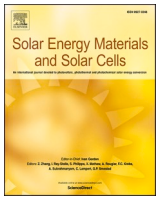
Solar Energy Materials & Solar Cells 260 (2023) 112483

Contents lists available at ScienceDirect

Solar Energy Materials and Solar Cells

journal homepage: www.elsevier.com/locate/solmat



Automated classification of electroluminescence images using artificial neural networks in correlation to solar cell performance parameters

Marko Turek *, Manuel Meusel

Fraunhofer Center for Silicon Photovoltaics CSP, Otto-Eißfeldt-Str. 12, 06120, Halle (Saale), Germany

ARTICLE INFO

Keywords:
Electroluminescence
Image classification
Artificial intelligence

ABSTRACT

The quality control in solar cell production relies on the acquisition of quantitative performance data (I–V-data) together with imaging diagnostics such as electroluminescence (EL) imaging. The classification of the solar cells into different quality groups based on the performance parameters or the identification of specific defects in the EL-images, such as cracks, is rather straightforward and well established. These techniques are extended in our work as we analyze the quantitative correlation between the two approaches. As a specific example, a quality issue related to the contacting of the solar cells during the solar simulator measurement is considered. We demonstrate that a reliable and fully automated classification of the images with respect to the considered quality issue can be achieved by our artificial intelligence (AI) approach. This approach is different from the majority of state-of-the-art approaches as it does not consider specific image objects or features, such as cracks or spots, but rather predicts an image category for the entire image. The performance of the developed supervised image classification approach based on a neural network is analyzed quantitatively. We show that a clear quantitative correlation between cell I–V-data and AI-predicted EL-image classes can be established. In this way, an early-warning procedure can be implemented to detect this performance related issue using the AI-analysis of the EL-images even when its appearance is rather weakly pronounced and before it can be discovered in the I–V-data. We also present results on the impact of the training data set on the performance of the neural network and key parameters describing the quality of the network.

1. Introduction

The throughput in solar cell production lines continuously increases while the solar cell technologies become more complex at the same time. This implies that new measurement technologies and data analysis approaches targeting at a fully automated quality control system need to be implemented. Ideally, such a digital quality control system considers all available data in a way that quality issues can be detected as early as possible. Additionally, the quality control system should also provide information on how a specific quality issue effects the performance parameters of the cells to quantitatively assess the severity of the detected issue. To this end, various machine learning algorithms have been developed in recent years, see for example [1–3].

In a solar cell production line, three types of data sources are available: cell process parameters along the production line (e.g. provided by equipment sensors), quantitative performance parameters of the solar cells (e.g. obtained from I–V-curve measurements), and data based on imaging diagnostics. The assessment of quantitative process parameters or performance values using, for example, thresholds is well established in this context. Furthermore, significant progress has also been made in the last few years regarding an automated analysis of quantitative data and images using AI-approaches [4–7]. However, the focus in the field of the image analysis has been in many cases on the detection of image features or objects [8] which can be implemented to detect defects such as cracks, finger interruptions or metallization issues. A direct correlation of the images to the associated performance parameters or even a complete and integrated assessment of all data sources in PV-production is not yet fully established.

In our work, we present an analysis of the applicability of neural networks to the classification of EL-images going beyond individual image feature detection. Our study investigates a specific quality issue that is related to a mis-aligned contacting of the solar cells during the measurement of the current-voltage-curve (I–V-curve) [9]. Ideally, such a quality issue is detected as early as possible in production so that any impact on the measured performance parameters can be avoided. This requires the development of a reliable image classification approach

* Corresponding author.
E-mail address: marko.turek@csp.fraunhofer.de (M. Turek).

https://doi.org/10.1016/j.solmat.2023.112483
Received 20 April 2023; Received in revised form 12 July 2023; Accepted 23 July 2023
Available online 27 July 2023

together with a thorough analysis of this classification in correlation to the obtained I–V-curve parameters, such as short circuit current I_{sc}, open circuit voltage V_{oc}, or fill-factor FF. A high correlation between image classification and performance parameters can only be expected with a proven reliability of the algorithmic approach. Thus, we investigate key performance parameters of the employed neural network and its robustness with respect to different training data sets. We show that the major classes separating well-contacted cells from cells with contacting issues are well identified by the network while image classes with minor differences in the images are less clear separated. We also find that the set of training data chosen plays a significant role in this context. It is shown that a random choice of training data leads to the best results. We discuss further an approach for a highly automated training independent of an operator.

