionality leads to a higher proportion of predictions with significant errors, as illustrated in Fig. 8a for GB. Fig. 8a and Table 1 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, $\text{GB}_{4:\text{PC}}^{940}$, $\text{GB}_{5:\text{PC}}^{940}$, and XGB_5^{940} yield the most precise results for this test

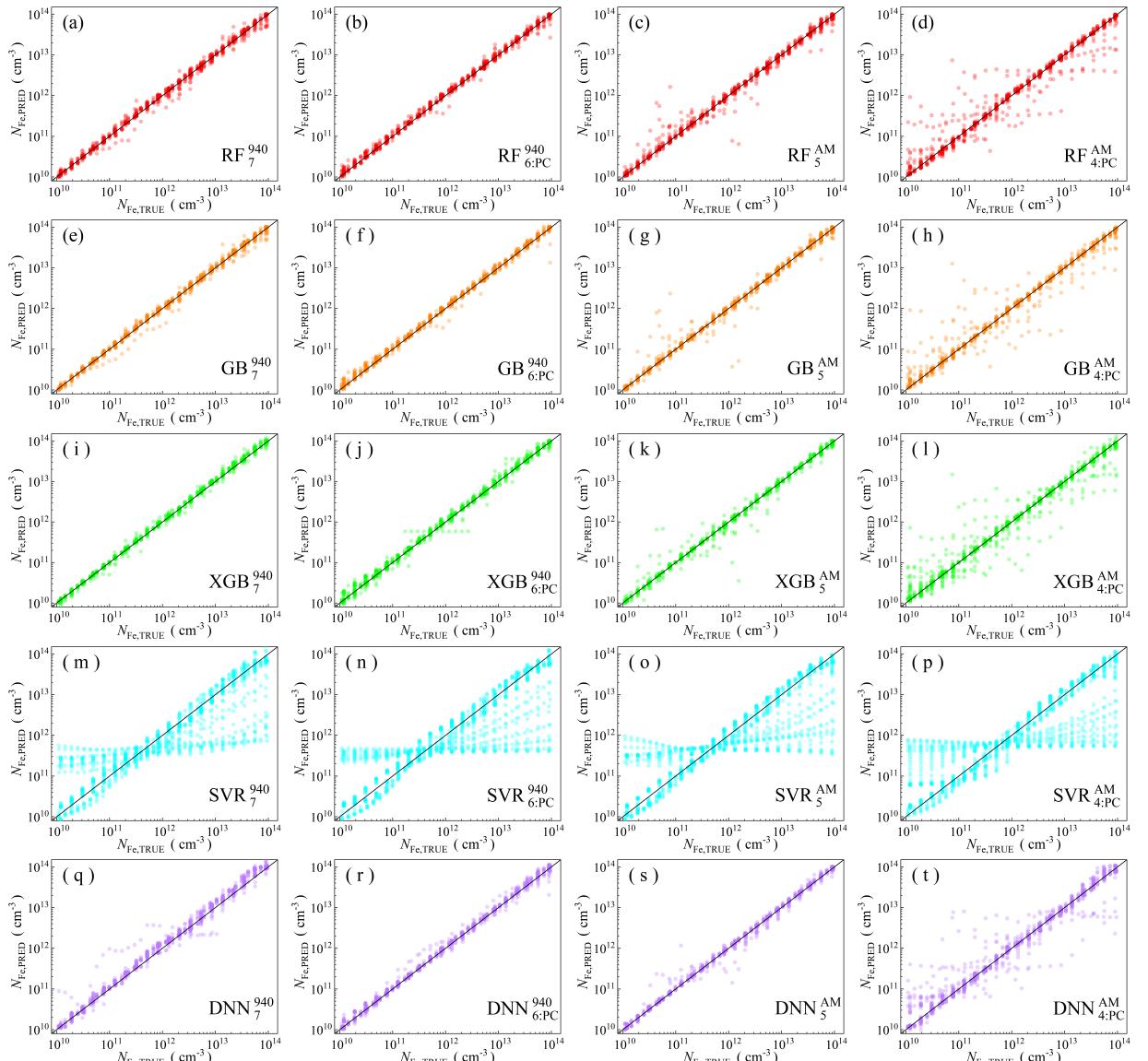


Figure 6: Scatter plots compare reference iron concentrations with ML-predicted values during the N_{Fe} -altered test phase. The ML algorithms include RF (a-d), GB (e-h), XGB (i-l), SVR (m-p), and DNN (q-t). The data come from simulation under monochromatic (a, b, e, f, i, j, m, n, q, r) and AM1.5 illumination (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.

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 N_{Fe} with MAPE slightly above 10%, though the results remain inferior to those under monochromatic illumination. The use of PCA significantly worsens the results (see Fig. 7f, Fig. 7n, Table 1). 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. 7p) and the proportion of predictions

