

Machine learning for screening and predicting the best surface modifiers for a rational optimization of efficient perovskite solar cells

Cite as: APL Energy 2, 036102 (2024); doi: 10.1063/5.0214736

Submitted: 19 April 2024 • Accepted: 28 August 2024 •

Published Online: 23 September 2024



View Online



Export Citation



CrossMark

Mattia Ragni,¹ Fabiola Faini,¹ Matteo Degani,¹ Silvia Cavalli,¹ Ian Postuma,^{2,3} and Giulia Grancini^{1,a)}

AFFILIATIONS

¹ Department of Chemistry and INSTM University of Pavia, Via T. Taramelli 14, 27100 Pavia, Italy

² Istituto Nazionale di Fisica Nucleare, Pavia unit, Via A. Bassi 6, 27100 Pavia, Italy

³ Department of Physics, University of Pavia, Via A. Bassi 6, 27100 Pavia, Italy

^{a)}Author to whom correspondence should be addressed: giulia.grancini@unipv.it

ABSTRACT

The key to keep the rising slope of perovskite solar cell performances is to reduce non-radiative losses by minimizing defect density. To this end, a large variety of strategies have been adopted spanning from the use of interfacial layers, surface modifiers, to interface engineering. Although winning concepts have been demonstrated, they result from a mere trial and error approach, which is time consuming and operator-dependent. To face this challenge, in this work, we propose the use of a machine learning approach for an educated and rational material screening with optimal characteristics in terms of surface passivation. In particular, we applied Shapley additive explanation to extract the specific chemical features of the passivator, which directly impact the device parameters, specifically the open circuit voltage (V_{oc}). By monitoring the different material parameters as input, we were able to list the most promising passivators and directly test them in working solar cells. By comparing the device performances with the results of the modeling and with additional optical and morphological characterization, we retrieved the most significant material properties linked to the highest efficiency, which are (i) the presence of chlorine and its strong binding capacity to positively charged defects on perovskite surface, reducing the non-radiative recombination and (ii) an increased flexibility of the molecule, resulting in better coverage of the surface. Finally, we tested the predictive power of the ML algorithm proposing a new passivator, which, implemented in a working device, leads to the predicted high V_{oc} confirming the results of the modeling.

© 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International (CC BY-NC-ND) license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>). <https://doi.org/10.1063/5.0214736>

04 January 2025 09:52:37

I. INTRODUCTION

In the past decade, hybrid halide perovskites (HPs) have stood out as frontrunners in the emerging field of photovoltaics (PVs) thanks to their remarkable optoelectronic characteristics, leading to a power conversion efficiency (PCE) record of 26.1%.^{1–4} This stems true despite the non-negligible amounts of defects ($>10^{16} \text{ cm}^{-3}$) affecting the non-radiative losses and device open circuit voltage (V_{oc}).^{5–8} To date, one of the most effective methods to limit the defective nature of HPs relies on surface defects passivation by means of surface modifiers. They include aromatic compounds,^{9,10} alkyl compounds,^{11,12} and organic salts with fluorine atoms or sulfur-rich functional groups.^{13–15} However, despite few examples

providing rational guidelines on the link between the chemical properties of the perovskite surface and the device V_{oc} ,^{16–18} the device optimization is still mainly driven by the trial-and-error approach. In this context, we move away from the current methods by developing and applying a machine learning (ML) approach to provide a rational screening of molecular species for an educated device optimization. On the one hand, physics-informed ML workflow is rising to speed up the discovery process.^{17,19} On the other hand, the experimental-driven ML workflow is essential to find correlation between variables in a database.²⁰ In particular, Hartono *et al.* and, more recently Zhi *et al.*, pioneered ML methods for the optimization of organic cations for the passivation of the perovskite active layer, Kouroudis *et al.* have used a ML pipeline to

predict the outdoor degradation behavior of PSCs, and finally, Kirman *et al.* have used ML to guide their robotic synthetic trials.^{20–23} However, experimental-driven ML workflow requires big efforts in terms of time and materials. Here, we try to close the gap between the big amount of data required for a suitable database for the ML approach and reliable ML results by considering both different organic salts for passivation and different experimental processing conditions.

In this work, we propose a multistep approach, where different ML models are trained for the choice of the best organic cations for the passivation of formamidinium lead iodide (FAPbI_3) perovskite solar cells, choosing as a target variable to enhance the open circuit voltage (V_{oc}). Through this procedure and by applying the Shapley additive explanations (SHAP) values method, we first identified the most influential processing variables and chemical properties that the organic cation should possess for the maximization of the device V_{oc} . Second, we obtained a reliable prediction on a new candidate cation, which we directly tested as an FAPbI_3 passivator in a working device. The proposed approach not only provides critical insight into the key properties of the organic cations crucial for achieving high V_{oc} but also enables a deeper understanding of the physics behind the passivation mechanism. Overall, this method paves the way for a smart selection of a new potential organic passivator by screening with ML their chemical properties.

II. RESULTS AND DISCUSSION

A. Model for organic cation analysis

For the material screening, we decided to consider nine organic salts among the most used as passivator materials for FAPbI_3 in n-i-p PSCs architecture, i.e., (I) 4-methylphenethylammonium iodide (MePEAI), (II) 4-methylphenethylammonium bromide (MePEABr), (III) 4-methylphenethylammonium chloride (MePEACl), (IV) n-butylammonium iodide (n-BAI), (V) iso-butylammonium iodide (iso-BAI), (VI) n-octylammonium iodide (n-OAI), (VII) benzylammonium bromide (BBr), (VIII) hexylammonium bromide (HBr), and (IX) octylammonium tosylate (OaTos). Figure 1(a) shows their molecular structures. Every compound was deposited on top of the FAPbI_3 using different concentrations (5, 10, and 15 mM) and temperatures (with or without annealing step at 100 °C), setting six different processing conditions. To evaluate the relationship between the V_{oc} , the processing variable, and the chemical properties of the salts, we applied the workflow shown in Fig. 1(b), testing the performance of various ML models coupled to a features importance ranking method. After the production of three samples for each combination of passivators and processing variables, a total of 162 samples were obtained and measured under a solar simulator. To validate the measurements, we develop, as control, solar cells with no passivation (namely, FAPbI_3) and cells passivated with MePEABr (our reference passivator, namely, $\text{FAPbI}_3 + \text{MePEABr}$) (Fig. S1). We chose the device V_{oc} as the target variable (considering only values exceeding 1000 mV). Subsequently, the filtered values were averaged, providing the average V_{oc} for the fifty-four combinations (Fig. S2). Then, the ML models were trained by using the chemical properties of the cations and two processing variables as input features, while the corresponding measured V_{oc} was set as the output variable. A set of sixteen chemical properties (computed via

CACTVS toolkit²⁴) concatenated with the two processing variables were initially selected. To enhance the model performance, mitigate model overfitting, decrease computational costs, and promote model interpretability, we conducted a preliminary selection of the input features by developing a correlation matrix before the training phase of the models. The square matrix, with its rows and columns representing input features, contains correlation coefficients (Pearson factors, P_f) in each cell. These coefficients measure the linear correlation and anticorrelation between features, denoted by values between +1 and -1, respectively. The computed matrix facilitated the screening of redundant variables by focusing on pairs with a Pearson factor lower than 0.9 (Fig. S3), allowing us to reduce the final set of chemical features to nine: molecular weight, rotatable bond count, rigid bond count, complexity (ensemble complexity according to modified Bertz/Hendrickson algorithm^{25,26}), halide fraction (fraction of molecular weight contributed by halogen atoms), heteroatom carbon ratio (number of atoms other than carbon and hydrogen, divided by the number of carbon atoms), iodine, bromine, and chlorine atom counts (Table S1). Furthermore, the selected input features were normalized between 0 and 1 (Fig. S4). This is a standard procedure to mitigate the potential influence of varying amplitudes of input features on the model training. Six models were independently employed in this study, including linear regression (LR) serving as the minimum performance threshold, two geometric models, i.e., k-nearest neighbor regression (KNN) and support vector machine regression (SVM), and two ensemble models based on decision trees, i.e., random forest regression (RF) and gradient boosting regression (GB). In addition, a neural network (NN) was implemented. To determine the optimal hyperparameters for each model, we employed a grid search (GS) with fivefold cross-validation (Fig. S5), ensuring the full utilization of the available dataset and mitigating the risk of overfitting on the test set, as it would occur with a direct split into training and test sets. To select the model with the best generalization, we employed an all-but-one (AbO) cross-validation approach (Fig. S6). Model training was conducted on $n - 1$ organic cations data, applying GS to identify the best hyperparameters while keeping the data of the n th cation as test set. At the end of each iteration, the root mean square error (RMSE) and standard deviation of each model were saved. This process continued iteratively, excluding from the training set a different cation at each cycle, until all n cations were evaluated. Finally, the model exhibiting the lowest average RMSE and standard deviation, i.e., RF was identified as the best-performing model, as shown in Fig. 1(c). Upon selecting the best model and training it on the full dataset, we applied the SHAP values method to extract the weights that the model assigns to the values of the input features in computing its predictions.²⁷ The method offers an interpretable strategy, unveiling the influence of each feature on the target variable.

B. Features interpretation

Figure 2(a) shows the SHAP values derived from the interpretation of the selected optimal model, particularly the RF. In particular, a positive/negative SHAP value indicates that the input property with the given value (indicated by color: yellow for high values and purple for low values) positively/negatively affects the output