2. Experimental approach, data acquisition and data analysis

In our study, the data set consists of I–V-data and EL-images. The data has been acquired using an automated measurement platform consisting of industrial production tools. As test samples, commercially available mono-facial M3 PERC solar cells with four busbars have been selected. During the experiment, the current-voltage-characteristics (I–V-data) of 792 solar cells were measured using a WAVELABS SINUS220 sun simulator without any additional cell processing. During the experiment, a first test run has been performed where the contacting unit has been intentionally and slightly mis-aligned to simulate a process monitoring quality issue. The test run has been designed such that only a small fraction of cells, i.e. 30%, is clearly affected by the mis-aligned contacting. This corresponds, for example, to a situation in a cell production line, where the positioning tolerances of the cell alignment in the contacting unit have increased during line operation. Subsequent to this first run, a second run has been performed with optimized contacting in order to generate a reference data set. A first, statistical analysis of the performance data has been presented in Ref. [9].

The data set includes the I–V-curves obtained during the flasher measurement and the associated cell performance parameters. In the same cell characterization module, EL-images were taken for each solar cell while being contacted by the contacting unit. All data has then been transferred to a MES-database (FabEagle by Kontron AIS). The entire data set was then exported as raw data and processed into structured data using Python. The structured data set contains the cell performance parameters extracted from the I–V-curves, e.g. short circuit current I_{sc}, open circuit voltage V_{oc} and fill-factor FF together with the corresponding EL-image.

In a second step, all EL-images have been sorted by an operator into categories and labeled with classes EL-cat1 . EL-cat5 with respect to the contacting quality, see Fig. 1. This initial supervised image classification is intended to address and quantify the severeness of the mis-aligned contacting. It resulted in the following distribution of cells into the defined categories: EL-cat1 with 201 cells, EL-cat2 with 297 cells, EL-cat3 with 88 cells, EL-cat4 with 182 cells, and EL-cat5 with 14 cells. The remaining 10 cells were sorted into a completely different class as they exhibited cracks or other issues unrelated to the contacting. These 10 cells were part of AI classification but have been excluded from further data evaluation.

In a further step, sub-sets of about 200 EL-images have been selected as training data sets for neural networks. The network architecture relies on the DENKweit VISION AI Hub which has also been used as a platform to train the image classification algorithm. While the number of training images has been between 30 and 50 for the first four EL-cat classes, the fifth category EL-cat5 has been trained with about 5–7 images only. Also, it is important to note, that the objective of this training has not been the setup of a network for the detection of specific image objects or features such as cracks or spot. The focus was rather on the development of a classification model that can classify an entire image into one of the five classes thus describing the contacting quality for each cell individually.

The trained network has then been applied to the full set of EL-images. As a result, each image is labeled by two EL-cat values: a first value given by the operator and a second value predicted by the network model. The predicted labels being a result of this supervised classification were then compared to the given manual classification by means of a confusion matrix analysis. Furthermore, the correlation of both labels, the originally given operator values and AI-predicted EL-cat values, to the performance parameters I_{sc}, V_{oc}, and FF is investigated.

