Journal of Applied Research and Technology

www.jart.icat.unam.mx

Instituto de Ciencias Aplicadas y Tecnología

Journal of Applied Research and Technology 21 (2023) 858-865

Original

Absorber layer thickness as a new feature in statistical learning tools of Perovskite solar cells

J. Vélez* • F. Sepúlveda • M. Botero • C. Otalora • C. Camacho

Escuela de Ingenierías Eléctrica, Electrónica y de Telecomunicaciones,
Universidad Industrial de Santander, Bucaramanga, Colombia

Received 09 02 2022; accepted 12 05 2022
Available 10 31 2023

Abstract: Recently, the development of Perovskite-based solar cells has emerged as a technological alternative to photovoltaic generation with a higher efficiency/cost ratio. Many contributions have been made in recent years, as evidenced by many academic publications with worldwide experimental results in this area. Machine learning as a tool can support the development of this technology by predicting new materials and discovering novel solar cell configurations. However, the implementation of these methods implies the selection of suitable descriptors. In the present work, we analyze the statistical relationship between the thickness of the absorber layer and solar cell performance parameters. We evaluated the use of the absorber layer thickness as a descriptor in a linear regression model using a database of 221 literature records containing information on the bandgap, the ΔHOMO (Perovskite-HTL), and ΔLUMO (Perovskite-ETL) of different Perovskite cells, together with the thickness of the absorber layer. By building two multiple linear regression models, including or not the thickness of the absorber layer, a reduction in the root means square error of 4.4% and 2.8% was found in the prediction of the Jsc and PCE, respectively. By applying a linear regression model, an improvement in the prediction of Jsc can be seen due to the inclusion of thickness as a descriptor, which is in line with the high value of the mutual information measure we found between the thickness and Jsc.

Keywords: Perovskite solar cell, Machine learning, Thickness, Mutual information

*Corresponding author.
E-mail address: jeisson2169094@correo.uis.edu.co (J. Vélez).
Peer Review under the responsibility of Universidad Nacional Autónoma de México.

1. Introduction

Photovoltaic devices are one of the most important technologies for renewable, clean, and low-cost energy generation. Perovskite-based cells have emerged as a promising, low-cost alternative with higher efficiencies among the different photovoltaic device technologies. In the last decade, Perovskite solar cells have attracted the attention of materials science researchers, and significant advances in terms of efficiency have been observed. In 2009, Perovskite cells reported efficiencies of around 3.8%; since then, this has grown to 25.5% (Best Research-Cell Efficiency Chart, n.d.), surpassing even the performance of consolidated photovoltaic technologies such as those based on CdTe, CIG, and even polycrystalline silicon. These crucial advances suggest a role for Perovskite cells in the future of the PV industry. Despite the significant advances in this area, numerous contributions are still being made. They are reflected in many publications per year, which contain valuable information and data that can be used to develop new materials or structures of Perovskite-based solar cells. As researchers in this area generate data, new approaches are open for discovering and designing materials with improved properties using data-driven methods for knowledge discovery (machine learning) (Odabaşı et al., 2019; Yılmaz & Yıldırım, 2021; Zhou et al., 2019).

Machine learning is an increasingly explored alternative in materials science for developing Perovskite-based solar cells. But exploiting this wealth of latest information requires collecting data for training and evaluating the proposed models. These data may come from experimental results or theoretical calculations. The generation of theoretical data does not consider experimental factors inherent to the synthesis processes of thin films; therefore, experimental information can be considered the most convenient source of data for the generation of models. Previous studies using machine learning to predict the performance of solar cells have used data from theoretical calculations (Balachandran, Kowalski, et al., 2018; Gladkikh et al., 2020; Takahashi et al., 2018) and from experimental studies (Lu et al., 2018, 2019; Odabaşı et al., 2019; Stanley & Gagliardi, 2019; Wu & Wang, 2019; Yu et al., 2019). However, collecting experimental data is costly unless we take such data from information available in academic literature.

