Supplementary Information

Data-Driven Design Strategies for Chemically Stabilizing Cathode-Coating Interfaces Employing Interpretable Machine Learning

Sang-Hyeon Park[1], Donghee Kang[2], Hyun-Ki Yoon[1], Yusong Choi[1] and Jongseo Lee[1]*

[1]Defense Materials and Energy Development Directorate, Agency for Defense Development, Yuseong P.O. Box 35, Daejeon, 34060, Korea. [*]E-mail: jongse7420@gmail.com

[2]Chem-Bio Technology Directorate, Agency for Defense Development, Yuseong P.O. Box 35, Daejeon, 34060, Korea

Table of Contents

Table S1. Number of coating-cathode pairs for data construction. Pairs with TRUE-labeled undergo decomposition reaction with consistent reaction energies calculated by two different methods: Total energy and formation energy considerations. Pairs with NONRXT-labeled indicates that there is no thermodynamically favorable reaction pathway to decompose. Pairs with FALSE-labeled are not considered in this study, since two different reaction energy calculations result in different values of reaction energy.

Coating type	TRUE	NONRXT	FALSE
M-O oxides	17818	1576 (7.77%)	896 (4.42%)
M-O-F oxyfluorides	10020	969 (8.30%)	687 (5.88%)
M-F fluorides	7890	3099 (24.5%)	1661 (13.1%)
Total	35728	5644	3244 (7.27%)

Table S2. Mean and median reaction energies of cathode-coating pairs.

Coating	Cathodes	# pairs	Mean E_{rxn} (meV/atom)	Mean E_{rxn} (meV/atom)	Median E_{rxn} (meV/atom)
Oxides	charged	7,728	-96.1	-120.4	-90.9
	discharged	10,090		-77.5	-57.6
Oxyfluorides	charged	4,234	-63.1	-76.2	-55.2
	discharged	5,786		-53.5	-42.5
Fluorides	charged	3,094	-32.9	-31.7	-22.6
	discharged	4,796		-33.8	-22.5

Table S3. Detailed feature information considered in this study. Group 1, 2, and 3 are classified as intrinsic materials properties, composition-derived descriptors, and cathode-coating interaction-based features, respectively. Abbreviations for top twenty important features are provided.

Group	No.	Feature name	Description (Abbreviation, rank)	Type
1	1	working_ion	Li, Mg, Ca, Na, and K	Categorial
	2	coating_system	Oxide, oxyfluoride, fluoride (Coat_sys, #10)	
	3	cathode_type	e.g., phosphates, fluorides (Cat_type, #16)	
	4	E_hull_cathodes (meV/atom)	E_{hull} (meV/atom) of cathodes (E_hull, #7)	Numeric
	5	formE_cathode (meV/atom)	Formation energies (FormE_cat, #3)	
	6	formE_coatings (meV/atom)	(FormE_coat, #9)	
	7[ex]	coating denstiy (g/cm^3)	Densities	
	8	cathode density (g/cm^3)	(Density_cat, #13)	
	9	coating bandgap (eV)	Bandgaps (BG_coat, #5)	
	10	cathode bandgap (eV)	(BG_cat, #14)	
	11	ox_coat_cation_mean	Average oxidation state of cations in coating materials (Oxstate_coat, #20)	
	12	coating_subtype	e.g., transition-metal oxides, rare-earth fluorides (Coat_sub, #18)	Categorial
2	13	en_cat_mean_en	Average electronegativity (EN) in cathode (cat) materials	Numeric
	14	en_cat_wmean_en	Weighted average EN (EN_cat_wm, #11)	
	15	en_cat_wstd_en	Weighted standard deviation of EN (EN_cat_wstd, #4)	
	16[ex]	en_cat_min_en	Minimum EN in cathode materials	
	17	en_cat_max_en	Maximum EN (EN_cat_max, #19)	
	18	en_cat_range_en	EN range (maximum-minimum)	
	19	en_coat_mean_en	Average EN in coating (coat) materials (EN_coat_mean, #2)	
	20	en_coat_wmean_en		
	21[ex]	en_coat_wstd_en		
	22[ex]	en_coat_min_en		
	23	en_coat_max_en		
	24	en_coat_range_en		
	25	sto_cat_ncomp	Stoichiometry (sto) based, Number of elements in cathode (cat)	
	26[ex]	sto_cat_max_frac	Maximum fraction of elements in given composition	
	27	sto_cat_min_frac	Minimum fraction of elements in given composition	
	28	sto_cat_l2	L2 norm (L2_cat, #15)	
	29	sto_coat_ncomp		
	30	sto_coat_max_frac	(Minfrac_cat, #6)	
	31	sto_coat_min_frac		
	32[ex]	sto_coat_l2		
3	33	sto_int_ncomp_ratio	No. 25 / No. 29	
	34	sto_int_l2_diff	No. 28 – No. 32 (ΔL2_int, #17)	
	35[ex]	sto_int_max_frac_diff	No. 26 – No. 30	
	36	diff_formE	No. 5 – No. 6	
	37	diff_bandgap	No. 9 – No. 10	
	38	en_int_mean_diff	Interactive(int), No. 13 – No. 19	
	39	en_int_wmean_diff	No. 14 – No. 20 (ΔEN_wm, #1)	
	40	en_int_wstd_diff	No. 15 – No 21 (ΔEN_wstd, #12)	
	41	en_mismatch_score	Specially designed electronegativity mismatch score (EN_mism, #8)	

[ex]Excluded features from the two-step feature selection procedure.

Table S4. Summary for five different classification models performance. Mean and test values are obtained from five-fold cross-validation and hold-out (8:2 split) test results. Standard deviation (std. dev.), mean f1 score, and AUC (area under the ROC curve) difference between validation and test results are tabulated.

Classification models						
Model	AUC_{mean}	Std. dev.	AUC_{test}	ΔAUC	$f1_{mean}$	Std. dev.
Logistic Regression	0.8737	0.0066	0.8804	0.0067	0.8656	0.0023
Random Forest	0.9786	0.0019	0.9809	0.0023	0.9672	0.0013
XGBoost	0.9914	0.0012	0.9923	0.0009	0.9783	0.0010
LightGBM	0.9913	0.0011	0.9929	0.0016	0.9772	0.0011
MLP	0.9815	0.0021	0.9844	0.0029	0.9748	0.0018

Table S5. Summary for five different regression models performance. Mean and test values are obtained from five-fold cross-validation and hold-out (8:2 split) test results. R^2 scores, tandard deviation (std. dev.), mean and test MAE (mean absolute error) values are tabulated.

Regression models						
Model	R^2_{mean}	Std. dev.	R^2_{test}	ΔR^2	MAE_{mean} (meV/atom)	MAE_{test} (meV/atom)
Ridge	0.6025	0.0067	0.6076	0.0051	35.13	35.02
Random Forest	0.8984	0.0045	0.9004	0.0020	14.71	14.51
XGBoost	0.9174	0.0039	0.9176	0.0002	14.48	14.54
LightGBM	0.9103	0.0022	0.9129	0.0026	15.35	15.07
MLP	0.9151	0.0069	0.9068	-0.0083	14.94	15.25

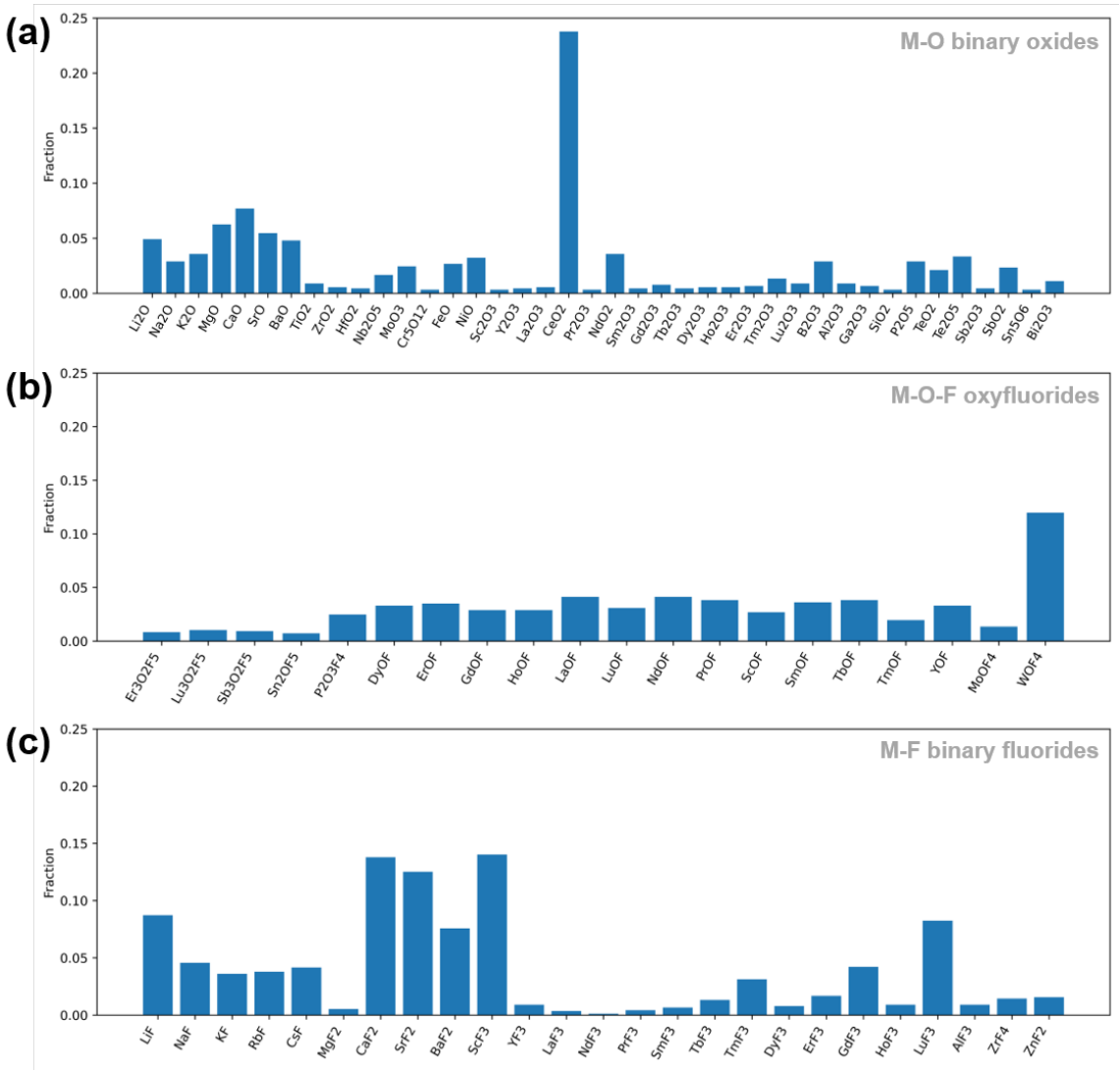

Fig. S1. Normalized occurrence (fraction) of coating materials excluded due to inconsistency between two computational methods, categorized into (a) M-O binary oxides, (b) M-O-F oxyfluorides, and (c) M-F binary fluorides. Each bar represents the fraction of a given coating material within the excluded dataset.

