Supporting Information

Machine Learning-Guided Screening of Phase-Stable High-Entropy Na Cathodes to Enable EVs and Low-Cost Charging Storage Systems

Ahhyun Jeong[a], Seungpyo Kang[a], Taeseo Ko[a], JunHo Song[b] Yerim Jung[b], Woojin Shin[a], Jiwon Sun[b,c], and Kyoungmin Min [a,*]

[a] School of Mechanical Engineering, Yonsei University, 50 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

[b] School of Mechanical Engineering, Soongsil University, 369 Sangdo-ro, Dongjak-gu, Seoul 06978, Republic of Korea

[c] Department of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60208, United States

[*] Corresponding author: Kyoungmin Min (kmin.min@yonsei.ac.kr)

S.1. Dopants selection

In this study, twelve dopant elements (B, Mg, Al, Si, Ti, Mn, Cu, Zn, Ga, Zr, Sn and Sb) were selected to enhance the structural stability and electrochemical performance of Na-based layered oxides, specifically those based on the $NaFeO_2$ structure. Each element contributes through distinct functional roles such as improving structural robustness, enhancing electronic conductivity, suppressing oxygen loss, and activating electrochemical redox reactions. Several of these dopants were selected with reference to previous findings reported in lithium-ion battery (LIB) research. Particular attention was given to properties such as high-temperature stability, suppression of phase transitions, improvement of ion/electron transport, and the ability to form stable solid solutions. This table summarizes the key functions and expected benefits of each dopant based on their physicochemical characteristics.

Dopants	Advantages
B[1][2]	Play important roles in stabilizing the lattice, suppressing phase transitions, and enhancing long-term durability in layered oxide structures.
Al[3]	
Zr[4]	Help prevent voltage fade and improve long-term cycling performance.
Mg[5]	Increases structural reversibility. Alleviates phase transitions during charge–discharge processes by promoting the formation of solid solutions.
Si[6]	Enhances electronic mobility, thereby improving electrochemical performance.
Ti[7]	Act as electrochemically inactive cations that suppress structural cracking.
Mn[8]	Enhance structural stability by increasing resistance to phase transitions.
Cu[9]	Enhances electronic mobility, thereby improving electrochemical performance. Aids in maintaining structural integrity under external conditions such as moisture exposure.
Zn[10]	Increases electronic conductivity, enabling high-rate capability. Enhances overall electrochemical performance.
Ga[11]	Improves structural stability and durability, while allowing retention of high energy density over extended cycling.
Sn[12]	Strengthens interlayer bonding. Suppresses deformation of the transition metal oxide layers, contributing to improved cycling stability.
Sb[13]	Forms strong bonds with oxygen ions, which suppress oxygen loss and improve structural stability during high-voltage cycling.

Table S1. Roles and Functions of the Twelve Selected Dopant Elements

S.2. Data construction details

In this study, approximately 150,000 materials data entries were collected from the Materials Project (MP) database, from which 145 chemical descriptors (CDs) were generated.[14] These CDs are categorized into four major types based on their underlying physical meaning. The resulting database was constructed to facilitate effective machine learning model training, providing comprehensive information on elemental and compound-level properties and enabling improved prediction accuracy and learning efficiency.

- **Stoichiometric Attributes**

  : L_p (p=2,3,5,7,10) norms of a vector representing the atomic fraction of the material corresponding to each element. p=0 norm is equivalent to the number of components.

- **Elemental-property-based attributes**

  : The minimum, maximum and range of the values of the properties of each element present in the material for each property.

- **Valence Orbital Occupation Features**

  : The fraction-weighted average of the number of valance electrons in each orbital divided by the fraction-weighted average of the total number of valance electrons.

- **Ionic compound attributes**

  : Factors related to ionic bonding, including the ability to form ionic compounds (CanFormIonic), the maximum ionic character (MaxIonicChar), and the mean ionic character (MeanIonicChar).

