

Identifying Lead-free Perovskites Using Machine Learning for High Efficiency Solar Cells

Sanjana S. Jilla

Monta Vista High School

June 2021

Table of Contents

Introduction	3
1.1: Motivations:	3
1.2: Background/Current Research:	3
1.3: Project Goal:	3
Materials:	4
2.1: Dataset:	4
Methods:	5
3.1: Structure of a Perovskite Solar Cell:	5
3.2: Machine Learning Models:	6
3.3: Density Functional Theory:	6
3.4: Engineering Methodology:	7
3.5: Predicted Values:	8
Evaluations	8
4.1: Testing Methodology	8
4.2: Evaluation Criteria	8
4.3: Results	9
Discussion	10
5.1: Summary	10
5.2: Summary of Importance Rankings	10
5.3: Future Advancements	10
Works Cited	12

Introduction

1.1: Motivations:

One of the biggest threats our planet faces is the threat of global warming. Electrical power has become crucial to our modern lifestyle, but today, electricity is generated by burning fossil fuels and coal, which has many harms and disadvantages associated with it. Fortunately, there are multiple alternative energy sources, including sunlight, a powerful, inexhaustible, and clean resource.

1.2: Background/Current Research:

There are several different types of solar cells, but one of the most efficient and low-cost types is the perovskite solar cell. Perovskite solar cells (PSCs) have recently received considerable attention due to the high energy conversion efficiency achieved within a few years of their inception. However, today, the most common perovskite is methyl ammonium iodide (MAPbI₃), which contains levels of toxic lead. The science community has been searching for lower-toxicity perovskite-type materials, but testing all of the possible lead-free perovskites requires a huge amount of time and funding. Recent advances in computing power have enabled the generation of large datasets for materials and data-driven approaches to problem-solving in materials science, including materials discovery. Machine learning is the primary tool for manipulating such large datasets, predicting unknown material properties and uncovering relationships between structure and property.

1.3: Project Goal:

The goal of this project is to create a Machine Learning (ML) driven software system that increases the efficiency of the solar cell design process. I will do this by identifying the best perovskites by optimizing material composition and determining the importance of the features

of each element in the perovskite to the overall efficiency of the PSC. The Machine Learning (ML) driven software system needs to accurately predict key information such as the heat of formation (ΔH_f) and band gap (E_g) and accurately use the training data to form an accurate prediction of the best materials to form a perovskite solar cell (η). The models must predict with 90% accuracy of prediction for the project to be successful.

This program is written in Python, and uses Machine Learning to make predictions about the heat of formation and bandgap of various double halide perovskites.

Materials:

2.1: Dataset:

I used the Castelli Dataset in order to train my models. The parameters of the dataset include: Pauling's electronegativity (χ), ionization potential (IP), highest occupied atomic level (h), lowest unoccupied atomic level (l), s-, p-, and d- valence orbital radii r_s , r_p , and r_d of isolated neutral atoms A, B¹⁺, B³⁺, and X, atomic distance (D) between cations and the nearest halogen atom, space group of crystal (SG), etc. (Castelli)

Methods:

3.1: Structure of a Perovskite Solar Cell:

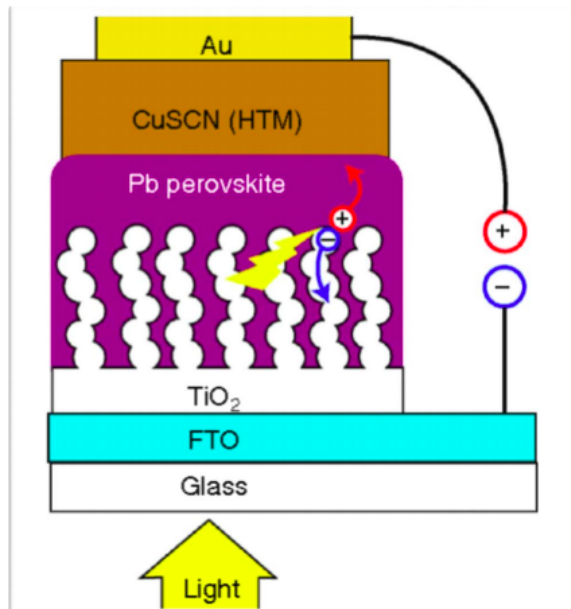


Figure 1: Perovskite Solar Cell Materials Structure (Qin)

Each solar cell contains specific materials layered in order to maximize efficiency. The Perovskite solar cells I developed consist of glass, FTO, TiO₂, Perovskite, CuSCN, and silver.

