Supplementary Information for

Accelerating Materials Discovery for Electrocatalytic Water Oxidation via Center-Environment Deep Learning in Spinel Oxides

Yihang Li[a], Xinying Zhang[a], Tao Li[a], Yingying Chen[a], Yi Liu[a] & Lingyan Feng[a,b,c,*]

[a] Materials Genome Institute, Shanghai University, Shanghai 200444, China

[b] Shanghai Engineering Research Center of Organ Repair, School of Medicine, Shanghai University, Shanghai 200444, China

[c] Joint International Research Laboratory of Biomaterials and Biotechnology in Organ Repair, Ministry of Education, School of Medicine, Shanghai University, Shanghai, 200444, China

1. Machine Learning

1.1 Center-Environment (CE) feature

By conducting a comparative analysis between our developed Center-Environment (CE) feature construction method and several popular feature construction techniques currently employed in the field, we have demonstrated the unique aspects and advantages of the CE method in encoding the complex interplay of chemical and structural properties in materials science. This comparison not only highlights the innovative approach of CE in capturing intricate material characteristics but also underscores its efficacy in enhancing the predictive accuracy of machine learning models in materials science.

The CE method innovatively integrates the essential properties of chemical elements and pure substances, along with the compositional and structural nuances of crystal structures. For each element within the compound, we have meticulously selected a set of 56 fundamental properties, such as ion radius, atomic number, and electronegativity (detailed in Supplementary Table 1). These properties form the foundation of our machine learning training dataset, constructed using the CE feature model. Chemical descriptors, particularly that encapsulating crystal geometry information, play a pivotal role in augmenting the predictive accuracy of our machine learning model. The essence of the CE model lies in its approach to defining a central atom and its environmental atoms within a given structure. The model calculates the linearly weighted sum of the elementary feature properties of these environmental atoms, effectively capturing and encoding the chemical composition and structural information into the input feature vectors for machine learning. The weighting is generally inversely proportional to the distance between each environmental atom and the central atom. The detailed formulation and construction of the CE features are presented in equations (1)-(5).

$$D = [F_1, F_2, ..., F_n, T], n = 56 \qquad \text{\\* MERGEFORMAT (1)}$$

$$F_i = [f_{C,i}, f_{E,i}], i = 1, 2, ..., 56 \qquad \text{\\* MERGEFORMAT (2)}$$

$$f_{C,i} = p_{C,i} \qquad \text{\\* MERGEFORMAT (3)}$$

$$f_{E,i} = \sum_j \omega_j p_{j,i} \qquad \text{\\* MERGEFORMAT (4)}$$

$$\omega_j = \frac{\dfrac{1}{r_j}}{\displaystyle\sum_j^{N_j} \dfrac{1}{r_j}} \qquad \text{\\* MERGEFORMAT (5)}$$

where D represents a high-dimensional vector that characterizes the training/testing dataset. Within this dataset, F_i ($i = 1$-56) denotes one of the 56 elementary properties of elements or pure substances.

Each F_i is structured as a two-dimensional vector, representing the property components of the central atom $f_{C,i}$ and the environmental atom $f_{E,i}$, respectively. The target properties, denoted by

T, are the focus of our predictive analysis. Here, $p_{C,i}$ represents the property i of the central atom,

2

while $p_{j,i}$ signifies the property i of the environmental atom j. The weight of each environmental atom, ω_j, is determined by the reciprocal of r_j, the distance from the central atom to the environmental atom j, where j indexes the environmental atoms ($j = 1$- N_j).

During the construction of CE features, atoms within a crystal structure are categorized into two distinct sets: central atoms and their surrounding environmental atoms. In spinel structures, A and B cation inequivalent sites are identified as the central atoms. The environmental atoms encompass those from the first nearest neighbors up to the nth nearest neighbors of the central atom. For this study, we limited our focus to the first nearest neighbors, a choice proven robust for predicting target properties in these systems. The space group symmetry dictates that A and B cations in spinel oxide structures occupy inequivalent sites. Consequently, we considered two types of central atoms and their corresponding environmental atoms (oxygen atoms in this study) in the CE feature construction process, namely dual center CE features.

To mitigate the impact of highly correlated features on the training of machine learning models, we computed the Pearson correlation coefficients among the features. Subsequently, we eliminated features from pairs with strong correlations, specifically those with a lower average importance ranking. This approach ensures that our model is trained on features that offer unique and non-redundant information, thereby improving the predictive performance and interpretability of the model. The Pearson correlation coefficient is defined as:

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \quad \backslash* \quad \text{MERGEFORMAT}$$

(6)

The Pearson correlation coefficient $\rho_{X,Y}$ is a statistical metric that measures the degree of linear association between two variables, X and Y. It is computed as the covariance of X and Y, denoted by $cov(X,Y)$, divided by the product of their respective standard deviations, ρ_X and ρ_Y. The formula expresses this coefficient as the sum of the product of deviations of each observed value from their respective sample means ($\overline{X}$ and $\overline{Y}$) over all n samples, normalized by the product of the sums of the squares of these deviations for X and Y. In essence, $\rho_{X,Y}$ provides a measure of how much Y changes with X on a standardized scale from -1 to 1, where -1 represents a perfect negative linear relationship, 0 indicates no linear relationship, and 1 signifies a perfect positive linear relationship.

1.2 Feature engineering

To accurately predict the MAX(D_T, D_O) values in spinel structures, we considered four different types of features as inputs for the machine learning model: a feature set {E} based on elemental

fractions (the ElemNet model), a Magpie feature set {M} considering only composition information, a Voronoi tessellations feature set {V} accounting for both composition and structural information, and a feature set {CE} we proposed that simultaneously includes both compositional and structural information. The feature set {E} based on elemental fractions includes the elemental fractions and stoichiometric ratios of each component in the crystal structure. The Magpie feature set {M} consists of 87 elemental descriptors such as atomic number, atomic radius, and electronegativity. The Voronoi tessellations feature set {V} describes the local environment around each atom, represented using Voronoi polyhedra, and includes the volume, surface area, and number of faces for each polyhedron. Lastly, our proposed feature set {CE} encodes both compositional and structural information into the machine learning input features by defining a set of central atoms and surrounding atoms in the crystal structure, as well as the weights for each surrounding atom. By comparing the performance of different types of features, we can evaluate their effectiveness in predicting the properties of spinel structures. Furthermore, our proposed compositional and structural feature set {CE} is expected to provide additional information about the crystal structure, thereby improving the accuracy of the machine learning model.

Given that each feature set contains a large number of features ({E}, {M}, {V}, and {CE} feature sets contain 71, 96, 271, and 224 features respectively), selecting the most representative features is crucial for optimizing the predictive performance of the machine learning model and avoiding overfitting. To this end, several types of ML algorithms were employed to calculate the importance scores for all features to perform average feature ranking. These ML algorithms include Ridge Regression, Least Absolute Shrinkage and Selection Operator (LASSO), Random Forests (RF), Recursive Feature Elimination (RFE), Support Vector Regression (SVR), and Extreme Gradient Boosting (XGBoost). This approach better samples the important features, as each type of ML algorithm calculates the correlation between each feature and the target variable in a unique way. Each algorithm can provide different rankings for all features, and the features with the highest average ranking should be selected. The Python library Pycaret[1] was used to train each feature set and the best features were selected based on the learning curve. To avoid linear correlation between features, Pearson correlation coefficients between all feature pairs were also calculated. From feature pairs with a Pearson correlation coefficient > 0.90, the feature with the higher average ranking was selected.

After a series of meticulous feature selection and optimization, we identified key features suitable for subsequent machine learning research. Additionally, using the same optimization methods, we selected 18 features from the {M} feature set (out of a total of 28), and 45 features from the {V} feature set (out of a total of 55). Notably, because the feature set {E} includes complete information on the elemental composition for each structure, the absence of any feature will affect the accurate representation of the corresponding elemental composition by the model. Hence, we chose to retain all 71 features in the {E} feature set, as detailed in Supplementary Table 3, Supplementary Table 4, and Supplementary Table 5 in Supporting Information.

2. DFT calculations

2.1 High-throughput electronic structure calculation

We employed Density Functional Theory (DFT) and used the Vienna ab initio Simulation Package (VASP)[2] version 5.4.4 for calculations. To ensure the accuracy and rigor of the calculations, we used the Projector Augmented Wave (PAW) basis set, set the cut-off energy to 520 eV, and

employed the Perdew-Burke-Ernzerhof (PBE)[3] Generalized Gradient Approximation (GGA) exchange-correlation functional. The rotationally invariant GGA + U approach was applied using the U-J value of 4.0 eV on the Mo[4] 4d orbitals and 6 eV to the Ag[5] 4d orbitals. A k-point convergence test was carried out to determine the appropriate integration scheme in the Brillouin zone, finally selecting a $7 \times 7 \times 7$ k-point grid. All geometric structures were fully relaxed under the guidance of the conjugate gradient method until the total energy change between two self-consistent calculations was less than 10^{-5} eV. Additionally, spin polarization was considered in all calculations.
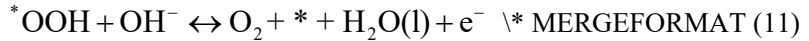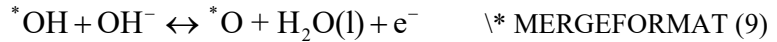
The energy of the metal d-band or the oxygen 2p-band center is determined by the following equation:

$$\varepsilon = \frac{\int E \rho(E) dE}{\int \rho(E) dE} \qquad \text{* MERGEFORMAT (7)}$$

where ε is the energy center of the metal d band or the oxygen 2p-band. $\rho(E)$ and E are the density of the state and the energy value, respectively.

2.2 OER mechanism

For OER intermediate (such as OH*, O*, and OOH*) adsorption studies, the low-index (111) surface was used to model the intermediate adsorption on the $MoAg_2O_4$ and $MONa_2O_4$ surfaces. The employed slab model was designed with multiple atomic layers, incorporating a sufficient number of atoms to ensure a representative surface structure. During structural optimization, a segment of the slab closer to the surface was allowed to relax to capture the dynamic nature of adsorption, while the remaining part was kept fixed to maintain structural integrity and emulate the bulk properties. A 12 Å vacuum space was added to two continuous slabs to eliminate the possible interactions from adjacent periodic images. In the condition of alkaline OER, the well-accepted process involves three adsorbed intermediates (*OH, *O, and *OOH) and four steps as follows[6]:

$$OH^- + * \leftrightarrow {}^*OH + e^- \qquad \text{* MERGEFORMAT (8)}$$

$${}^*OH + OH^- \leftrightarrow {}^*O + H_2O(l) + e^- \qquad \text{* MERGEFORMAT (9)}$$

$${}^*O + OH^- \leftrightarrow {}^*OOH + e^- \qquad \text{* MERGEFORMAT (10)}$$

$${}^*OOH + OH^- \leftrightarrow O_2 + * + H_2O(l) + e^- \quad \text{* MERGEFORMAT (11)}$$

where * stands for the active sites in catalysts.

