



# Machine learning assisted prediction and synthesis of new high-entropy fluorite oxide<sup>☆</sup>

Arthur da Silva Sousa Santos<sup>a</sup>, James Moraes de Almeida<sup>b</sup>, Fabiane de Jesus Trindade<sup>c</sup>,  
Marissol Rodrigues Felez<sup>a</sup>, Larissa da Silva Marques<sup>a</sup>, Daniel Zanetti de Florio<sup>a</sup>,\*

<sup>a</sup> Center for Engineering, Modeling and Applied Social Sciences, Federal University of ABC (UFABC), Avenida dos Estados, 5001, Santo André, 09210-580, SP, Brazil

<sup>b</sup> Ilum School of Science, Brazilian Center for Research in Energy and Materials (CNPEM), Rua Lauro Vannucci, 1020, Campinas, 13083-970, SP, Brazil

<sup>c</sup> Brazilian Synchrotron Light Laboratory, Brazilian Center for Research in Energy and Materials (CNPEM), R. Giuseppe Máximo Scolfaro, 10000, Campinas, 13083-100, SP, Brazil

## ARTICLE INFO

Dataset link: <https://github.com/Arthurns16/Machine-Learning-Assisted-Prediction-and-Synthesis-of-New-Single-Phase-High-Entropy-Oxides>,  
<https://doi.org/10.3390/jcs5120311>

### Keywords:

High-entropy oxides  
Machine learning

## ABSTRACT

High-entropy oxides are a novel class of materials with promising applications in energy conversion and storage; however, their rational design remains challenging due to the immense compositional space. Here, we propose a machine-learning-based methodology to design stable, single-phase HEOs. We trained predictive models to identify candidate fluorite-structured compositions. The ensemble achieved reasonable performance in a six-class classification task, as evaluated using nested stratified cross-validation and external validation (weighted-average F1-scores of 77% and 70%, respectively). We further applied SHAP analysis to assess the physical relevance of the predictors. Experimentally, we synthesized  $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{Al}_{0.2}\text{O}_{2-\delta}$ . X-ray diffraction confirmed this prediction, and transmission electron microscopy (TEM) combined with selected-area electron diffraction (SAED) validated the phase assignment. Overall, these results demonstrate that machine learning is a powerful approach to navigate the complex HEO compositional landscape and accelerate the discovery of materials with targeted properties.

## 1. Introduction

Inspired by recent advances in high-entropy alloys (HEAs) [1–4], high-entropy oxides (HEOs) have emerged as a novel class of functional ceramics. Although definitions differ, high-entropy materials are generally characterized by at least five principal elements—each with an atomic fraction of 5–35%—and a configurational mixing entropy exceeding  $1.5 R$  (or, in more stringent formulations,  $1.61 R$ ), where  $R$  is the gas constant [1–4]. The near-random distribution of cations with diverse ionic radii and electronegativities within a single crystalline lattice can stabilize unconventional single-phase structures and enable distinctive physical and chemical properties. This framework has driven rapid growth in HEO research, including composition design [5,6], synthesis methodologies [6,7], and applications [6,8–10].

High-entropy materials are commonly associated with four effects—lattice distortion, entropy-driven stabilization of single phases, sluggish diffusion, and a synergistic “cocktail effect”, whereby the emergent properties deviate from those predicted by the rule of mixtures [11–13]. In HEOs, these effects, together with the resulting crystal phase, play

a central role in determining functional performance across diverse application domains [3,13].

Despite rapid progress, robust models linking HEO composition and structure to properties of interest remain limited; consequently, HEO design is still frequently guided by laborious trial-and-error synthesis [5]. Traditional rules for high-entropy ceramics often fail, and the vast compositional space makes *ab initio* simulations inefficient for predictive screening [5]. Machine learning (ML) offers a practical alternative by enabling more accurate prediction than simple mixing rules (or single descriptors) by efficiently extracting patterns from data to navigate large composition spaces [14,15].