- Metal Electrode: (Cathode) Gold
- HTM: Spiro
- Absorber Layer: Methylammonium Lead Iodide
- Electron Transport Material: TiO₂
- Anode: FTO
- Base: Glass

3.2: Machine Learning Models:

I used several Machine Learning models to determine which had the most accurate predictions of bandgap and heat of formation. The first model I used was Logistic Regression; this is a model based on the logistic sigmoid function. Next was Linear SVC (Support Vector Classifier) which fits the data and returns a “best fit” line to categorize the data. K Nearest Neighbors is a nonparametric algorithm that divides the data into clusters to predict the classification of a new point. Random Forest uses a multitude of decision trees that form a hierarchy of the data to classify. Naïve Gaussian Bayes is a probability based model that calculates the probability of an outcome based on each parameter. Lastly, I used an ensemble classifier which is a combination of all of the other classifiers where the data is spread among the other models.

3.3: Density Functional Theory:

DFT is a method of obtaining an approximate solution to Schrödinger's equation of a many-body system. DFT is used in order to predict the heat of formation and bandgap (our desired outputs).

The following formulas were used:

$$\Delta H_F = E_{\text{tot}} (A_2 B^{1+} B^{3+} X_6) - (2E_{\text{ref}} (A) + E_{\text{rel}} (B^{1+}) + E_{\text{rel}} (B^{3+}) + 6E_{\text{rel}} (X))$$

where E_{tot} is the total energy and E_{ref} is the reference energy. For the band structures, spin-orbit interaction was considered.

3.4: Engineering Methodology:

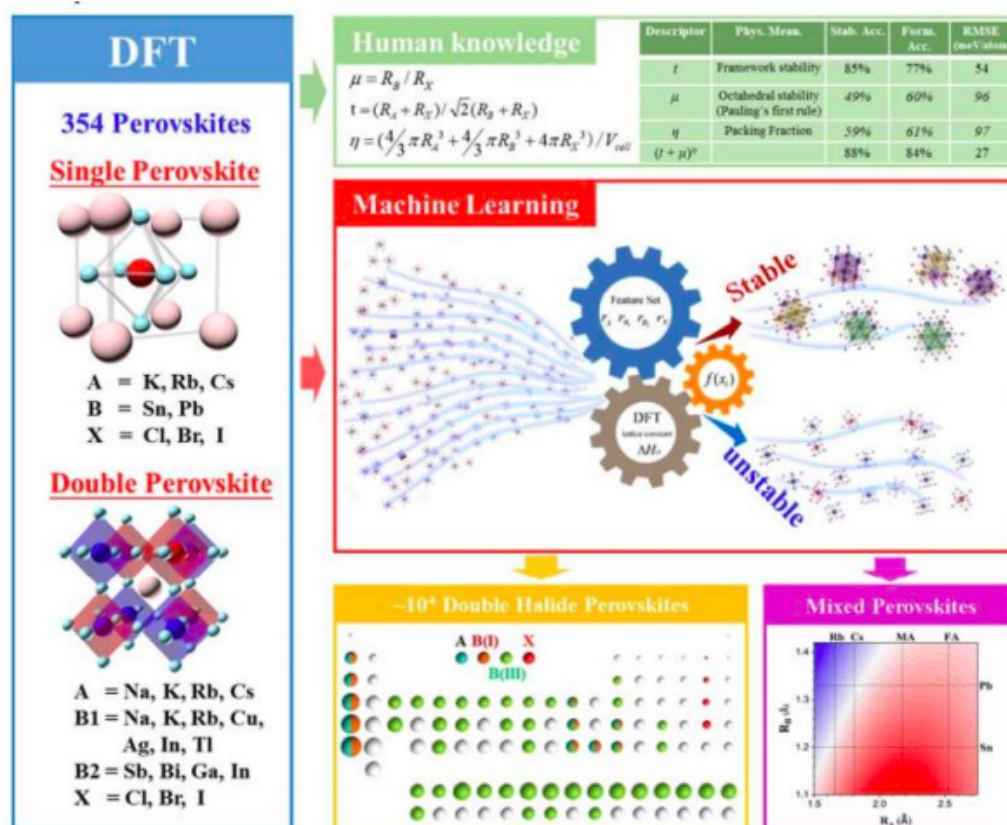


Figure 2: Flowchart and description of machine learning models based on Density Functional Theory (Research Gate)

First, I calculated the decomposition energies of 354 halide perovskites, using DFT. Then, I used the gradient-boosted regression tree (GBRT) algorithm to identify Pb-free double perovskite solar cell materials. The produces highly accurate predictions of the heat of formation (ΔH_f) and band gap (E_g) and importance scores for each feature of materials. Based on the scores, we extract crucial features to determine the values ΔH_f and E_g of halide double perovskites, enabling an overall understanding of the relationships between features and properties.