The free energy values of OER elementary steps were calculated by the equation:

$$\Delta G = \Delta E_{ads} + \Delta E_{ZPE} - T\Delta S_{ads} \qquad \text{* MERGEFORMAT (12)}$$

where ΔE_{ads} is the adsorption energy, ΔE_{ZPE} is the zero-point energy difference between adsorbed and gaseous species, and $T\Delta S_{ads}$ is the corresponding entropy difference between these two states (T was set to be 298.15 K).

The free energy values of OER intermediates (*OH, *O, and *OOH) were calculated as follows:

$$\Delta G_{(*OH)} = G_{*OH} - G_* - \left(G_{H_2O} - 1/2G_{H_2}\right) \text{* MERGEFORMAT (13)}$$

$$\Delta G_{(*O)} = G_{*O} - G_* - \left(G_{H_2O} - G_{H_2}\right) \quad \text{* MERGEFORMAT (14)}$$

$$\Delta G_{(*OOH)} = G_{*OOH} - G_* - \left(2G_{H_2O} - 3/2G_{H_2}\right) \text{*} \quad \text{MERGEFORMAT}$$

(15)

where, G_{*OH}, G_{*O} and G_{*OOH} were Gibbs free energies of adsorption of OH, O, and OOH species, respectively. The represented the Gibbs free energy of clean catalytic surface. In addition, G_{H_2O} and G_{H_2} were calculated Gibbs free energies of H_2O and H_2 molecules in gas phase.

For the OER process calculation, a standard four-electron reaction mechanism in alkaline condition was considered for the calculation of Gibbs free energy change for steps 8-11 according to the previous study[6]:

$$\Delta G_1 = \Delta G_{*OH} - eU + \Delta G_{H+}(\text{pH}) \quad \text{* MERGEFORMAT (16)}$$

$$\Delta G_2 = \Delta G_{*O} - \Delta G_{*OH} - eU + \Delta G_{H+}(\text{pH}) \text{* MERGEFORMAT (17)}$$

$$\Delta G_3 = \Delta G_{*OOH} - \Delta G_{*O} - eU + \Delta G_{H+}(\text{pH}) \text{* MERGEFORMAT (18)}$$

$$\Delta G_4 = 4.92[\text{eV}] - \Delta G_{*OOH} - eU + \Delta G_{H+}(\text{pH}) \text{*} \quad \text{MERGEFORMAT}$$

(19)

where U is the potential measured against normal hydrogen electrode (NHE) at standard conditions (T = 298.15 K, P = 1 bar).

The theoretical overpotential is then readily defined as:

$$\eta = \max\left(\Delta G_1, \Delta G_2, \Delta G_3, \Delta G_4\right)/e - 1.23 \text{* MERGEFORMAT (20)}$$

The free energy change of the protons relative to the above specified electrode at non-zero pH is represented by Nernst equation as:

$$\Delta G_{H+}(\text{pH}) = -k_B T \ln(10) \times \text{pH} \quad \text{* MERGEFORMAT (21)}$$

3. Experimental section

3.1 Chemical reagents

Ethanol (AR grade) was obtained from Sinopharm Chemical Reagent Co., Ltd. Bis(acetylacetonato)dioxomolybdenum(VI) ($C_{10}H_{14}MoO_6$, 97%), sodium nitrate ($NaNO_3$, AR, ≥99%), citric acid ($C_6H_8O_7·H_2O$, 99.995%), sodium molybdate dihy drate ($Na_2MoO_4·2H_2O$, AR, 99%), silver nitrate ($AgNO_3$, AR, 99%), Zinc nitrate ($Zn(NO_3)_2·6H_2O$, AR, 99%), and Aluminum nitrate ($Al(NO_3)_3·9H_2O$, AR, 99%) were acquired from Aladdin Industry Corporation. All of the chemicals were used without further purification. Deionized (DI)-water (with a specific resistance of 18.2 MΩ cm) was obtained from Millipore water purification system and was used during the whole experiments.

3.2 Sample preparation

For the synthesis of $MoNa_2O_4$ via the sol-gel method, a typical procedure was employed. Initially, 6 mmol of molybdenum acetylacetonate ($Mo(acac)_3$) and 12 mmol of sodium nitrate ($NaNO_3$) were homogeneously dissolved in a dilute nitric acid solution composed of 30 mL of ultrapure water and 5 mL of nitric acid under rigorous stirring conditions. Subsequently, 15 mmol of citric acid was added to serve as a chelating agent to enhance the uniformity and stability of the metal ions within the mixture. The blend was then subjected to constant stirring in an oil bath maintained between 80-100 °C to induce the formation of a highly viscous gel. This gel was transferred to a preheated oven set at 170 °C for a drying period of 12 hours. Finally, the dried gel was calcined in a muffle furnace at 300 °C with a carefully controlled heating rate of 5 °C/min for 3 hours, yielding high-purity $MoNa_2O_4$ powder.

For the hydrothermal synthesis of $MoAg_2O_4$, an initial solution (solution A) was prepared by dissolving 0.6440 g of $Na_2MoO_4 \cdot 2H_2O$ in 40 mL of deionized water. Concurrently, another solution (solution B) was prepared by dissolving 0.9043 g of silver nitrate ($AgNO_3$) in 40 mL of deionized water. Solution A was gradually titrated into Solution B under intense stirring, resulting in a stable colloidal suspension. This suspension was then transferred into a 100 mL stainless steel autoclave and subjected to a hydrothermal treatment at 120 °C for 2 hours. Upon completion, the solid product was isolated by centrifugation, thoroughly washed with deionized water and anhydrous ethanol, and subsequently dried overnight at 60 °C to obtain high-purity $MoAg_2O_4$ powder.

The $ZnAl_2O_4$ catalyst was synthesized using an isopropanol-mediated controlled hydrolysis method. Typically, 2.975 g of $Zn(NO_3)_2 \cdot 6H_2O$ and 7.5 g of $Al(NO_3)_3 \cdot 9H_2O$ were dissolved in 50 mL of isopropanol, with the Zn^{2+}/Al^{3+} ratio set to a stoichiometric value of 1:2 for $ZnAl_2O_4$. The solution was transferred to a 100 mL high-pressure vessel, stirred at ambient temperature for 2 hours, then heated in an oven to 200 °C and allowed to react for 10 hours. Subsequently, the resulting precipitate was collected by centrifugation, washed several times with ethanol and ultrapure water, dried at 90 °C for 5 hours, and finally calcined at 600 °C for 5 hours to obtain the $ZnAl_2O_4$ catalyst.

3.3 Characterization

Field emission scanning electron microscope (FE-SEM, Zeiss Sigma 500) and transmission electron microscopy (TEM, JEM-2100F, 200 kV) were employed to obtain the morphologies and elemental mapping images, respectively. The crystal structures were characterized by an X-ray diffraction (XRD, Bruker-D2 PHASE X-ray diffraction system, Cu kα radiation, λ=1.5418 Å). Fourier transform infrared spectra (FTIR) were recorded on a FTIR spectrometer (Nicolet iS 50, Thermo Scientific). UV-vis absorption spectra were collected on PerkinElmer Lambda 750 spectrophotometer. Raman was carried out by a Nanofinder 30A Raman spectrometer with a 532 nm laser source (Tokyo Instruments, Inc). X-ray photoelectron spectra (XPS) were recorded on an ESCALAB 250xi X-ray photoelectron spectrometer (Thermo Scientific) to investigate the surface chemical environment of samples.

3.4 Electrochemical measurements

The electrocatalytic performances of electrocatalysts were tested on a typical three-electrode system using an Autolab electrochemical workstation. A graphite rod and Hg/HgO (1 M KOH)

electrode were used as the counter electrode and reference electrode. In order to prepare the working electrode (WE), 4 mg of the synthesized electrocatalyst powder was dispersed in 500 μL of a mixed solution consisting of ultrapure water, isopropanol, and 5 wt% Nafion with volumes of 380 μL, 100 μL, and 20 μL, respectively. After sonification for 60 min, 80 μL of the resulting suspension was drop-cast onto a pre-cleaned nickel foam electrode (NF) with a surface area of 1 cm^2, achieving a mass loading of 0.64 mg cm^{-2}. Then, the WE could be further used after drying at room temperature. Prior all the electrochemical measurements, cyclic voltammetry (CV) method was carried out to activate the electrocatalysts at 50 mV/s. Linear sweep voltammetry (LSV) was performed to record the OER polarization curves with 95% iR-correction at a scan rate of 5 mV s^{-1} in O$_2$-saturated 1 M KOH electrolyte. The electrochemical impedance spectroscopy (EIS) was measured from 10^5 to 10^{-1} Hz with the amplitude of 0.005 V. Additionally, the stability tests were carried out by chronoamperometry (CP) method at a constant current density of 10 mA cm^{-2}. All the measured potential versus Hg/HgO were normalized to RHE by the following equation:

$$E_{RHE} = E_{Hg/HgO} + 0.059 \times pH + 0.098 \quad \text{\\* MERGEFORMAT (22)}$$

Supplementary Table 1. Atomic and structural properties of elements and pure substances as elementary properties.[7]