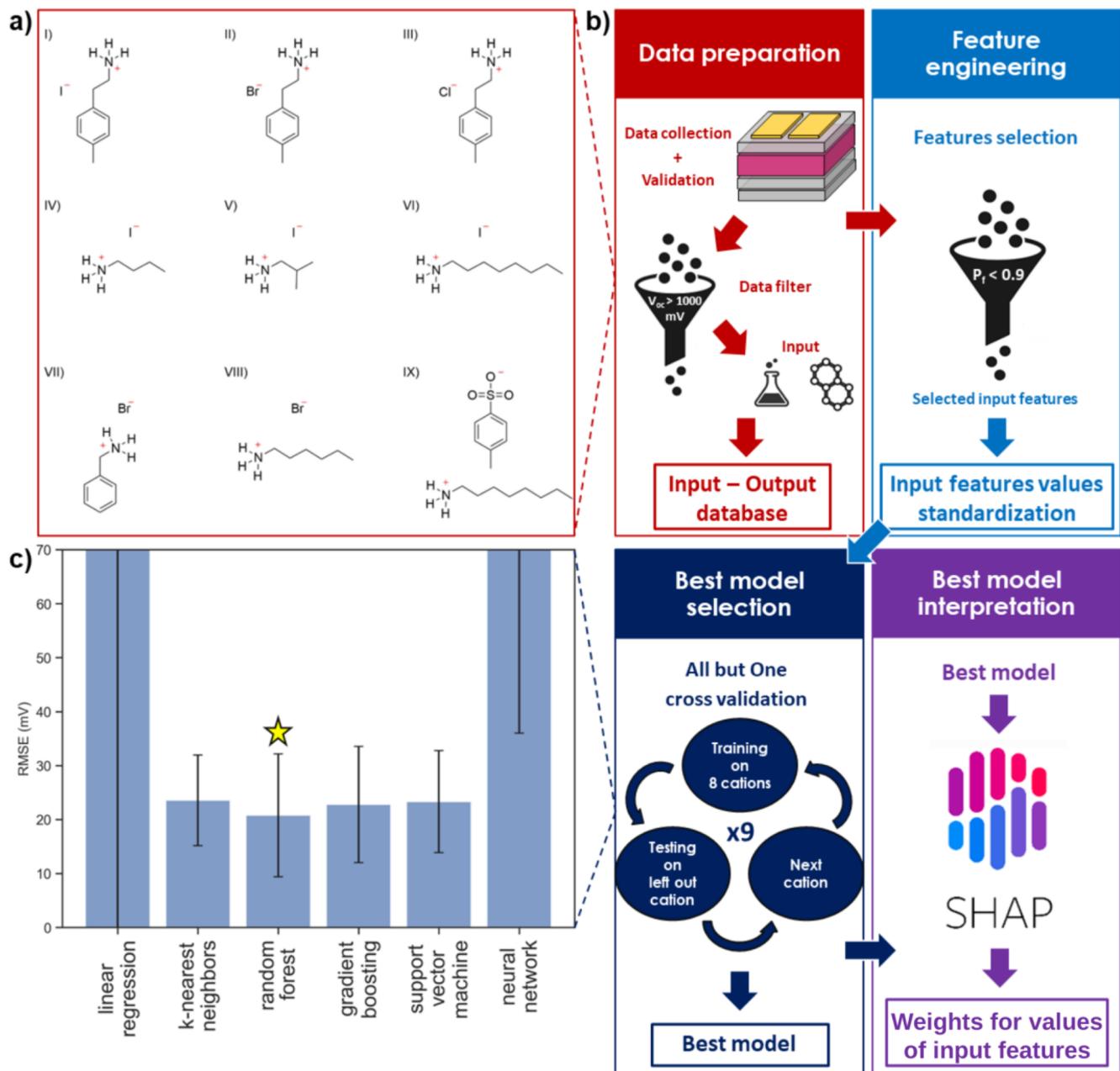


FIG. 1. (a) Organic salts used in the passivation of the FAPbI_3 active layer, i.e., (I) MePEAI, (II) MePEABr, (III) MePEACl, (IV) n-BAI, (V) iso-BAI, (VI) n-OAI, (VII) BBr, (VIII) HBr, and (IX) OaTos. (b) Flow chart of the applied machine learning approach with the different summarized steps: data preparation, feature engineering, best model selection, and best model interpretation. (c) RMSEs of trained models, i.e., linear regression, k-nearest neighbors, random forest, gradient boosting, support vector machine, and neural network, obtained from all but one cross-validation.

variable (V_{oc}), while the modulus indicates the weight of this influence. The most relevant and clear feature for the V_{oc} optimization is the negligible weight of the annealing step. On the other side, low/medium halide fraction, low molecular weight and heteroatom carbon ratio, and high rotatable bonds are predicted to be beneficial

for pushing the device V_{oc} . Ultimately, the presence of an atom of I seems to be detrimental, while the presence of Cl or Br halogen is beneficial to maximize the V_{oc} . The concentration parameter trend instead is more complex. To explain these results and find a link between the different variables, we considered the P_f matrix,

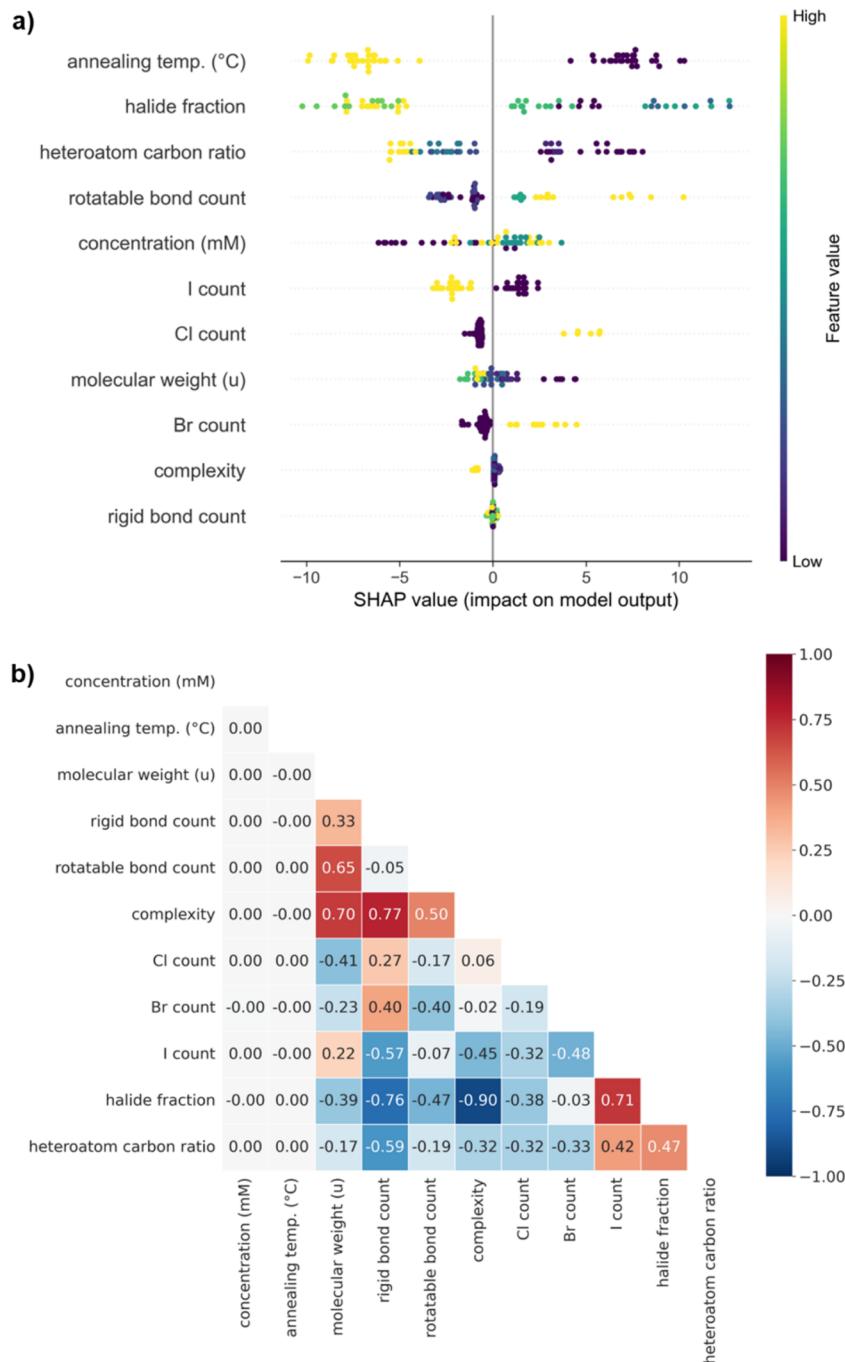


FIG. 2. (a) SHAP values of the best model, i.e., RF. (b) Pearson correlation factors between the key features of the organic cations.

as shown in Fig. 2(b). First, we observe that the processing conditions, namely, the annealing temperature and the concentration, evidence a P_f with any chemical features between -0.02 and $+0.02$ testifying that the processing variables are independent from the cation chemical properties. Regarding the chemical properties of the

cations, we combine them into two groups according to their correlation/anticorrelation values. The first group contains the main chemical properties, i.e., the absence/presence of I, Br, and Cl atoms; the molecular weight; and the halide fraction. In particular, iodine with its high atomic weight of 126.904 u (bromine and chlorine

have 79.904 and 35.453 u, respectively) strictly affects the overall molecular weight of the organic cation (for MePEAI, the weight of the halogen counts to ~48% to the total molecular weight), whose high value is detrimental for the V_{oc} . On the contrary, for Cl-based organic cations, the reduction in the weight of the halogen is reflected in the low molecular weight of the cation and the maximization of the V_{oc} (for MePEACl, the halogen/molecular weight is only 21%). These results allow us to explain that a low halide fraction, which is exactly the fraction of molecular weight contributed by halogen atoms, is beneficial for the V_{oc} . Finally, we grouped the heteroatom carbon ratio, the C and H counts, and the rotatable bond count (Fig. S3). Excluding OaTos, all the cations explored have a number of atoms other than C and H equal to two (halogen and N). So, the heteroatom carbon ratio, which is the number of heteroatoms divided by the number of carbon atoms, is strictly related to the number of C, consistent with the anticorrelation between the heteroatom carbon ratio and C count observable from $P_f = -0.62$. On the other hand, the count of rotatable bonds shows a high correlation with the counts of C and H ($P_f = 0.53$ and $P_f = 0.90$, respectively). The correlation/anticorrelation between rotatable bonds count/heteroatom carbon ratio with C (and H for the case of rotatable bonds) counts suggests that we can explain the hard to comprehend heteroatom carbon ratio through the rotatable bonds count. Following these considerations, we have three groups of variables shown in Fig. 2(c): processing variables (yellow circle), first chemical features group (light blue triangle), and second chemical features group (dark blue square). The annealing temperature emerges as the most influential factor, exhibiting a clear trend wherein no annealing step proves to be more efficient in yielding higher V_{oc} compared to annealing at 100 °C. Since the influence of concentration on V_{oc} appears unclear, further data collection is needed to validate this result. Then, lighter halogens yield better performance, corroborated by findings for specific cations with different halogens, properly TMAX with X = I, Br, and Cl while TMA is 2-thiophenemethylammonium.¹⁶ Indeed, from I to Cl, the increased electronegativity is accompanied by stronger binding capacity between the anion and positively charged defects at the perovskite surface with a beneficial result for V_{oc} .¹⁷ Finally, a higher number of C and H atoms, forming C–H bonds and contributing to rotatable bonds, results in a more flexible molecule structure. We believe that flexibility enhances the spreading and coverage of the FAPbI₃ surface, leading to more efficient passivation. This conclusion is consistent with prior studies correlating alkyl chain length with PSC performance improvement.¹⁸

C. Material characterizations

To link the most influential features of the organic cations, as defined by the SHAP values analysis and the film properties, we selected MePEACl and MePEABr as the best and iso-BAI as the worst performing material (10 mM and no annealing step). Indeed, we noticed that PSCs with FAPbI₃ + iso-BAI, FAPbI₃ + MePEABr, and FAPbI₃ + MePEACl evidence +1%, +6%, and +7%, respectively, in the mean V_{oc} compared to the V_{oc} of the FAPbI₃ PSCs (Fig. S2). X-ray diffraction (XRD) patterns shown in Fig. S7 evidence a dominant peak for all materials at $2\Theta = 14.3^\circ$ and higher order reflections assigned to α-FAPbI₃, while unstable δ-FAPbI₃ is present only for pure FAPbI₃ and the lead iodide peak, especially for

FAPbI₃ + iso-BAI.^{28–30} In addition, atomic force microscopy (AFM) images (Fig. S8) show an improved surface coverage for MePEABr and MePEACl, ascribed to the flexible MePEA.³¹ Photoluminescence (PL) spectra are shown in the inset of Fig. 3(a), evidencing a relative increase in the PL intensity from +16% to +85% and +84% for FAPbI₃ + iso-BAI, FAPbI₃ + MePEABr and FAPbI₃ + MePEACl, respectively, compared to bare FAPbI₃ (it should be noted that they all have the same absorption spectra, as shown in Figs. S9 and S10). Such a trend suggests that the surface treatment reduces non-radiative recombination, specifically for MePEA-based passivation.³² Time-resolved PL traces [Fig. 3(a) and Table S2] provide further confirmation of the reduced charge trapping, as a slower decay (from 30 to 70 ns) is registered for FAPbI₃ + MePEACl and FAPbI₃ + MePEABr with respect to bare FAPbI₃.^{33,34} Figure 3(b) shows the PL quantum yield (PLQY) and the experimental V_{oc} obtained for the respective PSCs. A clear increase in the PLQY, in agreement with the higher PL intensity and the longer lifetime as previously discussed, going from neat FAPbI₃ to FAPbI₃ + iso-BAI and, especially, to FAPbI₃ + MePEACl and FAPbI₃ + MePEABr, is observed, providing final confirmation of the reduction in the non-radiative recombination due to surface passivation with the evident beneficial impact on the device V_{oc} .⁶