Recently, machine learning tools have been used to estimate and predict the bandgap of the absorber layer (Chaube et al., 2020; Gladkikh et al., 2020); to search for new photovoltaic Perovskites (Balachandran, Emery et al., 2018; Lu et al., 2019; Pilania et al., 2016; Takahashi et al., 2018; Zhang et al., 2020); and, to find trends in the way different cell layer composites are grouped concerning the efficiency value (Li et al., 2019; Odabaşı et al., 2019; Xu et al., 2018). In these reported works, descriptors such as the types of compounds used in the different layers of the cell, the different methods used for the synthesis of the layers, and the annealing times and temperatures have been used. Another descriptor that can be particularly useful in prediction models corresponds to the thickness of the absorbing layer, which, having an inverse relationship with the absorption coefficient, determines to a significant extent, the capacity of a material to absorb photons and generate electrons. However, in previous works, the thickness of the absorbing layer as an input parameter has yet to be reported (Balachandran, Emery et al., 2018; Chaube et al., 2020; Gladkikh et al., 2020; Li et al., 2019; Lu et al., 2019; Odabaşı et al., 2019; Pilania et al., 2016; Takahashi et al., 2018; Velez Sanchez et al., 2022; Zhang et al., 2020).

The present work analyses (from a statistical point of view) the relationship between the thickness of the absorber layer and the main performance characteristics of Perovskite solar cells. In particular, the mutual information measure is employed to quantify the degree of nonlinear statistical relationship between the descriptors and the performance values of the cell. Then, we quantify the degree of contribution of the thickness in the decrease of the estimation error for predicting the electrical characteristics of Perovskite solar cells.

2. Method

Through the method described below, the impact of the thickness of the absorber layer on the estimation of the main electrical characteristics of Perovskite cells is analyzed. For this purpose, the automatic learning approach uses mutual information and multiple linear regression statistical analysis tools. The multiple linear regression was performed using cross-validation, for which the data were divided as follows: 90% of the data was for training and the remaining 10% for validation. The validation was performed successively until all the data were considered for training and validation, giving ten iterations for the multiple linear regression with different input data.

2.1. Data

In this work, we take the data reported and used in the supplementary information of (Li et al., 2019) as a source of information, whose database consists of 333 records of values taken from research articles. The descriptors used are the composition of the absorbing layer, the bandgap, the difference between the highest energy-occupied molecular orbitals (HOMO) of the HTL layer and the absorbing layer (ΔHOMO), and the difference between the lowest energy-unoccupied molecular orbitals (LUMO) of the absorbing layer and the ETL layer (ΔLUMO). Also, electrical characteristics are included, the open-circuit voltage Voc, the short-circuit current density Jsc and the fill factor FF. To analyze the relevance of thickness, this value was manually extracted from each scientific article. In those cases where this value was not reported, the authors were contacted directly via e-mail. In total, 221 thicknesses were obtained, so our dataset is limited to this number. This dataset is published in (Velez Sanchez et al., 2022).

2.2. Descriptors

Selecting descriptors is a task of significant importance when applying automatic learning methods. Descriptors must have an implicit meaning and be available, which requires that they be easily reportable and universally reported by authors in scientific papers to be considered. Ideally, the number of descriptors should be reduced to avoid over-fitting problems; and not contain redundant information. However, guaranteeing full compliance with the above conditions is difficult. Recent work has used variables such as bandgap, ΔHOMO, and ΔLUMO to represent the information from which cell characteristics could be predicted (Li et al., 2019; Odabaşı et al., 2019). In addition, in works such as those presented by Gladkikh et al. (2020), Pilania et al. (2016), Takahashi et al. (2018), Zhang et al. (2020), characteristics such as electronegativity, the number of atomic orbitals, the Goldschmidt tolerance factor of the elements that make up the absorbing layer are used to perform prediction or classification tasks in machine learning algorithms. However, no works were found in which the thickness of the absorbing layer is included as a descriptor. In the present work, in addition to the variables already mentioned in previous works, it is desired to use the thickness of the absorbing layer as an input characteristic of the model in charge of estimating the performance in the prediction of output electrical variables such as the short circuit voltage Voc, short circuit current Jsc, the filling factor FF and PCE. In addition to the inclusion of the thickness, a modification was made to the database used, which consisted of changing the coding for the compounds that form the Perovskite, which went from 8 to three variables that were as follows: $A = MA - FA - Cs$, which groups the compounds used in the Perovskite cation A, $B = Pb - Sn$ and $X = I - Br - Cl$. This modification was made to make the models used in the linear regression more flexible, thus lowering the complexity of the final model and reducing the effects of over-fitting on the results obtained.