Index	Symbol	Feature name	Feature type
1	$f_{0.5}$	Atomic electron scattering factor at 0.5	Elemental
2	N_{AE}	Atomic environment number (Villars, Daams)	Elemental
3	Z_{lt-lr}	Atomic number start counting left top, left-right sequence	Elemental
4	A_r	Atomic weight	Elemental
5	T_b	Boiling temperature	Elementary substance
6	Z_{eff}^{C}	Charge nuclear effective (Clementi)	Elemental
7	K	Compression modulus	Elementary substance
8	ρ	Density	Elementary substance
9	d_c	Distance from core electron (Schubert)	Elemental
10	d_v	Distance from valence electron (Schubert)	Elemental
11	Z_{eq}^{chem}	Electrochemical weight equivalent	Elemental
12	E_{ea}	Electron affinity	Elemental
13	χ_{AR}	Electronegativity (Alfred-Rochow)	Elemental
14	χ_{MB}	Electronegativity (Martynov&Batsanov)	Elemental
15	χ_P	Electronegativity (Pauling)	Elemental
16	χ_a	Electronegativity absolute	Elemental
17	E_c	Energy cohesive (Brewer)	Elemental
18	Z_i^{1st}	Energy of ionization first	Elemental
19	Z_i^{2nd}	Energy of ionization second	Elemental
20	Z_i^{3rd}	Energy of ionization third	Elemental
21	ΔH_{at}	Enthalpy of atomization	Elemental
22	ΔH_m	Enthalpy of melting	Elementary substance
23	ΔH_v	Enthalpy of vaporization	Elementary substance
24	S_s	Entropy of solid	Elementary substance
25	N_G	Group number	Elemental
26	ν	Magnetic frequency of nuclei	Elemental
27	μ	Magnetic resonance	Elemental
28	$\mu_{MoK\alpha}$	Mass attenuation coefficient for MoKα	Elemental
29	T_m	Melting temperature	Elementary substance
30	M_{chem}	Mendeleev chemists sequence	Elemental
31	M_{dtL}	Mendeleev d-t start left	Elemental
32	M_{dtR}	Mendeleev d-t start right	Elemental
33	M_{HdtL}	Mendeleev H d-t start left	Elemental
34	M_{HdtR}	Mendeleev H d-t start right	Elemental
35	M_{HtdL}	Mendeleev H t-d start left	Elemental
36	M_{HtdR}	Mendeleev H t-d start right	Elemental
37	M_{Pett}	Mendeleev Pettifor	Elemental
38	M_{PettR}	Mendeleev Pettifor regular	Elemental
39	M_{tdL}	Mendeleev t-d start left	Elemental
40	M_{tdR}	Mendeleev t-d start right	Elemental
41	C_m	Molar heat capacity	Elementary substance

Index	Symbol	Feature name	Feature type
42	μ_n	Moment nuclear magnetic	Elemental
43	Z_{eff}^S	Nuclear charge effective (Slater)	Elemental
44	OS_1	Oxidation state first	Elemental
45	N_{br-rl}	Periodic number start counting bottom right, right-left sequence	Elemental
46	N_{lb-lr}	Periodic number start counting left bottom, left-right sequence	Elemental
47	N_{tr-rl}	Periodic number start counting top right, right-left sequence	Elemental
48	N_q	Quantum number	Elemental
49	R_c	Radii covalent	Elemental
50	R_m	Radii metal (Waber)	Elemental
51	R_{ps}	Radii pseudo-potential (Zunger)	Elemental
52	L	Spectral lines n_o	Elemental
53	I	Spin nuclei	Elemental
54	σ_{th}	Thermal neutron capture cross section	Elemental
55	n_{val}	Valence electron number	Elemental
56	V_{atom}	Volume of atom (Villars, Daams)	Elemental

Supplementary Table 2: List of all Center-Environment {CE} features selected from mean feature ranking utilized in the present study

| Symbol	Feature name	Atomic environment type	Feature type
n_{val}	Valence electron number	O_C	Elemental& Structural
n_{val}	Valence electron number	T_C	Elemental& Structural
ρ	Compression modulus	T_C	Elemental& Structural
R_{ps}	Radii pseudo-potential	O_C	Elemental& Structural
M_{pett}	Mendeleev Pettifor	O_C	Elemental& Structural
M_{HdtL}	Mendeleev H d-t start left	O_C	Elemental& Structural
ν	Magnetic frequency of nuclei	O_C	Elemental& Structural
K	Compression modulus	T_C	Elemental& Structural
K	Compression modulus	O_C	Elemental& Structural
ΔH_{at}	Enthalpy of atomization	T_C	Elemental& Structural
E_i^{3rd}	Energy of ionization third	O_C	Elemental& Structural
T_b	Boiling temperature	O_C	Elemental& Structural
M_{lt-lr}	Atomic number start counting left top, left-right sequence	O_C	Elemental& Structural
R_m	Radii metal	T_C	Elemental& Structural
M_{pett}	Mendeleev Pettifor	T_C	Elemental& Structural
ν	Magnetic frequency of nuclei	T_C	Elemental& Structural
V_{atom}	Volume of atom	T_C	Elemental& Structural
χ_P	Electronegativity	T_C	Elemental& Structural
R_{ps}	Radii pseudo-potential	T_C	Elemental& Structural
M_{tdL}	Mendeleev t-d start right	T_C	Elemental& Structural
ΔH_m	Enthalpy of melting	T_C	Elemental& Structural
Z_{eff}^S	Nuclear charge effective	T_C	Elemental& Structural
d_v	Distance from core electron	T_C	Elemental& Structural
ρ	Compression modulus	O_C	Elemental& Structural
μ_n	Moment nuclear magnetic	O_C	Elemental& Structural
N_{lb-lr}	Periodic number start counting bottom right, right-left sequence	T_C	Elemental& Structural
E_i^{1st}	Energy of ionization first	O_C	Elemental& Structural
χ_{MB}	Electronegativity	O_C	Elemental& Structural
N_{AE}	Atomic environment number	T_C	Elemental& Structural
χ_a	Electronegativity absolute	T_C	Elemental& Structural
E_{ea}	Electron affinity	T_C	Elemental& Structural
M_{chem}	Mendeleev chemists sequence	O_C	Elemental& Structural
E_i^{2nd}	Energy of ionization second	T_C	Elemental& Structural
T_m	Melting temperature	O_C	Elemental& Structural
V_{atom}	Volume of atom	O_C	Elemental& Structural
ΔH_m	Enthalpy of melting	O_C	Elemental& Structural
S_s	Entropy of solid	O_C	Elemental& Structural
E_i^{3rd}	Energy of ionization third	T_C	Elemental& Structural
L	Spectral lines no	O_C	Elemental& Structural
L	Spectral lines no	T_C	Elemental& Structural

*T_C and O_C respectively represent the feature of the central atoms at tetrahedral and octahedral sites.

a

b

Hyperparameter Importances
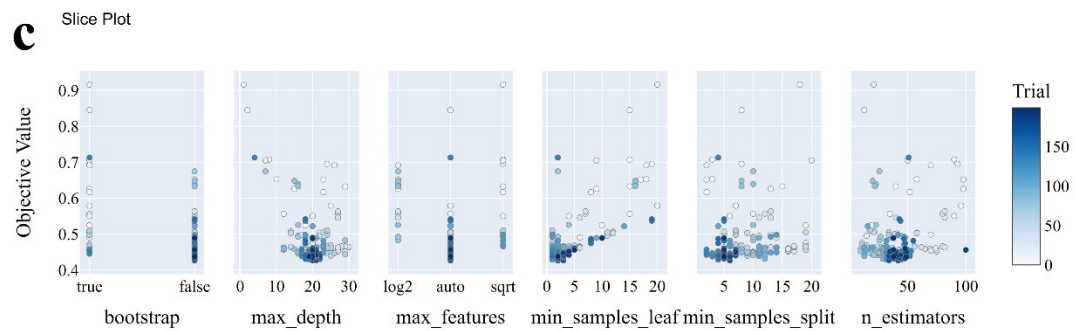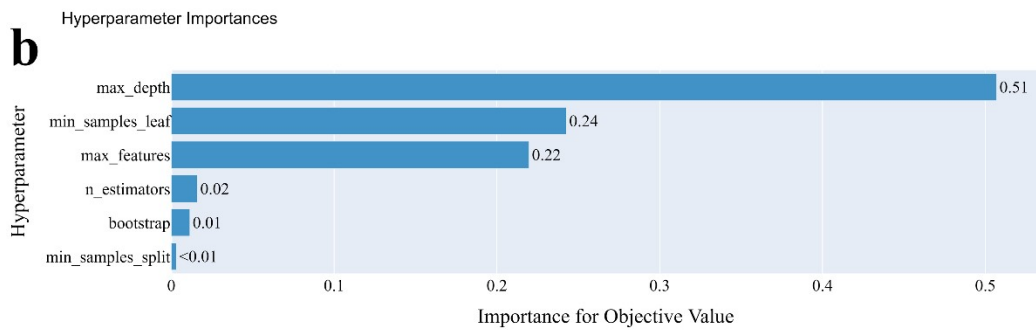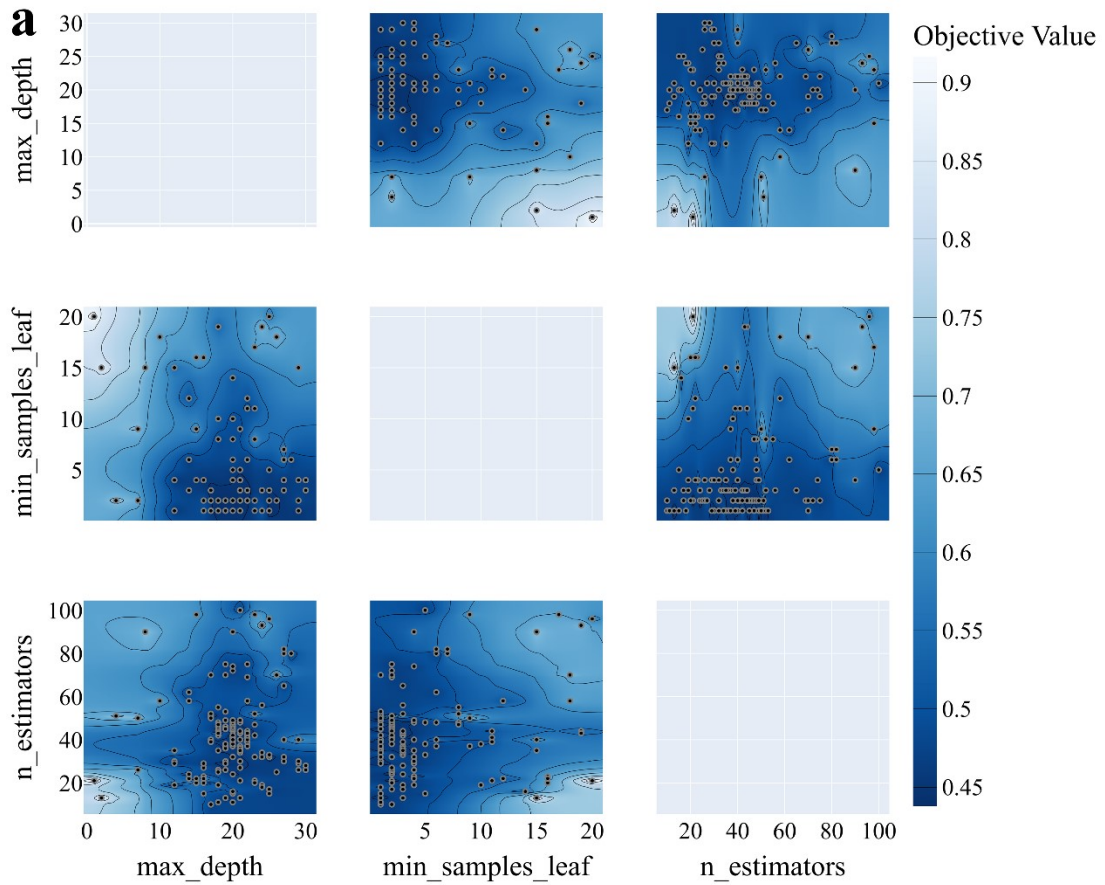
c