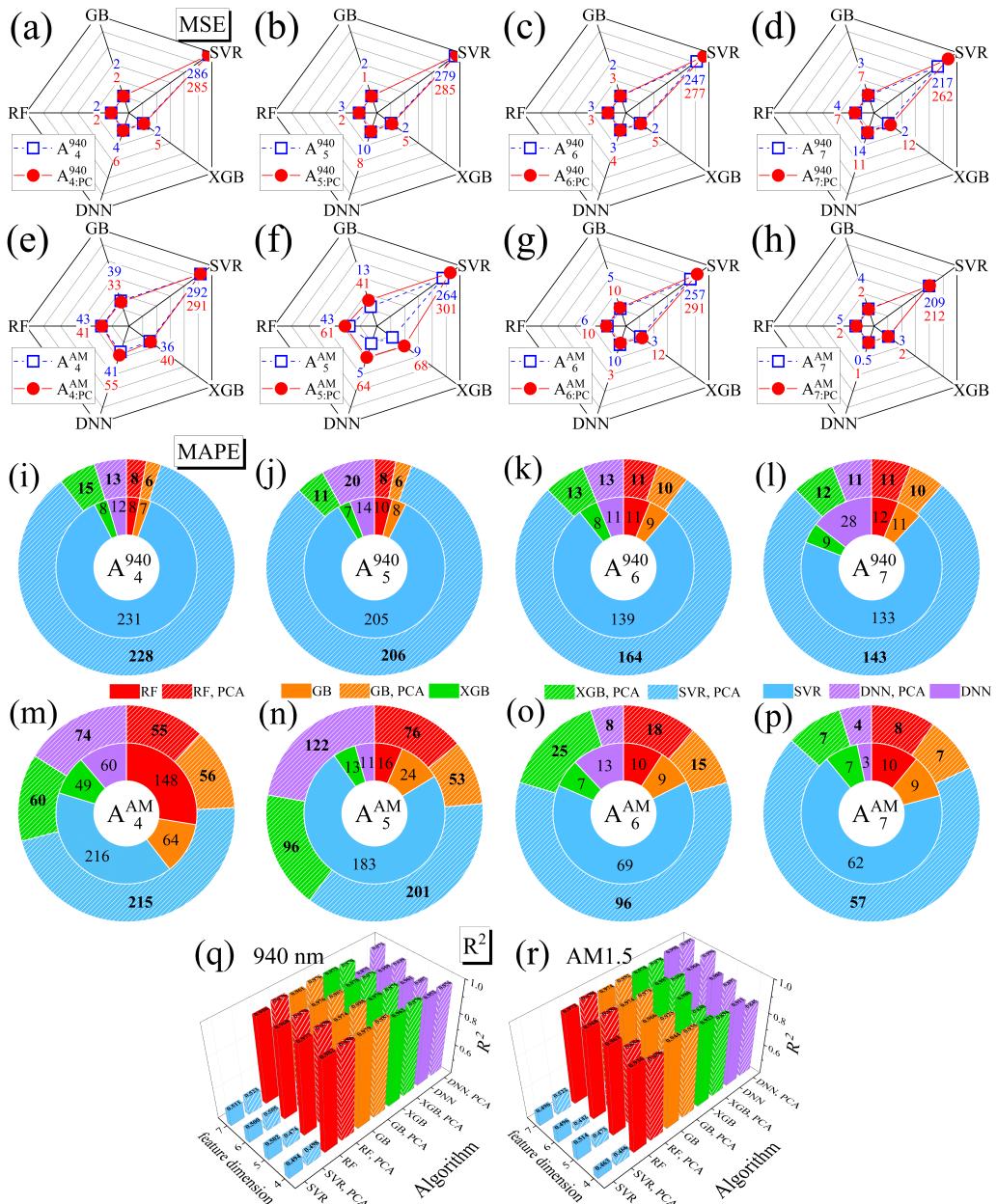


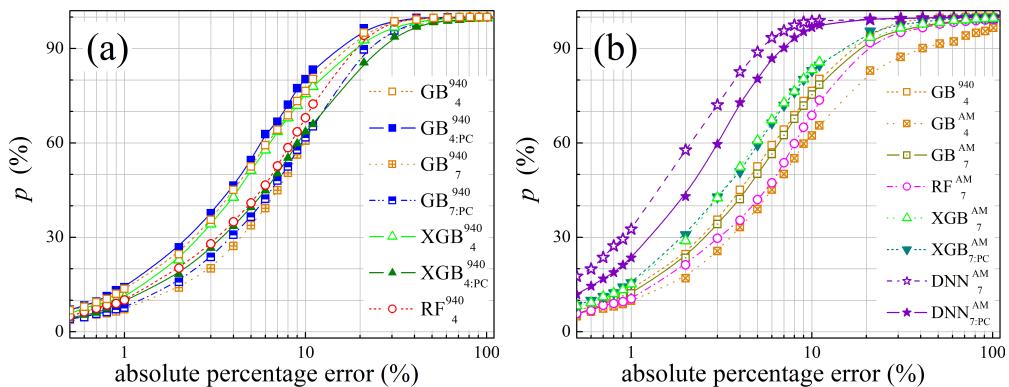
Figure 7: MSE (a–h), MAPE (i–p), and R^2 (q, r) scores for various models, feature combinations, and illumination conditions on the N_{Fe} -altered test dataset. Illumination: 940 nm (a–d, i–l, q), AM1.5 (e–h, m–p, r). Feature dimensions: 4 (a, e, i, m), 5 (b, f, j, n), 6 (c, g, k, o), and 7 (d, h, l, 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.

with large deviations from the true values (Table 1). For XGB and DNN — the most efficient models across the entire N_{Fe} -altered test dataset, including data under 940 nm illumination — PCA provides no benefit (Fig. 7p, Fig. 8b). 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.

Table 1Performance metrics of the ML models for N_{Fe} -altered dataset

Algorithm	Feature dimension	MdAPE (%)				p01 (%)				p10 (%)			
		940 nm		AM1.5		940 nm		AM1.5		940 nm		AM1.5	
		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	6.00	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	13.6	14.1	9.87	9.38	76.5	80.3	62.4	67.2
	5	4.96	4.63	6.55	8.05	13.6	12.5	11.2	7.25	72.7	82.4	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	4.46	6.20	4.26	9.04	11.7	10.2	13.8	6.58	78.6	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	1.63	2.40	4.06	8.90	32.6	23.5	33.1	64.7	98.7	96.7

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

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 2 summarize the performance metrics. The data show that the average metrics for all models worsen slightly compared to the N_{Fe} -altered dataset (Fig. 7, Fig. 9). In most cases, PCA reduces model performance. 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 1 and Table 2 reveals that, unlike the mean characteristics, MdAPE, $p01$, and $p10$ are often better than those in the N_{Fe} -altered dataset. For example, for XGB^{940}_6 , MdAPE and $p01$ reach 2.47% and 33.6%,