D. Predictions on new candidate organic cations

The obtained RF model can be used directly to make predictions on the device V_{oc} . To prove this capability, we listed a database containing the chemical properties of sixteen candidate organic cations (presented in Table S3), and we used the RF to calculate the prediction on the V_{oc} . To minimize the number of processing conditions, we exclude the annealing step (noA) and we use a single concentration (10 mM). The results are shown in Fig. 3(c). Generally, the predictions for the highest values are associated with cations with Br and Cl, while those with I have the lowest, in line with the obtained SHAP values of the model. In particular, the highest V_{oc} would be potentially obtained using PEACl. Notably, the prediction of 1127.69 mV for a PSC passivated with FAPbI₃ + PEACl, however, is slightly lower than the average experimental value of 1131.05 mV obtained with FAPbI₃ + MePEACl-(10mM-noA). This turns out to be consistent with their chemical properties. Indeed, although MePEACl has a much higher molecular weight than PEACl (+9%), it compensates with a lower halide fraction (~8%) and heteroatom carbon ratio (~11%). Taking into account that the RF prediction confidence range given by ± the AbO RMSE score is comparable with the range in which the predictions are contained, it is not possible to draw absolute conclusions about which of the candidate cations is the most effective. On the other hand, one must take into account that the candidate cations were chosen based on the conclusions drawn from the SHAP values, so we expect the model predictions to be close.³⁵ In fact, if we consider the range of the experimental data used for model training of 168 mV, we obtain a normalized AbO RMSE of 12% (Fig. S11). The AbO RMSE obtained is affected by both the limited dataset used and the high variability of experimental data for each combination of surface passivator, concentration, and temperature (see Fig. S12 for AbO cross-validation RMSEs with two, three, four, or nine cations with additional models). To verify whether the prediction is

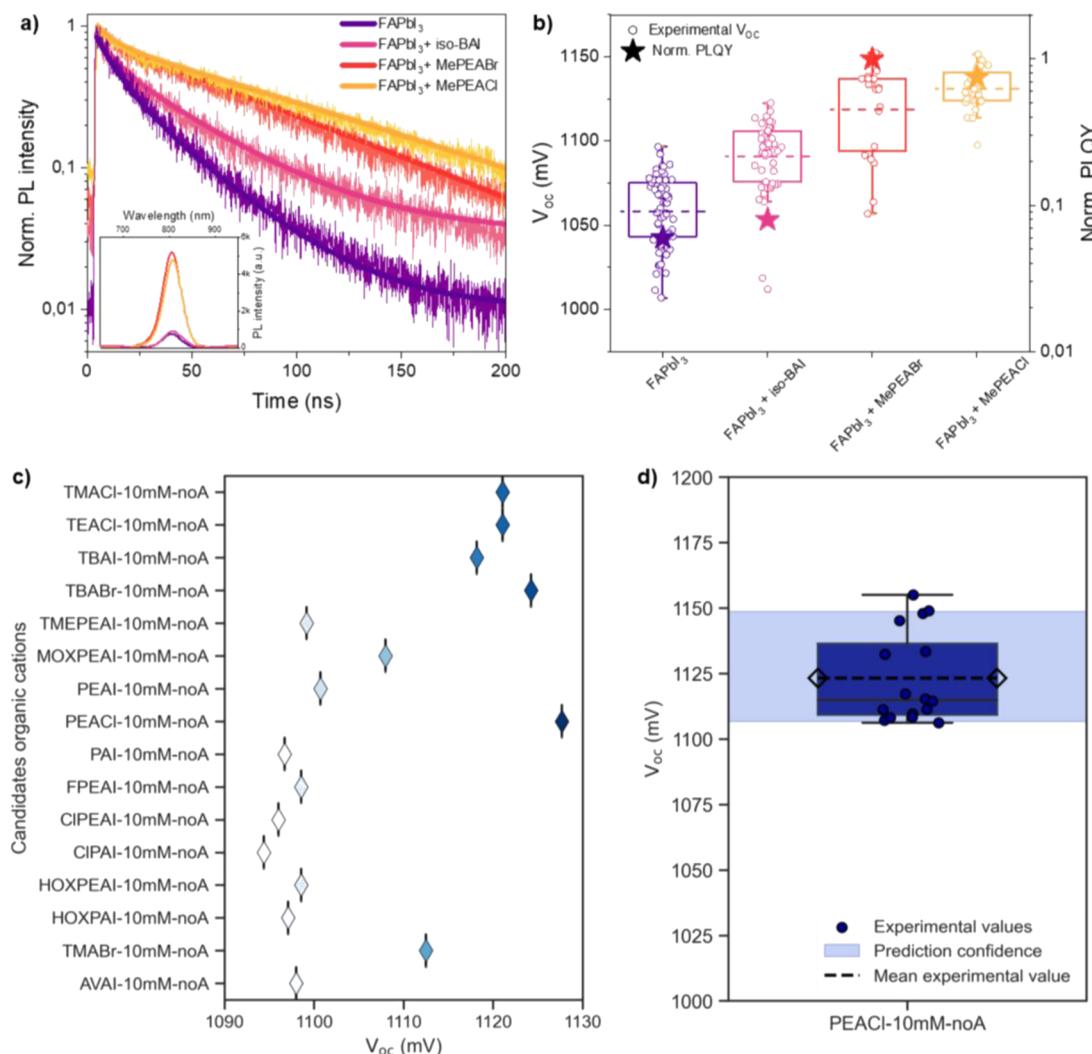


FIG. 3. (a) PL intensity as the inset and PL decays of non-passivated FAPbI₃ and FAPbI₃ passivated with iso-BAI/MePEABr/MePEACl. (b) Experimental V_{oc} and normalized PLQY for PSCs and films, respectively. (c) V_{oc} predictions on candidate cations by using the RF model, where “noA” indicates no annealing step. (d) Experimental V_{oc} values of the top two PSCs passivated with PEACl and RF AbO prediction confidence on PEACl.

empirically valid, we fabricated and measured a batch of five PSCs passivated with FAPbI₃ + PEACl. The measured V_{oc} values of the best two PSCs are shown in Fig. 3(d) with the RF prediction confidence range (light blue), given by the predicted V_{oc} value \pm the AbO RMSE score. The prediction obtained is very close to the mean value of the experimental measurements, and the confidence range is consistent with the variability of the experimental data. Although a more robust argument would require a larger test of the reliability of the predictions compared to experimental data, this result corroborates the potential of the proposed method to perform dependable *a priori* predictions for a smart and guided optimization of the experimental device optimization.

III. CONCLUSIONS

In conclusion, we have developed a robust machine learning-assisted investigation revealing the critical chemical features of different surface passivators used to push PSCs efficiency, mainly through the increase in the device V_{oc} . Fifty-four combinations of surface passivated FAPbI₃ solar cells have been evaluated for the training of six machine learning models, while AbO cross-validation was applied to select the best model, from which RF was selected. We applied a SHAP values method to identify which features most strongly correlate with the improved device V_{oc} , and we found that the most impacting features are (I) no need for the annealing step, (II) the low halide fraction explained through the beneficial role of

Cl in binding iodine vacancies, and (III) the low heteroatom carbon ratio, which correlates with the increased spreading of a flexible molecule. By comparing the organic salts that exhibited the highest improvement of V_{oc} , i.e., MePEACl and MePEABr, and the lowest, i.e., iso-BAI, we correlated the enhancement in the V_{oc} , specifically to both (II) and (III). Vacancies passivation induced by Cl (and Br) is further verified by PLQY, while the increased flexibility of MePEACl (and MePEABr) with respect to iso-BAI affects the crystalline size uniformity and the coverage in the perovskite film as testified by morphological mapping. Moreover, we further pushed the validity of our algorithm by performing *a priori* predictions to guide experimental efforts. We used the RF model to predict the V_{oc} upon considering sixteen new candidates as surface passivators. The highest predictive value was obtained with PEACl, which found an excellent validation with the measured V_{oc} values (i.e., the measured V_{oc} values fell within the confidence band given by the prediction of the RF \pm its ABO RMSE score). This proves the reliability and the capability of the predictive power of the ML model. We demonstrate a robust and rational ML approach, which can facilitate a deeper understanding of the chemical properties and an easier screening of a vast library of materials, getting rid of a time-consuming trial and error approach. In addition, the predictive power is demonstrated, essential to conduct *a priori* selection of candidate materials. We foresee that follow-up work will benefit from using active learning processes incorporating large additional data, to corroborate the model and generalize the observations. Beyond that, the model is extremely versatile, and we envisage that it could be applied to explore other chemical features/target variables and to further accelerate the optimization of perovskite solar cells.

SUPPLEMENTARY MATERIAL

The [supplementary material](#) includes experimental section, organic cations properties, PSCs performances, main ML results, XRD, AFM, ABS, Tauc plot, TRPL lifetimes, and the whole ML code at: https://github.com/MattiaSpider-SIMaP/PVSquared2/tree/main/ML_for_FAPbI3_passivation.

ACKNOWLEDGMENTS

The authors acknowledge the “HY-NANO” project (Grant Agreement No. 802862) and the “ERC Proof of Concept SPIKE” (Grant Agreement No. 101068936) that received funding from the European Research Council (ERC) Starting Grant 2018 under the European Union’s Horizon 2020 research and innovation programme. The authors acknowledge the project for infrastructures funded by Regione Lombardia RL3776, the Ministero dell’Università e della Ricerca (MUR), and the University of Pavia through the program “Dipartimenti di Eccellenza 2023–2027.” F.F. and G.G. acknowledge Edison for the collaboration in the PhD fellowship co-founded by the European Union–FSE, Programma Operativo Nazionale (PON) Ricerca e Innovazione 2014–2020 (CCI 2014IT16M2OP005). M.R. acknowledges MIAI@Grenoble Alpes (ANR-19-P3IA-0003). I.P. acknowledges the National Institute for Nuclear Physics (INFN) within the next_AIM

(Artificial Intelligence in Medicine: next steps) research Project No. (INFN-CSN5).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

M.R. and F.F. contributed equally to this work.

Mattia Ragni: Data curation (lead); Formal analysis (equal); Methodology (equal); Software (lead); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Fabiola Faini:** Conceptualization (equal); Data curation (supporting); Formal analysis (equal); Methodology (supporting); Validation (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Matteo Degani:** Formal analysis (equal); Methodology (equal); Visualization (equal); Writing – original draft (equal); Writing – review & editing (equal). **Silvia Cavalli:** Formal analysis (equal); Methodology (equal); Writing – original draft (equal); Writing – review & editing (equal). **Ian Postuma:** Methodology (supporting); Software (supporting); Writing – original draft (equal); Writing – review & editing (equal). **Giulia Grancini:** Conceptualization (equal); Funding acquisition (lead); Project administration (lead); Resources (lead); Supervision (lead); Visualization (supporting); Writing – original draft (supporting); Writing – review & editing (supporting).

DATA AVAILABILITY

The data that support the findings of this study are available within the article and its [supplementary material](#).