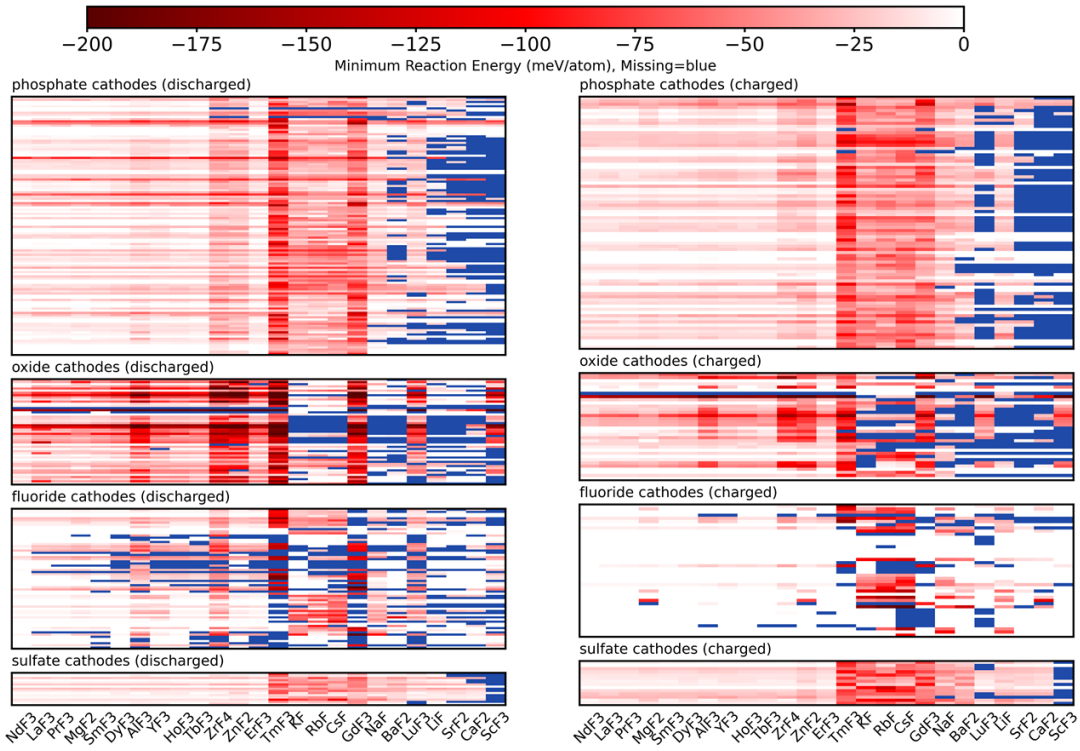

Fig. S2. Semi-global heatmap for reaction energies of group-separated cathodes with M-F binary metal fluorides. The darker color indicates the more negative E_{rxn} value. Left and right panels correspond to discharged and charged cathode materials, respectively.

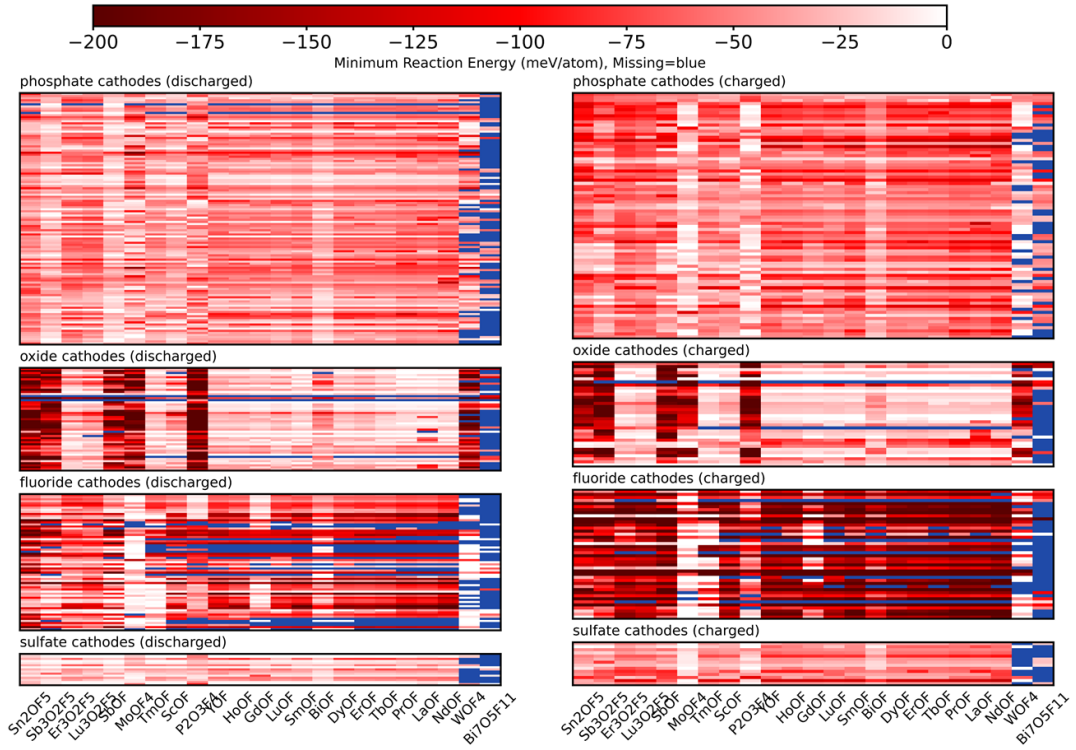

Fig. S3. Semi-global heatmap for reaction energies of group-separated cathodes with M-O-F ternary metal oxyfluorides. The darker color indicates the more negative E_{rxn} value. Left and right panels correspond to discharged and charged cathode materials, respectively.

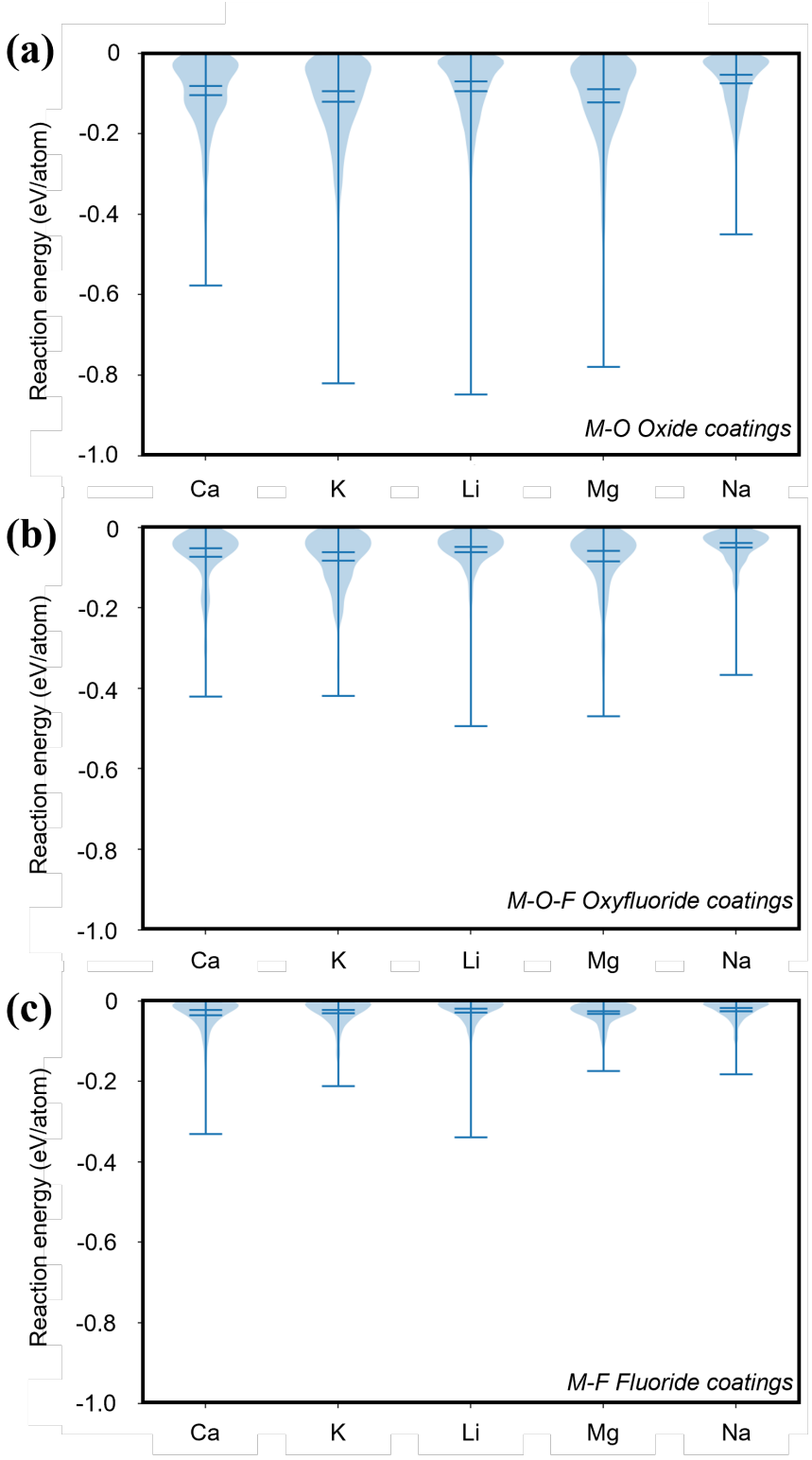

Fig. S4. Reaction energy insensitivity for working-ion species (Ca, K, Li, Mg, and Na) in cathode materials. (a), (b), and (c) represents reaction (decomposition) energy (eV/atom) distribution between working-ion specified cathodes and oxides, oxyfluorides, and fluorides coating materials, respectively.