The use of ML to address challenges in high-entropy ceramics gained traction in 2020 [16], where random forest models were trained to predict the entropy formation ability (EFA) descriptor for high-entropy carbides (HECs), a stability-related quantity that is costly to calculate via density functional theory (DFT); the study also demonstrated predictive utility via *ab initio* validation and synthesis of new single-phase compositions. Subsequent works expanded ML screening to additional

<sup>☆</sup> This article is part of a Special issue entitled: ‘High Entropy Ceramics’ published in Open Ceramics.

\* Corresponding author.

E-mail address: [daniel.florio@ufabc.edu.br](mailto:daniel.florio@ufabc.edu.br) (D.Z. de Florio).

systems and model families: J. Zhang et al. [17] combined artificial neural networks and support vector machines for high-entropy carbides candidate identification, while Mitra et al. [18] applied ML to predict single-phase high-entropy diborides and reported high classification performance.

More recently, the suitability of high-cost descriptors (e.g., EFA) for large-scale exploration has been questioned [19,20]. Leverant et al. [19] showed that low-cost, composition-based features can efficiently predict EFA in high-entropy carbides and support discovery of previously unreported high-EFA compositions. Guo et al. [20] achieved performance comparable to DFT-trained models for (TiZrHf)C-based stability using low-cost features, and Mitra et al. [21] used a compact set of low-cost descriptors to predict stability across borides, carbides, and nitrides for ultra-high-temperature applications with experimental validation. Overall, low-cost features appear promising for high-entropy ceramics, but they remain underexplored for predicting high-entropy oxide phases.

Beyond screening, ML can yield interpretable heuristics when explainability tools are applied. Yan et al. [22] employed Partial Dependence Plot (PDP) and Individual Conditional Expectation (ICE) analyses for single-phase high-entropy transition metal diborides and proposed a phase-stability rule based on bond mismatch. Mitra et al. [23] used SHapley Additive exPlanations (SHAP) for stable-phase high-entropy carbides, highlighting, among other findings, that medium-to-high differences in Mulliken electronegativity favor single-phase formation.

Explainability-driven ML has also been used to relate descriptors to functional properties across high-entropy ceramics. Zhou et al. [24] predicted mechanical properties of high-entropy carbides with high accuracy, while Wen et al. [25] predicted thermal conductivity in  $A_2B_2O_7$ -type oxides with small deviations from experiments. In catalysis-focused studies, Mints et al. [26] explored noble-metal HEO spaces for oxygen evolution reaction (OER), and Nie et al. [27] employed Extreme Gradient Boosting (XGBoost) to classify pure spinel formation and to regress the  $T_{90}$  descriptor (as defined in Ref. [27]) for hydrogen evolution reaction activity, identifying high-performing high-entropy spinel oxides (HESOs) using only atomic fractions as input features. Lin et al. [28] investigated spinel stability in high-entropy oxides by integrating first-principles simulations with ML, reporting low prediction deviations relative to validation simulations.

Compared with other high-entropy materials, ML studies targeting HEOs remain scarce, largely due to limited data availability [5]. For the specific task of classifying HEO compositions among crystal systems, only Liu et al. [29] have explicitly addressed the problem. While Liu et al. achieved high accuracy, the study considered only three crystal-structure classes and tested a limited set of descriptors; moreover, it did not include external-dataset validation, experimental synthesis of model inferences, or neural-network-based architectures, despite the potential benefits of broader feature engineering and model diversity for both performance and interpretability.