Slice Plot

d
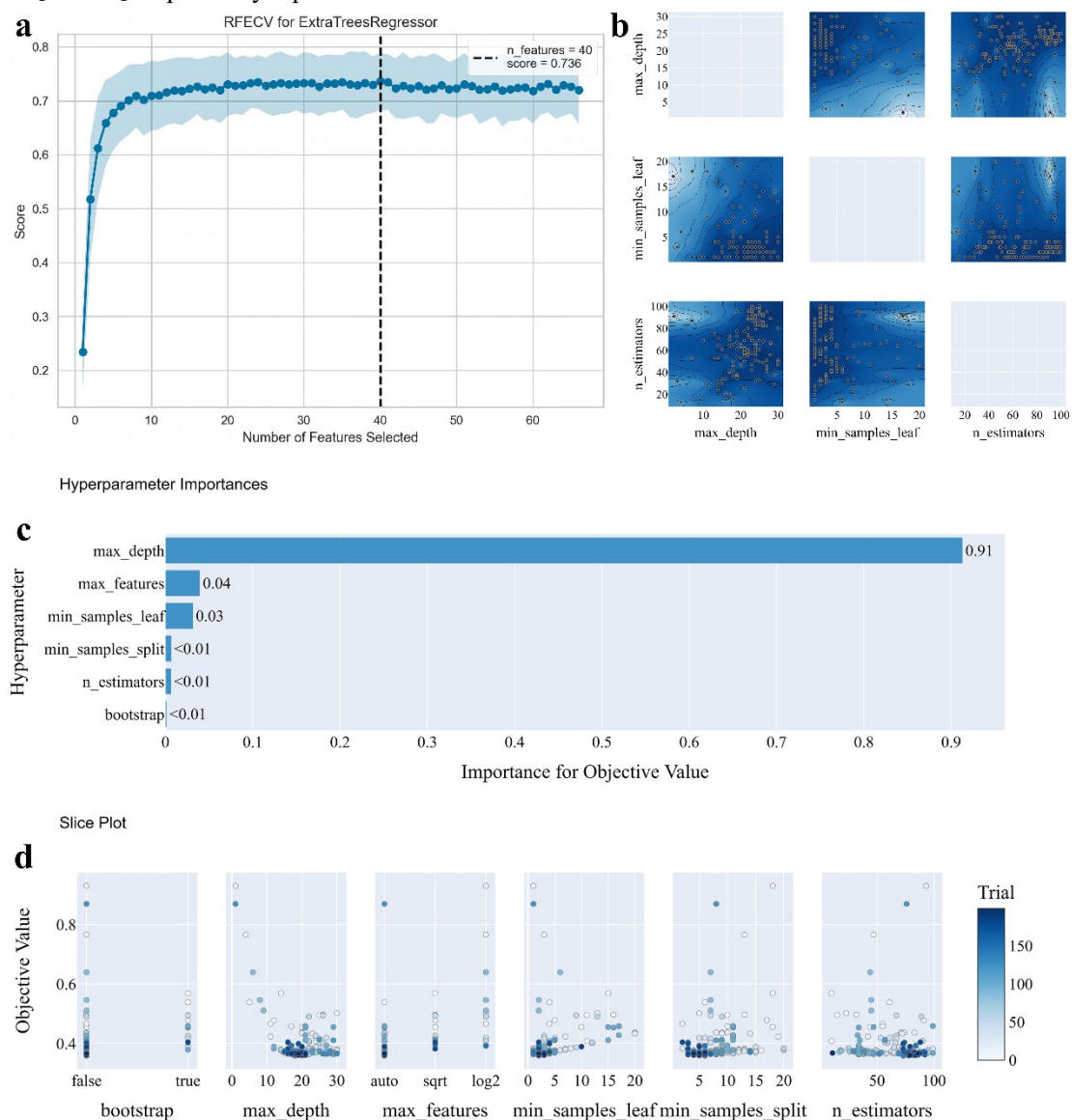
Supplementary Figure 1. (a) Learning curve for feature selection from {CE} feature Set, (b) Contour plot of best three hyperparameters utilized in the ML(ET) model using the selected {CE} features for MAX(D_T, D_O) regression, (c) Hyperparameter importance plot for ML(Extra Trees Regressor) model, (d) Slice plots for bootstrap, max_depth, max_features, min_samples_leaf, min_samples_split, and n_estimators hyperparameters for the ML(Extra Trees Regressor) model. The legend bar shows number of trials.

Supplementary Table 3: List of all Magpie {M} features selected from mean feature ranking utilized in the present study

Symbol	Feature name	Feature type
$\min \nu_{unfilled}$	MagpieData minimum NUnfilled	Elemental
$\min \nu_s$	MagpieData minimum NsValence	Elemental
$\max \nu_d$	MagpieData maximum NdValence	Elemental
$\bar{W}_{atomic}$	MagpieData mean AtomicWeight	Elemental
$\bar{R}_{covalent}$	MagpieData mean CovalentRadius	Elemental
δ_{column}	MagpieData avg_dev Column	Elemental
$R_{valence}$	MagpieData range NValence	Elemental
$\bar{n}_{d,unfilled}$	MagpieData mean NdUnfilled	Elemental
$\bar{n}_{valence}$	MagpieData mean NValence	Elemental
$\bar{n}_{unfilled}$	MagpieData mean NUnfilled	Elemental
δV_{GS}	MagpieData avg_dev GSvolume_pa	Elemental
δS_{group}	MagpieData avg_dev SpaceGroupNumber	Elemental
δT_{melt}	MagpieData avg_dev MeltingT	Elemental
$\bar{n}_d$	MagpieData mean NdValence	Elemental
$\min C_{column}$	MagpieData minimum Column	Elemental
$R_{p,unfilled}$	MagpieData range NpUnfilled	Elemental
$\min M_{Mendeleev}$	MagpieData minimum MendeleevNumber	Elemental
$\bar{n}_{p,unfilled}$	MagpieData mean NpUnfilled	Elemental
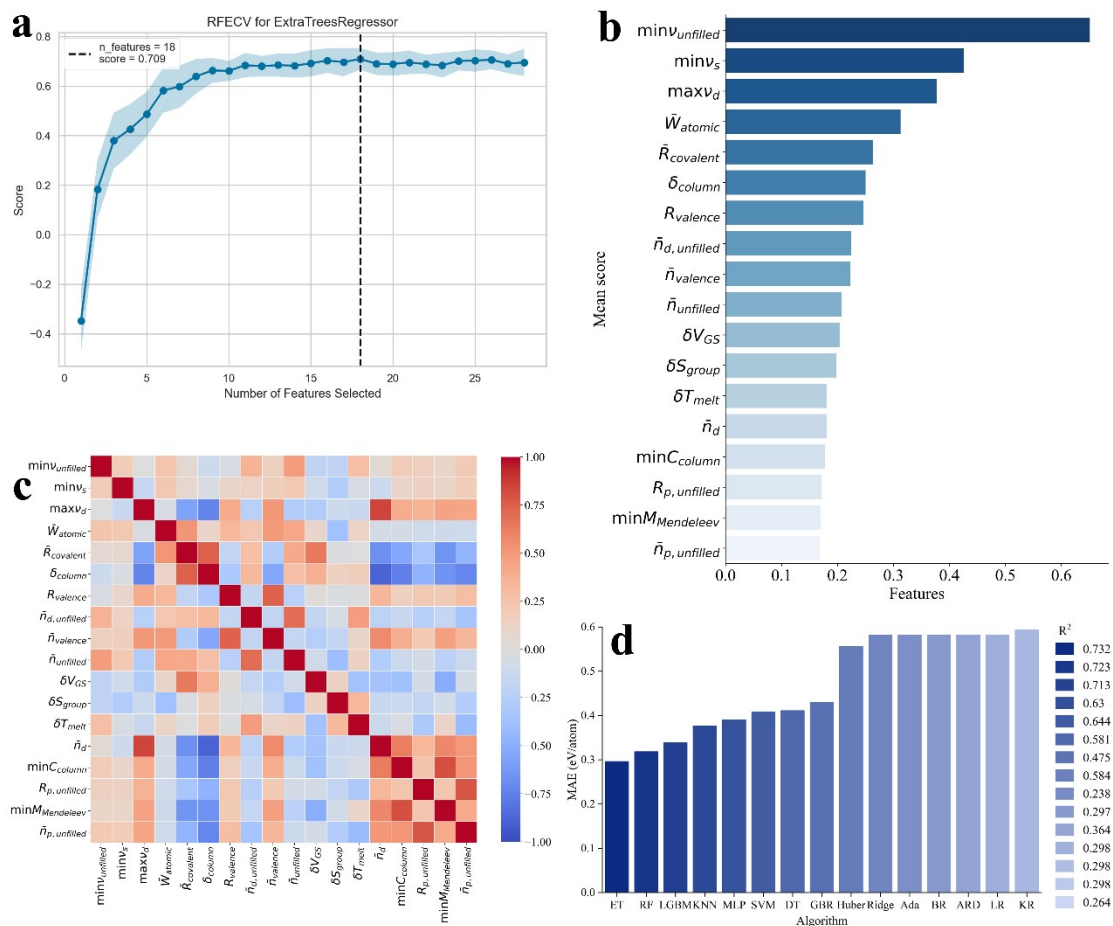
Supplementary Figure 2. (a) Learning curve for feature selection from Magpie {M} feature Set, (b) Mean score of best features selected from the {M} feature set (highly correlated features removed) for MAX(D_T, D_O). (c) Pearson correlation between the best {M} features (after removing highly correlated features) selected from feature ranking. (d) Test MAE and R^2 for MAX(D_T, D_O) regression corresponding to all ML algorithms using {M} feature set.