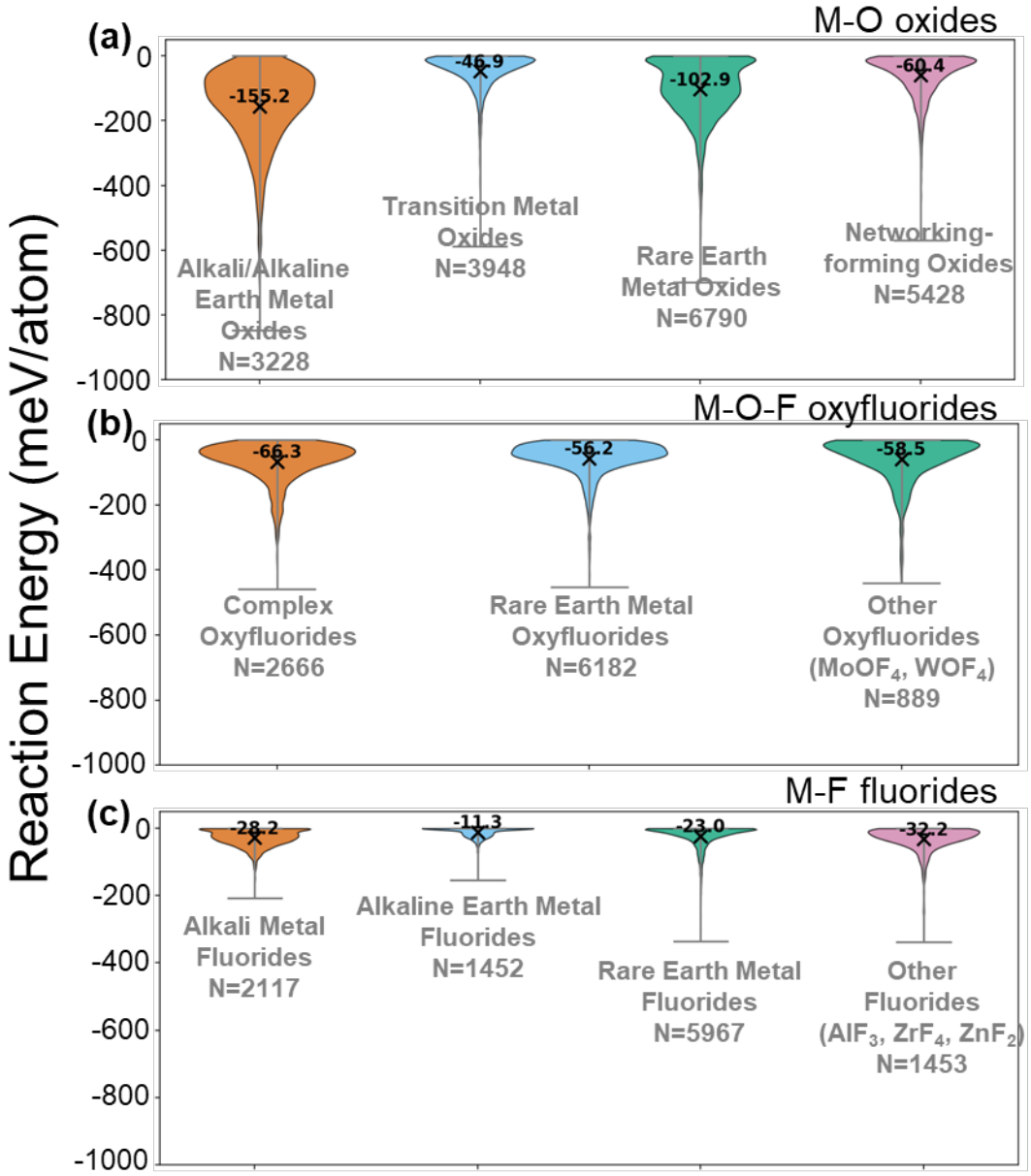

Fig. S5. Subtype-divided reaction energy distributions for (a) M-O oxides, (b) M-F fluorides, and (c) M-O-F oxyfluorides. Mean values are marked in cross signs, and specific compositions corresponding to each subtype is provided in figure 2b and figure 3 in the manuscript. N refers to total number of cathode-coating combinations for visualizing the distribution.

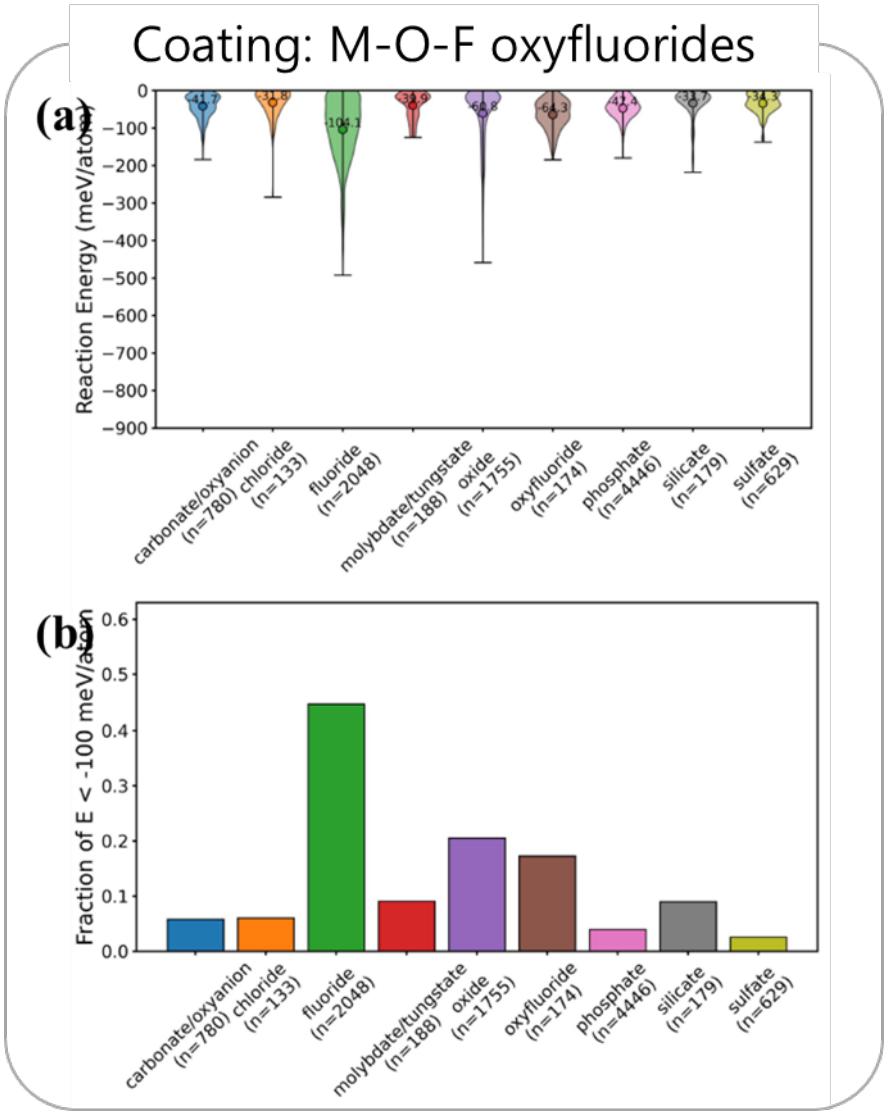

Fig. S6. Oxyfluoride coating materials reaction energy distributions according to cathode subtypes. (a) Distributions of reaction energy between subtype-classified cathodes and oxyfluoride coatings. (b) Fraction of cathode and oxyfluoride coating combinations whose reaction energy is less than -100 meV/atom.

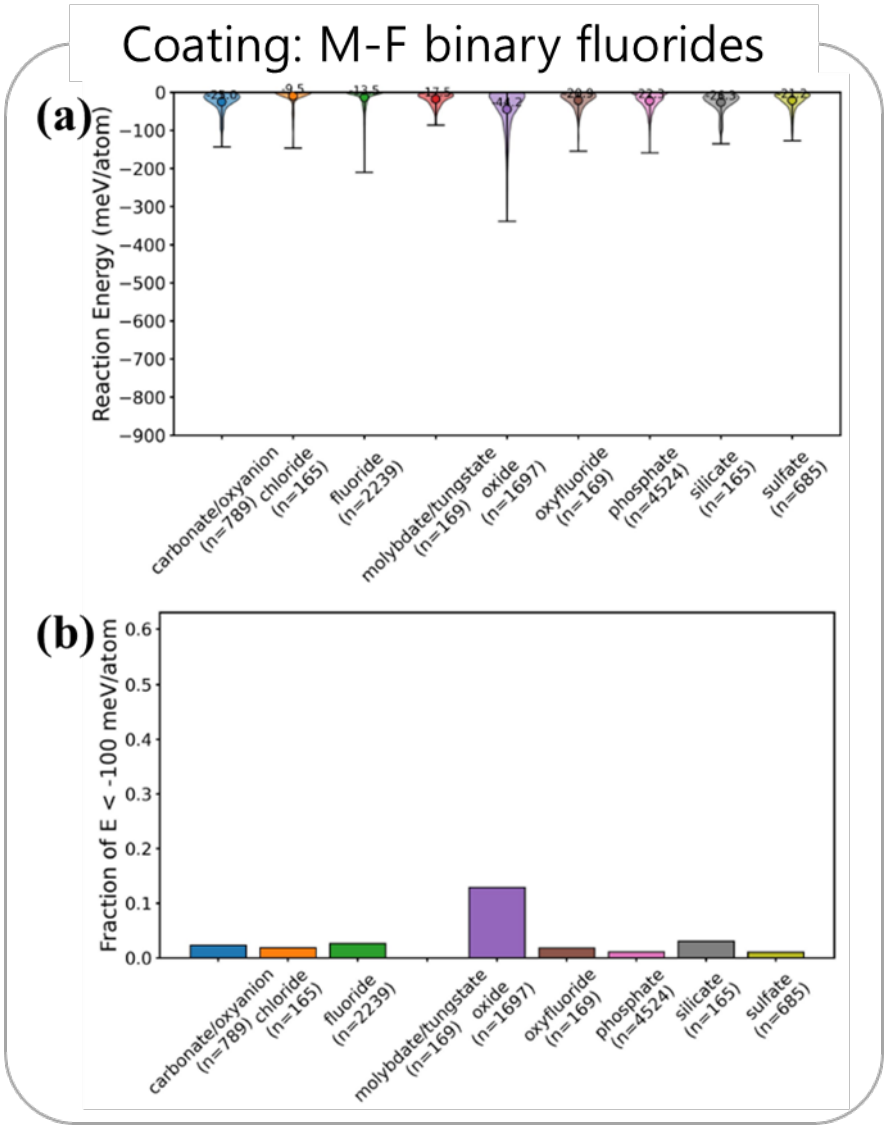

Fig. S7. Fluoride coating materials reaction energy distributions according to cathode subtypes. (a) Distributions of reaction energy between subtype-classified cathodes and fluoride coatings. (b) Fraction of cathode and fluoride coating combinations whose reaction energy is less than -100 meV/atom.