REFERENCES

- 1 L. Jonathan, L. J. Diguna, O. Samy, M. Muqyyanah, S. Abu Bakar, M. D. Birowosuto, and A. El Moutaouakil, “Hybrid organic–inorganic perovskite halide materials for photovoltaics towards their commercialization,” *Polymers* **14**(5), 1059 (2022).
- 2 L. M. Herz, “Charge-carrier mobilities in metal halide perovskites: Fundamental mechanisms and limits,” *ACS Energy Lett.* **2**(7), 1539–1548 (2017).
- 3 Y. Chen, H. T. Yi, X. Wu, R. Haroldson, Y. N. Gartstein, Y. I. Rodionov, K. S. Tikhonov, A. Zakhidov, X.-Y. Zhu, and V. Podzorov, “Extended carrier lifetimes and diffusion in hybrid perovskites revealed by Hall effect and photoconductivity measurements,” *Nat. Commun.* **7**(1), 12253 (2016).
- 4 G. Szabó, N.-G. Park, F. De Angelis, and P. V. Kamat, “Are perovskite solar cells reaching the efficiency and voltage limits?,” *ACS Energy Lett.* **8**(9), 3829–3831 (2023).
- 5 G. KH, “Advances in surface passivation of perovskites using organic halide salts for efficient and stable solar cells,” *Surf. Interfaces* **26**, 101420 (2021).
- 6 N. K. Noel, A. Abate, S. D. Stranks, E. S. Parrott, V. M. Burlakov, A. Goriely, and H. J. Snaith, “Enhanced photoluminescence and solar cell performance via Lewis base passivation of organic–inorganic lead halide perovskites,” *ACS Nano* **8**(10), 9815–9821 (2014).
- 7 Y. Li, H. Wu, W. Qi, X. Zhou, J. Li, J. Cheng, Y. Zhao, Y. Li, and X. Zhang, “Passivation of defects in perovskite solar cell: From a chemistry point of view,” *Nano Energy* **77**, 105237 (2020).

- ⁸W. Shockley and W. T. Read, "Statistics of the recombinations of holes and electrons," *Phys. Rev.* **87**(5), 835–842 (1952).
- ⁹Y. Lv, Y. Shi, X. Song, J. Liu, M. Wang, S. Wang, Y. Feng, S. Jin, and C. Hao, "Bromine doping as an efficient strategy to reduce the interfacial defects in hybrid two-dimensional/three-dimensional stacking perovskite solar cells," *ACS Appl. Mater. Interfaces* **10**(37), 31755–31764 (2018).
- ¹⁰F. Wang, W. Geng, Y. Zhou, H. Fang, C. Tong, M. A. Loi, L. Liu, and N. Zhao, "Phenylalkylamine passivation of organolead halide perovskites enabling high-efficiency and air-stable photovoltaic cells," *Adv. Mater.* **28**(45), 9986–9992 (2016).
- ¹¹Y. Lin, Y. Bai, Y. Fang, Z. Chen, S. Yang, X. Zheng, S. Tang, Y. Liu, J. Zhao, and J. Huang, "Enhanced thermal stability in perovskite solar cells by assembling 2D/3D stacking structures," *J. Phys. Chem. Lett.* **9**(3), 654–658 (2018).
- ¹²T. M. Koh, V. Shanmugam, X. Guo, S. S. Lim, O. Filonik, E. M. Herzig, P. Müller-Buschbaum, V. Swamy, S. T. Chien, S. G. Mhaisalkar, and N. Mathews, "Enhancing moisture tolerance in efficient hybrid 3D/2D perovskite photovoltaics," *J. Mater. Chem. A* **6**(5), 2122–2128 (2018).
- ¹³Q. Zhou, L. Liang, J. Hu, B. Cao, L. Yang, T. Wu, X. Li, B. Zhang, and P. Gao, "High-performance perovskite solar cells with enhanced environmental stability based on a (*p*-FC₆H₄C₂H₄NH₃)₂[PbI₄] capping layer," *Adv. Energy Mater.* **9**(12), 1802595 (2019).
- ¹⁴E. H. Jung, N. J. Jeon, E. Y. Park, C. S. Moon, T. J. Shin, T.-Y. Yang, J. H. Noh, and J. Seo, "Efficient, stable and scalable perovskite solar cells using poly(3-hexylthiophene)," *Nature* **567**(7749), 511–515 (2019).
- ¹⁵M. Kundar, S. Bhandari, S. Chung, K. Cho, S. K. Sharma, R. Singh, and S. K. Pal, "Surface passivation by sulfur-based 2D (TEA)₂PbI₄ for stable and efficient perovskite solar cells," *ACS Omega* **8**(14), 12842–12852 (2023).
- ¹⁶A. A. Sutanto, P. Caprioglio, N. Drigo, Y. J. Hofstetter, I. Garcia-Benito, V. I. E. Queloz, D. Neher, M. K. Nazeeruddin, M. Stolterfoht, Y. Vaynzof, and G. Grancini, "2D/3D perovskite engineering eliminates interfacial recombination losses in hybrid perovskite solar cells," *Chem* **7**(7), 1903–1916 (2021).
- ¹⁷J. Xu, H. Chen, L. Grater, C. Liu, Y. Yang, S. Teale, A. Maxwell, S. Mahesh, H. Wan, Y. Chang, B. Chen, B. Rehl, S. M. Park, M. G. Kanatzidis, and E. H. Sargent, "Anion optimization for bifunctional surface passivation in perovskite solar cells," *Nat. Mater.* **22**(12), 1507–1514 (2023).
- ¹⁸A. H. Howlader and A. Uddin, "Progress and challenges of chloride–iodide perovskite solar cells: A critical review," *Nanomanufacturing* **3**(2), 177–216 (2023).
- ¹⁹J. Xu *et al.*, "Anion optimization for bifunctional surface passivation in perovskite solar cells," *Nat. Mater.* (published online 2023).
- ²⁰C. Zhi, S. Wang, S. Sun, C. Li, Z. Li, Z. Wan, H. Wang, Z. Li, and Z. Liu, "Machine-learning-assisted screening of interface passivation materials for perovskite solar cells," *ACS Energy Lett.* **8**(3), 1424–1433 (2023).
- ²¹N. T. P. Hartono, J. Thapa, A. Tiihonen, F. Oviedo, C. Batali, J. J. Yoo, Z. Liu, R. Li, D. F. Marrón, M. G. Bawendi *et al.*, "How machine learning can help select capping layers to suppress perovskite degradation," *Nat. Commun.* **11**(1), 4172 (2020).
- ²²I. Kouroudis, K. T. Tanko, M. Karimipour, A. B. Ali, D. K. Kumar, V. Sudhakar, R. K. Gupta, I. Visoly-Fisher, M. Lira-Cantu, and A. Gagliardi, "Artificial intelligence-based, wavelet-aided prediction of long-term outdoor performance of perovskite solar cells," *ACS Energy Lett.* **9**, 1581–1586 (2024).
- ²³J. Kirman, A. Johnston, D. A. Kuntz, M. Askerka, Y. Gao, P. Todorović, D. Ma, G. G. Privé, and E. H. Sargent, "Machine-learning-accelerated perovskite crystallization," *Matter* **2**(4), 938–947 (2020).
- ²⁴W. D. Ihlenfeldt, Y. Takahashi, H. Abe, and S. Sasaki, "Computation and management of chemical properties in CACTVS: An extensible networked approach toward modularity and compatibility," *J. Chem. Inf. Comput. Sci.* **34**(1), 109–116 (1994).
- ²⁵S. H. Bertz, "The first general index of molecular complexity," *J. Am. Chem. Soc.* **103**(12), 3599–3601 (1981).
- ²⁶J. B. Hendrickson, P. Huang, and A. G. Toczek, "Molecular complexity: A simplified formula adapted to individual atoms," *J. Chem. Inf. Comput. Sci.* **27**(2), 63–67 (1987).
- ²⁷S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *NIPS'17 Proceedings of the 31st International Conference on Neural Information Processing Systems 2017* (Curran Associates Inc., 2017), pp. 4768–4777, [arXiv:1705.07874](https://arxiv.org/abs/1705.07874).
- ²⁸S. Prathapani, D. Choudhary, S. Mallick, P. Bhargava, and A. Yella, "Experimental evaluation of room temperature crystallization and phase evolution of hybrid perovskite materials," *CrystEngComm* **19**(27), 3834–3843 (2017).
- ²⁹Y. Liu, S. Akin, A. Hinderhofer, F. T. Eickemeyer, H. Zhu, J. Seo, J. Zhang, F. Schreiber, H. Zhang, S. M. Zakeeruddin, A. Hagfeldt, M. I. Dar, and M. Grätzel, "Stabilization of highly efficient and stable phase-pure FAPbI₃ perovskite solar cells by molecularly tailored 2D-overlays," *Angew. Chem., Int. Ed.* **59**(36), 15688–15694 (2020).
- ³⁰L. Wang, C. McCleese, A. Kovalsky, Y. Zhao, and C. Burda, "Femtosecond time-resolved transient absorption spectroscopy of CH₃NH₃PbI₃ perovskite films: Evidence for passivation effect of PbI₂," *J. Am. Chem. Soc.* **136**(35), 12205–12208 (2014).
- ³¹X. Shang, C. Chen, F. Meng, Z. Zhang, M. Li, D. Gao, and C. Chen, "Grain boundary defects passivation by bridging diammonium toward stable and efficient perovskite solar cells," *J. Colloid Interface Sci.* **649**, 528–534 (2023).
- ³²G. Pica, D. Bajoni, and G. Grancini, "A step beyond in steady-state and time-resolved electro-optical spectroscopy: Demonstration of a customized simple, compact, low-cost, fiber-based interferometer system," *Struct. Dyn.* **9**(1), 011101 (2022).
- ³³T. Kirchartz, J. A. Márquez, M. Stolterfoht, and T. Unold, "Photoluminescence-based characterization of halide perovskites for photovoltaics," *Adv. Energy Mater.* **10**(26), 1904134 (2020).
- ³⁴D. Zhang, H. Zhang, H. Guo, F. Ye, S. Liu, and Y. Wu, "Stable α-FAPbI₃ in inverted perovskite solar cells with efficiency exceeding 22% via a self-passivation strategy," *Adv. Funct. Mater.* **32**(27), 2200174 (2022).
- ³⁵S. Teale, M. Degani, B. Chen, E. H. Sargent, and G. Grancini, "Molecular cation and low-dimensional perovskite surface passivation in perovskite solar cells," *Nat. Energy* **9**, 779–792 (2024).

Supplementary Information

Machine learning for screening and predicting the best surface modifiers for a rational optimization of efficient perovskite solar cells.

Mattia Ragni^{a)}, Fabiola Faini^{a)}, Matteo Degani, Silvia Cavalli, Ian Postuma, Giulia Grancini^{b)}

Experimental section

A. Perovskite thin film preparation

Perovskite thin films with different passivation were manufactured with the following procedure. Glass substrates were consecutively cleaned in acetone and isopropyl alcohol (IPA) by ultrasonication for a period of 15 minutes for each solvent. The substrates were then dried with a stream of nitrogen (N_2) and treated with UV-ozone for 30 minutes. Solutions were prepared in an argon-filled glovebox, while the deposition of each layer was performed in a nitrogen-filled glovebox. The perovskite precursor solution (1.2 M) was prepared by dissolving lead iodide (PbI_2), formamidinium iodide (FAI), and methylammonium chloride (MACl) (35 mol%) in dimethylformamide/dimethyl sulfoxide (DMF/DMSO) (4/1) with a 5 mol% excess of PbI_2 . 25 μ L solution was deposited on the substrates and spin-coated using a three-step procedure: 1) substrates were spun at 1000 rpm for 12 seconds, 2) spin speed was increased to 5000 rpm for 27 seconds, and 3) speed reduction for 4 seconds. Ten seconds after the start of the second step, chlorobenzene (CB, 150 μ L) was dropped onto the spinning substrate for an antisolvent procedure. Subsequently, substrates were annealed at 150 °C for 10 minutes. For the passivation layer, different tested concentrations of 5-10-15 mM solutions of nine different cations (MePEAI, MePEABr, MePEACl, n-BAI, iso-BAI, n-OAI, BBr, HBr, and OaTos) in IPA were spin-coated onto the perovskite layer at 5000 rpm for 30 seconds. Annealed substrates were subjected to a temperature of 100°C for 10 minutes. For not-annealed samples, this last step is not expected.