Supplementary Figure 3. (a) Contour plot of best three hyperparameters utilized in the ML(ET) model using the selected {M} features for MAX(D_T, D_O) regression. (b) Hyperparameter importance plot for ML(Extra Trees Regressor) model, (c) Slice plots for bootstrap, max_depth, max_features, min_samples_leaf, min_samples_split, and n_estimators hyperparameters for the ML(Extra Trees Regressor) model. The legend bar shows number of trials.
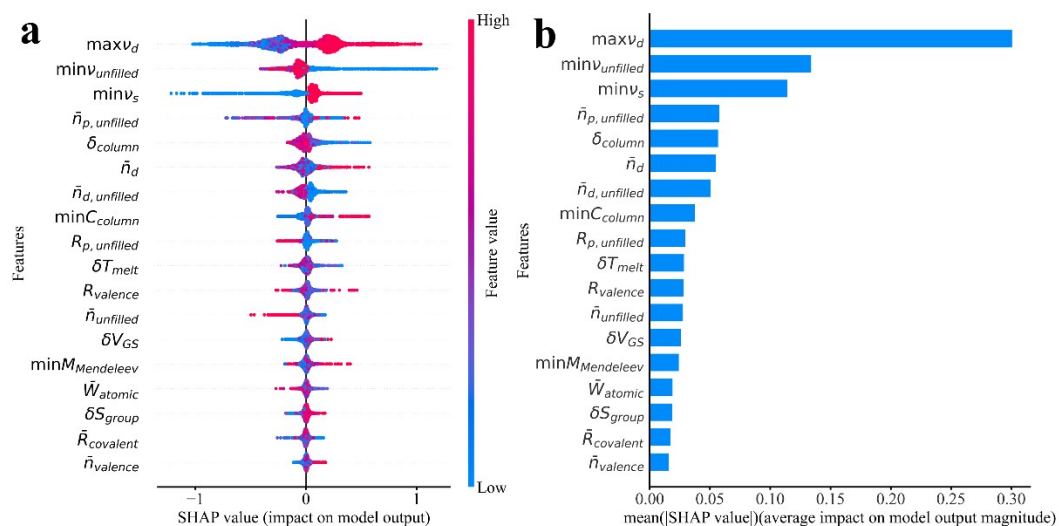
Supplementary Figure 4. (a) Global SHAP feature importance plot and (b) Simplified version of the SHAP feature importance summary plot for MAX(D_T, D_O) ({M}, ET model), arranged in the order of their decreasing importance.

Supplementary Table 4: List of all Voronoi{V} features selected from mean feature ranking utilized in the present study

Symbol	Feature name	Feature type
$\bar{v}_{NpV,s1}$	mean_NeighDiff_shell1_NpValence	Elemental& Structural
$\max(\omega_{s1})$	max_NeighDiff_shell1_AtomicWeight	Elemental& Structural
$\min(c_{s1})$	min_NeighDiff_shell1_Column	Elemental& Structural
$\min(m)$	min_MendeleevNumber	Elemental& Structural
$\min(u)$	min_NUnfilled	Elemental& Structural
$\max(c_{s1})$	max_NeighDiff_shell1_Column	Elemental& Structural
$\bar{v}$	mean_NValance	Elemental& Structural
$\sigma^2(c_{s1})$	var_NeighDiff_shell1_Column	Elemental& Structural
$\bar{r}_{s1}$	mean_NeighDiff_shell1_Row	Elemental& Structural
Δv	maxdiff_NValance	Elemental& Structural
δv	dev_NValance	Elemental& Structural
$\bar{c}_{s1}$	mean_NeighDiff_shell1_Column	Elemental& Structural
$\min(\varphi_{s1})$	min_NeighDiff_shell1_GSvolume_pa	Elemental& Structural
$\bar{T}_m$	mean_MeltingT	Elemental& Structural
$\bar{R}_c$	mean_CovalentRadius	Elemental& Structural
$\max(v_{NdV,s1})$	max_NeighDiff_shell1_NdValence	Elemental& Structural
$\sigma^2(R_{c,s1})$	var_NeighDiff_shell1_CovalentRadius	Elemental& Structural
$\max(\varphi_{s1})$	max_NeighDiff_shell1_GSvolume_pa	Elemental& Structural
δg	dev_SpaceGroupNumber	Elemental& Structural
$R(v_{s1})$	range_NeighDiff_shell1_NValance	Elemental& Structural
$R(v_{NdV,s1})$	range_NeighDiff_shell1_NdValence	Elemental& Structural
$\sigma^2(T_{m,s1})$	var_NeighDiff_shell1_MeltingT	Elemental& Structural
$\min(n_{s1})$	min_NeighDiff_shell1_Number	Elemental& Structural
$\bar{u}_{Nd,s1}$	mean_NeighDiff_shell1_NdUnfilled	Elemental& Structural
$\bar{u}$	mean_NUnfilled	Elemental& Structural
$\max(g_{s1})$	max_NeighDiff_shell1_SpaceGroupNumber	Elemental& Structural
$\bar{v}_{NdV,s1}$	mean_NeighDiff_shell1_NdValence	Elemental& Structural
$R(\omega_{s1})$	range_NeighDiff_shell1_AtomicWeight	Elemental& Structural
$\min(T_{m,s1})$	min_NeighDiff_shell1_MeltingT	Elemental& Structural
$\min(v_{s1})$	min_NeighDiff_shell1_NValance	Elemental& Structural
$\min(v_s)$	min_NsValence	Elemental& Structural
Δu	maxdiff_NUnfilled	Elemental& Structural
$R(g_{s1})$	range_NeighDiff_shell1_SpaceGroupNumber	Elemental& Structural
$\sigma^2(g_{s1})$	var_NeighDiff_shell1_SpaceGroupNumber	Elemental& Structural
$\max(T_{m,s1})$	max_NeighDiff_shell1_MeltingT	Elemental& Structural
$\min(v_{NdV,s1})$	min_NeighDiff_shell1_NdValence	Elemental& Structural
$\sigma^2(u_{Nd,s1})$	var_NeighDiff_shell1_NdUnfilled	Elemental& Structural
$\min(e)$	min_EffectiveCoordination	Elemental& Structural
$R(m_{s1})$	range_NeighDiff_shell1_MendeleevNumber	Elemental& Structural
$R(R_{c,s1})$	range_NeighDiff_shell1_CovalentRadius	Elemental& Structural
$\min(u_{s1})$	min_NeighDiff_shell1_NUnfilled	Elemental& Structural