Here we propose a machine-learning-assisted methodology for the design and synthesis of new single-phase HEO compositions. The workflow integrates (i) composition-based feature engineering for phase prediction, in which we generate more than 200 composition-derived descriptors and select informative subsets using two strategies (Pearson correlation filtering and permutation feature importance, PFI); (ii) training and benchmarking multiple base learners (Multi-Layer Perceptron, Extreme Learning Machine, Random Forest, and Gradient Boosting) under a leakage-safe nested stratified cross-validation protocol, followed by evaluation on an external dataset; (iii) a “cation combinatorial analysis” step to impose design constraints and generate more promising inferences; and (iv) a synthesis-oriented step. The final predictions are obtained with a soft-voting Ensemble Classifier that combines the calibrated probabilistic outputs of the individual models. Finally, to extract chemically meaningful insights, we interpret the ensemble using SHAP and use the trained framework to propose fluorite-targeted candidates; one predicted composition is synthesized and its phase is confirmed by X-ray diffraction (XRD) with Rietveld refinement and transmission electron microscopy (TEM).

## 2. Methodology

The complete methodology for synthesizing new HEOs with machine-learning assistance is summarized in Fig. 1. The workflow begins with data processing, where the raw data are formatted into a structure suitable for training machine-learning models. Next, the models are trained and evaluated using standard performance metrics, and the selected features are further examined using interpretability techniques such as SHAP. To bridge the trained models with the experimental synthesis stage, we introduce an intermediate cation combinatorial analysis step, in which the design objectives and constraints for the target material are defined.

### 2.1. From descriptors to features

In many studies that propose new HEO compositions, researchers focus on a limited region of the periodic table and rely heavily on trial-and-error synthesis, often without a detailed discussion of the underlying design rationale [30]. When HEO design is discussed explicitly, it is typically framed through a descriptor-based approach [30]. Here, descriptors refer to numerical quantities that probabilistically capture a material's physical or chemical attributes [13]. Historically, an early and influential example is the work of William Hume-Rothery, who proposed the rules – now known as the Hume-Rothery rules – describing conditions for forming solid-solution alloys [5]. For HEOs, commonly used descriptors (and those adopted here as machine-learning features) include elemental/atomic properties, outputs from thermodynamic experiments or simulations, and precursor-related properties [5].

### 2.2. Feature engineering

Feature engineering comprises two main stages – feature generation and feature selection – each encompassing a range of techniques that have been comprehensively reviewed by Morgan and Jacobs [31].

In this study, we consulted three data sources to assemble the information required to compute a large set of original features that, to the best of our knowledge, have not been systematically explored in the high-entropy materials machine-learning literature: (1) the atomic-features package provided by NOMAD Lab (Novel Materials Discovery Laboratory) [32]; (2) thermodynamic reaction data obtained from FactSage [33]; and (3) key attributes associated with the formation of stable phases in high-entropy oxides reported in prior studies [3,5,10,13,33], with the required elemental data retrieved from the Mendelev library [34].

To convert the raw inputs into features, we adopted the vectorization scheme proposed by Isayev et al. [35], which constructs composition-based quantities from elemental property vectors. Specifically, for each element in a compound, the relevant property value is assigned and then aggregated using operations such as the maximum, minimum, standard deviation, mean, or sum to yield a single scalar feature. For FactSage-derived quantities, analogous vectorization operations were applied, with values computed from oxide formation reactions under standard temperature and pressure conditions. This choice is motivated by common practice in the HEO literature: unless stated otherwise, descriptions of a compound as “single-phase” and “stable” typically refer to stability under standard conditions, and phase labels are generally assigned from X-ray diffraction measurements collected at room temperature. This convention also applies to the phase labels used to build the dataset in the present study [36]. For each reaction listed in Table 2, the software returned the entropy change, enthalpy change, Gibbs free energy change, and specific heat. The full set of 246 constructed features is shown in Figure S2 (Supplementary Material).