Chemical descriptors	The number of attributes		Attributes
Stoichiometric attributes	6		Ncomp, Comp_(L2Norm, L3Norm, L5Norm, L7Norm, L10Norm)
Elemental-property-based attributes	132	Mean (22)	Atomic Number, MendeleevNumber, AtomicWeight, MeltingT, Column, Row, CovalentRadius, Electronegativity, Ns(Valence, Unfilled), Np(Valence, Unfilled), Nd(Valence, Unfilled), Nf(Valence, Unfilled), N(Valence, Unfilled), GS(volume_pa, bandap, magmom), SpaceGropuNumber
		Range (22)	
		Dev (22)	
		Max (22)	
		Min (22)	
		Most (22)	
Valence orbital occupation attributes	4		frac_(sValence, pValence, dValence, fValence)
Ionic compound attributes	3		CanFormIonic, (Max, Mean)IonicChar

Table S2. CDs used for the model construction

S.3. Pycaret model performance details

To predict the formation energy (E_{form}) of materials, we employed PyCaret, an open-source automated machine learning framework. A total of 20 regression models were evaluated, and their performance was compared based on key metrics such as R^2 and RMSE. Among all models, the CatBoost Regressor demonstrated the best overall performance with an R^2 of 0.9375 and RMSE of 0.3007 eV/atom. The Random Forest Regressor and Light Gradient Boosting Machine (LightGBM) followed with R^2 values of 0.9311 and 0.9270, and RMSE values of 0.3158 eV/atom and 0.3249 eV/atom, respectively. The remaining models in the top 10 include Extra Trees (R^2 = 0.9199, RMSE = 0.3404), Gradient Boosting (R^2 = 0.8841, RMSE = 0.4096 eV/atom), Decision Tree (R^2 = 0.8785, RMSE = 0.4192 eV/atom), K Neighbors (R^2 = 0.8756, RMSE = 0.4243 eV/atom), Linear Regression (R^2 = 0.8073, RMSE = 0.5281 eV/atom), Bayesian Ridge (R^2 = 0.8070, RMSE = 0.5286 eV/atom), and Ridge Regression (R^2 = 0.8070, RMSE = 0.5285 eV/atom).