| $R(r_{s1})$	range_NeighDiff_shell1_Row	Elemental& Structural
$R(c_{s1})$	range_NeighDiff_shell1_Column	Elemental& Structural
$\sigma^2(m_{s1})$	var_NeighDiff_shell1_MendeleevNumber	Elemental& Structural
δf	dev_NfUnfilled	Elemental& Structural
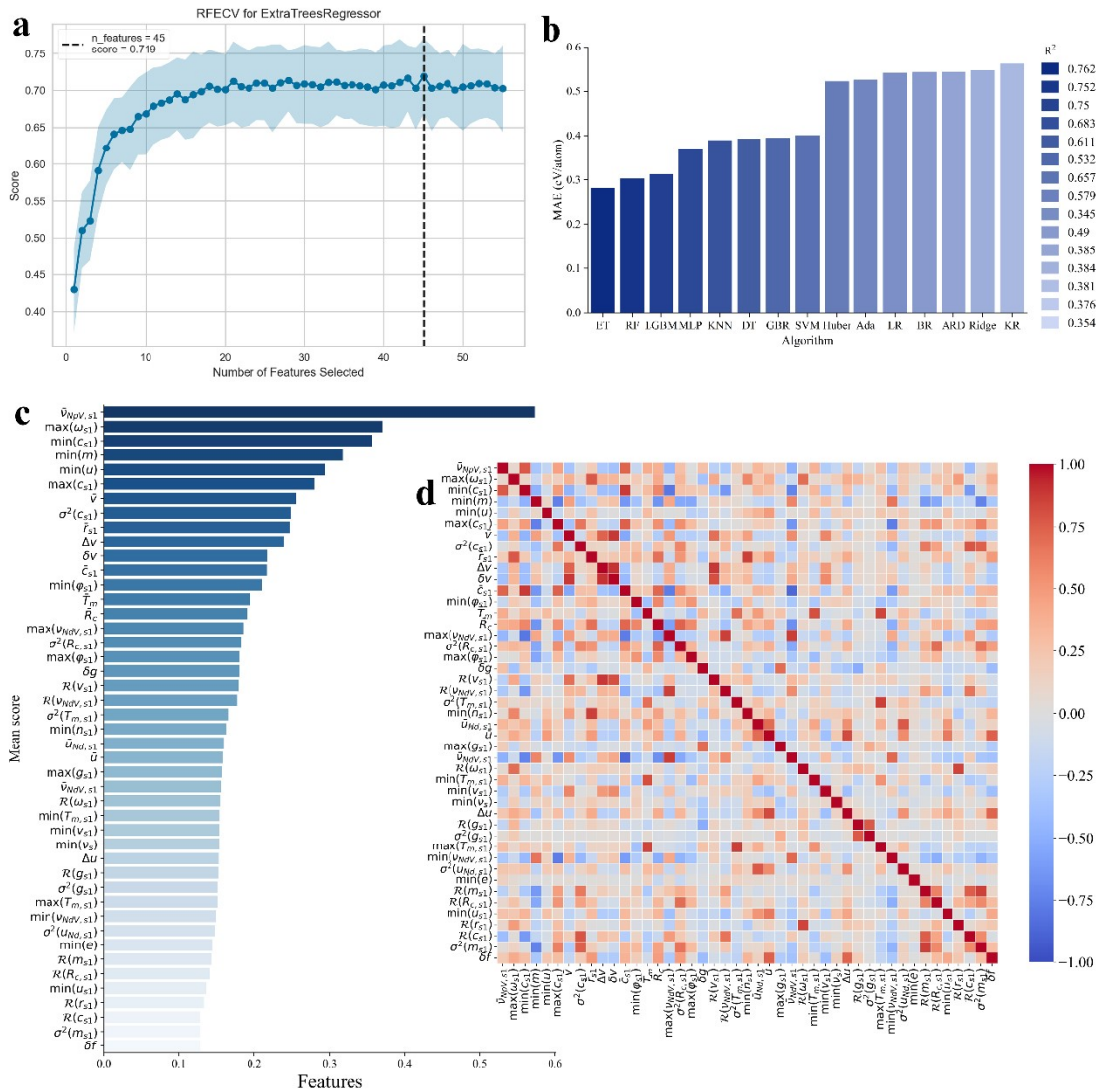
Supplementary Figure 5. (a) Learning curve for feature selection from Voronoi{V} feature Set, (b) Test MAE and R^2 for MAX(D_T, D_O) regression corresponding to all ML algorithms using {M} feature set. (c) Mean score of best features selected from the {M} feature set (highly correlated features removed) for MAX(D_T, D_O). (d) Pearson correlation between the best {M} features (after removing highly correlated features) selected from feature ranking.
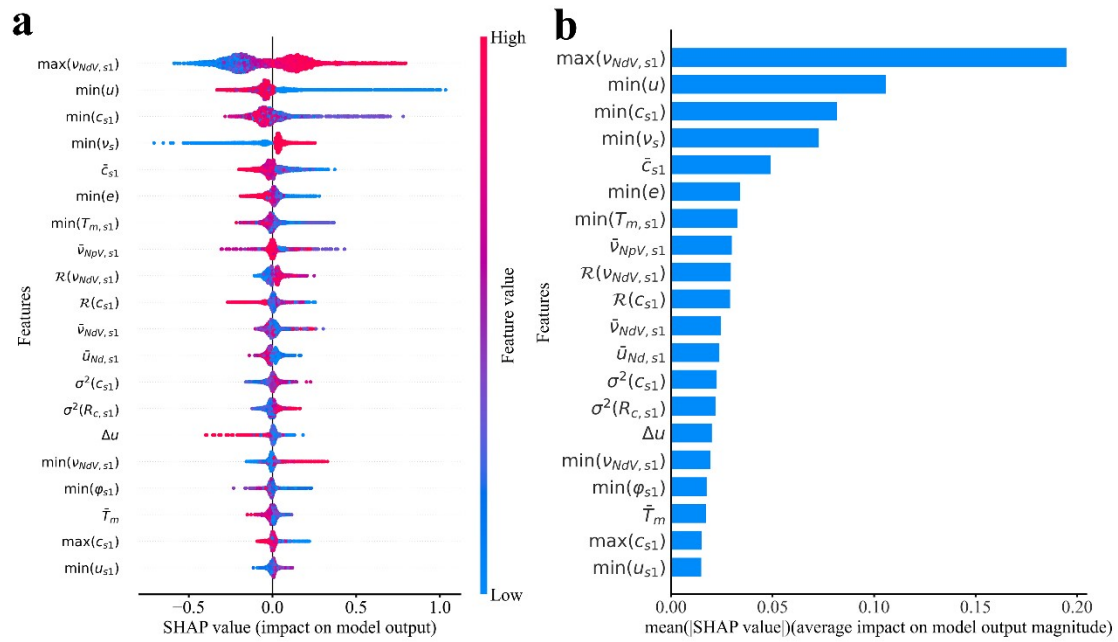
Supplementary Figure 6. (a) Global SHAP feature importance plot and (b) Simplified version of the SHAP feature importance summary plot for MAX(D_T, D_O) ({V}, ET model), arranged in the order of their decreasing importance.
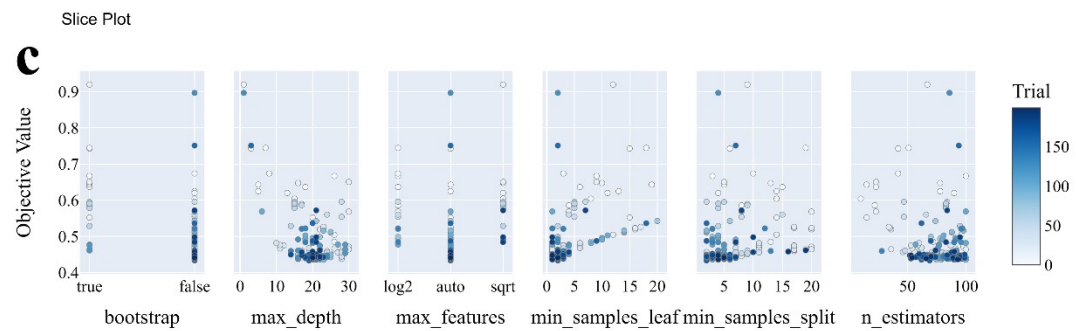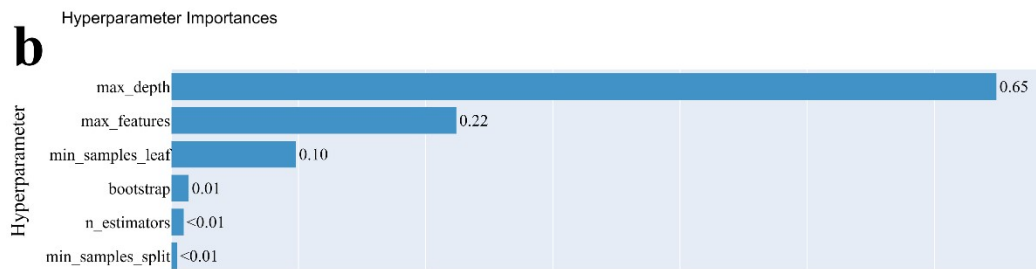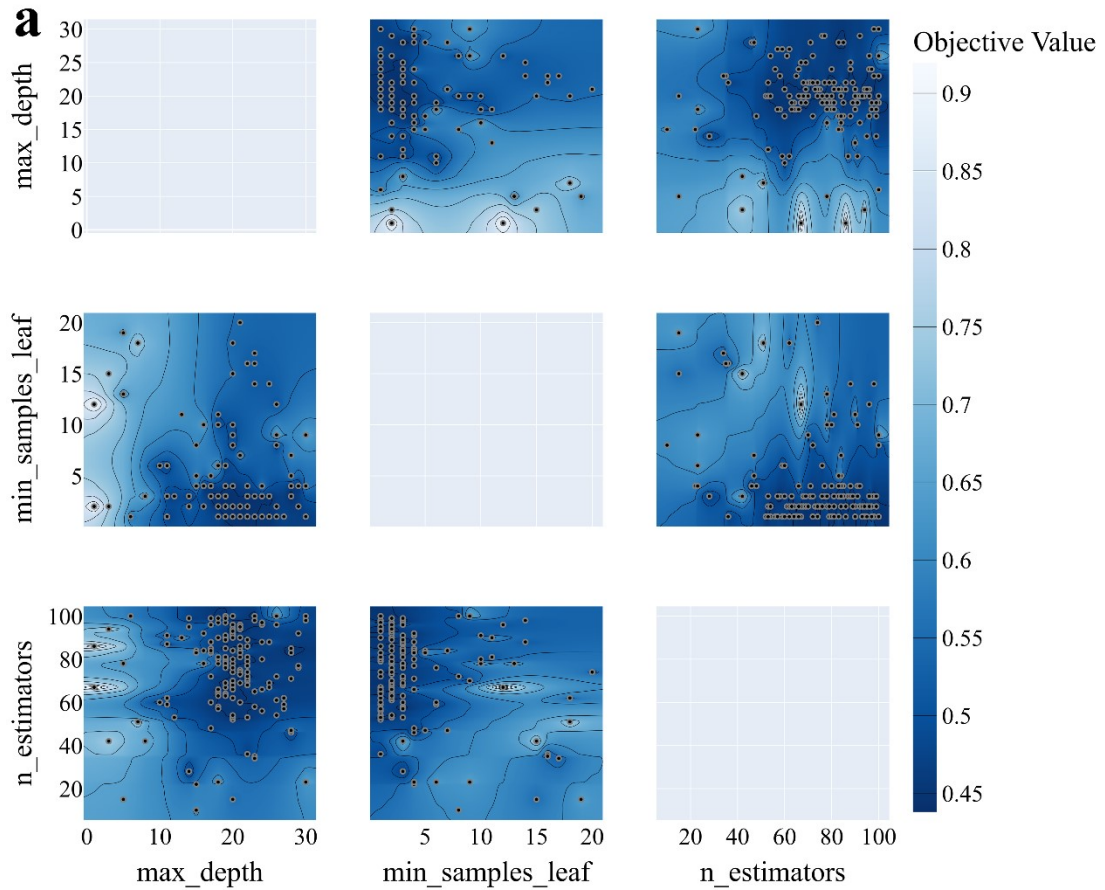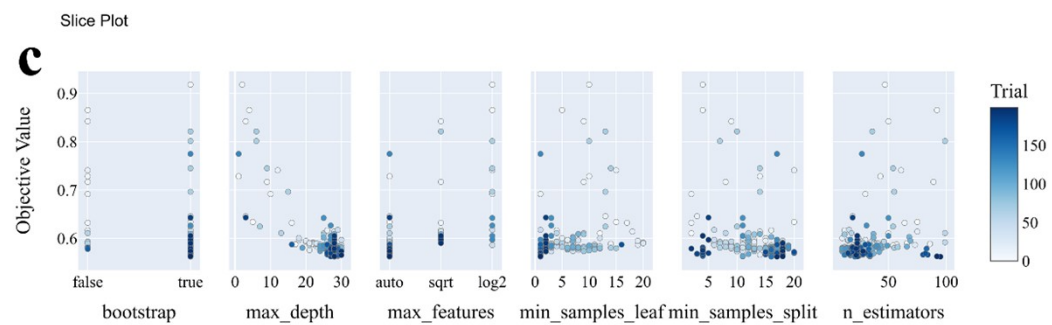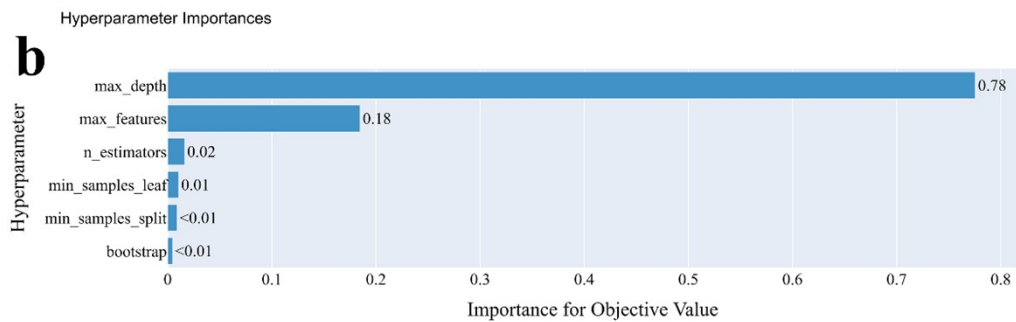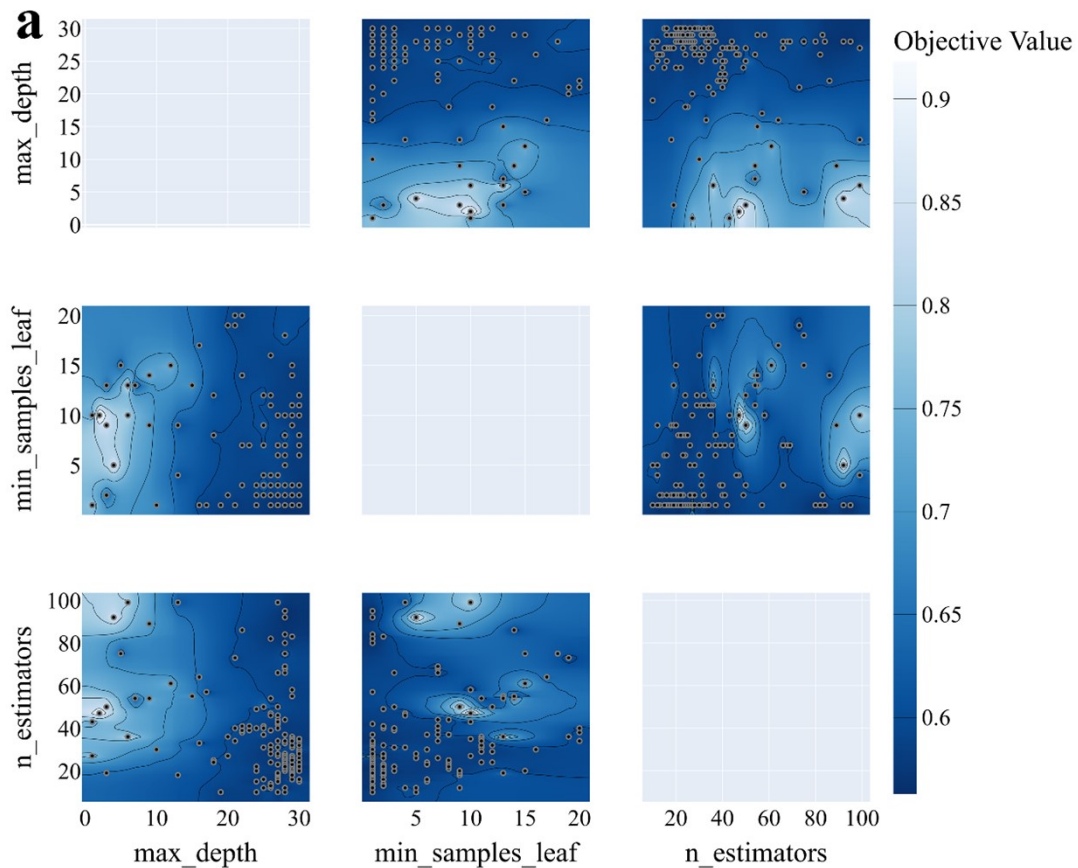
Supplementary Figure 7. (a) Contour plot of best three hyperparameters utilized in the ML(ET) model using the selected {V} features for MAX(D_T, D_O) regression. (b) Hyperparameter importance plot for ML(Extra Trees Regressor) model, (c) Slice plots for bootstrap, max_depth, max_features, min_samples_leaf, min_samples_split, and n_estimators hyperparameters for the ML(Extra Trees Regressor) model. The legend bar shows number of trials.