2.6:

3.5: Predicted Values:

There are two values I predict in order to determine which perovskite is most effective: heat of formation and bandgap. ΔH_f indicates the stability of a compound compared to those of the elemental phases of its chemical constituents. Generally, a more negative value of ΔH_f indicates a more stable compound. On the other hand, E_g can represent the capability to absorb solar energy, which is critical to achieving high performance of a solar cell.

Evaluations

4.1: Testing Methodology

I implement a 5 fold cross validation testing model. For each ML model, I used a randomly selected 80% of the data to train the model and the other 20% to test the model. Each model averaged 200 evaluations, as I repeated the fivefold training/test split 40 times. Each time, the data was randomly shuffled and the training data was varied each time.

4.2: Evaluation Criteria

There were several evaluation criteria that the project needed to meet in order to be possible. Each feature needed to have a valid and feasible importance score, which was generated from the Random Forest ML model. The models should vary based on the data entered and provide accurate results for the data provided. Additionally, the goal is a categorical accuracy of 90% comparing the predicted and expected value from the dataset.

4.3: Results

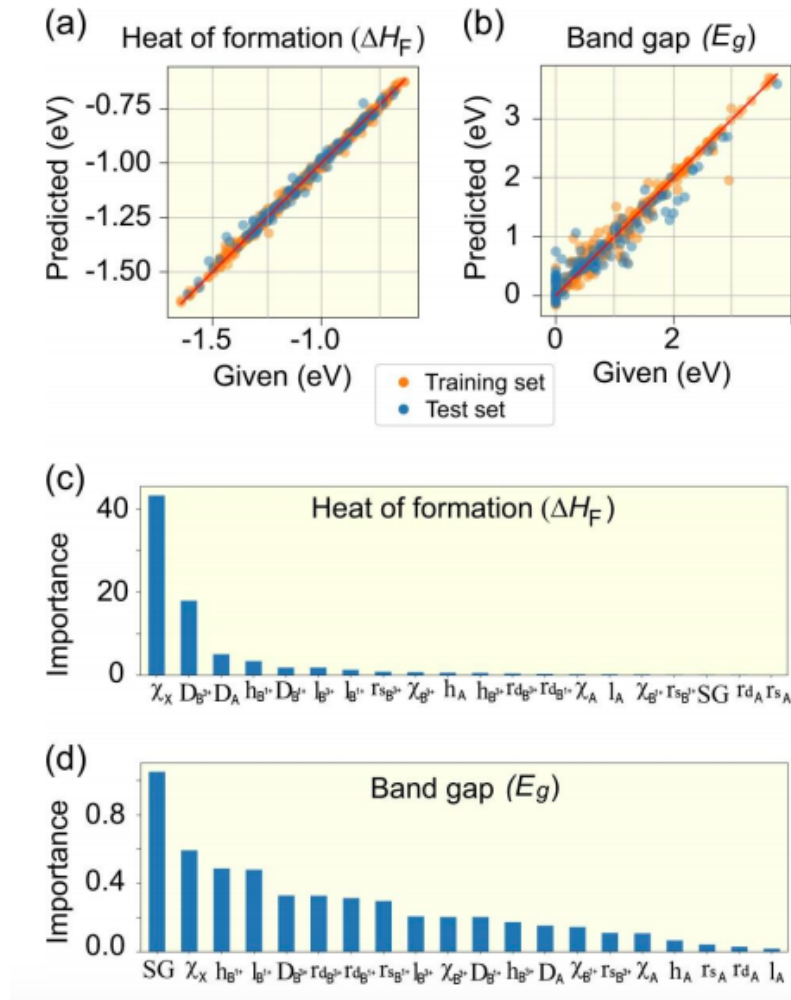


Figure 3: Perovskite Test Data Results and Importance Scores

Figure 3a: demonstrates the relationship between predicted and given heat of formation for the training and test datasets.

Figure 3b: demonstrates the relationship between predicted and given bandgap for the training and test datasets.

Figure 3c: demonstrates the 19 variables present in the datasets with their corresponding importance in the heat of formation calculation.

Figure 3d: demonstrates the 19 variables present in the datasets with their corresponding importance in the bandgap calculation.

The results show that averaged root-mean-square-error (RMSE) of test sets for ΔHF is 0.021 eV/atom. The accuracy of the predictive model of ΔHF from GBRT is 91.43%. In the case of E_g , the averaged RMSE of the test sets is equal to 0.223 eV. However, as the effective range

(1.1–1.8 eV) of the bandgap of solar cell material is much larger than the RMSE, such low accuracy of E_g could be acceptable for solar cell applications.