2.3. Mutual information as a measure of nonlinear statistical association

The mutual information´on between two random variables x, y, denoted as $I(x, y)$, measures the mutual dependence between the two variables; that is, it quantifies the amount of information obtained from one random variable through the observation of the other random variable (Bishop, 2006).

$$I(x, y) = -\int\int p(x, y)\, log\left(\frac{p(x)p(y)}{p(x,y)}\right) dxdy \qquad (1)$$

Where $I(, y) \geq 0$; and $I(x, y) = 0$ for the case where x, y are independent variables. The units in which this measure is expressed depend on the type of logarithm; in particular, if the logarithm is in base two, the mutual information is in units of bits. There are several methods for estimating $I(-,-)$; however, in the present work, the k-neighbors-based method, reported in (Kraskov et al., 2004; Ross, 2014), is used. $I(-,-)$ is calculated between each descriptor $(A, B, X, Eg, \Delta HOMO, \Delta LUMO, \delta)$ and each electric variable of the solar cell $(Voc,\ Jsc,\ FF,\ PCE)$. Those descriptors that offer a low value of mutual information are discarded.

2.4. Multiple linear regression

The regression function is assumed to be linear for the inputs in multiple linear regression. In our case for $A, B, X, Eg, \Delta HOMO, \Delta LUMO,$ and $f(\delta)$. To include the possibility of a nonlinear relationship $f(-)$, in the present work, we evaluated several options with the help of δ vs outputs plots of the collected data. We can obtain nonlinear models with a proper transformation of inputs or output. Although linear regression models are simple, they can sometimes give similar or better results than nonlinear models. Especially in models with a small number of training data (Hastie et al., 2009). Two prediction models are proposed for each $k = 1, \dots, 4$ performance measure ($Voc, Jsc, FF, and\ PCE$). In the first one, thickness is not considered,

$$\begin{aligned} y(k) = \alpha_0^k + \alpha_1^k A + \alpha_2^k X + \alpha_3^k Eg + \alpha_4^k \Delta HOMO \\ + \alpha_5^k \Delta LUMO + \varepsilon \\ = Z - \alpha^{(k)} + \varepsilon \end{aligned} \qquad (2)$$

Furthermore, in a second model, if the thickness is included.

$$\begin{aligned} y_\delta^{(k)} = \beta_0^k + \beta_1^k A + \beta_2^k X + \beta_3^k Eg + \beta_4^k \Delta HOMO \\ + \beta_5^k \Delta LUMO + \beta_6^k f(\delta) + \varepsilon_\delta \\ = Z\delta - \beta^{(k)} + \varepsilon_\delta \end{aligned} \qquad (3)$$

Where $Z = [1\ A\ X\ Eg\ \Delta HOMO\ \Delta LUMO]$; $Z_\delta = [1\ A\ X\ Eg\ \Delta HOMO\ \Delta LUMO\ f(\delta)]$; $f(-)$ is some nonlinear transformation´on of the descriptor δ, and $\varepsilon - \varepsilon_\delta$ are prediction errors´on. To estimate the degree of contribution of the variable δ, we compare the performance of the above two models $y^{(k)}$and $y_\delta^{(k)}$ for each kth electrical property. The parameters $\alpha^{(k)}$and $\beta^{(k)}$ are found using the least-squares criterion

3. Results

3.1. Estimation of the degree of statistical-linear association of descriptors

Table (1) reports the estimated bitwise values of the statistical association $I(-,-)$ for different pairs of descriptors vs electrical characteristics. This value was calculated for each pair of variables and 20 different subsets of 177 samples out of 221 (80% of the total), where samples were randomly selected without replacement. Each value reported in Table (1) results from the average of the 20 values obtained for each of the 20 subsets. This strategy allowed us to estimate the standard error associated with each of the estimates and thus develop a t-student hypothesis test to determine whether the estimated values are statistically different from zero. That is, to determine if statistically there is an association. As a result, it is obtained that all except 2 are statistically different from zero. The results show that the absorbing layer's thickness helps explain the behaviour of the Jsc inside the cell. They also allow us to infer that the results obtained for B, formed by $B = Pb - Cs$, are a product of a large amount of data in which B is lead. Because of this, the metric used cannot perceive the importance of B within the linear regression models.