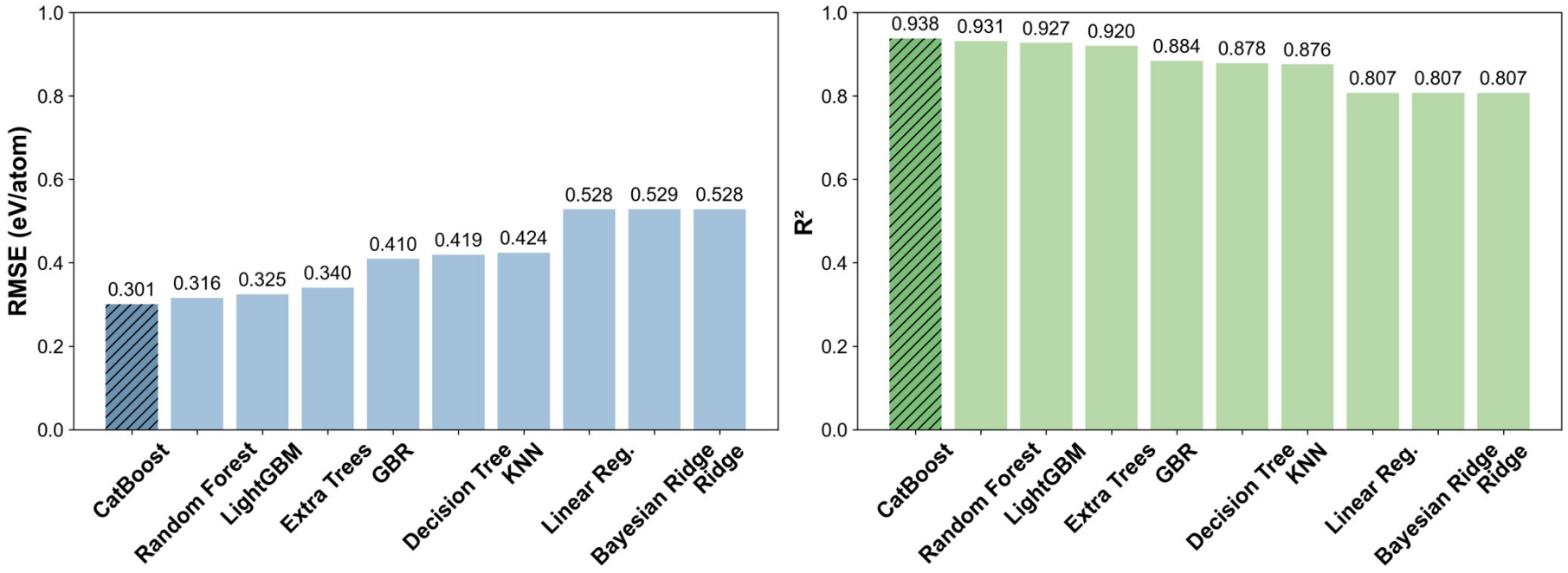

Figure S1. (a) RMSE performance comparison of the top 10 models for E_{form} prediction. (b) R^2 performance comparison of the top 10 models for E_{form} prediction.

S.4. Calculation of scalar descriptors X_1 and X_2 from GA-optimized elemental values

We generated elemental descriptors via the GA process based on 70 randomly selected compositions. To ensure robustness against sampling variability, this procedure was repeated five times, and the fixed sets of 15 elemental descriptors obtained from each run are provided in **Table S3**. These fixed descriptors were then applied to the compositions of the remaining dataset, allowing us to calculate the scalar features X_1 and X_2 using the two mismatch-based formulations summarized in **Table S4**. In this way, the two scalar features were derived for all 792 compositions.

The GA-derived scalar descriptors X_1 and X_2 were introduced to capture complementary physical aspects of multicomponent high-entropy Na-layered oxide systems that are not fully represented by conventional elemental descriptors. The scalar descriptor X_1 is a mismatch-based quantity that measures the degree of deviation of elemental descriptor values from the composition-weighted average. Physically, X_1 quantifies the extent of local compositional heterogeneity within a given compound. A larger X_1 value indicates stronger elemental mismatch, which is associated with increased local lattice strain, bond-length variability, and structural disorder in the TM–O framework. In layered TM oxides, such local heterogeneity is known to promote Jahn–Teller distortion, oxygen sublattice instability, and phase transitions between different stacking sequences (e.g., O3 to P3). In contrast, X_2 represents the mole-weighted average of the optimized elemental descriptor values and reflects the global, average bonding and electronic characteristics of the composition. This descriptor is therefore related to the overall thermodynamic stability of the TM–O framework, including average TM–O bond strength and electronic structure trends that influence E_{form} and phase preference.

Element	Trial 1	Trial 2	Trial 3	Trial 4	Trial 5
Na	(0.80, 0.61)	(0.41, 0.79)	(0.69, 0.53)	(0.43, 0.28)	(0.09, 0.41)
Fe	(0.34, 0.45)	(0.75, 0.26)	(0.29, 0.04)	(0.29, 0.4)	(0.29, 0.08)
O	(0.07, 0.50)	(0.22, 0.28)	(0.61, 0.73)	(0.63, 0.93)	(0.81, 0.78)
Mn	(0.76, 0.64)	(0.30, 0.63)	(0.44, 0.09)	(0.78, 0.61)	(0.56, 0.15)
Ti	(0.14, 0.85)	(0.46, 0.23)	(0.42, 0.62)	(0.39, 0.93)	(0.63, 0.66)
Sn	(0.38, 0.28)	(0.30, 0.79)	(0.15, 0.82)	(0.51, 0.51)	(0.26, 0.73)
Cu	(0.52, 0.71)	(0.93, 0.66)	(0.16, 0.49)	(0.09, 0.21)	(0.07, 0.23)
Mg	(0.03, 0.45)	(0.30, 0.73)	(0.24, 0.83)	(0.09, 0.39)	(0.04, 0.15)
Si	(0.17, 0.03)	(0.80, 0.98)	(0.67, 0.01)	(0.59, 0.46)	(0.51, 0.22)
Ga	(0.02, 0.39)	(0.89, 0.13)	(0.79, 0.73)	(0.17, 0.89)	(0.18, 0.94)
Zn	(0.68, 0.04)	(0.68, 0.62)	(0.87, 0.83)	(0.06, 0.65)	(0.20, 0.93)
Sb	(0.68, 0.12)	(0.82, 0.42)	(0.48, 0.75)	(0.94, 0.07)	(0.19, 0.40)
Zr	(0.07, 0.04)	(0.12, 0.43)	(0.91, 0.33)	(0.89, 0.53)	(0.55, 0.58)
Al	(0.54, 0.33)	(0.56, 0.06)	(0.77, 0.59)	(0.49, 0.67)	(0.38, 0.17)
B	(0.50, 0.06)	(0.75, 0.21)	(0.94, 0.62)	(0.30, 0.62)	(0.98, 0.97)