B. Perovskite solar cell fabrication

The substrates, glass covered with ITO, were cleaned with the procedure mentioned above. For the deposition of the electron-transporting layer (ETL), the colloidal solution of tin oxide (SnO_2) in water was used with a concentration of 10%. The spin-coating speed was 4000 rpm for 30 seconds, followed by an annealing step at 150 °C for 30 minutes. Subsequently, the SnO_2 -coated substrates underwent a 30-minute UV-ozone treatment, the perovskite and the passivation were manufactured with the same procedure described above, to deposit the hole-transporting layer (HTL), Spiro-OMeTAD was dissolved in CB (80 mg/mL). The solution was doped by adding Li salt (17.5 μL) dissolved in acetonitrile (ACN, 500 mg/mL) and 4-tert-butylpyridine (4-tBP, 28.8 μL) to 1 mL of Spiro-OMeTAD solution in CB. A 10 μL solution was spin-coated onto the passivated perovskite layer. Finally, an 80 nm layer of gold (Au) was thermally evaporated onto the device using a shadow mask with an area of 0.03 cm^2 , resulting in devices with an active area of 0.045 cm^2 due to the overlap between evaporated gold and ITO. The evaporation speed was adjusted to 0.01 nm/s for the first 5 nm, 0.02 nm/s from 5 to 15 nm, and 0.06 nm/s for the rest of the procedure.

C. Perovskite solar cell measurement

Current density–voltage (J–V) characteristics of the manufactured PSCs were conducted using a Keithley 2400 sourcemeter and an Abet solar simulator. The simulated 1Sun AM1.5G illumination was calibrated using a photodiode. The devices were measured both in reverse scan (1.2 V → 0 V, 130 mV/s) and forward scan (0 → 1.2 V, 130 mV/s). During every measurement, a metal shadow mask with an aperture area of 0.03 cm^2 was used to cover the device’s active area.

D. X-ray diffraction (XRD)

XRD patterns were recorded on a Bruker D2-Phaser diffractometer with θ/θ geometry and Cu 1.54184 Å. Patterns were recorded between 5° and 40° 2θ with a step size of 0.02° and an acquisition time of 0.5 s per degree.

E. Atomic force microscopy (AFM)

AFM images were recorded with an A.P.E. Research TriASNOM near-field light optical microscope in non-contact mode using the Mikro Mash HQ:NSC15 Al BS probe. AFM images editing was obtained with Gwyddion.

F. UV-Vis absorption (ABS)

Absorption spectra were taken with a PerkinElmer LAMBDA 1050+ UV-Vis-NIR spectrophotometer.

G. Steady state photoluminescence (PL) and time resolved photoluminescence (TRPL)

PL and TRPL signals were acquired on a custom-made setup, employing a pulsed laser (RR = 5 MHz) operating at 470 nm (PicoQuant, LDH-DC-470) with an energy fluence of \sim 70 nJ/cm². Decays were collected at the maximum of PL emission and curves have been fitted using a bi-exponential fit function: $y = A_1 e^{-t/\tau_1} + A_2 e^{-t/\tau_2}$ where τ_1 and τ_2 are the decay constants and A_1 and A_2 the amplitudes, respectively.

H. Photoluminescence quantum yield (PLQY)

PLQY was collected using a variation of the aforementioned optical setup by including an integration sphere (Labsphere) in which the excitation was provided by the same laser at 470 nm at RR = 80MHz and energy fluence of \sim 70 nJ/cm².

Table S1: Organic cations properties computed with Cactvs. The toolkit enabled the calculation of a comprehensive set of indicators for ensembles based on the simplified molecular input line entry system (SMILE) of the desired compound.

Name	Molecular weight (u)	Halide fraction	Heteroatom carbon ratio	Rotatable bond count	Rigid bond count	Complexity	I count	Br count	Cl count
MePEAI	171.6693	0.2065	0.2222	2	6	82.7291	0	0	1
MePEABr	216.1203	0.3697	0.2222	2	6	82.7291	0	1	0
MePEACl	263.1208	0.4823	0.2222	2	6	82.7291	1	0	0
n-BAI	201.05	0.6312	0.5	2	0	13.1194	1	0	0
iso-BAI	201.05	0.6312	0.5	1	0	17.6096	1	0	0
n-OAI	257.1572	0.4935	0.25	6	0	43.8323	1	0	0
BBr	188.0667	0.4249	0.2857	1	6	55.4072	0	1	0
HBr	182.1031	0.4388	0.3333	4	0	27.361	0	1	0
OaTos	301.4432	0	0.3333	6	8	236.8436	0	0	0

Table S2: τ_1 , A_1 , τ_2 , A_2 obtained from bi-exponential fit of TRPL decays on perovskite thin films deposited on glass. τ_1 and τ_2 are associated with radiative and trap-mediated (non-radiative) recombination mechanisms, respectively.

Name	τ_1 (ns)	A_1 (%)	τ_2 (ns)	A_2 (%)
Reference 1	9.7 ± 0.7	45 ± 3	33.2 ± 1.3	55 ± 3
Reference 2	10.9 ± 1.5	23 ± 3	42.7 ± 1.4	77 ± 3
Reference 3	10.5 ± 0.4	62 ± 2	36.7 ± 1.9	38 ± 2
iso-BAI 1	7.3 ± 0.2	62 ± 1	33.0 ± 1.0	38 ± 1
iso-BAI 2	9.5 ± 0.5	55 ± 2	35.7 ± 1.7	45 ± 2
iso-BAI 3	7.6 ± 0.6	33 ± 1	42.3 ± 1.1	67 ± 1
MePEABr 1	7.5 ± 0.8	19 ± 1	70.7 ± 1.5	81 ± 1
MePEABr 2	8.0 ± 1.1	20 ± 1	63.5 ± 1.9	80 ± 1
MePEABr 3	7.0 ± 0.7	20 ± 1	77.7 ± 1.7	80 ± 1
MePEACl 1	10.0 ± 1.4	17 ± 1	74.6 ± 2.3	83 ± 1
MePEACl 2	7.1 ± 0.8	18 ± 1	71.6 ± 1.7	82 ± 1
MePEACl 3	8.2 ± 0.6	23 ± 1	99.0 ± 2.7	77 ± 1

Table S3: List of the sixteen candidate organic cations.

Name	Code
2-thiopenemethylammonium chloride	TMACl
2-thiopheneethylammonium chloride	TEACl
tetrabutylammonium iodide	TBAI
tetrabutylammonium bromide	TBABr
trimethylphenylammonium iodide	TMEPAI
4-methoxyphenethylammonium iodide	MOXPEAI
2-phenethylammonium iodide	PEAI
2-phenethylammonium chloride	PEACl
phenylammonium iodide	PAI
4-fluorophenethylammonium iodide	FPEAI
4-chlorophenethylammonium iodide	ClPEAI
4-chlorophenylammonium iodide	ClPAI
4-hydroxyphenethylammonium iodide	HOXPEAI
4-hydroxyphenylammonium iodide	HOXPAI
2-thiopenemethylammonium bromide	TMABr
5-ammonium valeric acid iodide	AVAI

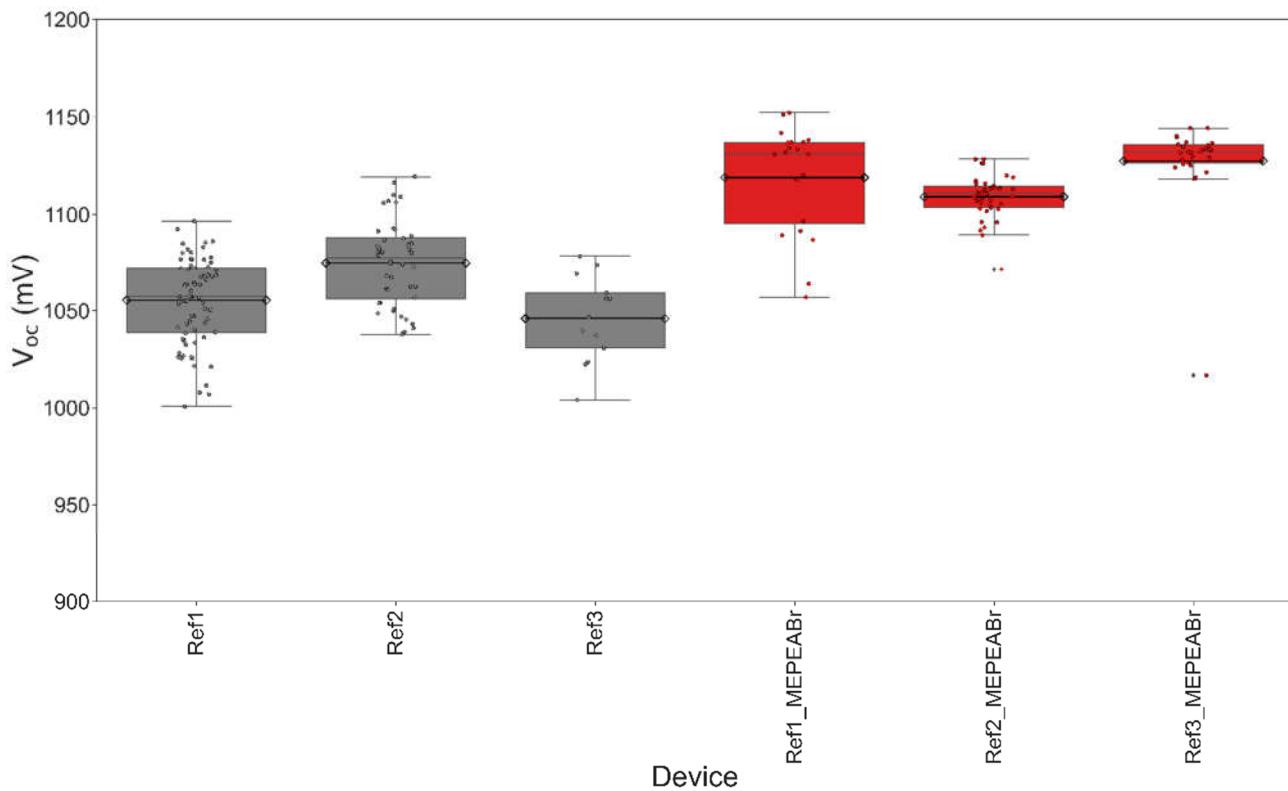


Figure S1: V_{oc} of FAPbI_3 and $\text{FAPbI}_3\text{+MePEABr-}10\text{mM-noA}$ PSCs.