Table 1. The rows correspond to film descriptors,
and the columns to performance indices.
The six most significant values (≥ 0.200) are shown in bold.

	PCE	Voc	Jsc	FF
A	**0.241**	**0.327**	0.172	0.016
B	0.0	0.042	0.038	0.0
X	**0.340**	**0.262**	0.157	0.060
E_g	0.071	**0.238**	0.157	0.069
$\Delta HOMO$	0.054	0.160	0.066	0.052
$\Delta LUMO$	0.081	0.051	0.111	0.022
Thickness	0.067	0.139	**0.264**	0.026

From the results obtained and summarized in Table (1), the following behaviors are observed, among others: i) thickness as a descriptor plays a significant role in describing the behavior of Jsc, being this the descriptor with the most significant contribution to this electrical characteristic. ii) Under this same working model, descriptors A, X and Eg are the ones that have the most significant contribution to Voc. iii) A comparable situation is found for efficiency (PCE), which presents a contribution from descriptors A and X. These results are consistent with the physical explanation of the phenomenon of photoconversion of radiation into electrical energy in a photovoltaic device. The thickness of the Perovskite layer, playing an indispensable role in photogeneration, influences the generation rate G described in the models reported in the literature (Le Corre et al., 2019). Similarly, the thickness of this absorbing layer affects the transport process of photogenerated carriers to the selective transport layers (ETL and HTL). Hence, as a parameter, it influences the carrier recombination rate (Le Corre et al., 2019). Likewise, Voc and PCE as electrical characteristics are strongly affected by the Eg of the absorber layer (Jarosz et al., 2020), which in turn depends strongly on the chemical composition of the Perovskite, which is described by A, B, and X (Kato et al., 2017).

3.2. Estimation of the performance values

Figure (1) shows the comparison graphs of the descriptor δ versus the main electrical characteristics of the devices. Figure (1c) shows the nonlinear relationship between the thickness δ and Jsc. This nonlinear relationship between thickness and parameters such as Jsc or PCE has been previously described in theoretical reports on this type of device (Le Corre et al., 2019; Sha et al., 2015). When calculating the linear correlation value (Pearson's correlation) between δ and Jsc, 0.28 is obtained; however, when applying the transformation of the form $f(\delta) = \sqrt{\delta}$, this linear correlation value becomes 0.35. Therefore, in the present work, it is preferred to use the transformed variable $\sqrt{\delta}$ as input.

The *Jsc*, *Voc*, *FF*, and *PCE* characteristics were estimated using multiple linear regression. For each electrical variable to be estimated, two models were applied by selecting those described in equations 2 and 3 belonging to the *Z* and Z_δ sets, respectively. Due to its low level of nonlinear correlation reported in Table 1, Descriptor B was not considered in any of these models. To estimate the performance of the models, typically measured in terms of root mean square error (RMSE), a cross-validation procedure of 10 partitions is used, where each partition is formed by a training subset of the linear regression model of 199 data, and the remaining 22 are used to measure the performance of the same model. This process is repeated ten times, once for each different participation. Table (2) reports the performance results regarding the mean square error.

On the other hand, to determine whether the differences in performance in terms of RMSE are statistically significant, a two-sample t-student test is applied. The results show that the inclusion of the thickness of the absorbing layer (√δ) as a descriptor improves the prediction of the *Jsc* and PCE values. The thickness information, transformed in the √δ form, provides relevant information for Jsc and *PCE* estimation purposes. As a result, it is found that for the case of *Jsc* and PCE, the improvement in performance by adding the thickness.