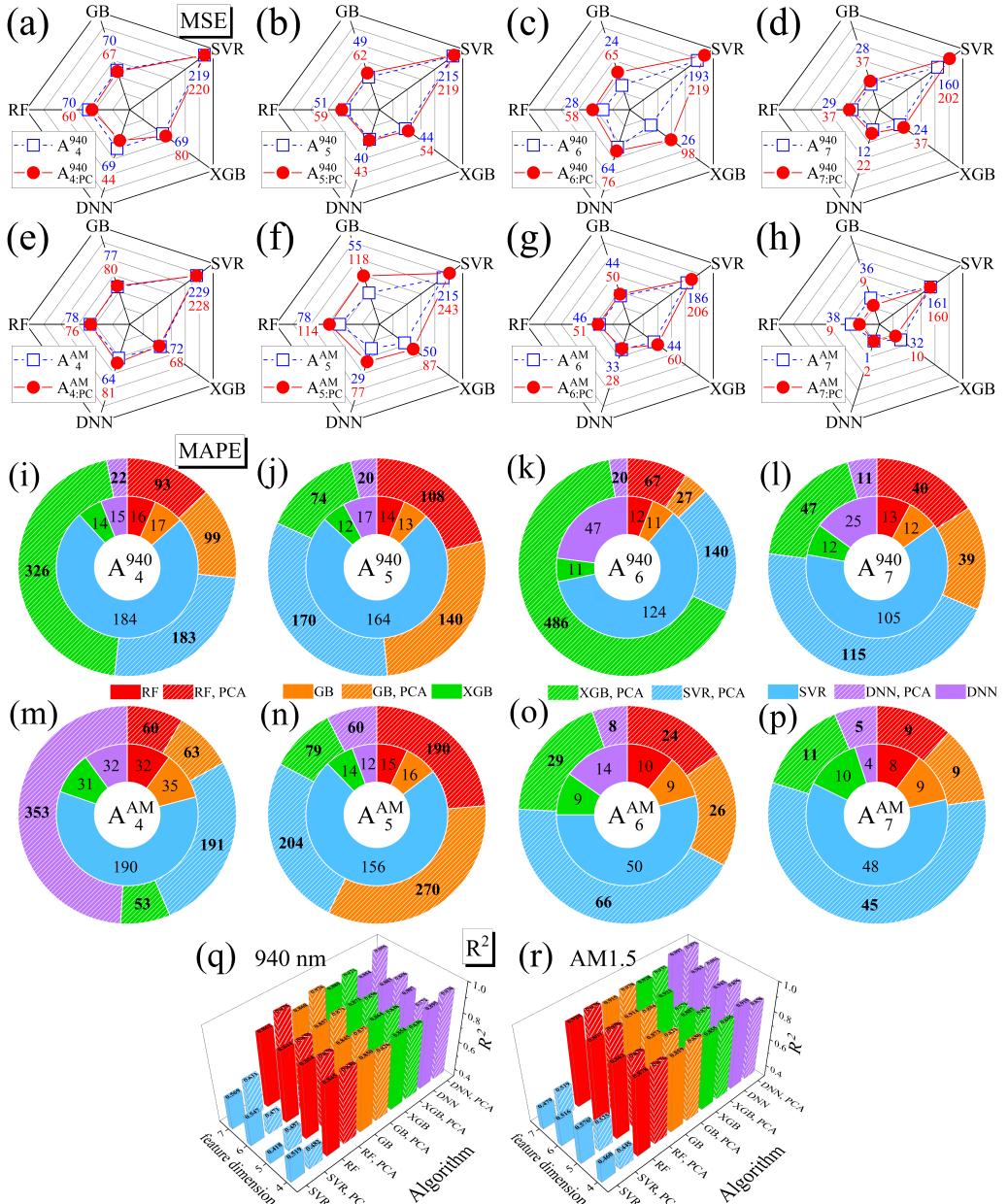


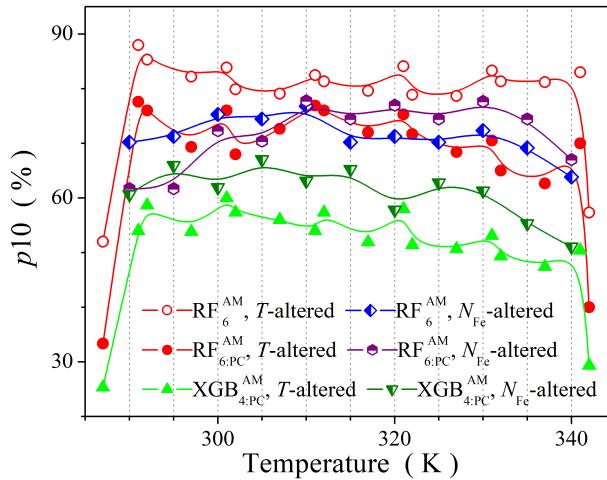
Figure 9: MSE (a–h), MAPE (i–p), and R^2 (q, r) scores for various models, feature combinations, and illumination conditions on the T -altered test dataset. Illumination: 940 nm (a–d, i–l, q), AM1.5 (e–h, m–p, r). Feature dimensions: 4 (a, e, i, m), 5 (b, f, j, n), 6 (c, g, k, o), and 7 (d, h, l, p). 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.

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 p_{10} values and overall trends, despite the temperature values in the T -altered test dataset differing from those used during

Table 2Performance 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								
RF	4	6.50	13.1	11.6	9.58	18.2	8.92	11.8	12.1	60.4	40.8	45.2	51.3
	5	5.59	13.2	5.90	14.2	22.4	6.75	16.6	9.5	62.9	40.7	62.7	42.1
	6	4.61	13.3	2.13	5.05	20.8	8.58	35.3	17.7	68.8	43.5	77.3	68.3
	7	5.07	9.82	1.94	3.47	21.5	12.0	38.7	21.5	67.4	50.4	79.8	80.6
GB	4	5.29	12.5	10.07	9.34	20.7	6.67	5.08	6.58	62.0	42.9	49.8	51.8
	5	4.91	12.6	4.89	12.7	19.1	7.00	22.2	4.67	66.2	44.7	65.0	70.7
	6	3.41	9.12	1.81	5.14	22.3	11.2	40.6	13.92	73.1	52.4	78.3	70.7
	7	5.03	9.93	2.44	3.40	15.6	9.92	31.3	20.8	68.3	50.3	79.3	80.8
XGB	4	5.03	10.9	8.72	9.88	22.8	8.92	8.42	7.00	64.6	46.9	53.8	50.3
	5	3.65	9.00	5.14	10.5	28.4	8.75	23.7	6.58	67.3	53.0	66.3	48.6
	6	2.47	10.4	2.08	5.78	33.6	6.75	34.4	12.3	73.7	49.2	78.0	69.2
	7	5.37	9.72	3.38	5.11	14.9	10.2	19.3	13.6	67.9	51.0	78.3	72.9
SVR	4	30.9	27.8	36.4	35.7	1.50	1.58	0.67	1.17	15.3	18.3	12.8	13.5
	5	33.7	38.0	37.5	47.6	1.75	1.00	0.92	0.67	15.3	12.9	12.8	9.8
	6	38.3	40.9	37.1	46.2	1.50	1.50	1.25	0.75	14.3	15.0	12.8	9.58
	7	39.3	35.91	36.3	32.2	1.25	1.50	1.83	1.00	10.7	13.3	15.4	14.7
DNN	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	18.6	6.85	1.88	2.53	3.83	7.5	28.1	23.2	29.8	66.3	95.58	92.5

**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. The vertical lines indicate the temperature values used in the training dataset.

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.