Phase-stability discussions in high-entropy oxides frequently invoke the Gibbs free energy of mixing,  $\Delta G_{\text{mix}} = \Delta H_{\text{mix}} - T\Delta S_{\text{mix}}$  [37]. However, explicitly evaluating these thermodynamic contributions for

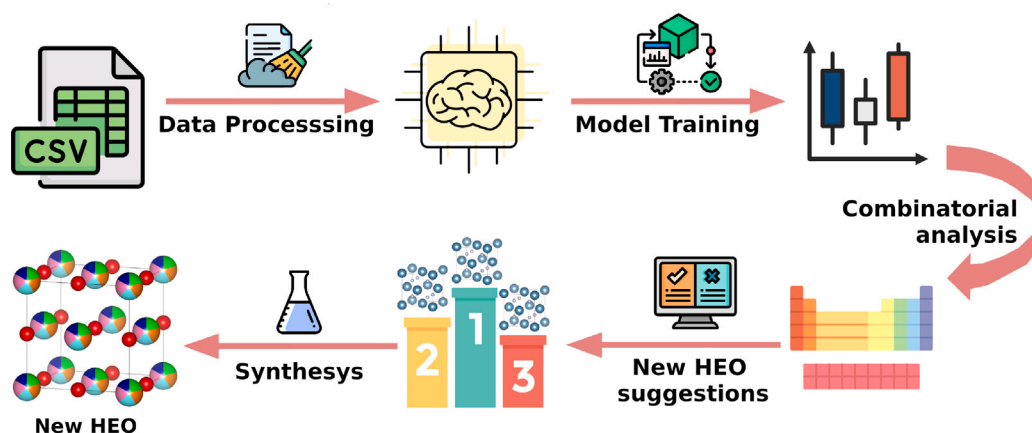


Fig. 1. Workflow for machine-learning-assisted synthesis of new HEOs.

multicomponent oxides is nontrivial and can be computationally demanding. It can therefore be advantageous to construct low-cost, thermodynamically motivated descriptors that are expected to correlate – albeit indirectly – with the driving forces that govern phase stability, and to use them as candidate inputs to machine-learning models within an exploratory feature framework. Indeed, FactSage-derived thermodynamic data have already been leveraged in machine-learning studies to predict phase stability in high-entropy alloys [38].

Within the HEA literature, several widely used approaches provide inexpensive approximations to thermodynamic quantities; for example, the enthalpy of mixing is often estimated with the Miedema model [39]. This approximation is not expected to be reliable for HEOs because oxide bonding differs substantially from that in metallic systems, for which the Miedema framework was originally developed [36]. To retain a thermodynamic basis while keeping the descriptors inexpensive to compute, we therefore use FactSage-derived thermodynamic quantities for binary oxides as proxy features. These quantities reflect the intrinsic cation-oxygen affinity of each constituent and can influence the energetic penalties associated with multi-cation mixing, as well as the competition with alternative oxide assemblages. According to the feature taxonomy of Ramprasad et al. [30], oxide-derived descriptors fall under *block-level* features (i.e., they are distinct from traditional elemental features). In this context, the present study is, to our knowledge, among the first to incorporate block-level features into HEO phase prediction, an approach more commonly adopted in polymer informatics. This type of extensive low-cost feature exploration remains largely absent from the high-entropy ceramics literature. Even in the work of Guo et al. [20], the workflow remains conventional, relying on a small set of features chosen largely by “researcher intuition”. More broadly, a common practice in the field is to select a limited feature set in an ad hoc manner. In contrast, our study presents a more systematic methodology for generating and selecting low-cost features. Given that the vast compositional space is a central challenge in high-entropy materials research [5], low-cost descriptors are essential for efficiently exploring the hypothesis space. If expensive features are required to train the models, then each new inference would also require recomputing these costly quantities, undermining one of the main advantages of machine learning: efficient exploration of large and complex search spaces.

### 2.3. Database

At present, there are no specialized databases dedicated to high-entropy oxides, and even major materials repositories contain relatively limited coverage of these systems [5]. As a result, applying machine learning to HEOs typically requires constructing datasets directly from the scientific literature. This limitation is a key reason why comparatively few ML studies on HEOs exist to date, with Liu et al. [29] being

a notable example; in that work, the authors leveraged the dataset compiled in Akrami’s review [40] (112 compositions). In the present study, we assembled a dataset of 137 compositions with corresponding phase labels from the supplementary material of Leong et al. [36].