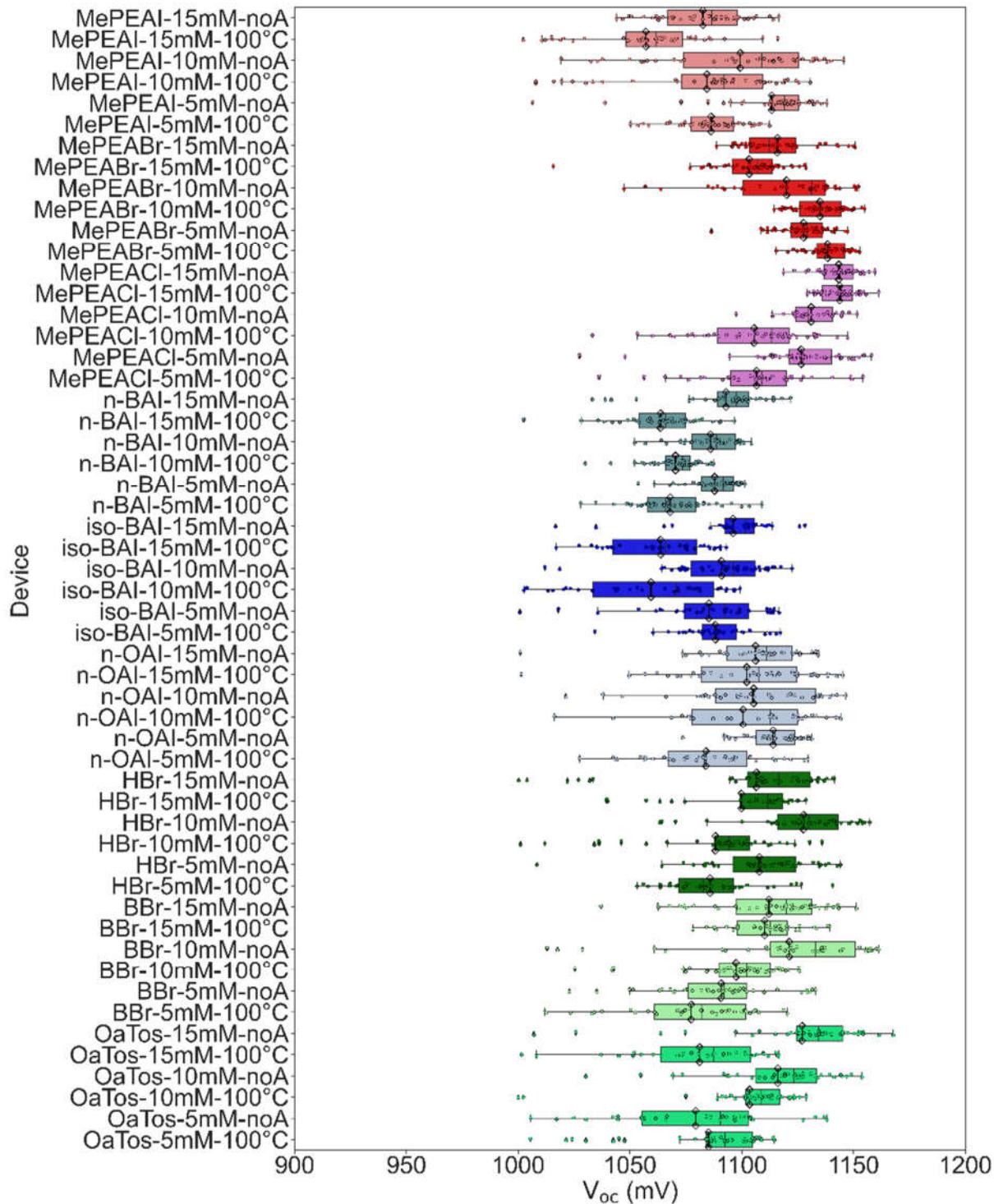


Figure S2: Experimental V_{oc} measurements of the PSCs.

PSCs are fabricated with the fiftyfour combinations of variables (nine organic cations and six processing conditions) studied, where “noA” indicates no annealing step.

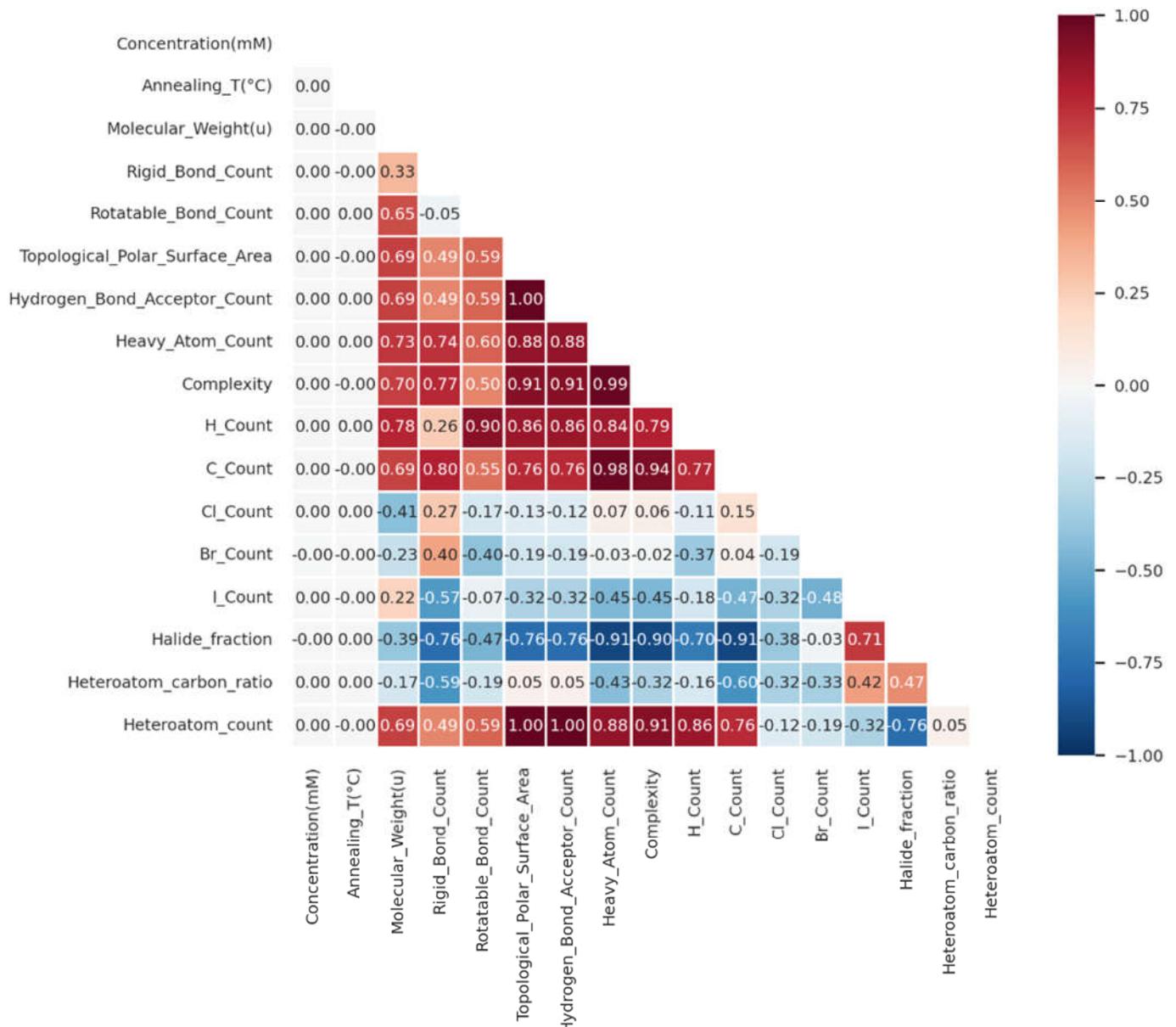


Figure S3: Matrix of the Pearson correlation factors including the sixteen organic cations properties and the two processing variables.

The coefficients measure the linear correlation and anticorrelation between features, denoted by values between +1 and -1, respectively.

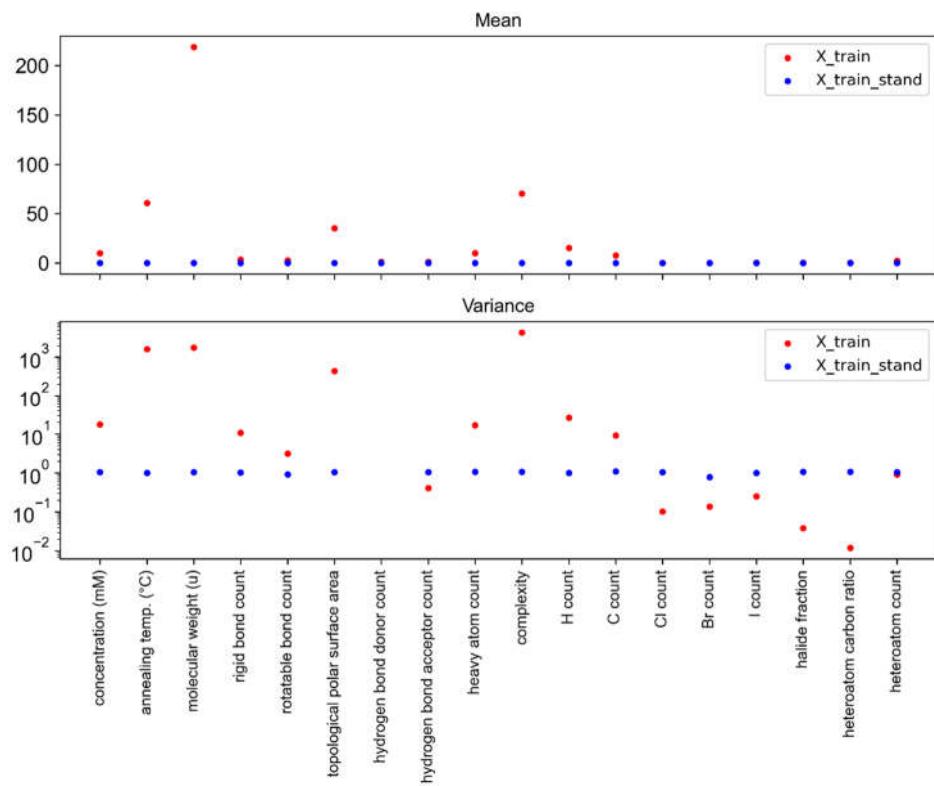


Figure S4: Mean and variance of input features values before (red) and after (blue) standardization.

Standardization involves scaling variable values to achieve a zero mean and unit variance.

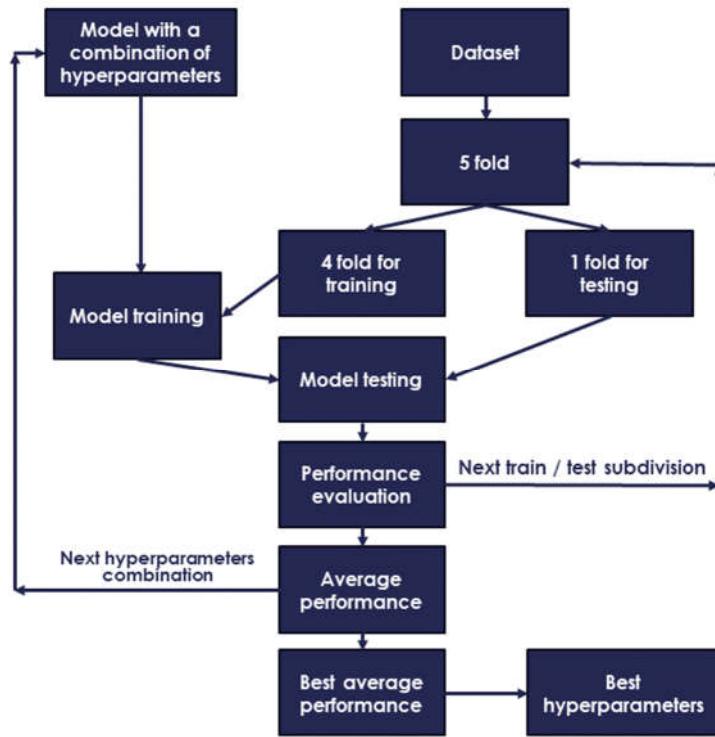


Figure S5: Flowchart of the grid search with 5-fold cross-validation applied to obtain the best hyperparameters for each applied model.