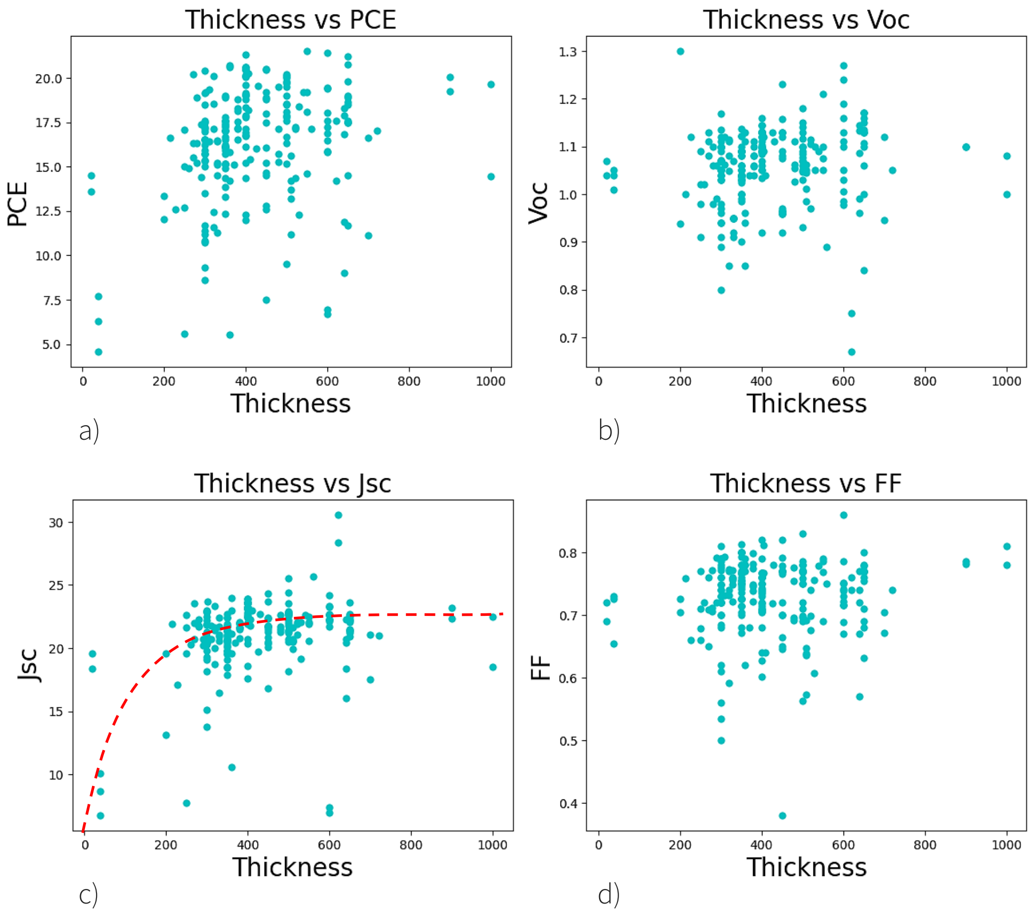

Figure 1. Example of scatter plots of some inputs versus some outputs.
a) Thickness vs PCE, b) thickness vs Voc, c) thickness vs Jsc, d) thickness vs FF.

The *Jsc*, *Voc*, *FF*, and *PCE* characteristics were estimated using multiple linear regression. For each electrical variable to be estimated, two models were applied by selecting those described in equations 2 and 3 belonging to the *Z* and Z_δ sets, respectively. Due to its low level of nonlinear correlation reported in Table 1, Descriptor B was not considered in any of these models. To estimate the performance of the models, typically measured in terms of root mean square error (RMSE), a cross-validation procedure of 10 partitions is used, where each partition is formed by a training subset of the linear regression model of 199 data, and the remaining 22 are used to measure the performance of the same model. This process is repeated ten times, once for each different participation. Table (2) reports the performance results regarding the mean square error.

On the other hand, to determine whether the differences in performance in terms of RMSE are statistically significant, a two-sample t-student test is applied. The results show that the inclusion of the thickness of the absorbing layer ($\sqrt{\delta}$) as a descriptor improves the prediction of the *Jsc* and PCE values. The thickness information, transformed in the $\sqrt{\delta}$ form, provides relevant information for Jsc and *PCE* estimation purposes. As a result, it is found that for the case of *Jsc* and PCE, the improvement in performance by adding the thickness.

To visualize this improvement, we have compared the predictions obtained vs the real values for the four models. Figure 2 shows the behaviors of the estimated values. Of these four models, *Jsc's* prediction is the best performing.