3. Results

The objective of this work is to identify quantities or parameters that can serve as sensitive, early indicators for quality related production issues which are in our case represented by the intentionally induced mis-aligned contacting. It has been described in a previous work, that individual I–V-parameters of single cells, such as the fill factor, are rather insensitive to a contacting issue. Only for very severe issues, a clear change in the FF-value of a single cell can be observed. Also, the mean or median fill-factor value of the entire test run with the mis-aligned contacting unit is only slightly different from the reference run with optimal contacting, see Fig. 2 (left). A better and more sensitive indicator can be obtained by determination of other distribution parameters such as the width or skewness of the FF-distribution. However, this approach would also require the data of several hundreds of cells before some reliable conclusions can be drawn using the distribution of fill-factor values if no information based on the EL-images is included. Even in our case of almost 800 cells, the estimated distributions and derived parameters such as the width or skewness are a rather qualitative fits only, see Fig. 2 (left). Furthermore, changes in statistical parameters might not be specific enough for the identification of a contacting issue as investigated in this work. Hence, we conclude that the statistical analysis of I–V-parameters without including any EL-image information is too insensitive for an early detection of such a contacting issue.

Therefore, the focus of this work has been set on a quantitative analysis of the EL-images. The resulting distribution of fill-factor values for each of the introduced categories EL-cat1 . EL-cat5 clearly indicates that the overall qualitative impression obtained from the EL-image quality correlates quantitatively well with the fill-factor values, see Fig. 2 (right). In particular, we find that the manual (operator) categorization is well reflected in the distribution of fill-factors: El-cat1 and EL-cat2 correspond to large FF-values with narrow distributions while EL-cat3-5 exhibit clearly reduced median FF-values with increased FF-

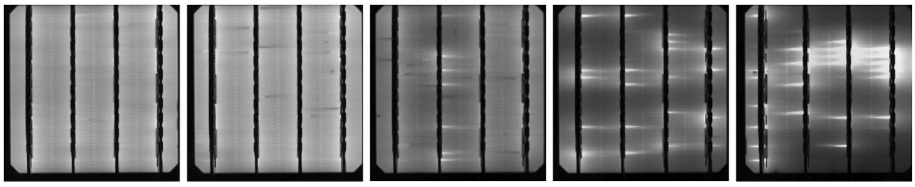


Fig. 1. Examples of EL-images (from left to right) within the defined EL-cat classes: optimal contacting (EL-cat1), minor contacting issue (EL-cat2), contacting issue clearly visible (EL-cat3), pronounced contacting issues (EL-cat4), and severe contacting issue (EL-cat5).

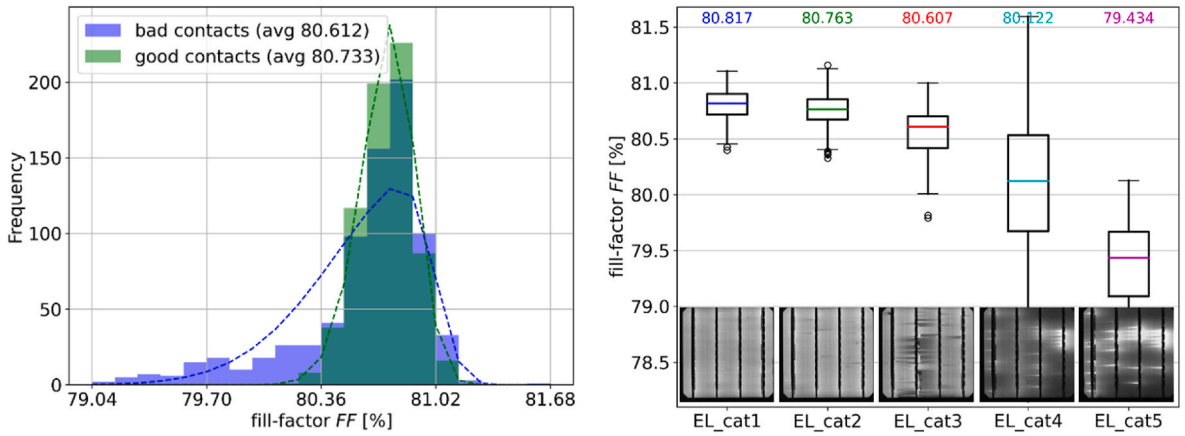


Fig. 2. (left) Distribution of fill-factor values for the test run with intentionally mis-aligned contacting unit (labeled "bad contacts", blue bars and line obtained by a fit) in comparison to the reference run (labeled "good contacts", green bars and line obtained by a fit); (right) Boxplots of fill-factor values corresponding to the EL-cat classes.