Supplementary Table 5: List of all ElemNet {E} features selected from mean feature ranking utilized in the present study

| Feature name	Feature type	Feature name	Feature type
Zn	Elemental	Re	Elemental
Cd	Elemental	Th	Elemental
Tl	Elemental	Pu	Elemental
Hg	Elemental	Sb	Elemental
U	Elemental	Zr	Elemental
Be	Elemental	Eu	Elemental
Mg	Elemental	Os	Elemental
Sr	Elemental	Ag	Elemental
Cr	Elemental	Pb	Elemental
Ca	Elemental	La	Elemental
Ni	Elemental	Sn	Elemental
Mo	Elemental	Rh	Elemental
Ba	Elemental	Rb	Elemental
W	Elemental	Pd	Elemental
Mn	Elemental	Ga	Elemental
Yb	Elemental	Au	Elemental
Fe	Elemental	In	Elemental
Pa	Elemental	Ir	Elemental
Si	Elemental	Bi	Elemental
Ta	Elemental	Pt	Elemental
Ru	Elemental	Nd	Elemental
Al	Elemental	Cs	Elemental
Li	Elemental	Pr	Elemental
Ce	Elemental	Sc	Elemental
Tc	Elemental	Ac	Elemental
Np	Elemental	Pm	Elemental
Ti	Elemental	Sm	Elemental
Co	Elemental	Gd	Elemental
Cu	Elemental	Y	Elemental
Na	Elemental	Tb	Elemental
K	Elemental	Lu	Elemental
As	Elemental	Ho	Elemental
Hf	Elemental	Er	Elemental
Ge	Elemental	Dy	Elemental
V	Elemental	Tm	Elemental
Te	Elemental		
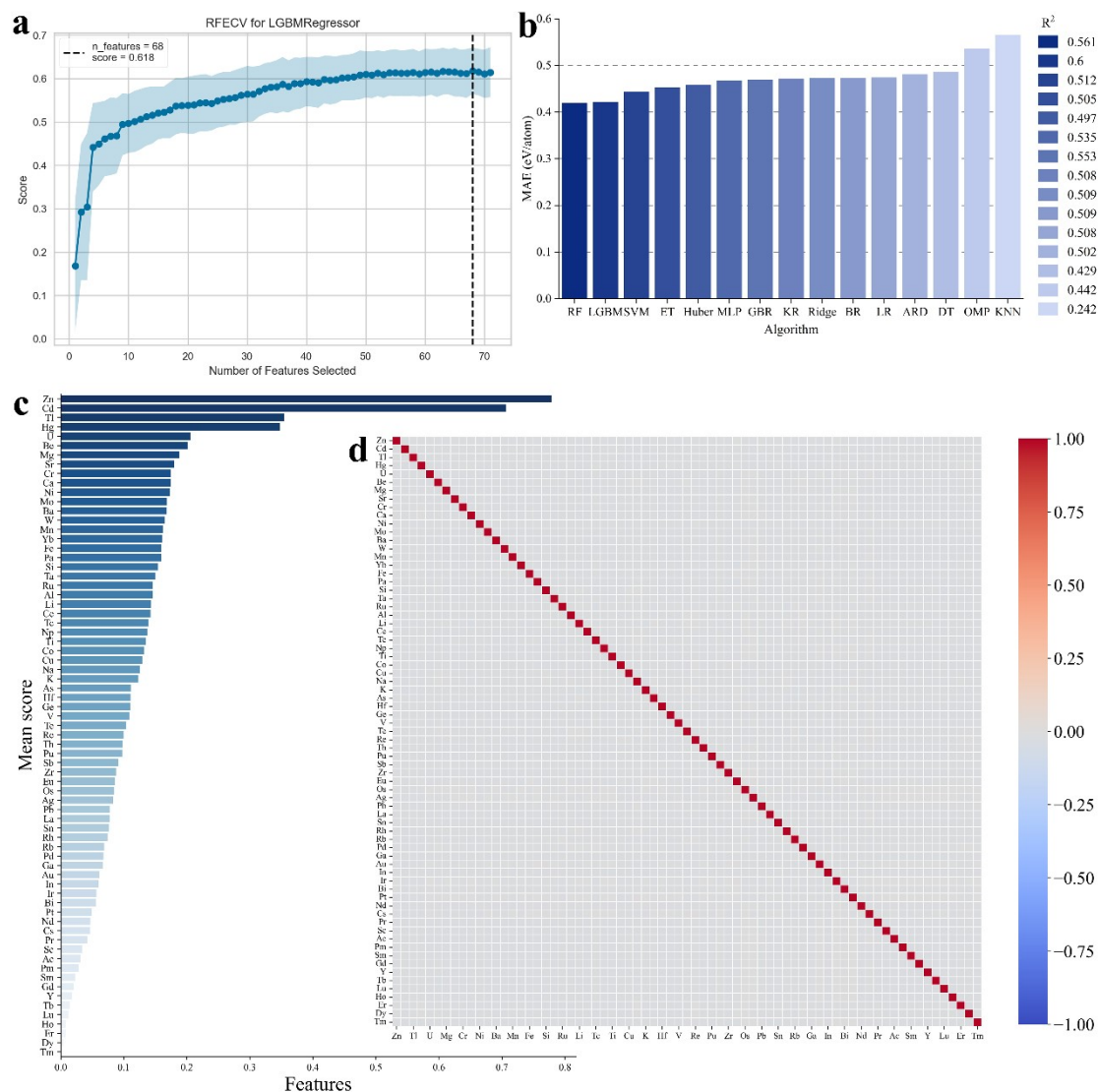
Supplementary Figure 8. (a) Learning curve for feature selection from ElemNet {E} feature Set, (b) Test MAE and R^2 for MAX(D_T, D_O) regression corresponding to all ML algorithms using {E} feature set. (c) Mean score of best features selected from the {E} feature set (highly correlated features removed) for MAX(D_T, D_O). (d) Pearson correlation between the best {E} features (after removing highly correlated features) selected from feature ranking.

Supplementary Figure 9. (a) Contour plot of best three hyperparameters utilized in the ML(ET) model using the selected {E} features for MAX(D_T, D_O) regression. (b) Hyperparameter importance plot for ML(Extra Trees Regressor) model, (c) Slice plots for bootstrap, max_depth, max_features, min_samples_leaf, min_samples_split, and n_estimators hyperparameters for the ML(Extra Trees Regressor) model. The legend bar shows number of trials.

Supplementary Table S6. Predicted 14 stable spinel structures with excellent OER catalytic activity