Table 2. Estimation results of Jsc, Voc, FF, and PCE using linear regression. We also report the improvement in performance by including the squared root of thickness √δ as a descriptor. The value in parenthesis corresponds to the standard deviation of the measurement.

	Using Z	Using Z_δ	Improvement
Voc	0.07	0.07	- -
Jsc	22.69 (0.0028)	21.68 (0.0013)	4.44%
FF	6.08	6.07	- -
PCE	2.88 (0.004)	2.79 (0.002)	2.88%

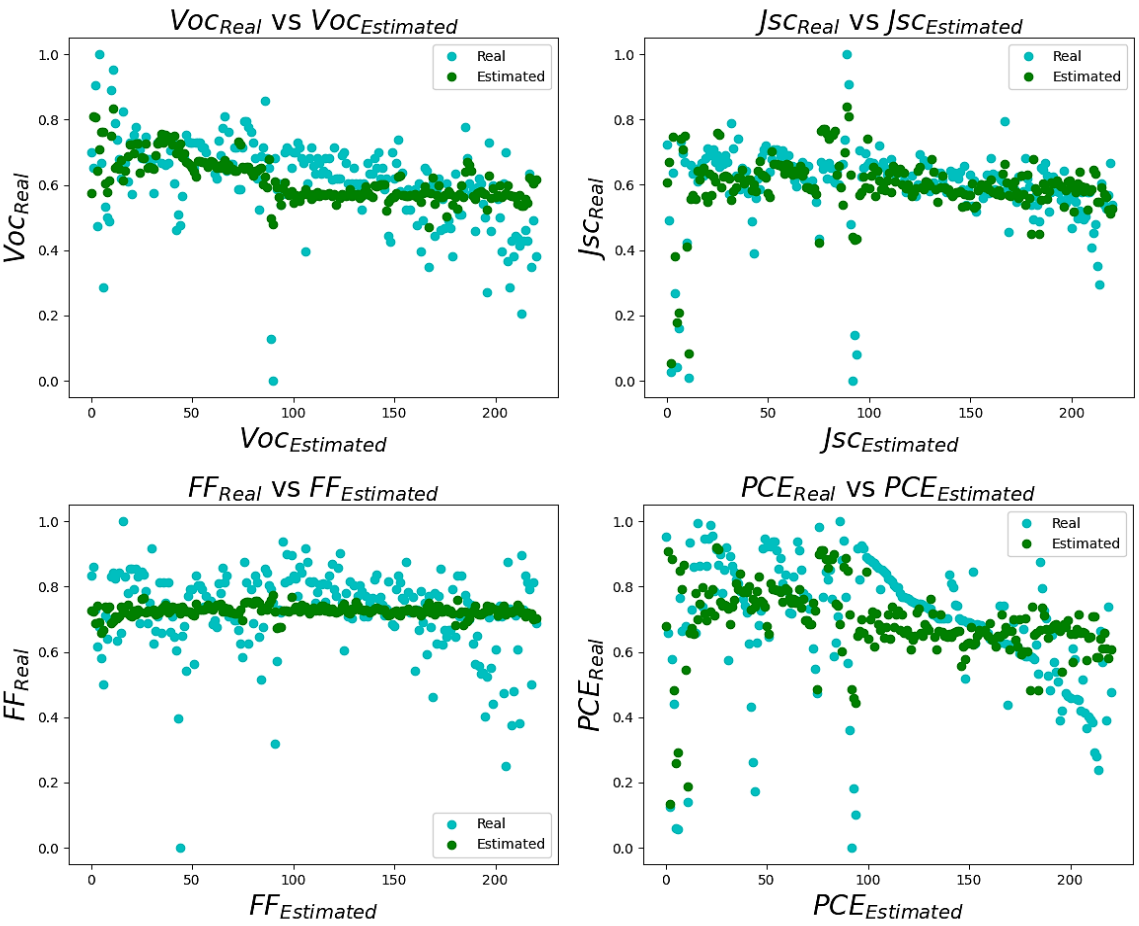

Figure 2. Estimated values vs real values plot for linear regression model outputs. a) Voc_{real} vs $Voc_{estimated}$, b) Jsc_{real} vs $Jsc_{estimated}$, c) FF_{real} vs $FF_{estimated}$, d) PCE_{real} vs $PCE_{estimated}$.