Table 1 summarizes the number of instances per class, the class proportions, and representative compositions for each phase. A key observation is that the majority class accounts for 27.7% of the dataset. Under a naive frequency-based baseline, labeling every sample as the majority class would therefore yield an expected accuracy of 27.7%. In contrast, our models achieve substantially higher accuracies (see Results and discussion), indicating that they capture latent and physically meaningful patterns encoded in the feature space and exploit these patterns for phase classification. As expected, fewer robust patterns may be learned for minority classes, particularly the Mixed and Pyrochlore categories.

Leong’s dataset also exhibits reasonable compositional diversity, as reflected by the representative examples. Specifically, it contains 43 unique cation species, and their occurrence frequencies are summarized in the Pareto chart in Fig. 2. The 80% cumulative frequency threshold is reached at samarium (rank 16). Accordingly, after training, the model is expected to make more reliable inferences within the compositional subspace spanned by the 16 most frequent cation species.

Regarding label definition, Leong et al. first separated compositions into single-phase and multiphase groups [36]. Single-phase compositions were defined as those with no detectable traces of secondary/minority phases, whereas the Mixed class comprises multiphase compositions spanning both low and high fractions of secondary phases. Notably, Leong et al. did not systematically adopt the single-phase labels reported in the original source articles; instead, when minority phases could be identified from the reported characterization (e.g., [41–43]), compositions were reclassified accordingly. This indicates that the dataset construction included an additional relabeling step beyond straightforward data collection. Finally, for each single-phase composition, the specific crystal-structure class (e.g., fluorite, spinel) was assigned based on the information reported in the source publications, most likely relying primarily on X-ray diffraction evidence.

### 2.4. From phases to classes

High-entropy oxides can stabilize complex crystal structures, in part because the high configurational entropy favors cation disorder. Nevertheless, most reported HEO compositions crystallize into a relatively restricted set of phases [3,10,13,36]. We therefore defined the target labels in our machine-learning models using these commonly observed phases: pyrochlore, spinel, perovskite, fluorite, and rock salt. In addition, we included a Mixed class to represent compositions that do not stabilize as a single phase.

**Table 1**

Class proportions and representative compositions in the Leong dataset ( $N = 137$ ).

| Class      | Count | Proportion (%) | Representative samples                                                                                                                                                                                                                                                                                                          |
|------------|-------|----------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Rock salt  | 38    | 27.7           | $(\text{Co}_{0.2}\text{Cu}_{0.2}\text{Mg}_{0.2}\text{Ni}_{0.2}\text{Zn}_{0.2})\text{O}$ ;<br>$(\text{Mg}_{0.167}\text{Co}_{0.167}\text{Cu}_{0.167}\text{Ni}_{0.167}\text{Zn}_{0.167}\text{Sc}_{0.167})\text{O}$ ;<br>$\text{Pt}_{0.3}\text{Mg}_{0.2}\text{Co}_{0.2}\text{Cu}_{0.2}\text{Ni}_{0.2}\text{Zn}_{0.2}\text{O} \dots$ |
| Spinel     | 26    | 19.0           | $(\text{CoCuMgMnNi})\text{Cr}_2\text{O}_4$ ;<br>$(\text{Zn}_{0.2}\text{Fe}_{0.2}\text{Ni}_{0.2}\text{Mg}_{0.2}\text{Cd}_{0.2})\text{Fe}_2\text{O}_4$ ;<br>$(\text{CoMgMnNiZn})(\text{AlCoCrFeMn})_2\text{O}_4 \dots$                                                                                                            |
| Fluorite   | 25    | 18.2           | $(\text{CeZrYLaGd})\text{O}_2$ ;<br>$\text{Gd}_{0.2}\text{La}_{0.2}\text{Y}_{0.2}\text{Hf}_{0.2}\text{Zr}_{0.2}\text{O}_2$ ;<br>$(\text{HfZrCeYb})\text{O} \dots$                                                                                                                                                               |
| Pyrochlore | 16    | 11.7           | $\text{Y}_2(\text{TiZrHfMoV})_2\text{O}_7$ ;<br>$(\text{LaCeNdSmEu})_2\text{Zr}_2\text{O}_7$ ;<br>$\text{Gd}_2(\text{TiZrHfSn})_2\text{O}_7 \dots$                                                                                                                                                                              |
| Mixed      | 11    | 8.0            | $\text{Hf}_{0.314}\text{Zr}_{0.314}\text{Ce}_{0.314}\text{Y}_{0.029}\text{Yb}_{0.029}\text{O}_2$ ;<br>$\text{TiHfZrNbTaO}_{11}$ ;<br>$(\text{CeZrYbHoEr})\text{O} \dots$                                                                                                                                                        |
| Perovskite | 21    | 15.3           | $\text{Ba}_{0.5}\text{Sr}_{0.5}(\text{Zr}_{0.4}\text{Hf}_{0.3}\text{Ti}_{0.2}\text{Fe}_{0.3})\text{O}_3$ ;<br>$(\text{Ce}_{0.2}\text{La}_{0.2}\text{Pr}_{0.2}\text{Sm}_{0.2}\text{Y}_{0.2})\text{O}_3$ ;<br>$(\text{Na}_{0.2}\text{K}_{0.2}\text{Ca}_{0.2}\text{La}_{0.2}\text{Ce}_{0.2})\text{TiO}_3 \dots$                    |