spread. Thus, an early indicator on whether there is an issue with a mis-aligned contacting unit could already be drawn based on an individual EL-image if a reliable classification of each image into the proposed classes can be achieved. In particular, the occurrence of a certain number of cells associated with EL-cat3 or above can thus serve as an early indicator of the investigated quality problem.

In highly automated production lines, the image classification into the proposed classes cannot be performed by operators but should rather be relying on an automated image analysis algorithm. Thus, the performance of such a classification algorithm needs to be evaluated first before applying it to larger production data sets. A measure of the prediction quality relative to the operator-based categories is given by the confusion matrix. This matrix represents the numbers of images correctly and falsely classified into the categories EL-cat1 . 5, see Fig. 3 (left). We find that the major classes are well distinguished, i.e. category EL-cat3 from EL-cat1 and EL-cat2. Categories with minor differences are to some extend mixed, i.e. EL-cat1 is mixed with EL-cat2.

When a certain class is considered, then there are true-positive TP, false-positive FP, false-negative FN and true-negative TN classifications. The true-positives TP are given by the diagonal entries of the confusion matrix. Images of the class under consideration that are sorted into other classes are termed false-negatives FN, i.e. off-diagonal entries in each line. Images that are sorted into the considered class but have been originally a different class are termed false-positive FP. All other values are false-negatives FN. These values can be further used to provide the key indicators *precision* being PR = TP/(TP + FP) and *recall* being RE = TP/(TP + FN). Here, precision PR and recall RE values close to one would indicate a very reliable prediction with low numbers of false-positives FP and false-negatives FN.

To validate our model, we have tested several training sets. Comparing confusions matrices obtained by varying training sets leads to the conclusion that the training set itself also plays a significant role. The resulting network performance of two different exemplary training sets, i.e. randomly-chosen TS 1 and non-random TS 2 made of the first cells of the data set, is represented in Fig. 3 (left) and (right). Using the confusion matrix, one can compare the two key parameters precision PR and recall RE for class EL-cat3. This class is chosen as it is intended to serve as an early indicator separating the case of well aligned contacting from mis-aligned contacting. While the first training set TS 1 results in a precision- and recall-value of PR = 0.69 and RE = 0.95, respectively, the second non-random training set TS 2 yields smaller values of PR = 0.58 and RE = 0.84, respectively. The analysis of five additional variations of randomly chosen training sets TS 1b – TS 1f shows that all these training sets result in higher PR and RE values than the second type of non-random training set TS 2, see Fig. 4. The PR values scatter from 0.66 to 0.76 with a mean of 0.71 and the RE values scatters from the lower limit of 0.84, which is also achieved by the second training set, to 0.95 with a mean of 0.88. These result underlines the importance of choosing an appropriate training set carefully. In our case, we found that a random selection of training images yields better networks than, for example, taking the first few images that occur for each class within the data set.