4. Conclusions

The results obtained show that the thickness of the absorbing layer is a relevant variable in predicting electrical measures of performance. It could be inferred that the thickness is related to electrical variables, especially with the *Jsc*, for which higher improvement percentages were observed. An explanation for this fact is that the thickness of the absorbing layer influences the process of photogeneration and carrier transport. On the other hand, these results show that the inclusion of new features, such as thickness, can positively influence the models used in machine learning tasks. Unfortunately, it requires the scientific community to report these parameters in research articles.

On the other hand, extracting new variables from the available scientific literature is a time-consuming task that could be improved through natural language processing tools or by standardizing the reporting formats of solar cell parameters published in academic reports.

Conflict of interest

The author(s) has(have) no conflict of interest to declare.

Acknowledgements

The authors are grateful for the financial support the Universidad Industrial de Santander provided through project 2523 assigned to the Vicerrectoria de Investigación y Extension-VIE.

Funding

The Universidad Industrial de Santander provided financial support through project 2523, assigned to the Vicerrectoria de Investigación y Extension-VIE.

References

Balachandran, P. V., Emery, A. A., Gubernatis, J. E., Lookman, T., Wolverton, C., & Zunger, A. (2018). Predictions of new AB O 3 perovskite compounds by combining machine learning and density functional theory. *Physical Review Materials*, *2*(4), 043802.
https://doi.org/10.1103/PhysRevMaterials.2.043802

Balachandran, P. V., Kowalski, B., Sehirlioglu, A., & Lookman, T. (2018). Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning. *Nature Communications*, *9*(1), 1–9.
https://doi.org/10.1038/s41467-018-03821-9

Best Research-Cell Efficiency Chart | Photovoltaic Research | NREL. (n.d.). Retrieved September 3, 2023, from https://www.nrel.gov/pv/cell-efficiency.html

Bishop, C. M. (2006). Pattern recognition and machine learning (Vol. 4, No. 4, p. 738). New York: Springer.

Chaube, S., Khullar, P., Goverapet Srinivasan, S., & Rai, B. (2020). A Statistical Learning Framework for Accelerated Bandgap Prediction of Inorganic Compounds. *Journal of Electronic Materials*, *49*(1), 752–762.
https://doi.org/10.1007/s11664-019-07779-2

Gladkikh, V., Kim, D. Y., Hajibabaei, A., Jana, A., Myung, C. W., & Kim, K. S. (2020). Machine Learning for Predicting the Band Gaps of ABX3 Perovskites from Elemental Properties. *Journal of Physical Chemistry C*, *124*(16), 8905–8918. https://doi.org/10.1021/acs.jpcc.9b11768

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. Springer New York. https://doi.org/10.1007/978-0-387-84858-7

Jarosz, G., Marczyński, R., & Signerski, R. (2020). Effect of band gap on power conversion efficiency of single-junction semiconductor photovoltaic cells under white light phosphor-based LED illumination. *Materials Science in Semiconductor Processing*, *107*, 104812. https://doi.org/10.1016/j.mssp.2019.104812

Kato, M., Fujiseki, T., Miyadera, T., Sugita, T., Fujimoto, S., Tamakoshi, M., ... & Fujiwara, H. (2017). Universal rules for visible-light absorption in hybrid perovskite materials. *Journal of Applied Physics*, *121*(11).
https://doi.org/10.1063/1.4978071

Kraskov, A., Stögbauer, H., & Grassberger, P. (2004). Estimating mutual information. *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, *69*(6), 16.
https://doi.org/10.1103/PhysRevE.69.066138

Le Corre, V. M., Stolterfoht, M., Perdigon Toro, L., Feuerstein, M., Wolff, C., Gil-Escrig, L., ... & Koster, L. J. A. (2019). Charge transport layers limiting the efficiency of perovskite solar cells: how to optimize conductivity, doping, and thickness. *ACS Applied Energy Materials*, *2*(9), 6280-6287.
https://doi.org/10.1021/acsaem.9b00856