Table S3. Optimized elemental descriptors for 15 elements from five random sampling trials using GA.

Scalar features	Formula
X_1 [15]	$X_1 = \sqrt{\sum_{i=1}^{n} c_i \left(1 - \frac{\vec{x}_i}{X}\right)^2}$
X_2 [15]	$X_2 = \sum_{i=1}^{n} c_i \vec{x}_i$

Table S4. Equations for scalar features based on elemental descriptors.

S.5. Structural and Electrochemical Characteristics of Top 5 Screened Compositions

This section presents the final screening results for the top 5 compositions, summarizing their formation energy (E_{form}), energy above the convex hull (E_{hull}), O3–P3 phase energy difference (ΔE_{phase}), average voltage (V_{avg}), and volume change (ΔV) in **Table S5**. These properties highlight the overall electrochemical and thermodynamic suitability of each composition.

In addition, **Table S6** provides a detailed analysis of the local bonding environments in the optimized structures. It includes the average bond lengths and standard deviations between dopant elements and oxygen atoms, offering insight into local lattice distortion and potential structural stability.

Composition	E_{form} (eV/atom)	E_{hull} (eV/atom)	ΔE_{phase} (eV/atom)	V_{avg} (V)	$\lvert\Delta V\rvert$ (%)
$Na_{27}Mg_3Ti_3Zn_3Ga_3Fe_{12}Cu_3O_{54}$	-1.7080	0.2938	-0.1401	3.2461	2.0467
$Na_{27}Ti_3Mn_3Zn_3Ga_3Fe_{12}Cu_3O_{54}$	-1.7127	0.2704	-0.8633	3.0619	3.6077
$Na_{27}Ti_3Mn_3Zn_3Fe_{12}Cu_3Sb_3O_{54}$	-1.6850	0.2958	-1.2622	2.5023	2.0153
$Na_{27}Mg_3Mn_3Zn_3Ga_3Fe_{12}Cu_3O_{54}$	-1.5843	0.2826	-0.7700	3.1968	2.9968
$Na_{27}Mn_3Zn_3Ga_3Fe_{12}Cu_3B_3O_{54}$	-1.6465	0.2600	-0.3403	3.1206	3.6749

Table S5. Summary of calculated properties for the top 5 screened compositions that satisfy all stability and performance criteria.

Formulae	Mg-O bond length [Å]	Ti-O bond length [Å]	Mn-O bond length [Å]	Zn-O bond length [Å]	Ga-O bond length [Å]	Cu-O bond length [Å]	Sb-O bond length [Å]	B-O bond length [Å]
$Na_{27}Mg_3Ti_3Zn_3Ga_3Fe_{12}Cu_3O_{54}$	2.108 ± 0.022	2.001 ± 0.030	-	2.118 ± 0.028	2.050 ± 0.031	2.059 ± 0.028	-	-
$Na_{27}Ti_3Mn_3Zn_3Ga_3Fe_{12}Cu_3O_{54}$	-	2.015 ± 0.066	2.010 ± 0.081	2.158 ± 0.052	2.054 ± 0.029	2.125 ± 0.047	2.018 ± 0.034	-
$Na_{27}Ti_3Mn_3Zn_3Fe_{12}Cu_3Sb_3O_{54}$	-	2.013 ± 0.054	2.128 ± 0.086	2.151 ± 0.045	-	2.132 ± 0.065	-	-
$Na_{27}Mg_3Mn_3Zn_3Ga_3Fe_{12}Cu_3O_{54}$	2.120 ± 0.007	-	1.973 ± 0.020	2.137 ± 0.021	2.048 ± 0.018	2.067 ± 0.015	-	-
$Na_{27}Mn_3Zn_3Ga_3Fe_{12}Cu_3B_3O_{54}$	-	-	1.982 ± 0.060	2.159 ± 0.084	2.072 ± 0.103	2.105 ± 0.115	-	1.407 ± 0.009