However, one must also keep in mind that the labeling of the images by an operator also imposes some uncertainties. In particular, the operator-based sorting into EL-cat1 and EL-cat2 depends to a certain extent on the subjective operator decision. In our study, we attribute the

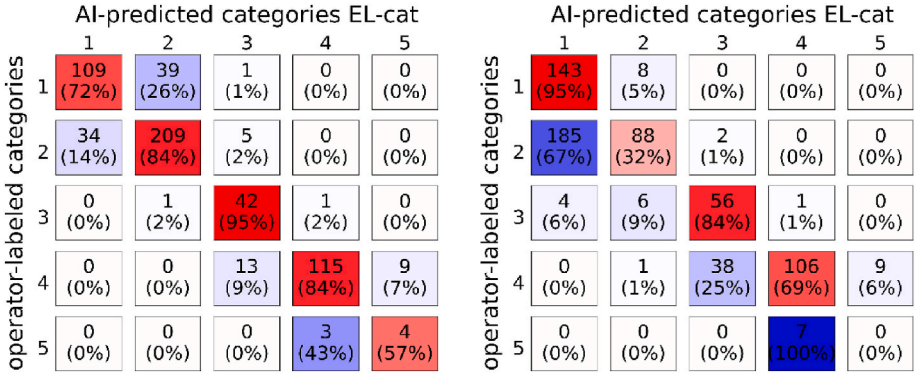


Fig. 3. Confusion matrix relating the given categories by the operator to the AI-predicted categories for two different training sets (left: random training set TS 1, right: training set TS 2 defined by first occurrences within the entire data set).

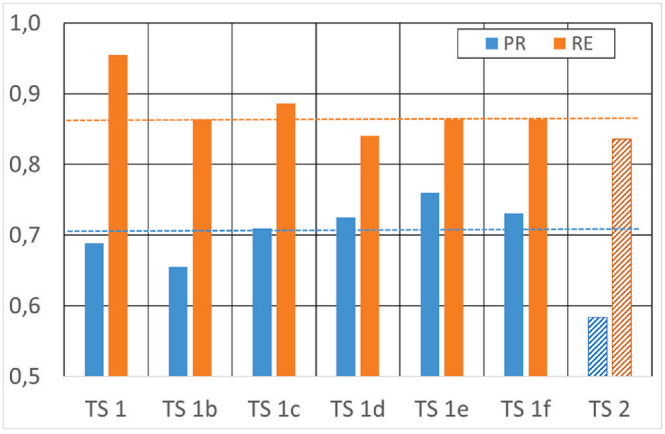


Fig. 4. Comparison of resulting precision (PR, blue color) and recall (RE, orange color) values for random training set TS 1, five additional variations of random training sets TS1b – TS1f, and training set TS 2 (indicated by dash-filled bars) defined by first occurrences within the data set. The dashed lines indicate the average PR-value of 0.71 (blue) and average RE-value of 0,87 (orange), respectively, for the six training sets with randomly chosen images.

reduced predictive power of the model regarding the distinction between EL-cat1 and EL-cat2 to some extend to an ambiguous operator labelling of the images. Combining the results of EL-cat1 and EL-cat2 into one single class yields a precision of PR = 0.997 and a recall-value of 0.985 for this combined EL-cat1&2 class. Similarly, one can combine EL-cat4 and EL-cat5 into one class obtaining PR = 0.992 and RE = 0.910, respectively, for the combined EL-cat4&5 class. This implies that the model can separate EL-cat3 extremely reliable from the other two groups of classes below or above.

As the central parameters in quality control are the performance parameters and not some artificially defined EL-classes, we have also investigated the distribution of major I–V-parameters, i.e. I_{sc}, V_{oc} and FF, in dependence of the predicted image classes. As the V_{oc} analysis did not show any correlation to the contacting issue, the corresponding results are not shown in this work. This relates to a quantitative correlation between the EL-cat values and the I–V-performance parameters. To establish an early detection of a contacting issue, a categorization of the EL-images itself would be less relevant if there is no impact on quantitative cell parameters such as the fill-factor. In Fig. 4, the distribution of the fill-factor values is shown for each EL category value in case of the manual operator-labels (left) and in case of the AI-predicted labels (right). It becomes apparent that the AI-based EL-image classification describes the relevant parameter distribution very well. Using network-predicted categories, the resulting distributions are very similar to the original one. One difference that can be observed is the width of the FF-distribution of EL-cat5. It is increased for the AI-predicted categories which is due to the low number of samples in these categories which lead to an imperfect training of the model. Despite this minor difference, the trained classification model which can be implemented into an automated data analysis approach replacing the operator decisions yields a very similar and high correlation to the performance parameters. Also, it supports the approach to identify the contacting issue by using the occurrence of EL-cat3 as an early indicator as the FF is clearly reduced for this and all higher categories. Cells being sorted into EL-cat3 . EL-cat5 do not only exhibit a clearly distinguishable EL-signature but they are also associated with a statistically relevant measurement error.