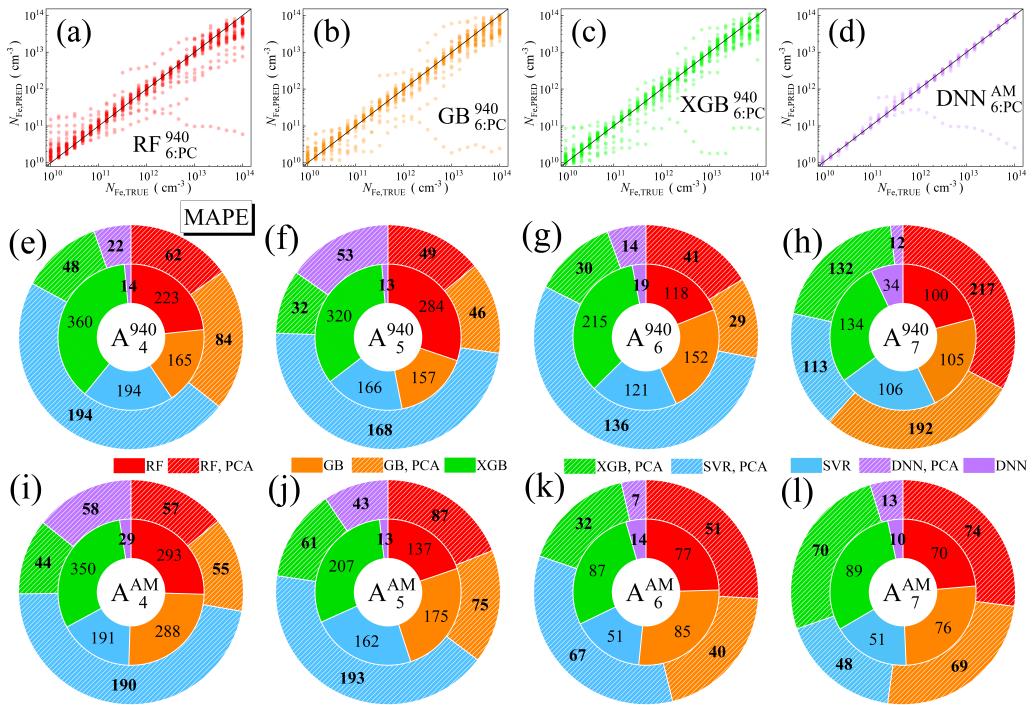


Figure 11: (a)–(d) Scatter plots comparing reference iron concentrations with ML-predicted values for the N_B -altered dataset for the best models based on RF (a), GB (b), XGB (c), and DNN (d). (e)–(l) MAPE scores for various models, feature combinations, and illumination conditions. Illumination: 940 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.

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.

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 descriptors 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} cm^{-3} . However, for

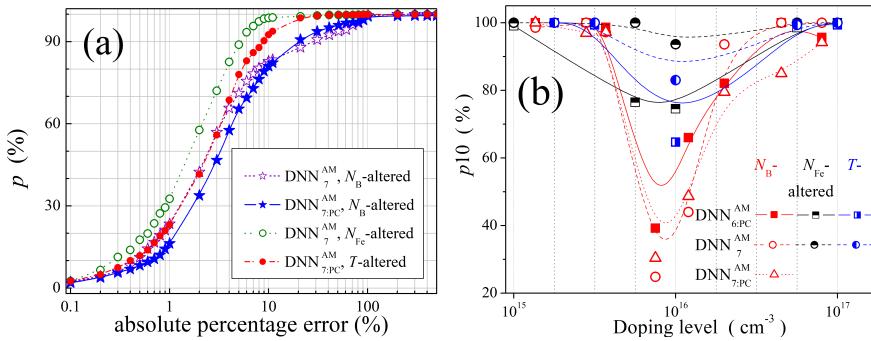


Figure 12: The proportion of samples in the test dataset with an absolute error below a given threshold as a function of the threshold value (a) and the dependencies of the fraction of predictions with an accuracy not exceeding 10% on the doping level for various test datasets (b). The vertical lines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

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.

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. Comparative Analysis

Before presenting the results based on experimental data, a few summarizing remarks are in order.

The presence of iron impurities in silicon increases the recombination rate of charge carriers. However, each photovoltaic parameter captures a different aspect of how iron-related recombination influences solar cell performance. Specifically, the short-circuit current reflects the photogenerated current (I_{ph}), which primarily depends on the diffusion length and recombination in the quasi-neutral region. In contrast, the open-circuit voltage is influenced not only by I_{ph} but also by the saturation current (I_0) and the ideality factor (n), both of which are determined by recombination processes in the space-charge region [87]: $V_{OC} = nkT [\ln(I_{ph}/I_0) + 1]$. The fill factor, in turn, is less sensitive to variations in n [88]:

$$FF \approx \frac{v_{OC} - \ln(0.72 + v_{OC})}{v_{OC} + 1}, \quad (5)$$

where v_{OC} being the normalized open-circuit voltage $v_{OC} = qV_{OC}/nkT$. Ultimately, the efficiency integrates all of these effects: $\eta \propto I_{SC} \cdot V_{OC} \cdot FF$. Incorporating multiple descriptors associated with variations in photovoltaic

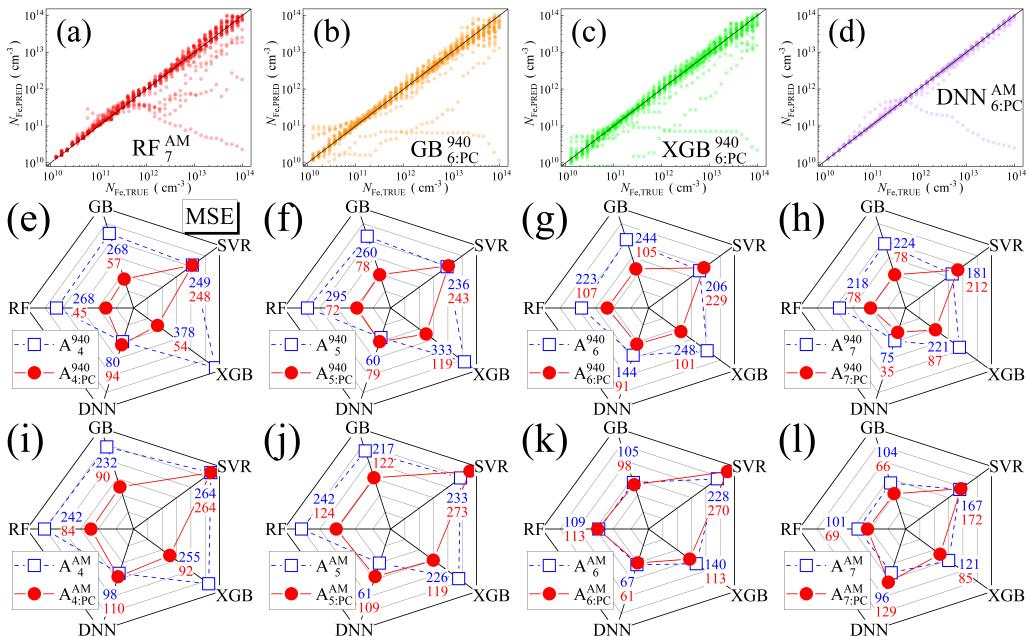


Figure 13: (a)–(d) Scatter plots comparing reference iron concentrations with ML-predicted values for the All-altered test dataset for the best models based on RF (a), GB (b), XGB (c), and DNN (d). (e)–(l) MSE scores for various models, feature combinations, and illumination conditions. Illumination: 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.

parameters provides a more comprehensive representation of the effects of iron impurities. This expanded descriptor set increases the informational diversity available for model training, which can enhance predictive accuracy.