Table S6. Average bond lengths and corresponding standard deviations for each dopant–O pair in the top 5 screened compositions.

S.6. Screening O3 cathode candidates for SIBs

To evaluate whether the final five candidates arise from physically meaningful screening criteria rather than arbitrary numerical choices, each cutoff was independently applied to the full set of 792 compositions. When examined separately, all 792 structures satisfied the formation-energy requirement (E_{form} < 0 eV/atom), confirming that this criterion imposes essentially no restriction on the dataset. In contrast, the metastability threshold (E_{hull} ≤ 0.3 eV/atom) was found to be the strongest filter, reducing the pool from 792 to 99 structures.

The O3/P3 phase-preference condition (ΔE_{phase} < 0) was comparatively permissive, with 596 compositions meeting this requirement. The voltage criterion (2 V ≤ V_{avg} ≤ 4 V) was also non-restrictive for most structures, satisfied by 746 compositions. Meanwhile, the volume-change constraint ($|\Delta V| \leq 5\%$) served as the second most restrictive step, with only 238 compositions passing this condition.

These results confirm that the two dominant factors governing the screening outcome are the metastability threshold (E_{hull}) and the structural robustness criterion ($|\Delta V|$). The final five candidates therefore represent robust selections guided by physically meaningful filters rather than artifacts of arbitrary cutoff choices.

Composition	S_{conf}	Reference
$Na_{0.67}(Mn_{0.45}Ni_{0.18}Co_{0.18}Ti_{0.10}Mg_{0.03}Al_{0.04}Fe_{0.02})O_2$	1.519R	[16]
$NaMn_{0.2}Fe_{0.2}Co_{0.2}Ni_{0.2}Ti_{0.2}O_2$	1.609R	[17]
$NaNi_{0.3}Cu_{0.05}Fe_{0.1}Mn_{0.3}Mg_{0.05}Ti_{0.2}O_2$	1.574R	[18]
$NaNi_{0.3}Cu_{0.1}Fe_{0.2}Mn_{0.3}Ti_{0.1}O_2$	1.505R	[19]
$Na_{0.95}Li_{0.06}Ni_{0.25}Cu_{0.05}Fe_{0.15}Mn_{0.49}O_2$	1.498R	[20]

Table S7. Experimentally synthesized high-entropy Na-layered oxide compositions reported in the literature. The calculated S_{conf} for each composition was obtained using Equation (1), confirming that all (or most) exceed the high-entropy criterion ($S_{conf} \geq 1.5$ R).