Besides the fill-factor, which is directly affected by the contacting quality, other cell parameters were also analyzed. As another example, the short circuit current distribution in dependence on the EL-cat values and the impact of the choice of the training set is presented in Fig. 5 The training set with the better recall and precision scores (corresponding to the left confusion matrix in Fig. 3) reproduces a very similar correlation to the short circuit values as the original manual labeling (not shown here). The decreasing current due to the mis-aligned contacting can be explained by a higher shadowing if the contact bars are not optimal aligned. However, the second training set represented by the second confusion matrix in Fig. 3 (right) clearly shows a difference for EL-cat5 by predicting larger current values than the operator-labeled cells of this category. This second training set has led to an AI-model that classifies solar cells into EL-cat5 that do not show this current reduction in this clear way (see Fig. 6).

To gain further insights into the decision making of the network, we have furthermore analyzed not only the EL-cat with the highest probability – according to the network – but also predicted probabilities for the other categories. For this analysis, the EL-cat2 has been chosen as an illustrative case since there were many false negatives in case of training set TS 2 while this was not the case for training set TS 1, see Fig. 3. As the trained networks predict the probabilities of all classes for each sample, one can calculate weighted category values for all samples which are then given by continuous values between 1 and 5 instead of discrete values defined only by the most probable class. As an example, the distribution of all weighted continuous category values for samples with operator-label EL-cat2, where many cells are categorized in EL-cat1 by the network, is shown in Fig. 7 (left). From this representation, one can conclude that the network classifies the vast majority of EL-cat2 samples correctly and furthermore with high probability. The same analysis for the network obtained by the second training set shows a different picture: The prediction of the images classes is less reliable as many of the EL-cat2 samples are associated with some significant probability to class EL-cat1. However, since the differences between EL-cat1 and EL-cat2 are very subtle, see example images in Fig. 1, there is a pronounced uncertainty already in the manual labelling of the samples as well as the choice of training data within these classes which makes an accurate prediction hard to achieve.

4. Conclusions and discussion

In our work, we analyze a method to detect quality issues by the combined and quantitative analysis of different data sources, i.e., the solar cell's I–V-data and their EL-images. As a first result, we find that distribution parameters of the FF-distribution such as width or skewness can be used to distinguish data sets affected by a contacting issue from others without this issue. However, this statistical approach requires the

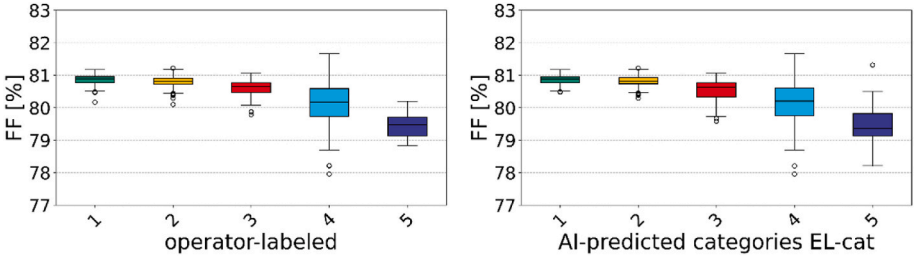


Fig. 5. Boxplots of fill-factor values for the five classes EL-cat1.5: (left) classes defined by manual operator-labelling, (right) classes obtained by the AI-prediction.