One can see that models trained on data obtained under AM1.5 illumination perform worse in almost all cases. We believe this can be attributed to the following reasons. It is well known that a regression model may achieve high accuracy by having the target variable be an unambiguous function of the descriptor. Moreover, the absolute value of the function's derivative should be as significant as possible. As shown earlier [61], the condition of uniqueness in the relationship between N_{Fe} and the changes in PVPs under AM1.5 illumination is often not fulfilled in contrast to the case of monochromatic illumination. In addition, the ranges of changes in εI_{SC} , εV_{OC} , and $\varepsilon \eta$ under monochromatic illumination exceed those observed under AM1.5 illumination (see Subsection 2.1). The physical origin of the differences in the relationship between iron concentration and parameter changes under AM1.5 and monochromatic illumination lies in the fact that, in the former case, the generation of non-equilibrium charge carriers occurs throughout the entire volume of the solar cell, rather than being limited to the base region. As a result, the processes involved in photovoltaic conversion are more complex. In our view, these differences explain why the performance of ML models trained on AM1.5 illumination data is generally poorer. One way to address the ambiguity in the relationship between features and the target variable is to incorporate additional features into the model. As it is shown, for AM1.5 illumination, increasing the number of descriptors often improves prediction accuracy, supporting our hypothesis.

Increasing the number of descriptors while keeping the sample size constant leads to higher sparsity in the data space due to increased dimensionality, which in turn makes it more difficult for the model to generalize. In such cases, the model may also learn spurious correlations that do not generalize to test data — a risk that is particularly pronounced in deep neural networks and deep tree-based ensemble methods. Moreover, a larger number of features results in more model parameters, increasing the risk of overfitting. These factors can collectively lead to a reduction in prediction accuracy. In our opinion, for $A^{AM1.5}$ models, both the positive effects achieved by increasing the number of descriptors (enhancing overall information variance and reducing ambiguity between the target variable and the input features) were dominant. However, for A^{940} models where this ambiguity was already low, the negative effects described above outweighed the benefits, which explains why prediction accuracy did not improve when using an expanded feature set. Therefore, simply increasing the number of features does not guarantee improved performance. Instead, optimal predictive accuracy is typically achieved with specific, well-chosen combinations of descriptors.

Another effect that may arise when increasing the number of features is multicollinearity. It occurs when new features are highly correlated with existing ones and can lead to instability in model estimates. However, in our case, it is unlikely that multicollinearity among the photovoltaic parameters (incidentally, their correlation is significantly higher in the 940 nm case than in the AM1.5 case — see Fig. S1) affects prediction accuracy. First, all algorithms used are sufficiently robust to multicollinearity. Second, a standard method for mitigating multicollinearity is PCA. Despite previous successes in applying principal component analysis to various photovoltaics-related machine learning tasks [10, 78, 79, 80, 81] — and the presence of favorable conditions for its use in our case, such as the high correlation between changes in photoelectric parameters caused by FeB pair dissociation — applying PCA as a data pre-processing step for evaluating iron impurity concentration did not improve the maximum prediction accuracy in most cases. One likely reason is that PCA preserves the overall variance in the data but does not account for the specific relationships between individual features and the target variable. For instance, some photovoltaic parameters may be more informative for predicting low iron concentrations, while others may be more relevant for high concentrations. So ϵ_{FF} are shown [61] to be most pronounced at low iron concentrations, unlike other PVPs. When these components are merged through PCA transformation, such distinctions may be lost, leading to reduced accuracy. Another possible reason is that principal components — being linear combinations of original features — do not support localized decision splits (i.e., those based on individual feature thresholds), which are particularly effective in tree-based models.

Rare cases where PCA pre-processing leads to improved prediction accuracy typically occur when the original model (without PCA) performs poorly. In this context, one might consider using PCA as a diagnostic indicator of model quality when evaluating iron concentration in silicon solar cells: if PCA improves prediction accuracy, it suggests that the baseline model is weak; if not, the model is likely to be strong. However, this approach relies on access to the true N_{Fe} values, significantly restricting its practical applicability.

When comparing the effectiveness of different algorithms, several points are worth noting. The poorest performance observed with SVR is likely due to its limited ability to model complex nonlinear relationships — such as those that apparently exist between iron concentration and variations in photoelectric parameters. In contrast, the superior performance of the XGB and DNN models is attributable to their ability to effectively capture nonlinear patterns, which in our case become more prominent with increasing descriptor dimensionality. Specifically, XGB performs best in relatively simpler scenarios, such as those involving 940 nm illumination (see Figs. 7i-7l, 9i-9l). This is because XGB requires less training data than DNN (12,000 examples are sufficient, though this is not a particularly large amount for a neural network), handles high-quality structured data well (simulated data are noise-free), and is more resilient to the influence of correlated features. At the same time, DNNs are more powerful tools for nonlinear modeling, capable of uncovering hidden structures in the data. This enables them to achieve better results on more complex tasks — such as the transition to A^{AM1.5}-models with high-dimensional features (see Figs. 7m-7p, 9m-9p). As shown by the results in Subsection 3.4, successful prediction within the N_B -altered also requires accounting for higher-order relationships, which explains why DNN models achieved the best performance (Figs. 11e-11l). Regarding the RF and GB algorithms, as indicated in Tables 1 and 2, they tend to perform best only on less complex data (i.e., small sets of descriptors), where more powerful models are at higher risk of overfitting.

3.6. 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_{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.