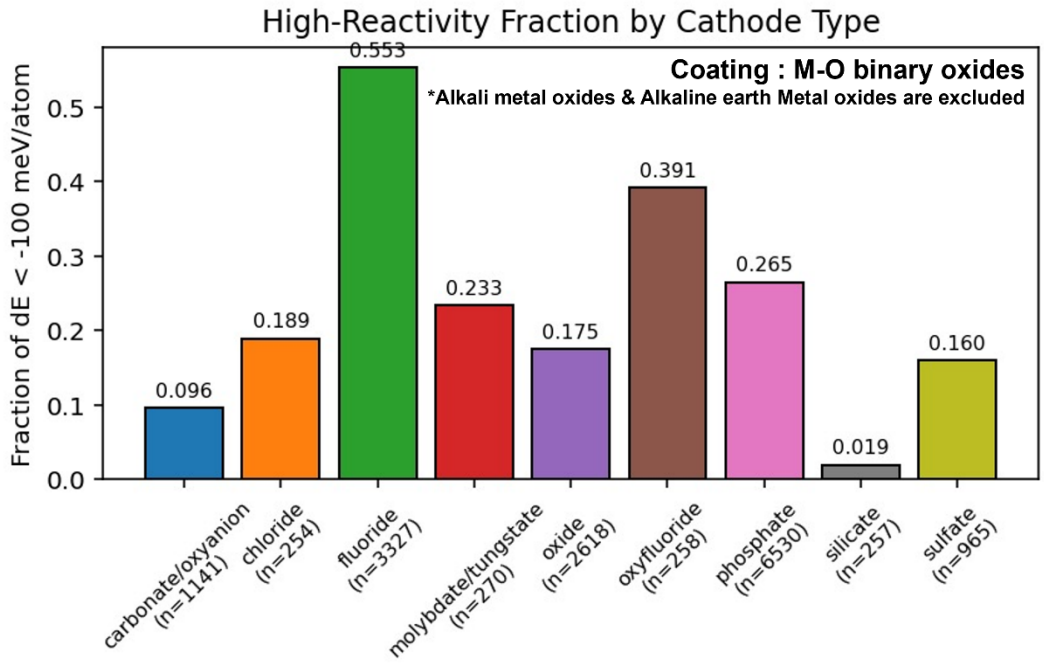

Fig. S8. Fraction of cathode and oxide coating combinations whose reaction energy is less than -100 meV/atom. M_2O-type alkali metal oxides and MO-type alkaline earth metal oxides, identified as highly reactive coating species, are excluded to avoid the trap of the average.

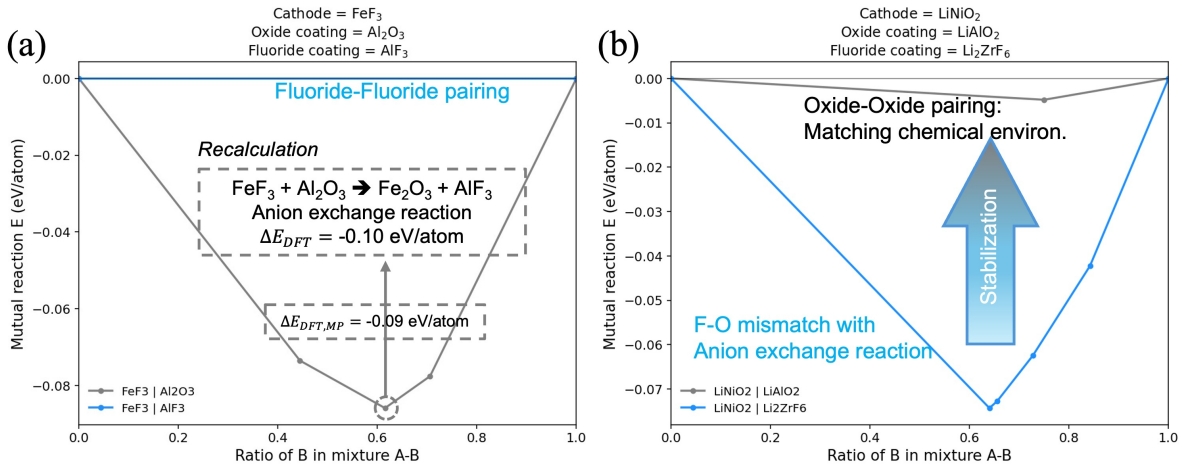

Fig. S9. Representative thermodynamic reaction energy landscapes for selected cathode–coating combinations, illustrating the relationship between anion chemistry and interfacial compatibility. (a) FeF_3-AlF_3 and FeF_3-Al_2O_3 pairs with independently conducted DFT recalculation result. (b) represents $LiNiO_2$-$LiAlO_2$ and $LiNiO_2$-Li_2ZrF_6 pairs. The mutual reaction energy was calculated as a function of the mixing ratio between cathode and coating compositions, where more negative values indicate a stronger thermodynamic driving force for interfacial decomposition.

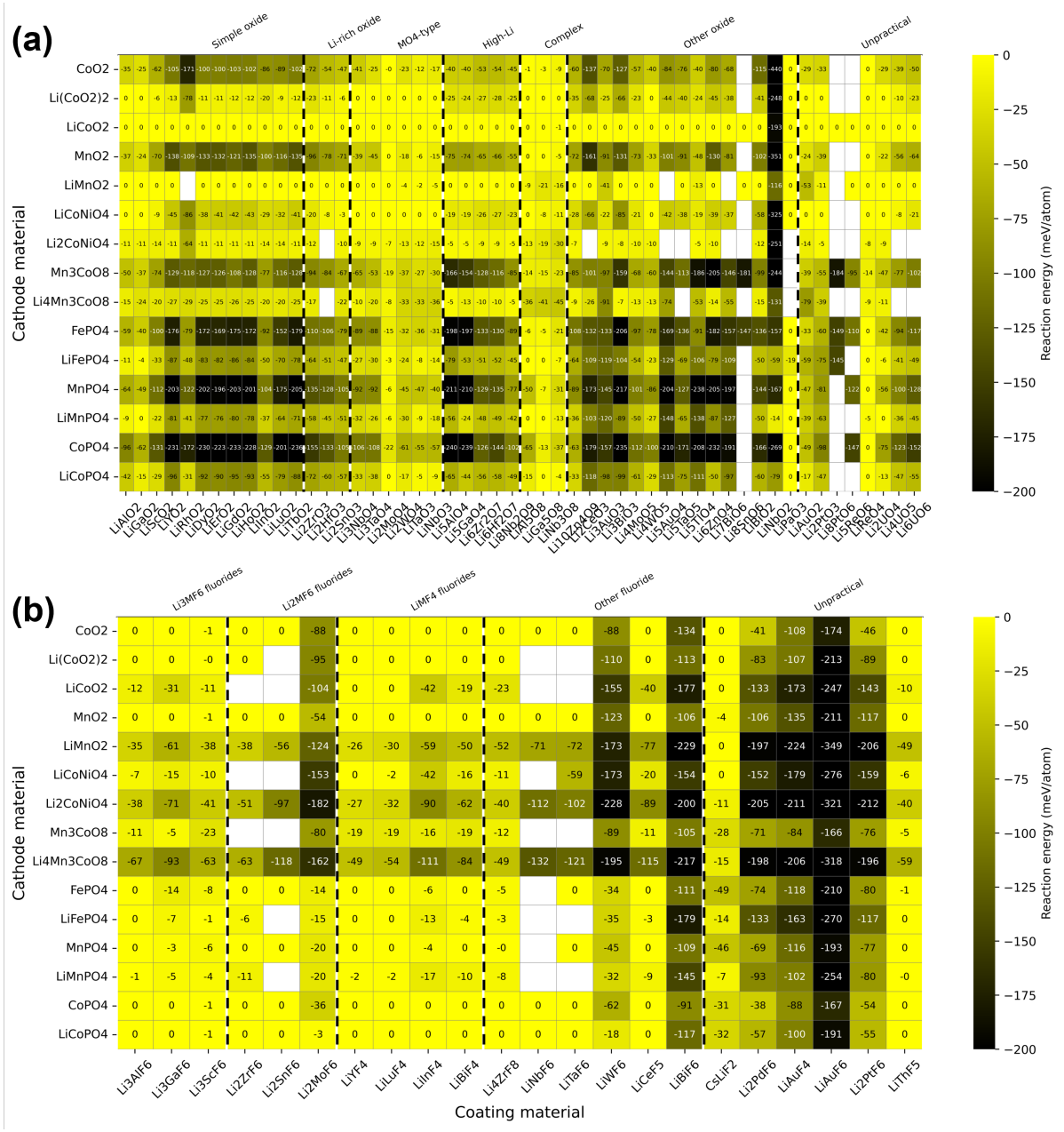

Fig. S10. Reaction energy landscapes presented as local heatmaps for promising/commercial cathodes, where coating material classes include (a) Li-M-O ternary oxides and (b) Li-M-F ternary fluorides. Darker colors indicate more negative E_{rxn} values.

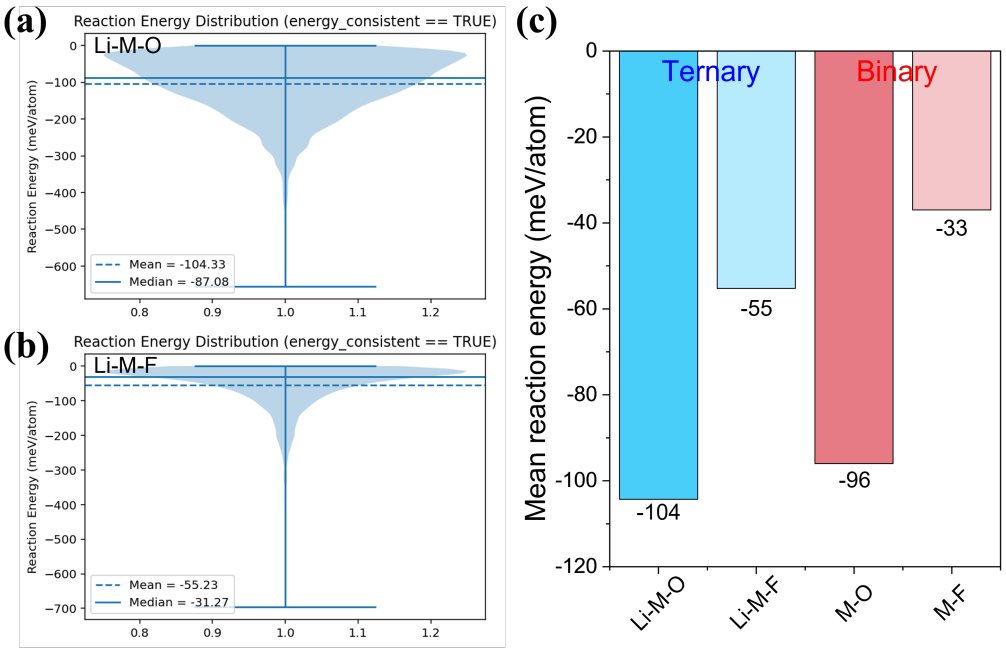

Fig. S11. Reaction energy distribution of Li-M-A ((a) A=oxygen and (b) A=fluorine) ternary compounds and lithium-ion battery cathode combinations. (c) Direct comparison of mean reaction energies with cathodes, between ternary oxide/fluoride compounds and binary oxide/fluoride coating materials.