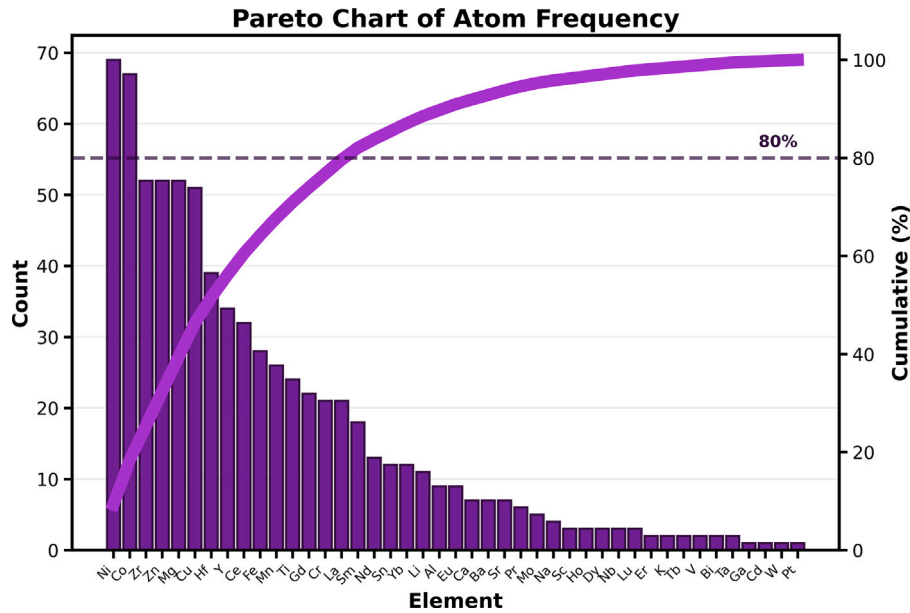


**Fig. 2.** Pareto chart of cation occurrence frequencies across the dataset.

Because the HEO compositions considered here are equimolar (or close to equimolar), we do not require the models to infer detailed stoichiometric proportions. Under this assumption, phase prediction can be formulated as a standard multi-class classification task.

## 2.5. Feature selection

We employed two feature-selection strategies: Pearson-correlation filtering and permutation feature importance. Given the dataset size and the complexity of the models considered, we targeted a compact feature set of approximately 10 variables, aiming to mitigate both underfitting and overfitting.