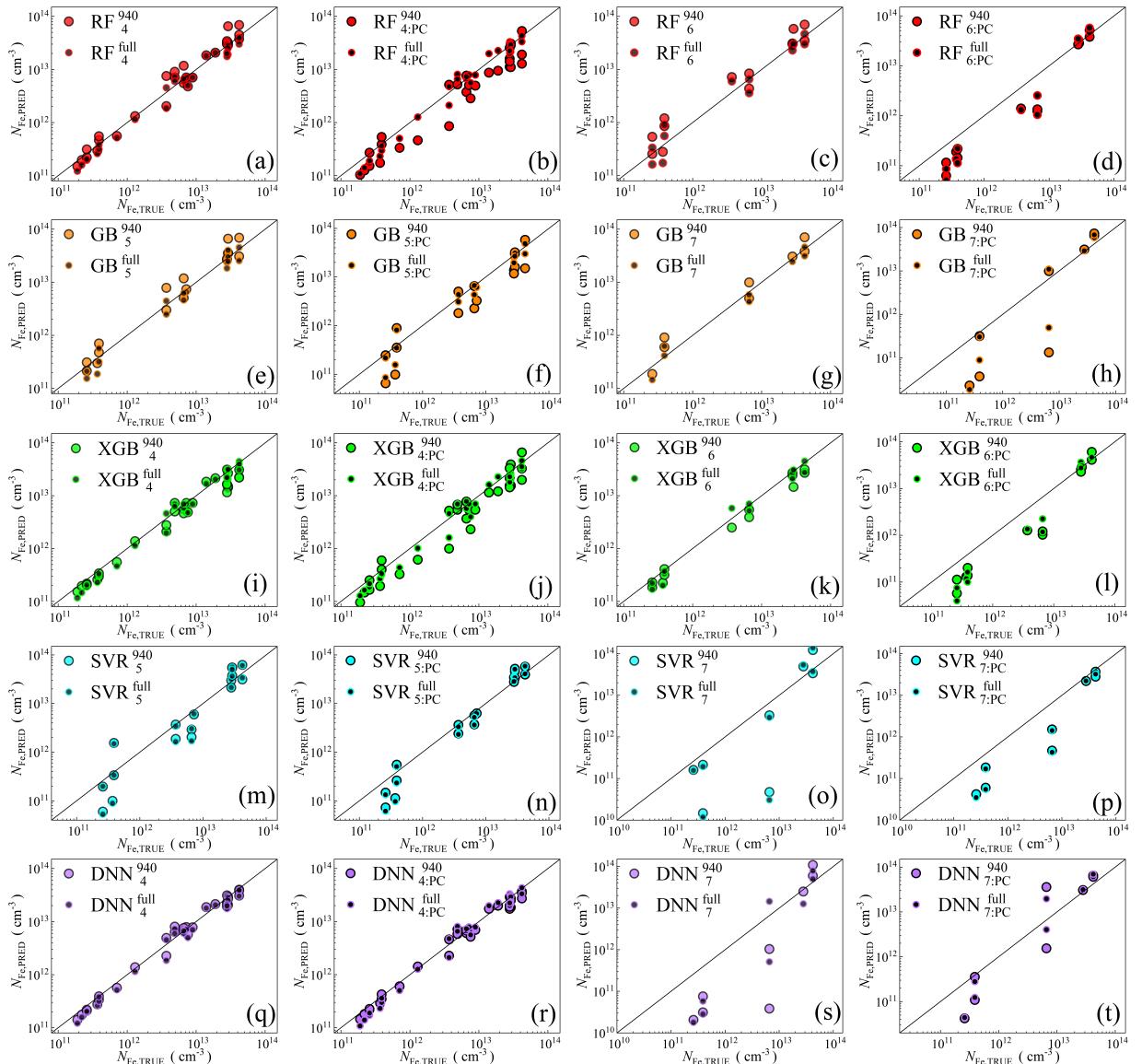


Figure 14: Scatter plots compare reference iron concentrations with ML-predicted values during the experimental test phase. The ML algorithms include RF (a–d), GB (e–h), XGB (i–l), SVR (m–p), and DNN (q–t). Panels (b, d, f, h, j, l, n, p, r, t) incorporate PCA pre-processing. The input feature dimensions 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.

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 ,

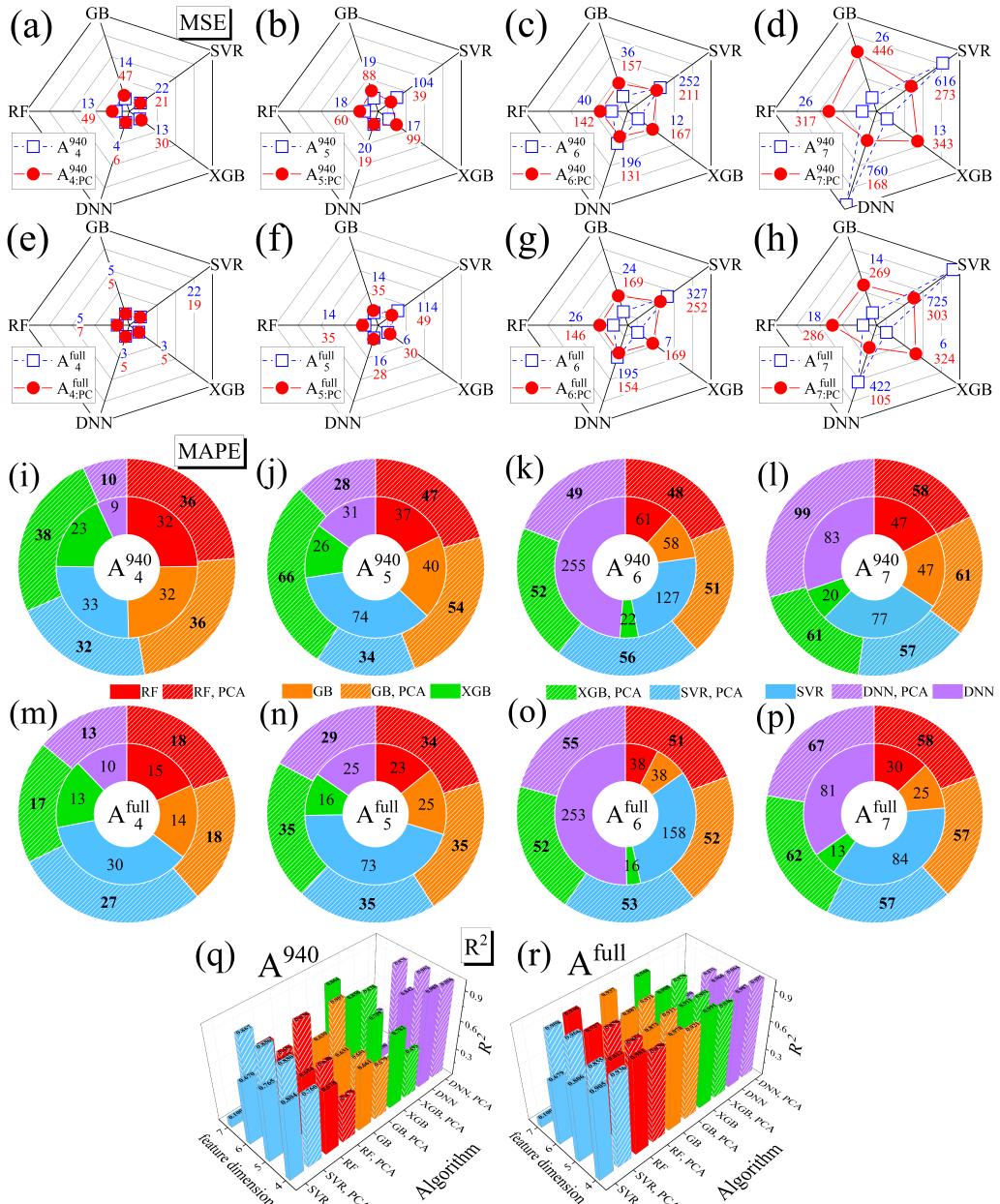


Figure 15: MSE (a–h), MAPE (i–p), and R^2 (q, r) scores for various models and feature sets on the experimental test set. The models used the training dataset (a–d, i–l, q) or the full synthetic dataset (e–h, m–p, r) for training. The feature dimensions are as follows: 4 (a, e, i, m), 5 (b, f, j, n), 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.

εI_{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_6^{940}) 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_4^{940} , 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_{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 3×10^{-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.