| Structure	E_{hull} (eV)	MAX(D_T, D_O) (eV)
$CdSc_2O_4$	0.038	3.408
$CdRh_2O_4$	0.030	2.830
$ZnAl_2O_4$	0.025	3.157
$IrLi_2O_4$	0.019	2.700
$RuNa_2O_4$	0.011	2.830
$IrRb_2O_4$	0	2.838
$MoAg_2O_4$	0	2.912
$MoNa_2O_4$	0	2.536
$RuLi_2O_4$	0	2.713
$TcCs_2O_4$	0	2.973
TcK_2O_4	0	3.089
$TcLi_2O_4$	0	2.742
$TcNa_2O_4$	0	3.134
$TcRb_2O_4$	0	3.060
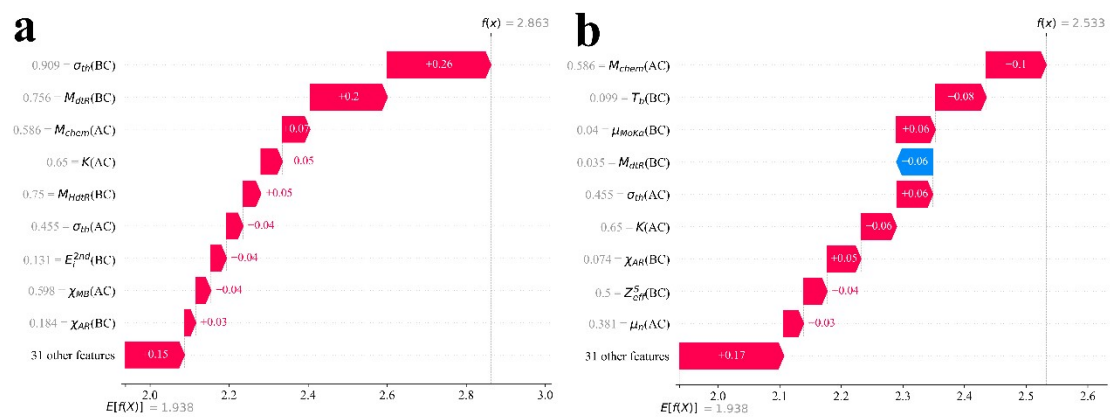
Supplementary Figure 10. The individual SHAP force plots for (a) MoAg$_2$O$_4$ and (b) MoNa$_2$O$_4$.
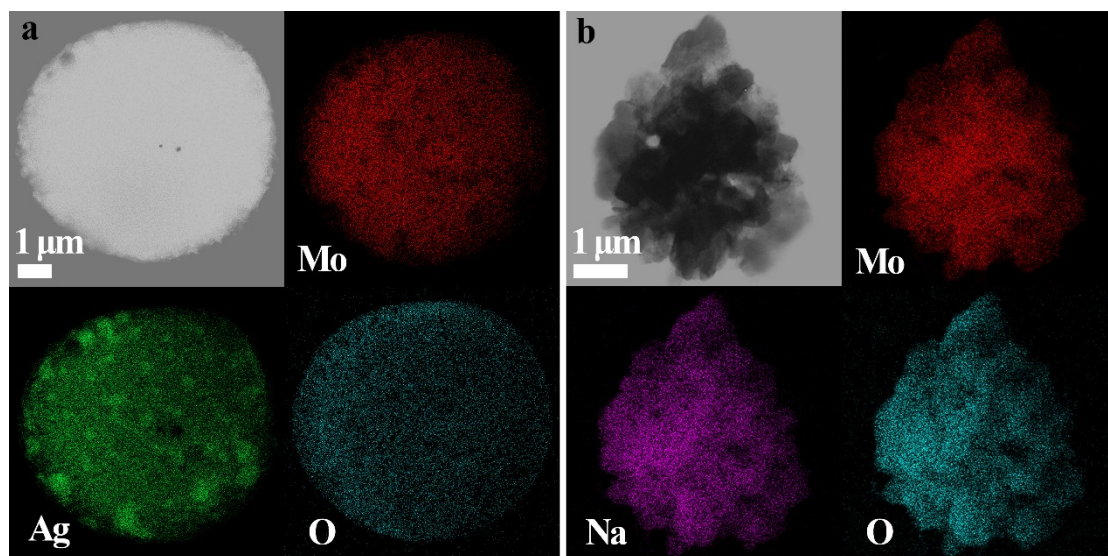
Supplementary Figure 11. HAADF-STEM image and EDS elemental mapping images for (a) $MoAg_2O_4$ and (b) $MoNa_2O_4$.
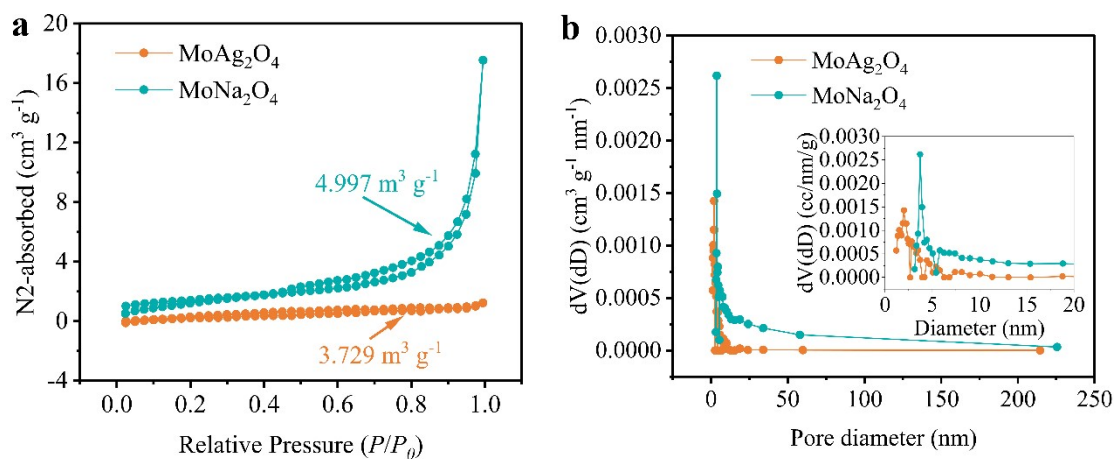
Supplementary Figure 12. (a) N_2-sorption isotherms and (b) pore size distribution of $MoAg_2O_4$ and $MoNa_2O_4$.
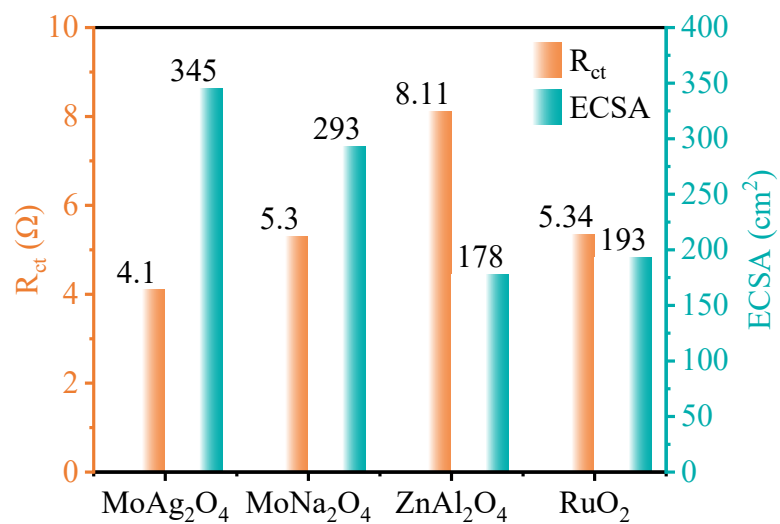
Supplementary Figure 13. The comparison chart of R_{ct} and ECSA of $MoAg_2O_4$, $MoNa_2O_4$, $ZnAl_2O_4$, and RuO_2 electrocatalysts, respectively.
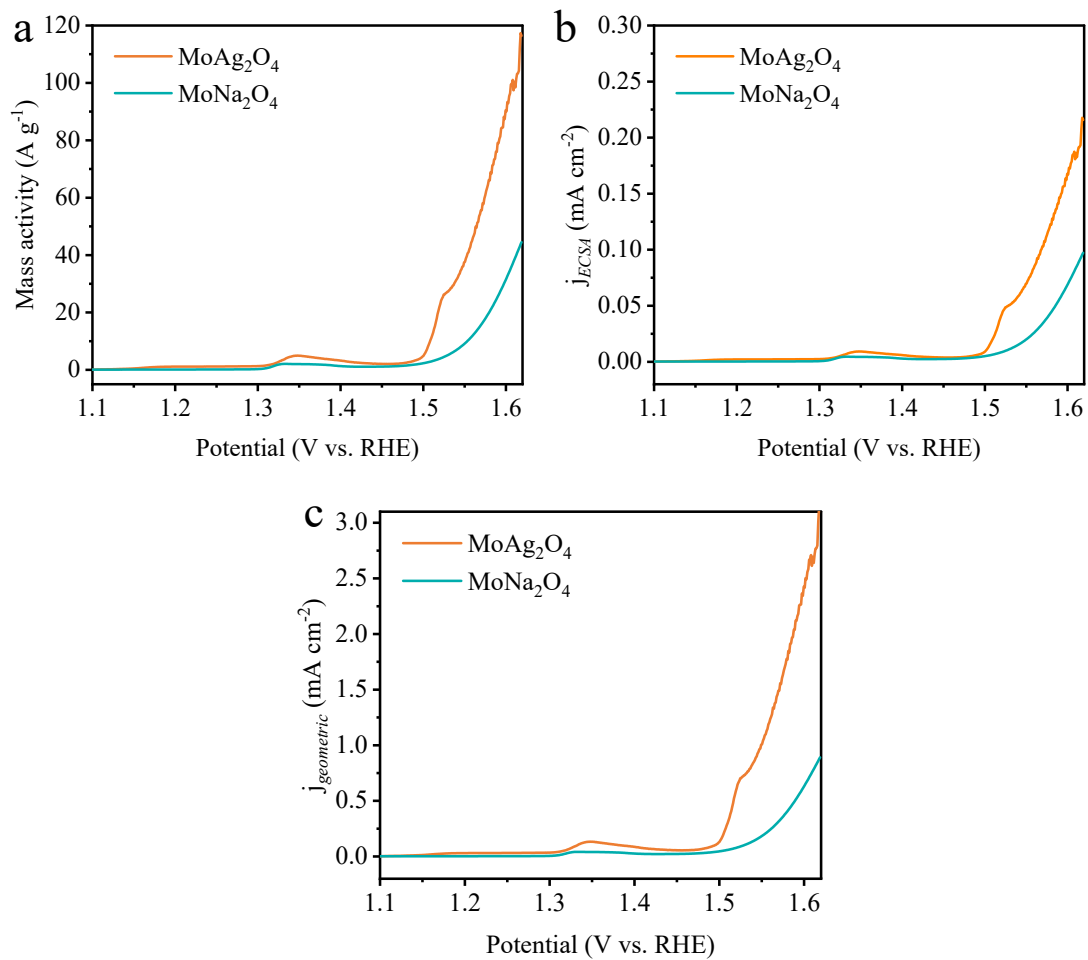
Supplementary Figure 14. The LSV curves of MoAg$_2$O$_4$ and MoNa$_2$O$_4$ samples normalized to (a) catalytic mass, (b) ECSA-, and (c) BET geometric surface area.

Supplementary References

1 A. Moez, 2020. https://pycaret.org

2 G. Kresse and J. Furthmüller, *Phys. Rev. B*, 1996, **54**, 11169.

3 J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.

4 M. Esmaeilirad, Z. Jiang, A. M. Harzandi, A. Kondori, M. Tamadoni Saray, C. U. Segre, R. Shahbazian-Yassar, A. M. Rappe and M. Asadi, *Nat. Energy*, 2023, **8**, 891–900.

5 N. Hamzah, M. H. Samat, N. A. Johari, A. F. A. Faizal, O. H. Hassan, A. M. M. Ali, R. Zakaria, N. H. Hussin, M. Z. A. Yahya and M. F. M. Taib, *Microelectron. Int.*, 2022, **40**, 53–62.

6 Bajdich Michal, García-Mota Mónica, Vojvodic Aleksandra, Nørskov Jens K., and Bell Alexis T., *J. Am. Chem. Soc.*, 2013, **135**, 13521–13530.

7 A. Stolyarenko, DATABASE ON PROPERTIES OF CHEMICAL ELEMENTS. A.A.BAIKOV INSTITUTE OF METALLURGY AND MATERIALS SCIENCE (2020).