Discussion

5.1: Summary

The project was a success as all of the success criteria were met. The origin of the high accuracy of the predictive models can be attributed to several reasons. First is the nonlinear nature of the GBRT algorithm, in contrast to that of most linear algorithms. Second is the structural similarity of the materials considered in the dataset. All the crystals in this study are perovskites with the same structure which allows for more feasible interpolation of properties in the predictive models

5.2: Summary of Importance Rankings

For the heat of formation, ΔH_f , it is revealed that the type of halogen anions is the most important feature, more than two times higher than that of the secondly-ranked feature. The bandgap, E_g , has a more complex relationship with the material features as the space group of crystal was the most important but several other factors came close such as the type of halogen anion. Overall it seems that the halogen anions were the most important factor that impacted both criteria and therefore has the greatest impact on perovskite solar cell efficiency rates.

5.3: Future Advancements

There are several potential avenues to take from the conclusion of this project. I believe there are three that are most promising. First, fabricating and testing the efficiencies of the perovskites that

have the highest predicted efficiencies could provide tangible insight as to whether the models should be implemented in the future. Additionally, I could improve the accuracy of the models used by creating and using more comprehensive and detailed data sets. Currently, the dataset has approximately 5,000 points with 19 characteristics. Using a more comprehensive dataset could provide even better results. The third avenue is using deep learning models. Currently, I am using linear models, but using Neural Networks or other deep learning approaches should perform with even higher accuracies than the models used in this project.

Works Cited

Castelli, I. E., Thygesen, K. S., and Jacobsen, K. W. Bandgap Engineering of Double Perovskites for One- and Two-photon Water Splitting. MRS Proc. 1523, mrsf121523qq0706.

Castelli, I. E., García-Lastra, J. M., Hüser, F., Thygesen, K. S., and Jacobsen, K. W. Stability and bandgaps of layered perovskites for one- and two-photon water splitting New Journal Of Physics 15, 105026.

Im, Jino et al. "Identifying Pb-Free Perovskites For Solar Cells By Machine Learning". Npj Computational Materials, vol 5, no. 1, 2019. Springer Science And Business Media LLC, [www.doi:10.1038/s41524-019-0177-0](https://doi.org/10.1038/s41524-019-0177-0)

Li, Jinxin, et al. "Predictions and Strategies Learned from Machine Learning to Develop High-Performing Perovskite Solar Cells." Wiley Online Library, John Wiley & Sons, Ltd, 1 Oct. 2019, [www.onlinelibrary.wiley.com/doi/epdf/10.102/aenm.201901891?purchase_referrer=onlinelibrary.wiley.com&tracking_action=preview_click&r3_referer=wol&show_checkout=](https://onlinelibrary.wiley.com/doi/epdf/10.102/aenm.201901891?purchase_referrer=onlinelibrary.wiley.com&tracking_action=preview_click&r3_referer=wol&show_checkout=)

Odabaşı, Çağla, and Ramazan Yıldırım. "Performance Analysis of Perovskite Solar Cells in 2013–2018 Using Machine-Learning Tools." Nano Energy, Elsevier, 28 Nov. 2018, www.sciencedirect.com/science/article/pii/S2211285518308802.

"Perovskite Solar Cells." Energy.gov, www.energy.gov/eere/solar/perovskitesolar-cells

Qin, Peng et al. "Inorganic Hole Conductor-Based Lead Halide Perovskite Solar Cells With 12.4% Conversion Efficiency". Nature Communications, vol 5, no. 1, 2014. Springer Science And Business Media LLC, doi:10.1038/ncomms4834. Accessed 6 June 2021.

Ray, Sunil, and Business Analytics. "Essentials of Machine Learning Algorithms (with Python and R Codes)." Analytics Vidhya, 3 Sept. 2019,

www.analyticsvidhya.com/blog/2017/09/common-machine-learningalgorithms/.

Ripalda, Jose M. "Solar Cells and Machine Learning." Nature News, Nature Publishing Group, 2019,

www.devicematerialscommunity.nature.com/users/198387-jose-mripalda/posts/41628-solar-cells-and-machine-learning.

Stability Engineering of Halide Perovskite via Machine Learning - Scientific Figure on ResearchGate. Available from:

https://www.researchgate.net/figure/Schematic-strategy-of-ML-based-on-DFT-for-high-throughput-stability-engineering-of-halide_fig1_323846273 [accessed 6 Jun, 2021]