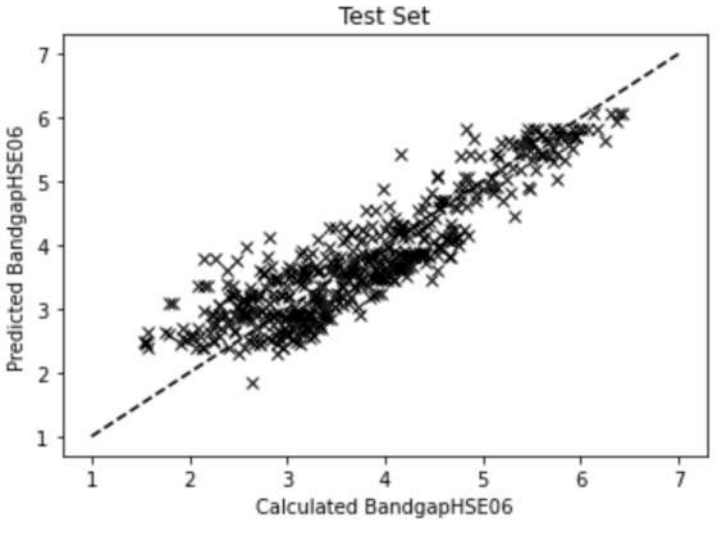
A multi-step, DFT-free screening of hybrid organic-inorganic perovskites using machine-learning and geometric limits

Vedant Bahadur, Dr. Jarvis Loh

Perovskites, particularly hybrid organic-inorganic perovskites (HOIPs), are an extremely promising material to increase the efficiency of modern solar photovoltaic panels. However, the vast HOIP space exceeds 500,000 possible combinations (Herbol, H.C., Hu, W., Frazier, P. et al), making experimental or Density functional theory (DFT) approaches overly time consuming and resource-consuming. Using a combination of first-principles geometric limits based on the work of Marina R. Filip and Feliciano Giustino, and machine-learning (ML) predictions, we develop a multi-step, DFT-free method to screen for new stable HOIPs with target electronic band-gap. Using features like the effective ionic radii of the respective species, we train and optimise a neural network on a diverse dataset consisting of over 1,300 HOIP structures and 192 chemically distinct HOIP formulas, applying the model to screen another dataset of over 1,200 new HOIPs consisting of 106 new organic cations. In this work, we further demonstrate the high accuracy of ML predictions, achieving an R^2 of 0.801 and MSE of 0.222 for band gap predictions, screening the candidate HOIPs down to 7 structures for further DFT or experimental analysis. Through our work and supporting literature, we show ML as a viable and efficient method to perform initial screenings and guide the targeted exploration of perovskite structures. Through the use of high-volume ML predictions, we hope to see the acceleration of research into perovskites, opening the opportunity for high-performance, durable perovskite solar cells in the future.

ID	A	B	X	Predicted Bandgap (eV)	Goldschmidt Tolerance Factor	Octahedral Factor
1		Sn	I	1.14	1.04	0.50
2		Sn	I	1.24	1.03	0.50
3		Sn	I	1.29	1.03	0.50
4		Sn	I	1.30	1.03	0.50
5		Sn	I	1.35	1.03	0.50
6		Sn	I	1.42	1.04	0.50
7		Pb	I	1.76	1.01	0.54

Left: Seven candidate HOIPs whose band gaps are predicted to be within the range of 1.00-2.00 eV and are inside the stable geometric region (given uncertainties of 0.1Å for A-site radii and 0.02Å for B-site and X-site radii). The perovskite ions, predicted band-gap, Goldschmidt tolerance factor and octahedral factor are also given. The A-site cations are imaged using VESTA with the C, H, and N atoms shown in dark brown, pink, and cyan, respectively.



Left: Band-gap predictions of the neural network on the test set of the Hybrid Organic-Inorganic Dataset. R^2 of 0.801 and MSE of 0.222. Horizontal lines of crosses can be seen as the dataset provides multiple band-gap calculations for the same perovskite formula due to the structural variations that can exist.