AI for Materials
https://www.sciltp.com/journals/aimat

Supplementary Materials

NJmatWeb: An AI-Driven Web Platform Automating Machine Learning and Symbolic Regression for Materials Science

Lei Zhang [1,2,*], Xinyi Wang [1,2] and Youle Wang [2]

[1] Key Laboratory of Spectral Imaging Technology, Xi'an Institute of Optics and Precision Mechanics of the Chinese Academy of Sciences, Xi'an 710119, China
[2] Department of Internet Engineering, School of Software, Nanjing University of Information Science & Technology, Nanjing 210044, China
* Correspondence: zhanglei1@opt.ac.cn

How To Cite: Zhang, L.; Wang, X.; Wang, Y. NJmatWeb: An AI-Driven Web Platform Automating Machine Learning and Symbolic Regression for Materials Science. *AI for Materials* **2026**, *1*(1), 14. https://doi.org/10.53941/aimat.2026.100014

Table S1. Performance comparison of machine learning and symbolic regression models for perovskite property prediction using the matbench_perovskites.csv dataset.

Model	Train R^2	Validation R^2	MAE	RMSE	Gap (Train-Validation)
DNN	0.452 ± 0.008	0.435 ± 0.007	0.421 ± 0.008	0.558 ± 0.009	0.017
Extra Trees	0.516 ± 0.005	0.336 ± 0.015	0.452 ± 0.011	0.605 ± 0.014	0.181
XGBoost	0.540 ± 0.005	0.305 ± 0.018	0.453 ± 0.011	0.619 ± 0.015	0.235
Random Forest	0.534 ± 0.005	0.294 ± 0.019	0.463 ± 0.011	0.624 ± 0.016	0.241
Symbolic Regression	0.250 ± 0.010	0.248 ± 0.009	0.490 ± 0.009	0.644 ± 0.012	0.002

Table S2. Performance comparison of machine learning and symbolic regression models for perovskite formation energy prediction [1].

Model	Train R^2	Validation R^2	MAE	RMSE	Gap (Train-Validation)
XGBoost	0.997 ± 0.001	0.699 ± 0.071	0.823 ± 0.039	1.217 ± 0.109	0.298
Extra Trees	0.882 ± 0.009	0.674 ± 0.041	0.890 ± 0.085	1.275 ± 0.092	0.208
Random Forest	0.855 ± 0.006	0.628 ± 0.065	0.969 ± 0.052	1.358 ± 0.086	0.227
DNN	0.775 ± 0.033	0.584 ± 0.097	0.992 ± 0.097	1.434 ± 0.161	0.191
Symbolic Regression	0.417 ± 0.027	0.255 ± 0.171	1.429 ± 0.123	1.918 ± 0.184	0.162

Table S3. Mapping between symbolic-regression variables and the corresponding descriptors.

Case	Variable	Descriptor	Original Unit
Case3	X_0	Specific Heat	J g^{-1} K^{-1}
Case3	X_1	Ion Density	g cm^3
Case3	X_2	A2 Firstionization	eV
Case3	X_3	Ion Firstionization	eV
Case3	X_4	AtomicMass	u
Case3	X_5	Ion Melting Point	K
Case4	X_0	A nO	count; dimensionless
Case4	X_1	A FMF	dimensionless fraction
Case4	X_2	B nH	count; dimensionless
Case4	X_3	B nO	count; dimensionless
Case4	X_4	B ETA_BetaP	dimensionless topological descriptor
Case4	X_5	Time	s

The units listed in this table refer to the original descriptor definitions before normalization. All descriptors were subsequently transformed using MinMaxScaler within each training fold, so the variables X_i used in the symbolic regression equations are dimensionless values scaled to the interval [0, 1].

Copyright: © 2026 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).
Publisher's Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reference

1. Hu, W.; Zhang, L. High-Throughput Calculation and Machine Learning of Two-Dimensional Halide Perovskite Materials: Formation Energy and Band Gap. *Mater. Today Commun.* **2023**, *35*, 105841. https://doi.org/10.1016/j.mtcomm.2023.105841.