For the Pearson-correlation approach, we retained only one feature from each pair exhibiting an absolute correlation coefficient greater

than 0.6. For the PFI-based approach, we first removed highly collinear features (absolute correlation coefficient greater than 0.8), which yielded a reduced set of 48 features. We then trained models using this set and applied PFI to rank features by importance, selecting the top-ranked variables for subsequent modeling.

## 2.6. Models

In machine-learning studies, it is common to benchmark multiple model families because no single algorithm is universally optimal across tasks and datasets. Here, we considered two artificial neural network architectures—the multilayer perceptron (MLP) [44] and the extreme learning machine (ELM) [45]—and two tree-based methods—random forest (RF) and gradient boosting (GB) [44].

**Table 2**  
Equations used in calculations with FactSage.

| Cation | Equation                                                   | Cation | Equation                                                   |
|--------|------------------------------------------------------------|--------|------------------------------------------------------------|
| Er     | 4 Er + 3 O <sub>2</sub> → 2 Er <sub>2</sub> O <sub>3</sub> | Ni     | 2 Ni + O <sub>2</sub> → 2 NiO                              |
| Be     | 2 Be + O <sub>2</sub> → 2 BeO                              | Na     | 4 Na + O <sub>2</sub> → 2 Na <sub>2</sub> O                |
| Sm     | 4 Sm + 3 O <sub>2</sub> → 2 Sm <sub>2</sub> O <sub>3</sub> | Nb     | 2 Nb + 5 O <sub>2</sub> → Nb <sub>2</sub> O <sub>5</sub>   |
| Sc     | 4 Sc + 3 O <sub>2</sub> → 2 Sc <sub>2</sub> O <sub>3</sub> | Ru     | 2 Ru + 7 O <sub>2</sub> → 2 RuO <sub>4</sub>               |
| Eu     | 4 Eu + 3 O <sub>2</sub> → 2 Eu <sub>2</sub> O <sub>3</sub> | Fe     | 4 Fe + 3 O <sub>2</sub> → 2 Fe <sub>2</sub> O <sub>3</sub> |
| Pb     | Pb + O <sub>2</sub> → PbO <sub>2</sub>                     | Ag     | 4 Ag + O <sub>2</sub> → 2 Ag <sub>2</sub> O                |
| Tm     | 4 Tm + 3 O <sub>2</sub> → 2 Tm <sub>2</sub> O <sub>3</sub> | K      | 4 K + O <sub>2</sub> → 2 K <sub>2</sub> O                  |
| Ba     | 2 Ba + O <sub>2</sub> → 2 BaO                              | Sr     | 2 Sr + O <sub>2</sub> → 2 SrO                              |
| Li     | 4 Li + O <sub>2</sub> → 2 Li <sub>2</sub> O                | V      | 2 V + 5 O <sub>2</sub> → V <sub>2</sub> O <sub>5</sub>     |
| Ga     | 4 Ga + 3 O <sub>2</sub> → 2 Ga <sub>2</sub> O <sub>3</sub> | Gd     | 4 Gd + 3 O <sub>2</sub> → 2 Gd <sub>2</sub> O <sub>3</sub> |
| Dy     | 4 Dy + 3 O <sub>2</sub> → 2 Dy <sub>2</sub> O <sub>3</sub> | Ho     | 4 Ho + 3 O <sub>2</sub> → 2 Ho <sub>2</sub> O <sub>3</sub> |
| Co     | 2 Co + O <sub>2</sub> → Co <sub>3</sub> O <sub>4</sub>     | Lu     | 4 Lu + 3 O <sub>2</sub> → 2 Lu <sub>2</sub> O <sub>3</sub> |
| Mn     | 4 Mn + 3 O <sub>2</sub> → 2 Mn <sub>2</sub> O <sub>3</sub> | Ca     | 2 Ca + O <sub>2</sub> → 2 CaO                              |
| Ta     | 4 Ta + 5 O <sub>2</sub> → 2 Ta <sub>2</sub> O <sub>5</sub> | Yb     | 4 Yb + 3 O <sub>2</sub> → 2 Yb <sub>2</sub> O <sub>3</sub> |
| Pr     | 4 Pr + 3 O <sub>2</sub> → 2 Pr <sub>2</sub> O <sub>3</sub> | Mo     | 2 Mo + 3 O <sub>2</sub> → 2 MoO <sub>3</sub>               |
| Cr     | 4 Cr + 3 O <sub>2</sub> → 2 Cr <sub>2</sub> O <sub>3</sub> | Mg     | 2 Mg + O <sub>2</sub> → 2 MgO                              |
| Cu     | 2 Cu + O <sub>2</sub> → 2 CuO                              | Pt     | 2 Pt + O <sub>2</sub> → 2 PtO <sub>2</sub>                 |
| Tb     | 4 Tb + 3 O <sub>2</sub> → 2 Tb <sub>2</sub> O <sub>3</sub> | Cd     | 2 Cd + O <sub>2</sub> → 2 CdO                              |
| Bi     | 4 Bi + 3 O <sub>2</sub> → 2 Bi <sub>2</sub> O <sub>3</sub> | Nd     | 4 Nd + 3 O <sub>2</sub> → 2 Nd <sub>2</sub> O <sub>3</sub> |
| Hf     | 2 Hf + O <sub>2</sub> → 2 HfO <sub>2</sub>                 | Y      | 4 Y + 3 O <sub>2</sub> → 2 Y <sub>2</sub> O <sub>3</sub>   |
| Ce     | 2 Ce + O <sub>2</sub> → 2 CeO <sub>2</sub>                 | Si     | Si + O <sub>2</sub> → SiO <sub>2</sub>                     |
| S      | S + 3 O <sub>2</sub> → SO <sub>3</sub>                     | La     | 4 La + 3 O <sub>2</sub> → 2 La <sub>2</sub> O <sub>3</sub> |
| Sn     | Sn + O <sub>2</sub> → SnO <sub>2</sub>                     | Zn     | 2 Zn + O <sub>2</sub> → 2 ZnO                              |
| Zr     | Zr + O <sub>2</sub> → ZrO <sub>2</sub>                     | Al     | 4 Al + 3 O <sub>2</sub> → 2 Al <sub>2</sub> O <sub>3</sub> |
| Ti     | Ti + O <sub>2</sub> → TiO <sub>2</sub>                     | W      | 2 W + 3 O <sub>2</sub> → 2 WO <sub>3</sub>                 |

