Interpretable Multi-Objective Machine Learning for the Discovery of Stable Halide Perovskites with Targeted Optoelectronic Properties

Khabat Zandkarimi [1, †], **Shahab Sobhani** [1]

[1] Department of Chemistry, Faculty of Science, University of Zanjan, Zanjan, Iran
corresponding author's email: zandkarimikhabat@gmail.com

Abstract

The discovery of halide perovskites that simultaneously exhibit thermodynamic stability and desirable optoelectronic properties remains challenging because of their vast compositional space and the computational cost of first-principles screening. Machine-learning approaches have recently demonstrated considerable potential for predicting stability and band gaps in ABX_3 perovskites, including high-throughput screening of thousands of compositions and inverse design of hypothetical halide-perovskite alloys (Yang2024; Biswas2024). However, reliable extrapolation toward compositionally unexplored materials and subsequent validation of their optical response remain important challenges.

Here, we propose an interpretable, multi-objective machine-learning framework for discovering stable halide perovskites with targeted optoelectronic properties. A curated dataset of inorganic ABX_3 perovskites will be assembled from first-principles materials databases, incorporating formation energies, energy above the convex hull, band gaps, and band-gap character. Composition-derived descriptors, including ionic radii, electronegativities, Goldschmidt tolerance factors, octahedral factors, and site-specific elemental properties, will be used to develop ensemble learning models. Composition-based and element-extrapolation validation will be employed to distinguish genuine materials-discovery capability from simple interpolation within the training chemical space. The trained models will then explore a substantially larger hypothetical composition space, while a Pareto-based multi-objective strategy will identify candidates combining low energy above the convex hull, targeted band gaps, and a high probability of direct-gap character. SHAP-based explainable AI will further be used to identify the chemical descriptors governing the stability–optoelectronic-property relationship (Yang2024; Manon2026).

Selected unexplored candidates will subsequently be validated using first-principles calculations, including spin–orbit coupling for heavy-element-containing compositions. For the most promising candidates, frequency-dependent dielectric functions and optical absorption spectra will be calculated in addition to band structures and density of states. The resulting ML–DFT workflow aims to provide an interpretable route from chemical composition to materials discovery and to identify halide perovskites combining favorable stability with targeted optoelectronic characteristics.

References

J. Yang, P. Manganaris, and A. Mannodi-Kanakkithodi, "Discovering novel halide perovskite alloys using multi-fidelity machine learning and genetic algorithm," *The Journal of Chemical Physics* **160**, 064114 (2024). https://doi.org/10.1063/5.0182543

M. Biswas, R. Desai, and A. Mannodi-Kanakkithodi, "Screening of novel halide perovskites for photocatalytic water splitting using multi-fidelity machine learning," *Physical Chemistry Chemical Physics* **26**, 23177–23188 (2024). https://doi.org/10.1039/D4CP02330G

A. Manon et al., "SHAP-guided Machine Learning for Interpretable Band Gap Prediction and Inverse Design in ABX_3 Perovskites," *International Journal of Computational Intelligence Systems* (2026). https://doi.org/10.1007/s44196-026-01454-1