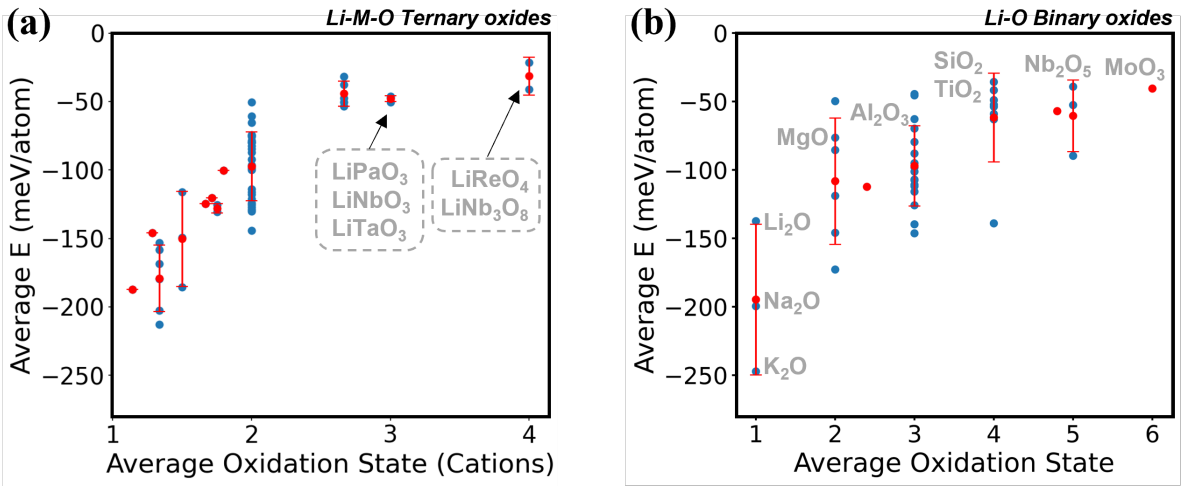

Fig. S12. Scatter plots (blue points) for mean reaction energy and average oxidation state of metal species in (a) Li-M-O ternary and (b) M-O binary oxide materials. Red dots and error bars represent total average and standard deviations in specific oxidation states.

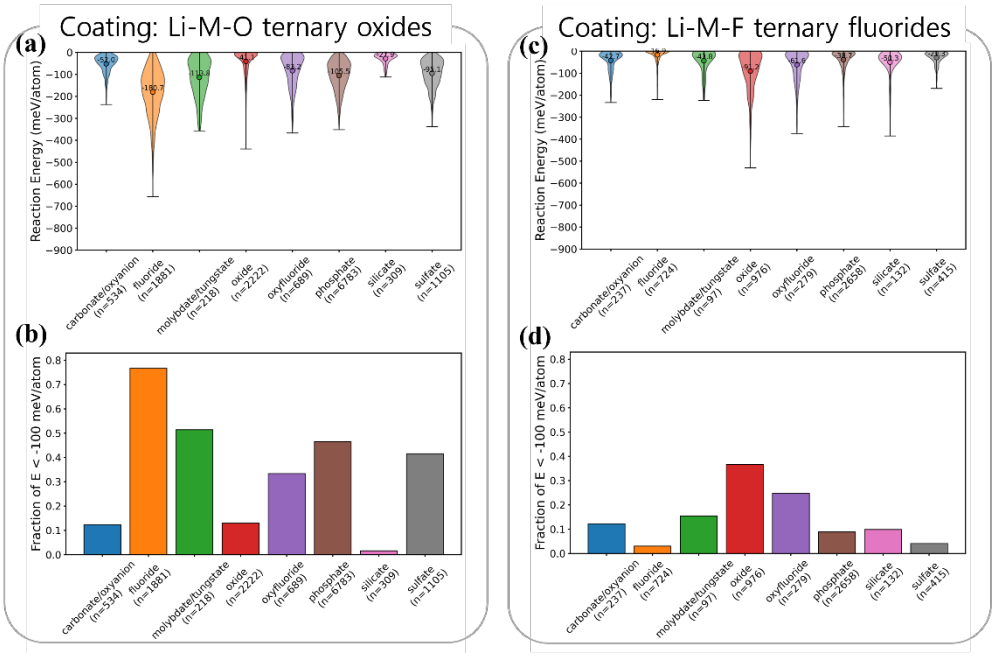

Fig. S13. (a) Distributions of reaction energy between subtype-classified cathodes and Li-M-O ternary coatings. (b) Fraction of cathode and Li-M-O coating combinations whose reaction energy is less than -100 meV/atom. (c) Distributions of reaction energy between subtype-classified cathodes and Li-M-F ternary coatings. (d) Fraction of cathode and Li-M-F coating combinations whose reaction energy is less than -100 meV/atom.

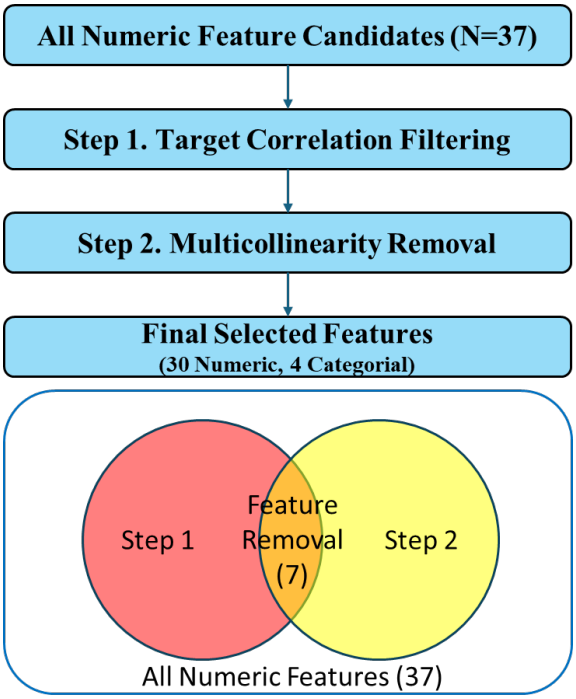

Fig. S14. Scheme for feature selection procedure in this work

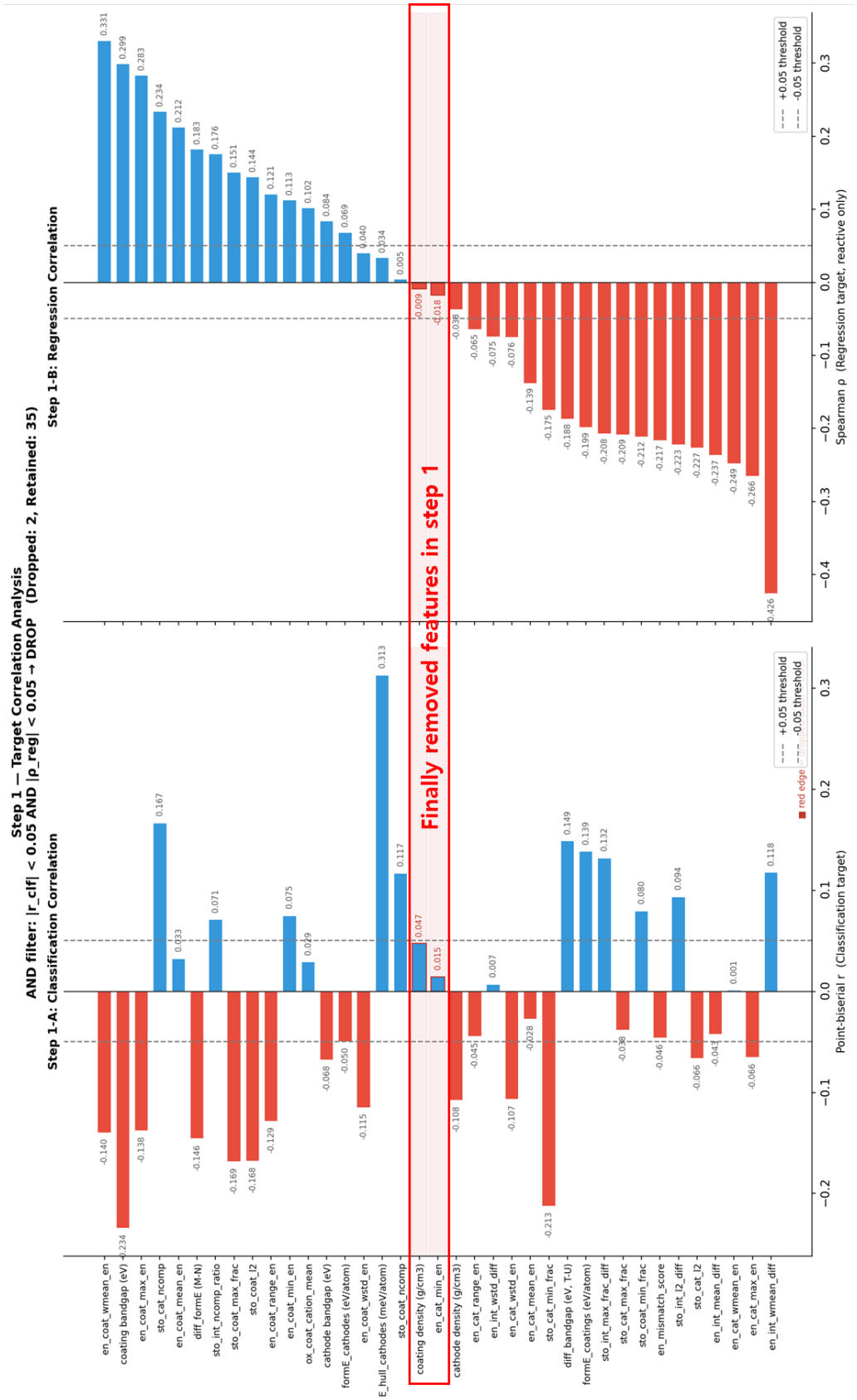

Fig. S15. Feature-target correlation analysis conducted in step 1 in feature selection procedure