The dataset was divided into five folds, and models with varying sets of hyperparameters were trained on four folds and tested on the fifth. The process was repeated five times, and the combination of hyperparameters with the lowest root mean square error (RMSE) was selected as optimal.

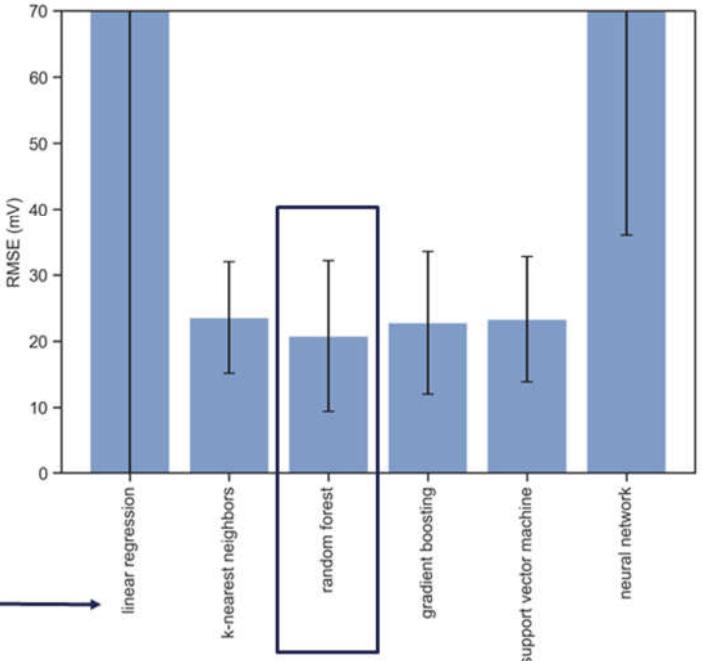
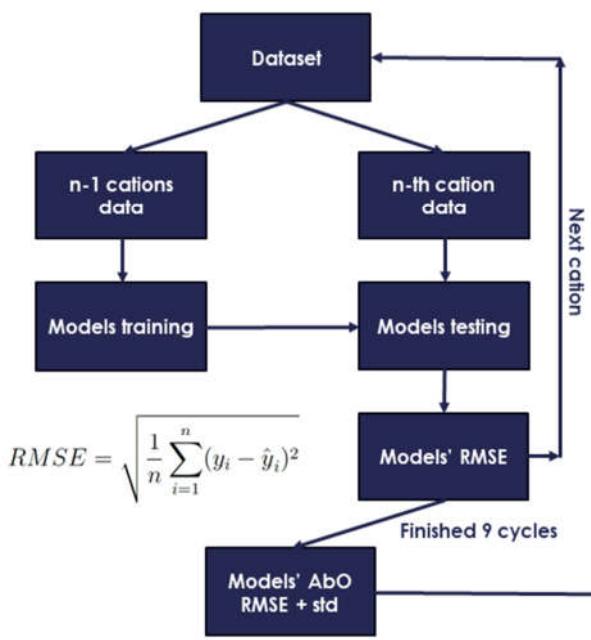


Figure S6: Flowchart of the all-but-one cross-validation applied to obtain the best performing model.

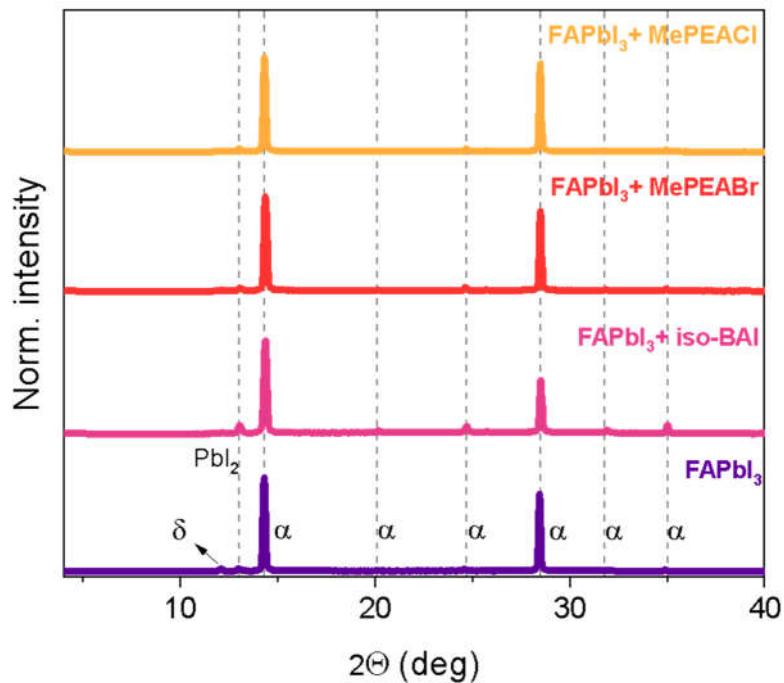


Figure S7: XRD patterns of perovskite thin films deposited on glass.

XRD patterns evidence a dominant peak for all materials at $2\Theta = 14.3^\circ$ and higher order reflections assigned to α -FAPbI₃. In addition, for the pure FAPbI₃ film, a peak at $2\Theta = 12.1^\circ$ is also present, ascribed to the unstable δ -FAPbI₃, while upon passivation procedure this peak disappears, in agreement with the literature. Concomitantly, the peak related to lead iodide at $2\Theta = 13.0^\circ$ grows by 72%, 18% and 23% for FAPbI₃+iso-BAI, FAPbI₃+MePEABr and FAPbI₃+MePEACl, with respect to the pure FAPbI₃, respectively, suggesting an increase in unreacted lead iodide during the passivation, especially for FAPbI₃+iso-BAI.

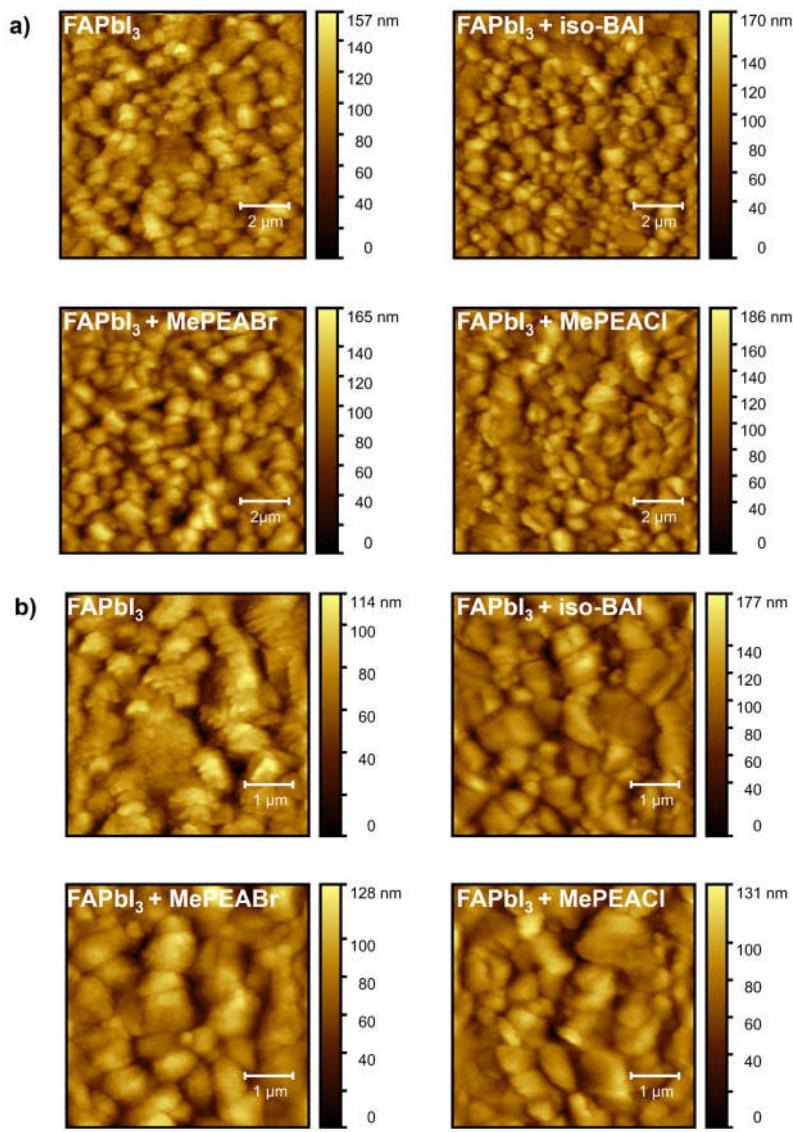


Figure S8: atomic force microscopy (AFM) 10 μm x 10 μm (a) and 5 μm x 5 μm (b) images collected for different perovskite thin films deposited on glass.

10 μm x 10 μm (AFM) images evidence micrometer-sized large regions consisting in turn of smaller domains for FAPbI₃, while for the passivated materials, a suppression of the nano-modulation is observed. We ascribed the variation in the nano-morphology to the beneficial role of the organic cation, which effectively passivates the grain boundaries locally. 5 μm x 5 μm AFM images, instead, evidence of an improved surface coverage when MePEABr and MePEACl are used, ascribed to the flexibility of MePEA.

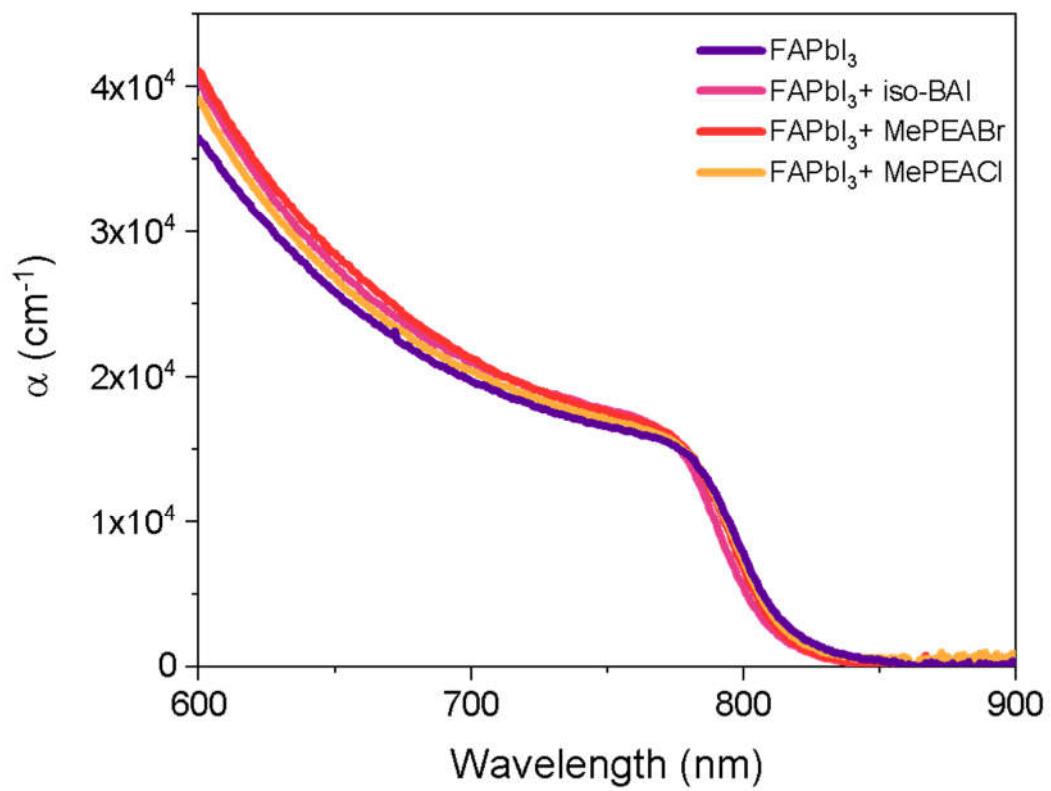


Figure S9: Spectra of the absorption coefficient of perovskite thin films deposited on glass.