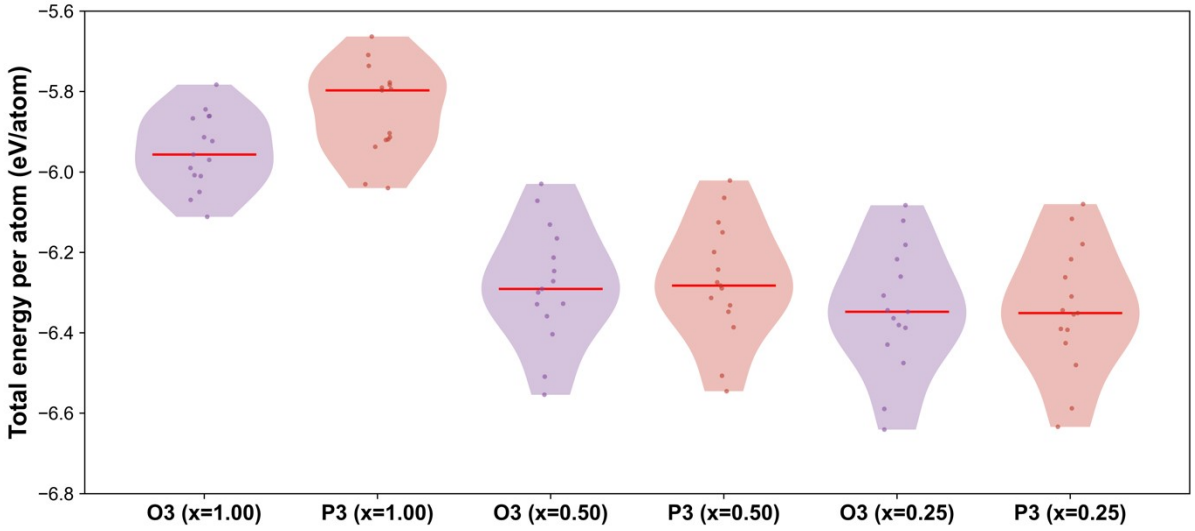

Figure S2. Violin plot of total energy per atom (eV/atom) for O3 and P3 phases in x=1.00, x=0.50, and x=0.25 states

Average voltage (Vavg), volume change ($^{\Delta}$V) were calculated by (1) and (2) equations, respectively, and $^{x}_{2}$=1, $^{x}_{1}$=0.

The variables in the Vavg calculation are as follows: μ[Na] = -1.312223005 eV (chemical potential of Na), E = total energy, and e = 1.602176634×10^{-19} (charge of one electron).

$$V_{avg}\,[V]\,(x_1 \le x \le x_2) = \frac{(x_2 - x_1)\mu[Na] - E[Na_{x2}MO_2] + E[Na_{x1}MO_2]}{(x_2 - x_1)e} \cdots (1)$$ [21]

Volume change' variable is as follows: v = volume.

$$\Delta v\,[\%]\,(x_1 \le x \le x_2) = \frac{v_{x2} - v_{x1}}{v_{x1}} \times 100 \cdots (2)$$ [21]

S.7. Active learning details

To evaluate the learning efficiency of the active learning (AL) framework, we applied two ensemble-based classifiers: Random Forest (RF) and Gradient Boosting Classifier (GBC). During the AL process, the model was iteratively updated by selecting informative samples based on uncertainty sampling. Specifically, model uncertainty for each unlabeled sample was quantified by analyzing the predicted class-probability distribution output by the classifier. This distribution, denoted as $P(y \mid x_i) = [p1, p_2, \cdots, p_K]$, represents the model's confidence across all possible classes. When the probabilities are concentrated in a single class(e.g., one $p_k \approx 1$), the model is confident, whereas a more uniform distribution(e.g., one $p_k \approx 1/K$) indicates greater uncertainty. To measure this uncertainty quantitatively, we computed the Shannon entropy of the probability distribution[22][23]:

$$U(x_i) = -\sum_{k=1}^{K} p_k log(p_k)$$

Samples with higher entropy values were regarded as more uncertain and thus more informative for the next labeling step. The validity of the entropy-based uncertainty metric was examined by quantifying its correlation with prediction errors(**Figure S3**). A statistically significant positive correlation ($r = 0.18$, $p < 0.01$) was observed, and the bin-wise analysis revealed that samples with higher entropy exhibited larger error rates, confirming that the uncertainty measure effectively reflects model confidence. The average classification accuracy at each iteration, calculated over 3 independent runs, is summarized in **Table S7**. These values reflect the overall model performance improvement as the labeled dataset grew during AL cycles.