Li, J., Pradhan, B., Gaur, S., & Thomas, J. (2019). Predictions and Strategies Learned from Machine Learning to Develop High-Performing Perovskite Solar Cells. *Advanced Energy Materials*, *9*(46), 1901891.
https://doi.org/10.1002/aenm.201901891

Lu, S., Zhou, Q., Ma, L., Guo, Y., & Wang, J. (2019). Rapid Discovery of Ferroelectric Photovoltaic Perovskites and Material Descriptors via Machine Learning. *Small Methods*, *3*(11), 1900360.
https://doi.org/10.1002/smtd.201900360

Lu, S., Zhou, Q., Ouyang, Y., Guo, Y., Li, Q., & Wang, J. (2018). Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nature Communications*, *9*(1).
https://doi.org/10.1038/s41467-018-05761-w

Odabaşı Özer, Ç., & Yıldırım, R. (2019). Performance analysis of perovskite solar cells in 2013–2018 using machine-learning tools. *Nano Energy*, *56*(November 2018), 770–791. https://doi.org/10.1016/j.nanoen.2018.11.069

Pilania, G., Balachandran, P. V., Kim, C., & Lookman, T. (2016). Finding new perovskite halides via machine learning. *Frontiers in Materials*, *3*.
https://doi.org/10.3389/fmats.2016.00019

Ross, B. C. (2014). Mutual information between discrete and continuous data sets. *PLoS ONE*, *9*(2), e87357. https://doi.org/10.1371/journal.pone.0087357

Sha, W. E. I., Ren, X., Chen, L., & Choy, W. C. H. (2015). The efficiency limit of CH3NH3PbI3 perovskite solar cells. *Applied Physics Letters*, *106*(22), 221104. https://doi.org/10.1063/1.4922150

Stanley, J., & Gagliardi, A. (2019). Machine Learning Bandgaps of Inorganic Mixed Halide Perovskites. *Proceedings of the IEEE Conference on Nanotechnology*, *2018-July*. https://doi.org/10.1109/NANO.2018.8626420

Takahashi, K., Takahashi, L., Miyazato, I., & Tanaka, Y. (2018). Searching for Hidden Perovskite Materials for Photovoltaic Systems by Combining Data Science and First Principle Calculations. *ACS Photonics*, *5*(3), 771–775. https://doi.org/10.1021/acsphotonics.7b01479

Velez Sanchez, J. E., Sepúlveda-Sepúlveda, A., & Botero, M. A. (2022). Absorber layer thickness as a new feature in statistical learning for predicting electrical measures of performance on perovskite solar cells, Mendeley Data, V1, https://doi.org/10.17632/bbybhrz4ny.1

Wu, T., & Wang, J. (2019). Global discovery of stable and non-toxic hybrid organic-inorganic perovskites for photovoltaic systems by combining machine learning method with first principle calculations. *Nano Energy*, *66*, 104070. https://doi.org/10.1016/j.nanoen.2019.104070

Xu, Q., Li, Z., Liu, M., & Yin, W. J. (2018). Rationalizing Perovskite Data for Machine Learning and Materials Design. *Journal of Physical Chemistry Letters*, *9*(24), 6948–6954. https://doi.org/10.1021/acs.jpclett.8b03232

Yılmaz, B., & Yıldırım, R. (2021). Critical review of machine learning applications in perovskite solar research. *Nano Energy*, *80*. https://doi.org/10.1016/J.NANOEN.2020.105546

Yu, Y., Tan, X., Ning, S., & Wu, Y. (2019). Machine Learning for Understanding Compatibility of Organic-Inorganic Hybrid Perovskites with Post-Treatment Amines. *ACS Energy Letters*, *4*(2), 397–404. https://doi.org/10.1021/acsenergylett.8b02451

Zhang, L., He, M., & Shao, S. (2020). Machine learning for halide perovskite materials. *Nano Energy*, *78*. https://doi.org/10.1016/j.nanoen.2020.105380

Zhou, T., Song, Z., & Sundmacher, K. (2019). Big Data Creates New Opportunities for Materials Research: A Review on Methods and Applications of Machine Learning for Materials Design. *Engineering*, *5*(6), 1017–1026. https://doi.org/10.1016/j.eng.2019.02.011