Plots are obtained through ABS measurements normalized for the thickness (~ 300 nm) of the film.

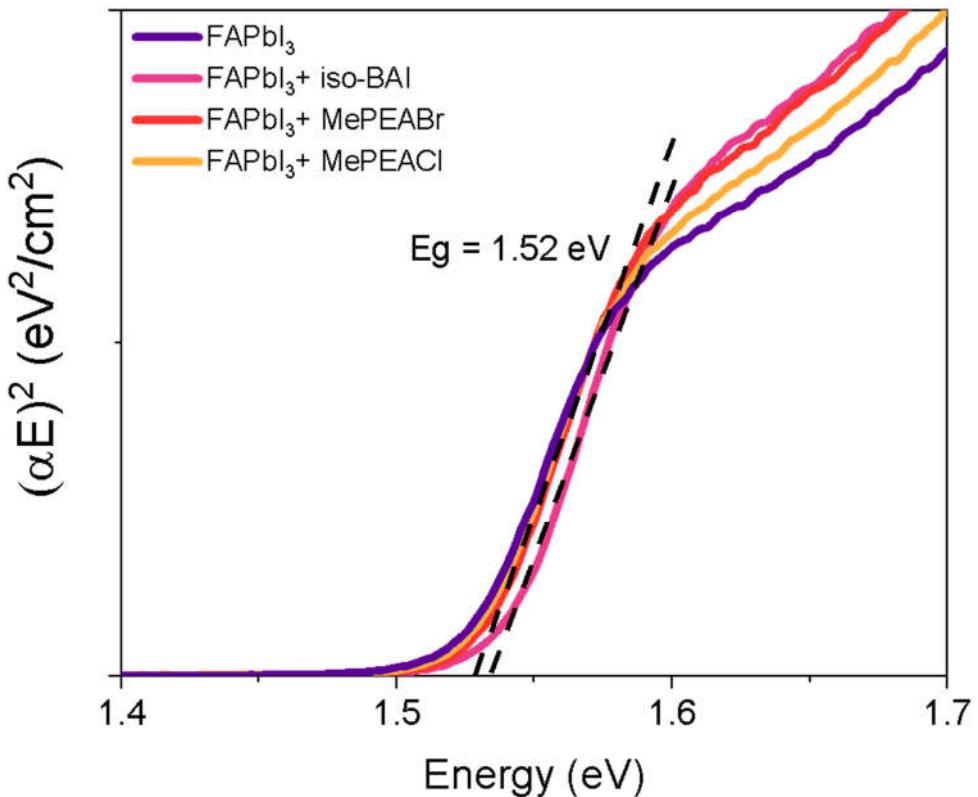


Figure S10: Tauc plots for direct band gap perovskite thin films deposited on glass.

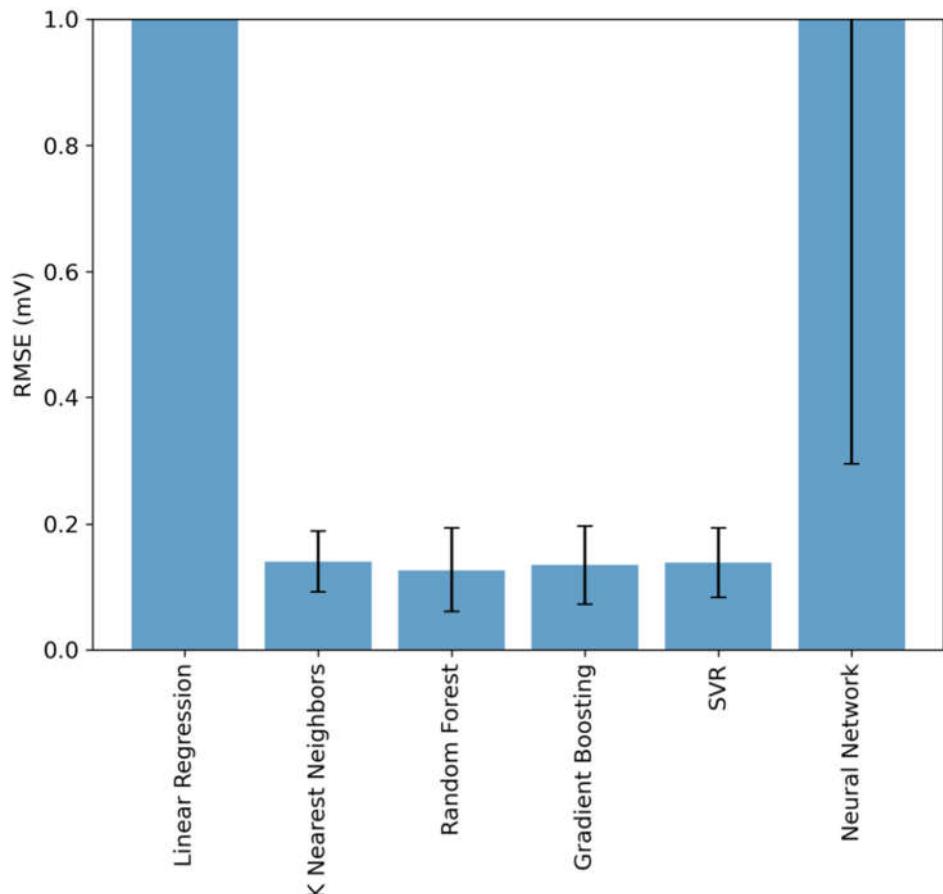


Figure S11: AbO RMSEs normalized on the experimental data range.

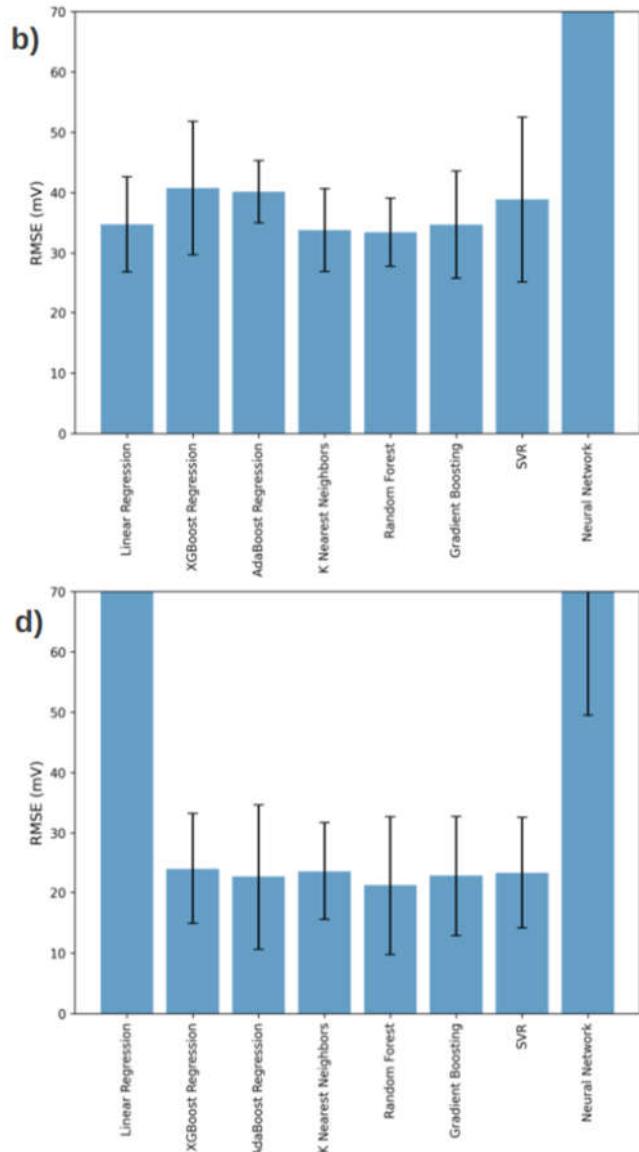
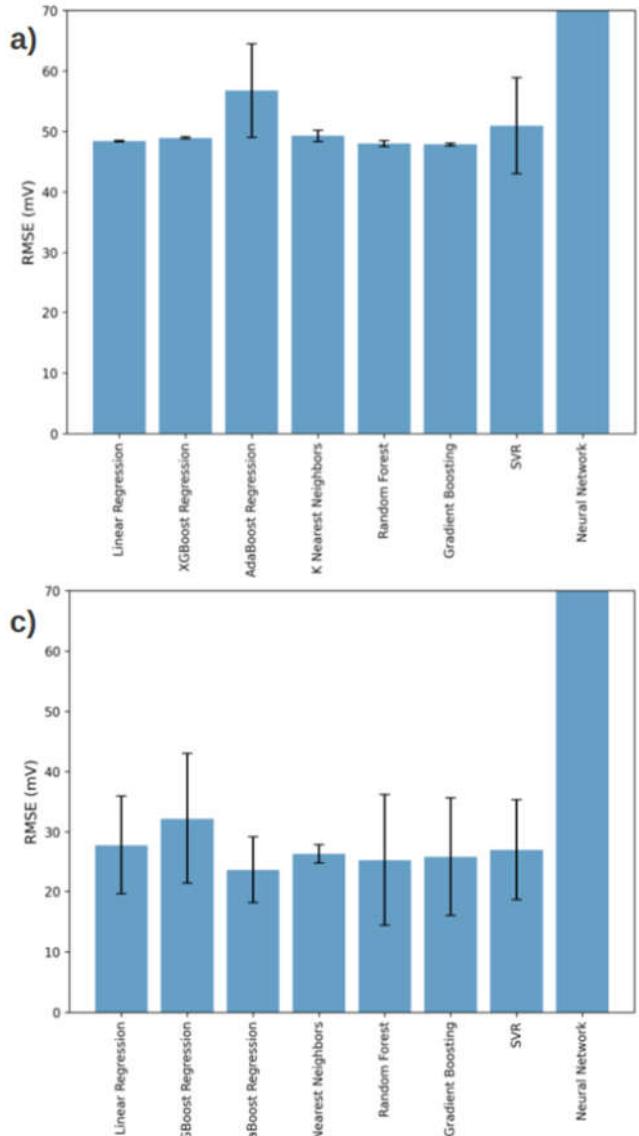


Figure S12: AbO cross validation RMSEs considering experimental data of 2 cations (a), 3 cations (b), 4 cations (c) and all 9 cations (d).