3.7. Potential directions for enhancing predictive accuracy

When exploring potential strategies for improving model accuracy, it is essential to consider the efforts already undertaken towards this goal. Specifically, data normalization was applied, multiple descriptor sets were tested, and the impact of reducing feature correlation using PCA was evaluated. In addition, the main model parameters were optimized (see Tables S1–S5 and Figure S2) through bayesian optimization using the Optuna package, and cross-validation was employed.

Among the remaining justified options, the following approaches can be highlighted:

1. Expanding the training dataset, primarily by incorporating a greater number of boron concentration values into the database. This strategy aligns with a fundamental principle of machine learning—the need for as much data as possible.
2. Deriving new descriptors from existing ones that are more directly and simply related to iron concentration. These descriptors should not be limited to linear combinations, such as those used in PCA; instead, their development should be guided by underlying physical processes. Alternatively, generative models, such as variational autoencoders, can be employed to discover latent variables. Notably, this approach may also facilitate easier dataset expansion.
3. Incorporating additional features, especially those obtainable without extra measurements, is desirable because the requirements of additional experiments would substantially reduce the practicality of the proposed method. For example, the minority carrier diffusion length could serve as a valuable descriptor; however, to our knowledge, no methods exist for its determination based solely on single or dual I - V characteristic measurements. The most apparent step in this direction involves the use of relative changes in voltage (εV_{MP}) and current (εI_{MP}), corresponding to the maximum power point, as supplementary features. However, it should be noted that these quantities exhibit correlation with other PVPs (see Fig. S31).

Separately, a strategy for improving the quality of the training dataset, which involves increasing the realism of the simulation model, can be distinguished. For example, instead of employing SCAPS, three-dimensional simulators such as Synopsys Sentaurus TCAD, Silvaco ATLAS TCAD, or Solcore can be used to generate labeled datasets that account for the presence of additional impurities commonly found in Cz-Si, such as oxygen and carbon. Furthermore, incorporating experimental uncertainty into the training data by introducing noise can enhance the model's robustness and predictive accuracy.

To some extent, the issue of accuracy is related to the applicability of the developed models beyond the selected range of N_B , d_p , and T . To tackle this problem, it is first necessary to consider each parameter separately. The ability of the models to make predictions beyond the range of the training dataset depends on the relationship between the target variable and the descriptors. As previously shown [61], the relationship between the relative changes in the photoelectric parameters and iron concentration can vary significantly depending on the temperature and the concentration of the doping impurity in the base material. This is because a number of parameters, including those of silicon such as bandgap, carrier mobilities, intrinsic recombination coefficients, and Fermi level position, as well as iron-related defects such as carrier capture cross-sections, depend on temperature and carrier concentration. These relationships, which were taken into account when creating the training set, are sufficiently complex, particularly with respect to N_B , as demonstrated by the results obtained for the N_B -altered dataset. As previously mentioned, even the task of interpolation within the range of N_B used during training was only satisfactorily addressed by the DNN models due to their high ability to identify hidden patterns. Predicting beyond the training range of N_B is even more challenging. Therefore, in our view, the capabilities of the developed models are quite limited.

The temperature-related relationships are simpler, as demonstrated by the reasonably good results obtained using tree-based models for the T -altered dataset (see Fig. 9, Table 2). However, for the training set that included T values beyond the original range, the performance of these models degraded significantly (see Fig. 10). Once again, the best average results were achieved by the deep neural network models (see Fig. 9), and only with their use can relatively accurate predictions beyond the training range of temperature be expected.

The situation is more straightforward concerning the base width, as its influence on the dependence of N_{Fe} on variations in PVPs value is much weaker and often approximately linear [61]. Therefore, in our opinion, the developed models can be expected to remain applicable for predictions beyond the range selected in this study. Naturally, the extension of the base thickness range is only possible within certain limits, until the ratio between d_p and the diffusion length of minority carriers changes significantly.

In summary, we note that if predictions are required for cases outside the range of parameter values used during training, it is advisable to rely exclusively on DNN models. Among the considered parameters, base thickness allows for the largest permissible deviations, whereas for temperature, deviations are unlikely to exceed 10 K. The use of the developed models with doping impurity concentrations different from those in the training set is not recommended. This requirement effectively serves as an additional condition that the training set must satisfy to ensure satisfactory prediction accuracy.

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.

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 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} 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 factors contributing to differences in the predictive accuracy of various models are analyzed, and potential measures for its further improvement are proposed.

From a practical standpoint, the developed models are applicable for iron quantification in monocrystalline silicon solar cells with an $n^+ - p - p^+$ structure, featuring a boron-doped base and a wide range of base thicknesses ($180 - 380 \mu\text{m}$) and doping concentrations ($10^{15} - 10^{17} \text{ cm}^{-3}$). The proposed ML-based approach provides a non-destructive, fast, and simple method for determining iron concentrations that does not require specialized equipment. The methodology used to develop models can be adapted for other types of solar cells and/or for quantifying other impurities.

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

Oleg Olikh 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

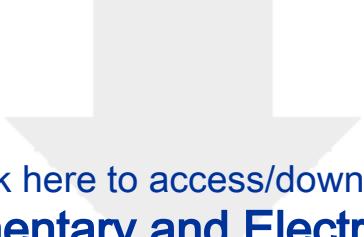
Data will be made available on request.