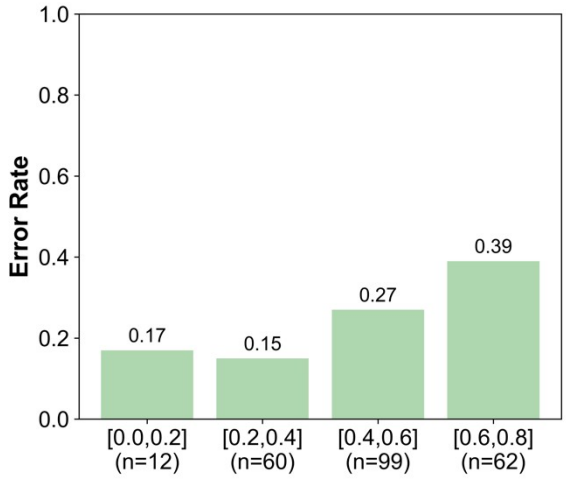

Figure S3. Average error rate across different uncertainty bins, showing an increasing trend with higher uncertainty levels.

Iteration	RF Accuracy	GBC Accuracy
0	0.7242	0.6920
1	0.7402	0.7104
2	0.7412	0.7124
3	0.7502	0.7214
4	0.7522	0.7248
5	0.7564	0.7340
6	0.7580	0.7412
7	0.7612	0.7460
8	0.7638	0.7496
9	0.7652	0.7504

Table S8. Average classification accuracy of RF and GBC models at each iteration of the active learning process.

Model	Initial	Final
GBC (AL)	0.6448	0.7962
GBC (PL)	0.6448	0.7577
RF (AL)	0.7208	0.8154
RF (PL)	0.7208	0.7500

Table S9. Quantitative comparison of mean accuracies at the initial and final iterations for GBC and RF classifiers under active (AL) and passive (PL) learning strategies.

Model	Target accuracy	Iteraion (AL)	Iteration (PL)	Labels (AL)	Labels (PL)	Relative cost saving
GBC	0.70	1	2	30	40	25%
GBC	0.72	3	-	50	-	-
GBC	0.775	9	-	110	-	-
RF	0.70	0	0	20	20	0%
RF	0.72	0	0	20	20	0%
RF	0.75	3	-	50	-	-

Table S10. Estimated computational cost savings of active learning (AL) compared to passive learning (PL) for GBC and RF classifiers.

REFERENCES

[1] S.-H. Lee, B.-S. Jin, H.-S. Kim, Superior Performances of B-doped LiNi0.84Co0.10Mn0.06O2 cathode for advanced LIBs, Sci Rep 9 (2019) 17541. https://doi.org/10.1038/s41598-019-54115-z.

[2] S. Jamil, L. Yue, C. Li, M. Fasehullah, M. Aizaz Ud Din, W. Yang, S. Bao, M. Xu, Significance of gallium doping for high Ni, low Co/Mn layered oxide cathode material, Chemical Engineering Journal 441 (2022) 135821. https://doi.org/10.1016/j.cej.2022.135821.

[3] L. Wang, Y. Wang, X. Yang, J. Wang, X. Yang, J. Tang, Excellent cyclability of P2-type Na–Co–Mn–Si–O cathode material for high-rate sodium-ion batteries, J Mater Sci 54 (2019) 12723–12736. https://doi.org/10.1007/s10853-019-03807-y.

[4] M. Leng, J. Bi, Z. Xing, W. Wang, X. Gao, J. Wang, Z. Qian, A new perspective on the composition-structure-property relationships on Nb/Mo/Cr-doped O3-type layered oxide as cathode materials for sodium-ion batteries, Chemical Engineering Journal 413 (2021) 127824. https://doi.org/10.1016/j.cej.2020.127824.

[5] H.-R. Yao, W.-J. Lv, Y.-X. Yin, H. Ye, X.-W. Wu, Y. Wang, Y. Gong, Q. Li, X. Yu, L. Gu, Z. Huang, Y.-G. Guo, Suppression of Monoclinic Phase Transitions of O3-Type Cathodes Based on Electronic Delocalization for Na-Ion Batteries, ACS Appl Mater Interfaces 11 (2019) 22067–22073. https://doi.org/10.1021/acsami.9b00186.