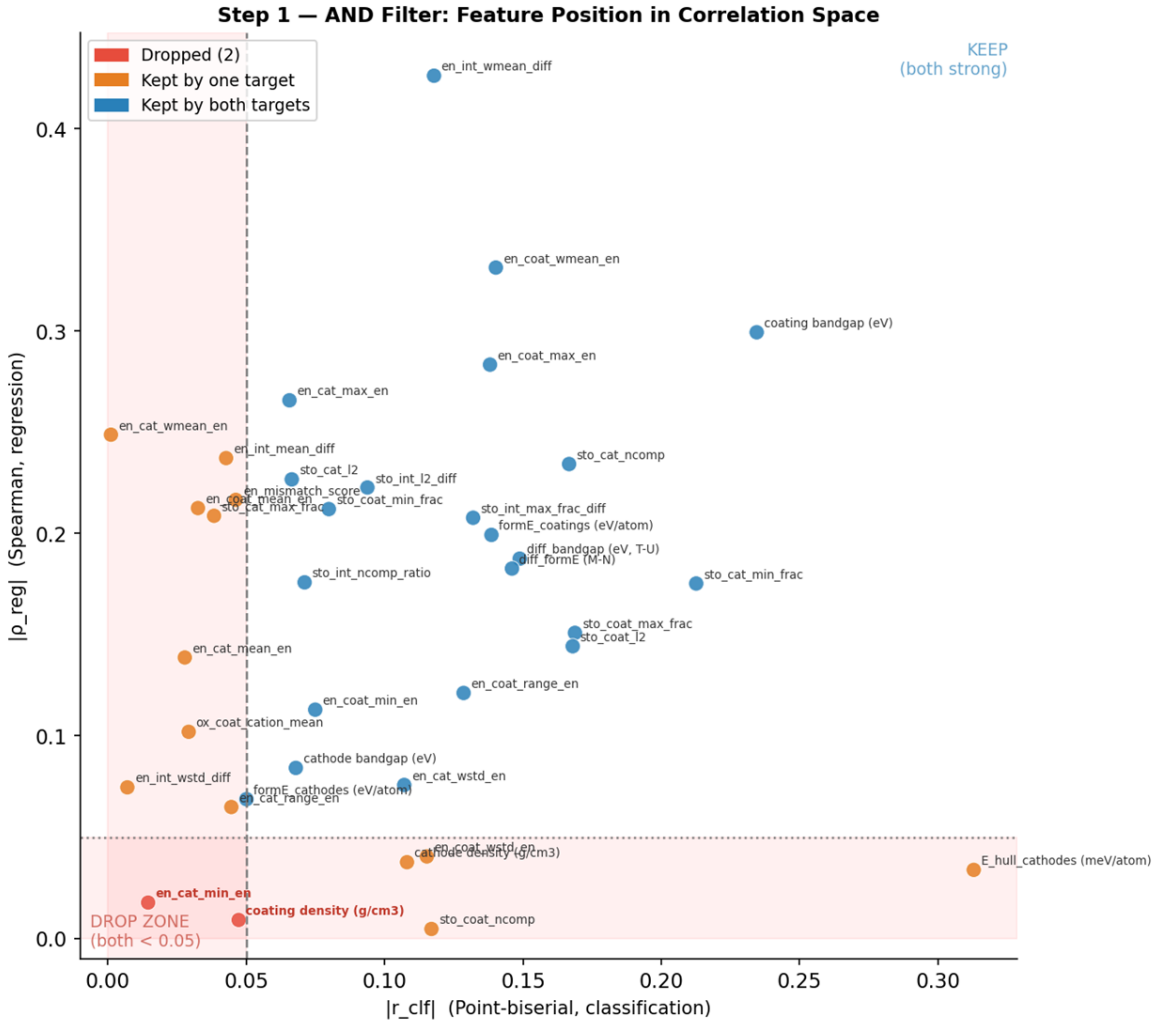

Fig. S16. Feature position in correlation space

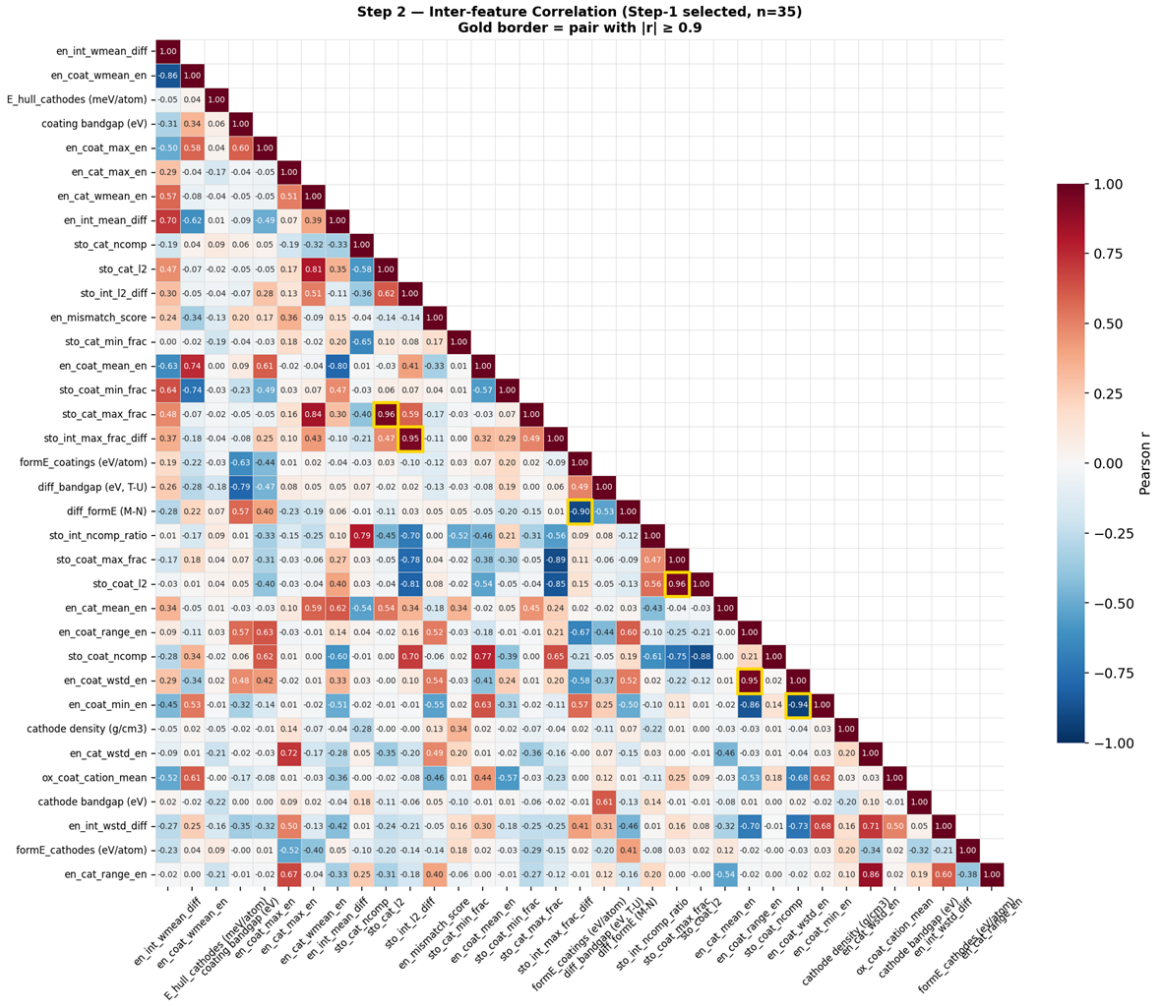

Fig. S17. Multicollinearity identification (|r| > 0.9) for step 2 in feature selection