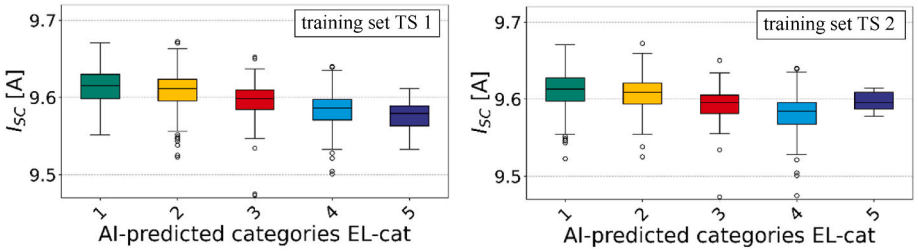


Fig. 6. Boxplots of short circuit current values for the five classes EL-cat1.5: (left) classes predicted by the neural network model obtained with the random training set TS 1, (right) same analysis for training set TS 2 defined by first occurrences within the entire data set.

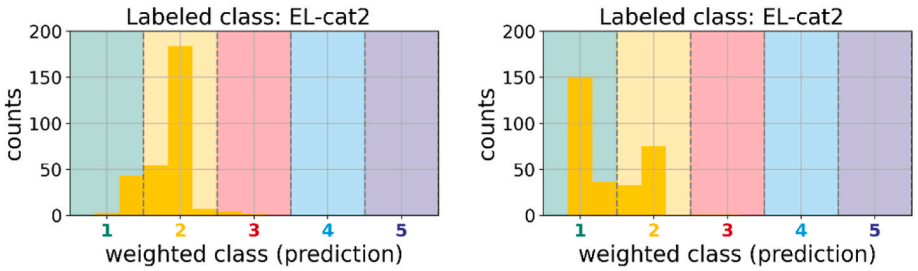


Fig. 7. Distribution of probability-weighted classes for EL-cat2 for two different training sets (left: random training set 1, right: training set 2 defined by first occurrences within the entire data set).

data of several hundreds or even more than one thousand solar cells. Furthermore, changes in the statistical parameters of the FF-distribution might also be caused by other reasons than a contacting issue. Hence, we conclude that an early and sensitive prediction based on the I–V-parameters only without the EL-images is not sensitive enough for an early detection and reliable identification of the investigated contacting issue.

Therefore, we included the EL-images into the analysis and have defined EL-categories describing the severeness of the contacting failure. These categories take the overall appearance of the EL-images into account instead of considering individual features such as for example cracks, dark spots, or finger interruptions. It is found that these EL-cat classes correlate well with the FF-values of the associated solar cells. We also show that artificial neural networks can reliably be applied to classify the EL-images according to these pre-defined categories. However, we also found that a careful choice of the training set is required to achieve reliable predictions. The predictive power of the trained models is quantified in terms of the confusion matrix or the recall and precision values for each category. In our application case, the occurrence of EL-images associated with EL-cat3 can serve as an early indicator for the discussed quality issue. The categorization into this class is achieved with very high precision and recall values. In this manner, a reliable classification model can be set up and then be applied to a larger data set as it would be the case for a production facility.