[6] H.-H. Ryu, G. Han, T.-Y. Yu, Y.-K. Sun, Enhanced Cycling Stability of O3-Type Na[Ni 0.5 Mn 0.5]O 2 Cathode through Sn Addition for Sodium-Ion Batteries, The Journal of Physical Chemistry C 125 (2021) 6593–6600. https://doi.org/10.1021/acs.jpcc.1c00522.

[7] Y. Sui, Y. Hao, X. Zhang, J. Li, G. Wen, S. Zhong, Z. Zhang, L. Wu, Improved electrochemical properties of vanadium substituted Na0·67Fe0·5Mn0·5O2 cathode material for sodium-ion batteries, Ceram Int 47 (2021) 5227–5234. https://doi.org/10.1016/j.ceramint.2020.10.102.

[8] Y. Zhang, J. Wang, L. Wang, L. Duan, G. Zhang, F. Zhao, X. Zhang, W. Lü, The multi-metal synergetic mechanism of O3-Na0.5Mn0.65Ni0.15Al0.1Mg0.05Co0.05O2 nanoflower for a high-voltage and long-cycle-life cathode material of sodium-ion batteries, J Mater Sci 55 (2020) 13102–13113. https://doi.org/10.1007/s10853-020-04940-9.

[9] Y.-J. Guo, P.-F. Wang, Y.-B. Niu, X.-D. Zhang, Q. Li, X. Yu, M. Fan, W.-P. Chen, Y. Yu, X. Liu, Q. Meng, S. Xin, Y.-X. Yin, Y.-G. Guo, Boron-doped sodium layered oxide for reversible oxygen redox reaction in Na-ion battery cathodes, Nat Commun 12 (2021) 5267. https://doi.org/10.1038/s41467-021-25610-7.

[10] H. Wang, R. Gao, Z. Li, L. Sun, Z. Hu, X. Liu, Different Effects of Al Substitution for Mn or Fe on the Structure and Electrochemical Properties of Na 0.67 Mn 0.5 Fe 0.5 O 2 as a Sodium Ion Battery Cathode Material, Inorg Chem 57 (2018) 5249–5257. https://doi.org/10.1021/acs.inorgchem.8b00284.

[11] N. Kuganathan, A. Chroneos, Na3V(PO4)2 cathode material for Na ion batteries: Defects, dopants and Na diffusion, Solid State Ion 336 (2019) 75–79. https://doi.org/10.1016/j.ssi.2019.03.025.

[12] P. Li, T. Yuan, J. Qiu, H. Che, Q. Ma, Y. Pang, Z.-F. Ma, S. Zheng, A comprehensive review of layered transition metal oxide cathodes for sodium-ion batteries: The latest advancements and future perspectives, Materials Science and Engineering: R: Reports 163 (2025) 100902. https://doi.org/10.1016/j.mser.2024.100902.

[13] N. Hong, K. Wu, Z. Peng, Z. Zhu, G. Jia, M. Wang, Improved High Rate Performance and Cycle Performance of Al-Doped O3-Type NaNi 0.5 Mn 0.5 O 2 Cathode Materials for Sodium-Ion Batteries, The Journal of Physical Chemistry C 124 (2020) 22925–22933. https://doi.org/10.1021/acs.jpcc.0c06032.
[14] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K.A. Persson, Commentary: The materials project: A materials genome approach to accelerating materials innovation, APL Mater 1 (2013). https://doi.org/10.1063/1.4812323.
[15] Y. Zhang, C. Wen, P. Dang, X. Jiang, D. Xue, Y. Su, Elemental numerical descriptions to enhance classification and regression model performance for high-entropy alloys, NPJ Comput Mater 11 (2025). https://doi.org/10.1038/s41524-025-01560-2.
[16] M. Kim, W.H. Yeo, K. Min, Co-free and low strain cathode materials for sodium-ion batteries: Machine learning-based materials discovery, Energy Storage Mater 69 (2024). https://doi.org/10.1016/j.ensm.2024.103405.
[17] D.D. Lewis, A Sequential Algorithm for Training Text Classifiers: Corrigendum and Additional Data, n.d.
[18] B. Settles, Computer Sciences Department Active Learning Literature Survey, 2009.