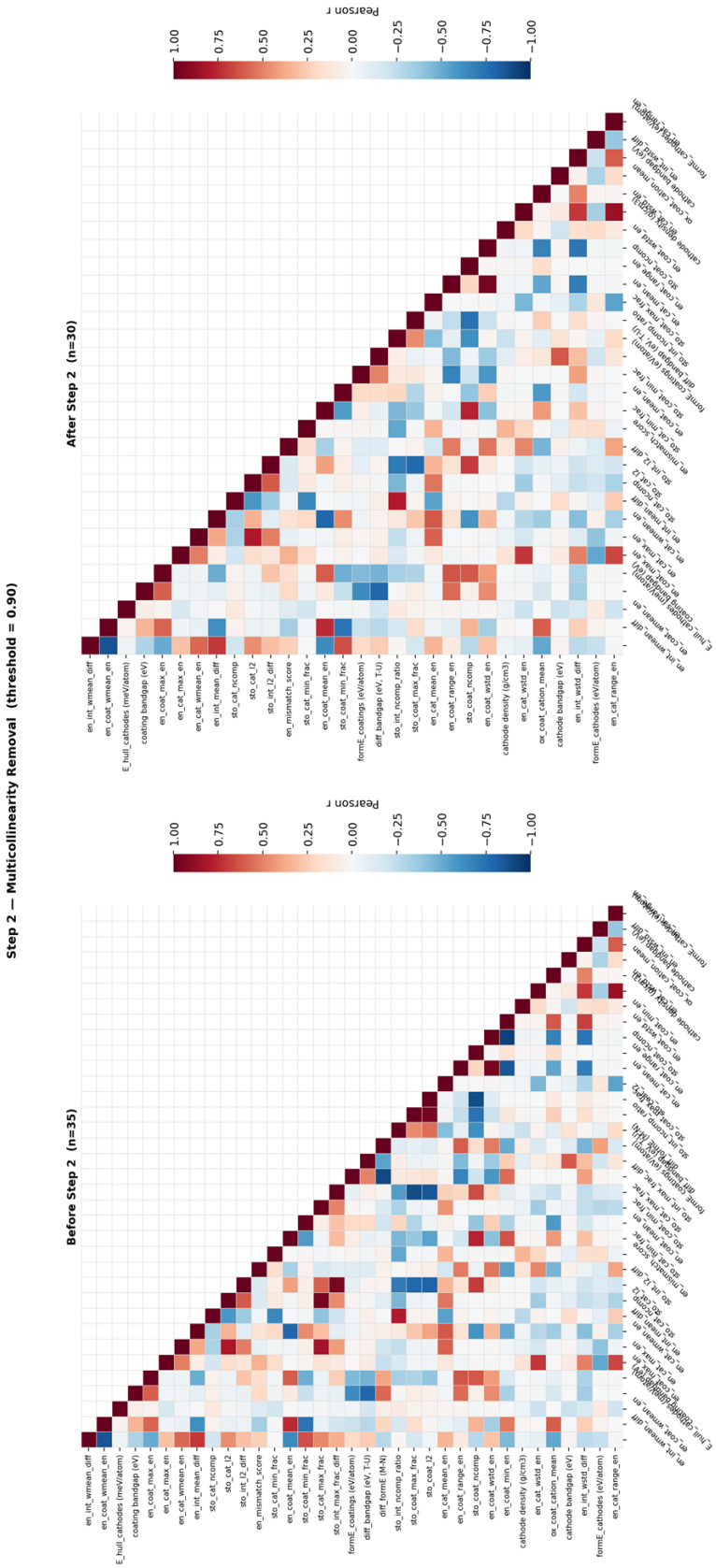

Fig. S18. Multicollinearity removal in step 2 of feature selection procedure

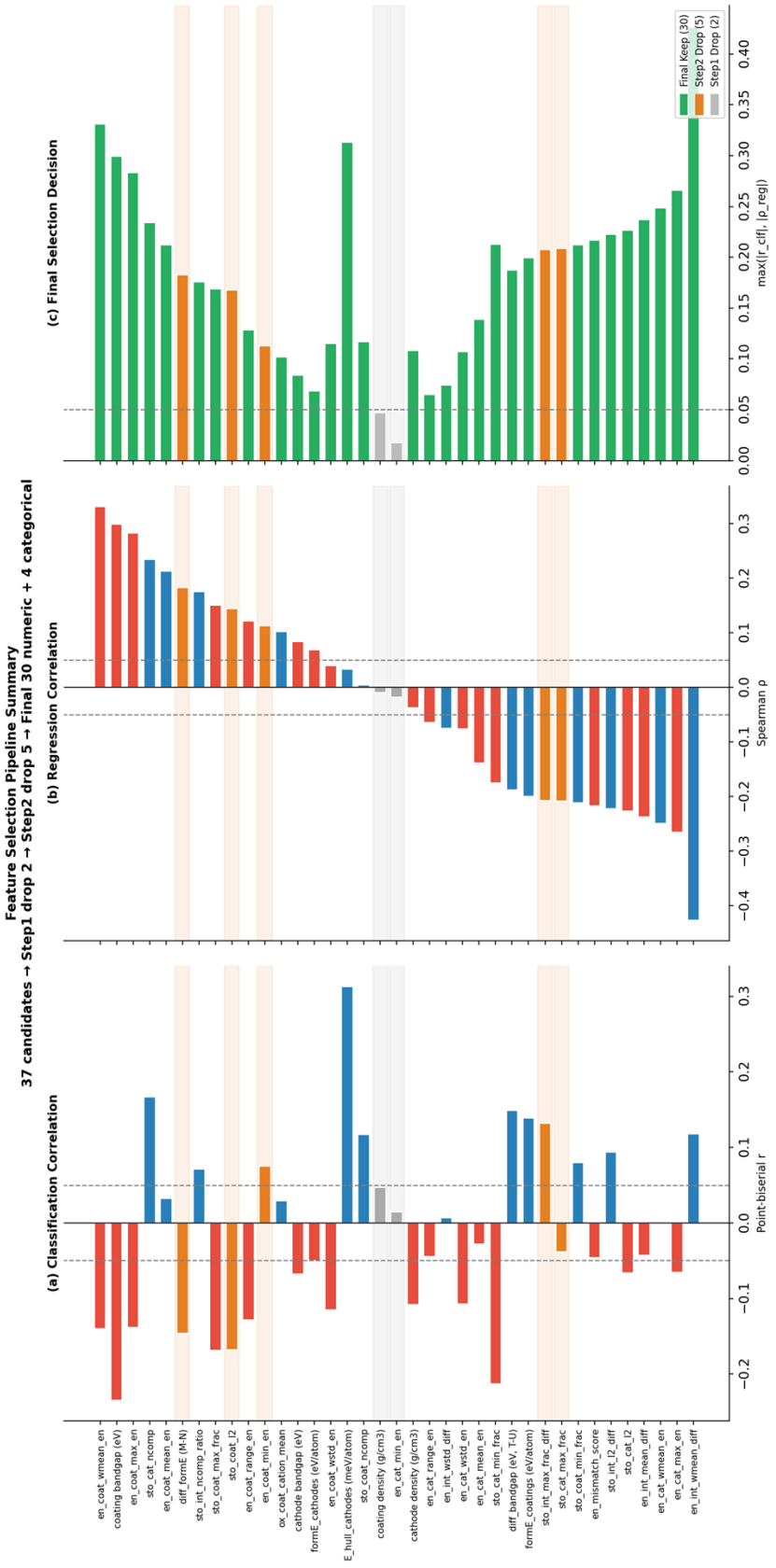

Fig. S19. Summary for two-step feature selection pipeline in this work

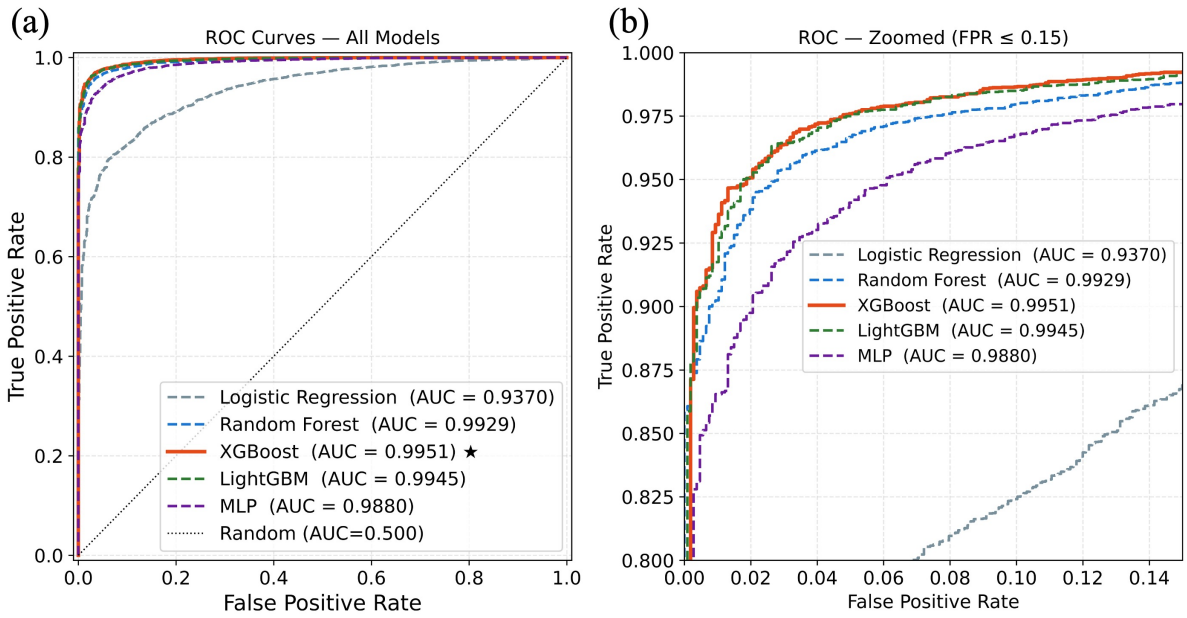

Fig. S20. Comparison of (a) original and (b) magnified ROC curves of trained classifiers

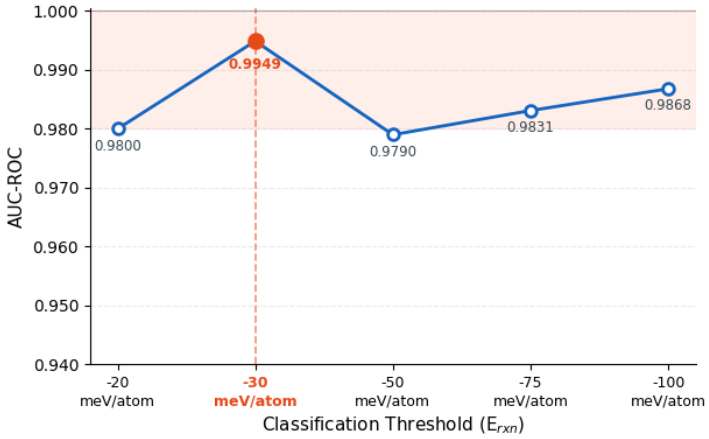

Fig. S21. Result of sensitivity analysis (AUC values) for varying classification threshold.

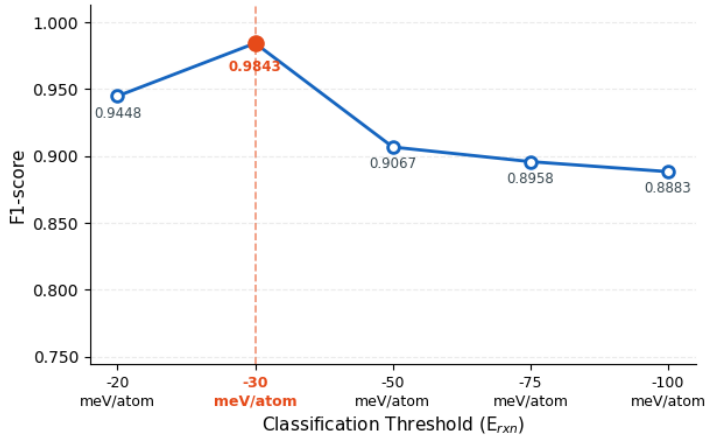

Fig. S22. Result of sensitivity analysis (F1-scores) for varying classification threshold.

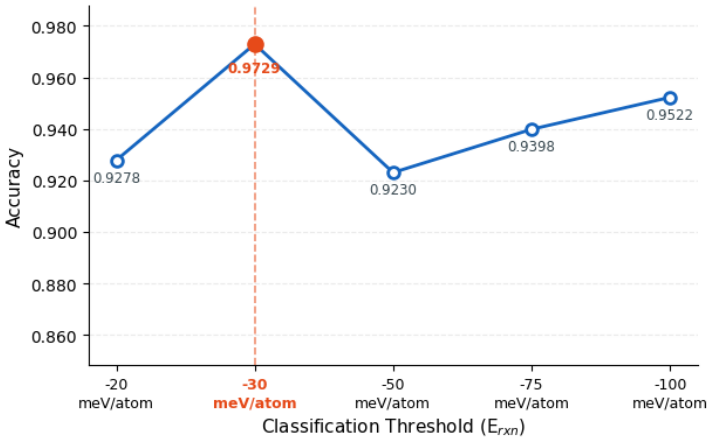

Fig. S23. Result of sensitivity analysis (accuracy values) for varying classification threshold.