In order to apply the proposed workflow to production data, the quality management would need to deduce the number of incorrectly measured cells in rather small-sized sub-groups, e.g. 100–500, of cells. In other words, the probability of a single cell being incorrectly measured needs to be estimated from the EL-cat analysis. If this probability exceeds a certain limiting value an early indicator warning could be triggered. If the trained AI-models were perfect, i.e. recall and precision values are both exactly equal to one, than one could obtained this probability directly by counting the number of AI-detected "not-ok" cells, i.e. cells classified in EL-cat3 . EL-cat5, among all other "ok"-cells. However, with imperfect AI-models, one can only estimate the real number of "not-ok" cells that have been incorrectly measured by the AI-detected number of "not-ok" cells, i.e. cells categorized in EL-cat3 to EL-cat5 by the network. The mathematical approach for estimating this probability that a cell is incorrectly measured is rather complex. It will be presented in a future publication as it is outside the scope of this work. However, it turns out that the error of the estimated probability using an AI-model compared to the real probability is given by the model's specific recall values RE as (1 - RE_{ok}) * (1 - $RE_{not\text{-}ok}$)/(RE_{ok} * $RE_{not\text{-}ok}$). Here, subscript "ok" stands for the combined EL-cat1&2 class while "not-ok" represents the combined class of remaining categories EL-cat3, EL-cat4, and EL-cat5. Thus, the error in estimating the real probability of a cell being incorrectly measured by use of the AI-models is very small if the recall values RE_{ok} and $RE_{not\text{-}ok}$ of the model are close to one.

Furthermore, a clear correlation between the AI-predicted EL-classification and the cell's I–V-parameters, such as the fill-factor or the short circuit current, can be established already for minor changes in these parameters. Hence, our AI-approach using EL-images can be employed as an early failure detection system with relevance to the measured I–V-values which is significantly more sensitive than an analysis of the cells' performance parameters alone could be.

CRediT authorship contribution statement

Marko Turek: Writing – original draft, Supervision, Methodology, Investigation, Conceptualization. **Manuel Meusel:** Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Marko Turek reports financial support was provided by German Federal Ministry for Economic Affairs and Energy. Manuel Meusel reports financial support was provided by German Federal Ministry for Economic Affairs and Energy.

Data availability

Data will be made available on request.

Acknowledgments

The authors thank the German Federal Ministry for Economic Affairs and Energy for financial support of the project "OptiLearn" (FKZ 03EE1108A).

References

[1] D.T. Pham, A.A. Afify, Machine-learning techniques and their applications in manufacturing, in: Proc. IMechE Vol. 219 Part B: J. Eng. Manufact., 2005, p. 395.
[2] J.A. Harding, M. Shahbaz, S. Srinivas, A. Kusika, Data mining in manufacturing: a review, J. Manuf. Sci. Eng. 128 (2006) 969.
[3] D. Weichert, P. Link, A. Stoll, S. Rüping, S. Ihlenfeldt, S. Wrobel, A review of machine learning for the optimization of production processes, Int. J. Adv. Des. Manuf. Technol. 104 (2019) 1889.
[4] Y. Buratti, A. Sowmya, R. Evans, T. Trupke, Z. Hameiri, End-of-Line binning of full and half-cut cells using deep learning on electroluminescence images, in: 47th Ieee Photovoltaic Specialists Conference (Pvsc) 2020, 2020, pp. 133–137.
[5] H. Wagner-Mohnsen, P.P. Altermatt, Machine learning for optimization of mass-produced industrial silicon solar cells, in: 2021 International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD), IEEE, 2021, https://doi.org/10.1109/NUSOD52207.2021.9541457.
[6] H. Wagner-Mohnson, P. Altermatt, A combined numerical modeling and machine learning approach for optimization of mass-produced industrial solar cells, IEEE J. Photovoltaics 10 (2020) 1441.
[7] M. Alt, S. Fischer, S. Schenk, S. Zimmermann, K. Ramspeck, M. Meixner, Electroluminescence imaging and automatic cell classification in mass production of silicon solar cells, in: IEEE 7th World Conference on Photovoltaic Energy Conversion, Wcpec, 2018, pp. 3298–3304.
[8] A. Dhillon, G.K. Verma, Convolutional neural network: a review of models, methodologies and applications to object detection, Prog. Artif. Intell. 9 (2020) 85–112.
[9] M. Turek, D. Hevisov, C. Hagendorf, Improved lab-to-fab solar cell performance assessment by statistical data analysis in an automated, high-throughput metrology line, AIP Conf. Proc. 2826 (2023), 030014, https://doi.org/10.1063/5.0141831.