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. Holecek, 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-Otum, 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, Psopruner: 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. Thamrong siripak, 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] 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.
- [41] A. A. Istratov, H. Hieslmair, E. Weber, Iron and its complexes in silicon, *Appl. Phys. A: Mater. Sci. Process.* 69 (1999) 13–44.
- [42] E. Wright, J. Coutinho, S. Öberg, V. J. B. Torres, Mössbauer parameters of fe-related defects in group-iv semiconductors: First principles calculations, *J. Appl. Phys.* 119 (2016) 181509.
- [43] J. R. Poindexter, R. L. Z. Hoye, L. Nienhaus, R. C. Kurchin, A. E. Morishige, E. E. Looney, A. Osherov, J.-P. Correa-Baena, B. Lai, V. Bulović, V. Stevanović, M. G. Bawendi, T. Buonassisi, High tolerance to iron contamination in lead halide perovskite solar cells, *ACS Nano* 11 (2017) 7101–7109.
- [44] L. Kimerling, J. Benton, Electronically controlled reactions of interstitial iron in silicon, *Physica B+C* 116 (1983) 297–300.
- [45] 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.
- [46] G. Zoth, W. Bergholz, A fast, preparation-free method to detect iron in silicon, *J. Appl. Phys.* 67 (1990) 6764–6771.
- [47] 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.
- [48] J. Schmidt, D. Macdonald, Recombination activity of iron-gallium and iron-indium pairs in silicon, *J. Appl. Phys.* 97 (2005) 113712.
- [49] 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.
- [50] O. Olikh, V. Kostylyov, V. Vlasiuk, 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.
- [51] 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.
- [52] A. Herguth, Quantification of iron in boron-doped silicon solar cells from open circuit voltage measurements, *IEEE J. Photovolt.* 12 (2022) 937–947.
- [53] M. Di Sabatino, Detection limits for glow discharge mass spectrometry (gdms) analyses of impurities in solar cell silicon, *Measurement* 50 (2014) 135–140.
- [54] D. K. Schroder, *Semiconductor Material and Device Characterization*, John Wiley & Sons, New Jersey, third edition, 2006.
- [55] R. C. Kurchin, J. R. Poindexter, V. Viähäniissi, 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.
- [56] A. R. Peaker, V. P. Markevich, J. Coutinho, Tutorial: Junction spectroscopy techniques and deep-level defects in semiconductors, *J. Appl. Phys.* 123 (2018) 161559.
- [57] 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.
- [58] P. Kunze, J. M. Greulich, A. Tummaliyah, W. Wirtz, H. Hoeffler, N. Woehrle, S. Glunz, S. Rein, M. Demant, Contactless inline iv measurement of solar cells using an empirical model, *Sol. RRL* 7 (2023) 2200599.
- [59] 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.
- [60] 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.
- [61] 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.
- [62] M. Burgelman, P. Nollet, S. Degrave, Modelling polycrystalline semiconductor solar cells, *Thin Solid Films* 361–362 (2000) 527–532.
- [63] M. Masum Mia, M. Faruk Hossain, M. Rahman, N. Badi, A. Irfan, M. Ferdous Rahman, Unveiling the impact of Se based HTM on BaZrSe₃ perovskites solar cell and improving the theoretical efficiency above 32%, *Materials Science and Engineering: B* 311 (2025) 117817.
- [64] T. K. Joshi, G. Sharma, Y. R. Sharma, A. S. Verma, Spectroscopic screening and performance parameters of hybrid perovskite (CH₃CH₂PH₃PbI₃) using WIEN2k and SCAPS-1d, *Phys. B Condens. Matter* 682 (2024) 415793.

- [65] 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/CsGeI₃ perovskite tandem solar cells using a combination of density functional theory and SCAPS-1D frameworks, *Sol. Energ. Mat. Sol.* 267 (2024) 112688.
- [66] 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.
- [67] L. You, X. Zhang, Q. Ma, W. Zhu, J. Wu, Optimization of electron transport layer-free Cs₂TiBr₆/MASnBr₃ laminated structure perovskite solar cells by SCAPS-1D simulation, *Phys. Status Solidi A* 220 (2023) 2300071.
- [68] M. Abdulmalik, E. Danladi, Influence of perovskite thickness on the performance of silver-doped NaZnBr₃ perovskite solar cells using SCAPS software, *Semiconductor Physics, Quantum Electronics & Optoelectronics* 26 (2023) 321–331.
- [69] O. Olikh, Relationship between the ideality factor and the iron concentration in silicon solar cells, *Superlattices Microstruct.* 136 (2019) 106309.
- [70] O. Olikh, V. Kostylyov, V. Vlasiuk, 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.
- [71] N. Khelifati, H. S. Laine, V. Vähäniemi, H. Savin, F. Z. Bouamama, D. Bouhafs, Dissociation and formation kinetics of iron–boron pairs in silicon after phosphorus implantation gettering, *Phys Status Solidi A* 216 (2019) 1900253.
- [72] B. Paviet-Salomon, J. Levrat, V. Fakhfouri, Y. Pelet, N. Rebeaud, M. Despesse, C. Ballif, New guidelines for a more accurate extraction of solar cells and modules key data from their current–voltage curves, *Prog. Photovoltaics Res. Appl.* 25 (2017) 623–635.
- [73] L. J. Geerligs, D. Macdonald, Dynamics of light-induced feb pair dissociation in crystalline silicon, *Appl. Phys. Lett.* 85 (2004) 5227–5229.
- [74] K. Lauer, C. Möller, D. Debbih, M. Auge, D. Schulze, Determination of activation energy of the iron acceptor pair association and dissociation reaction, in: *Gettering and Defect Engineering in Semiconductor Technology XVI*, volume 242 of *Solid State Phenomena*, Trans Tech Publications Ltd, 2016, pp. 230–235.
- [75] O. Olikh, O. Datsenko, S. Kondratenko, Influence of illumination spectrum on dissociation kinetics of iron–boron pairs in silicon, *Phys. Status Solidi A* 221 (2024) 2400351.
- [76] Y. Srivastava, A. Jain, End-to-end material thermal conductivity prediction through machine learning, *J. Appl. Phys.* 134 (2023) 225101.
- [77] 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.
- [78] S. Fadhel, C. Delpha, D. Diallo, I. Bahri, A. Migan, M. Trabelsi, M. Mimouni, PV shading fault detection and classification based on I-V curve using principal component analysis: Application to isolated PV system, *Sol. Energy* 179 (2019) 1–10.
- [79] W. Gao, R.-J. Wai, A novel fault identification method for photovoltaic array via convolutional neural network and residual gated recurrent unit, *IEEE Access* 8 (2020) 159493–159510.
- [80] T. W. David, G. A. Soares, N. Bristow, D. Bagnis, J. Kettle, Predicting diurnal outdoor performance and degradation of organic photovoltaics via machine learning: relating degradation to outdoor stress conditions, *Prog. Photovoltaics Res. Appl.* 29 (2021) 1274–1284.
- [81] Y. Liu, W. Yan, S. Han, H. Zhu, Y. Tu, L. Guan, X. Tan, How machine learning predicts and explains the performance of perovskite solar cells, *Sol. RRL* 6 (2022) 2101100.
- [82] L. Breiman, Random forests, *Mach. Learn.* 45 (2001) 5–32.
- [83] A. Natekin, A. Knoll, Gradient boosting machines, a tutorial, *Front. Neurorob.* 7 (2013).
- [84] S. Akinpelu, S. Abolade, E. Okafor, D. Obada, A. Ukpong, S. Kumar R., J. Healy, A. Akande, Interpretable machine learning methods to predict the mechanical properties of abx₃ perovskites, *Results Phys.* 65 (2024) 107978.
- [85] W. Cao, X. Liu, J. Ni, Parameter optimization of support vector regression using henry gas solubility optimization algorithm, *IEEE Access* 8 (2020) 88633–88642.
- [86] 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.
- [87] D. Yang (Ed.), *Handbook of Photovoltaic Silicon*, Springer Berlin, Heidelberg, first edition, 2019.
- [88] M. A. Green, Accuracy of analytical expressions for solar cell fill factors, *Sol. Cells* 7 (1982) 337–340.



Click here to access/download

Supplementary and Electronic files
SupplementaryMaterial.pdf



Click here to access/download
LaTeX Source Files
LaTeXSourceFiles.zip

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Oleg Olikh reports financial support was provided by National Research Foundation of Ukraine. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.