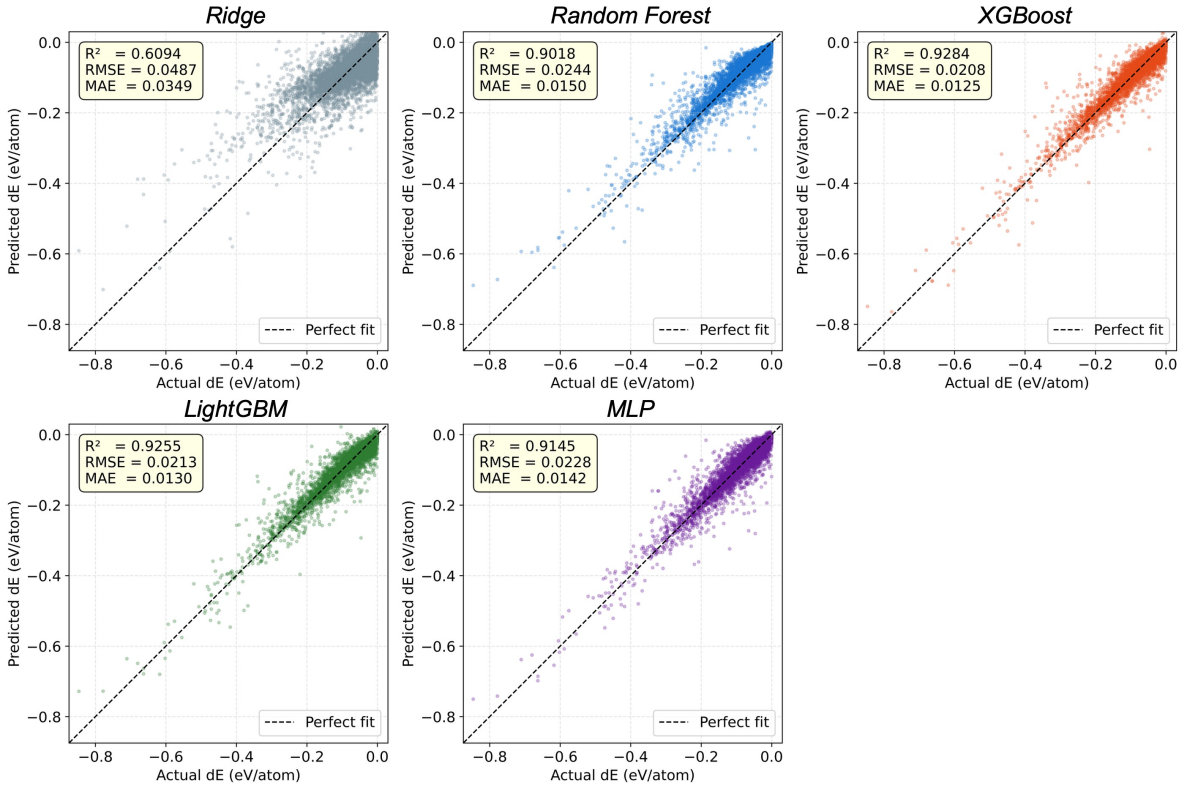

Fig. S24. Comparison of parity plots of five different trained regressors: Ridge, random forest, XGBoost, LightGBM, and MLP models. R^2 score, RMSE, and MAE values are provided. The units of both RMSE and MAE are eV/atom.

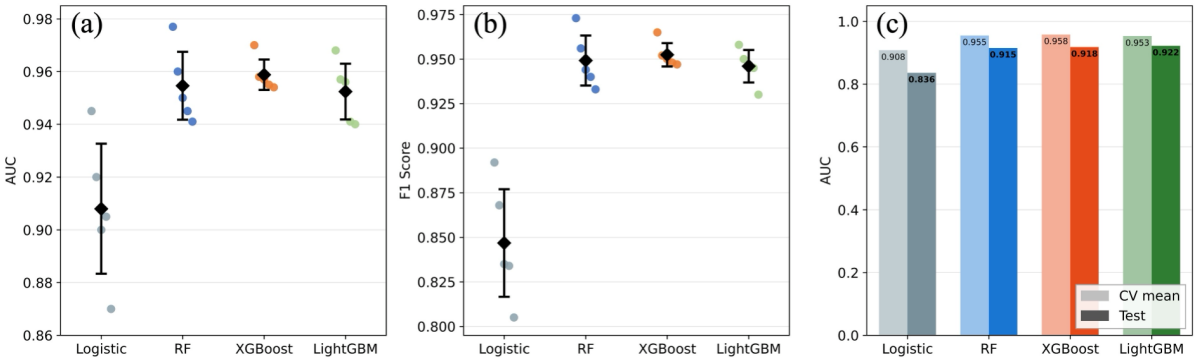

Fig. S25. Five-fold cross-validation (CV) test results for five classification models, where all cathode and coating materials in test set are completely not leaked to train set. (a) AUC values, (b) f1 scores, and (c) comparison between mean AUC values of CV tests and AUC of held-out test.

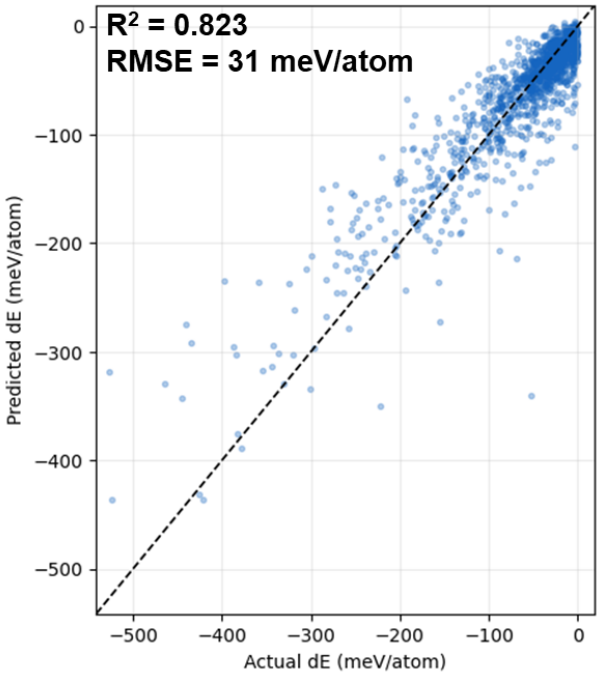

Fig. S26. Group-split test results of optimized XGBoost E_{rxn} regressor. Again, materials of test set were entirely excluded from the training set. R2 score and RMSE values are corresponding to 0.823 and 31 meV/atom, respectively.

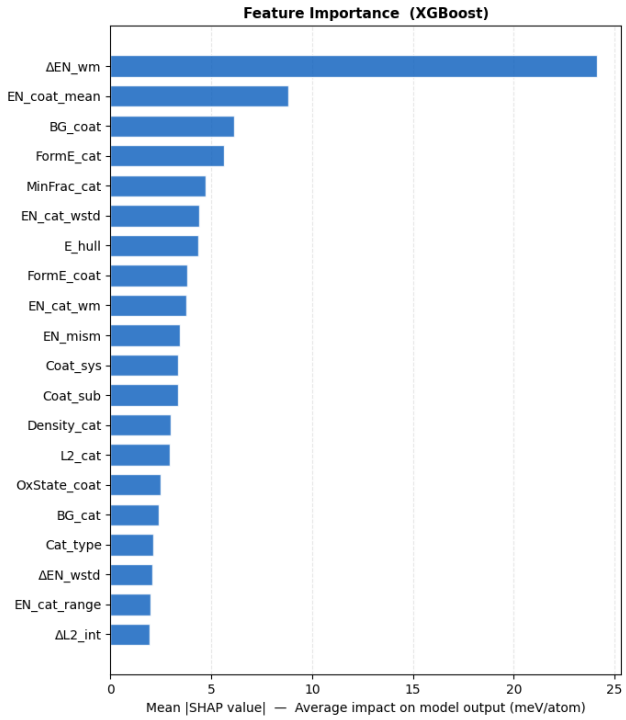

Fig. S27. Top 20 important features in optimized XGBoost-based reaction energy regressor for total dataset including all of oxide, oxyfluoride, fluoride coating materials.

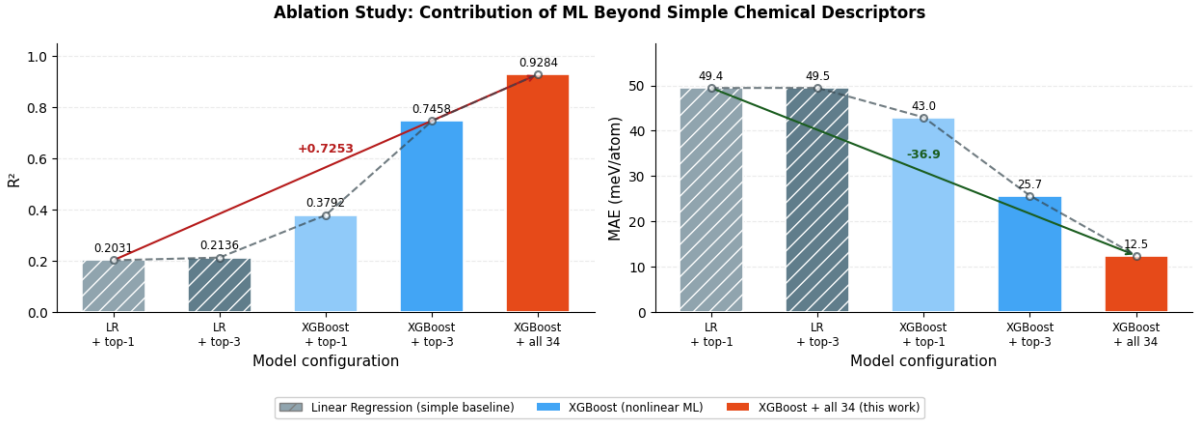

Fig. S28. Conducted ablation study where two logistic regression (LR) and XGBoost models, and use of top 1, 3, and all features are considered in regression task for predicting mutual reaction energy (E_{rxn}). Left and right panels show the evolution of R^2 scores and mean absolute error (MAE) values, respectively.

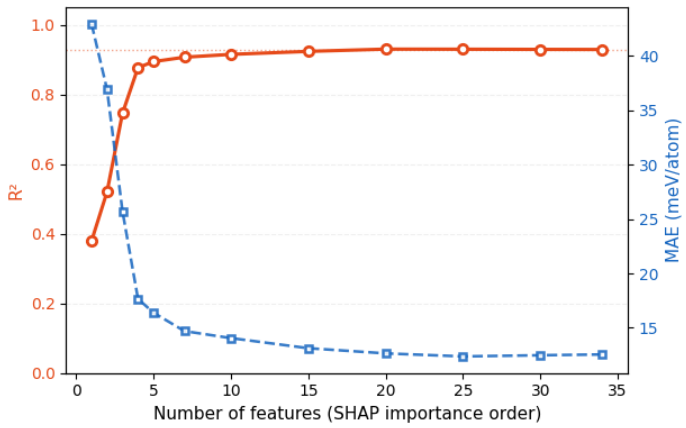

Fig. S29. Cumulative R^2 curve of trained XGBoost-based model for reaction energy prediction.