Neural networks were included because they often achieve strong predictive performance in materials informatics applications [46]. Tree-based models were selected because they are frequently competitive with neural networks, particularly on small datasets [47], as in the present study.

While many ML-based workflows evaluate more than one model, the models are often used independently. In contrast, we combine all four learners using a soft-voting ensemble classifier [48], which aggregates calibrated class probabilities to produce the final prediction, as discussed further in the Results and discussion section.

## 2.7. Model evaluation protocol and performance metrics

To reduce the variance and potential bias associated with a single hold-out split on a small and class-imbalanced dataset, we evaluated all models using a leakage-safe nested stratified cross-validation protocol. We used  $K_{\text{outer}} = 5$  outer folds to estimate generalization performance and  $K_{\text{inner}} = 5$  inner folds for hyperparameter selection.

Let  $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^N$  denote the full dataset, where  $\mathbf{x}_i$  is the feature vector of sample  $i$ ,  $y_i \in \{1, \dots, C\}$  is its ground-truth class label,  $N$  is the number of samples, and  $C$  is the number of classes. In each outer iteration  $k \in \{1, \dots, K_{\text{outer}}\}$ , we held out a stratified index set  $\mathcal{I}_{\text{test}}^{(k)}$  as the *outer-test* fold, while the remaining indices  $\mathcal{I}_{\text{train}}^{(k)}$  formed the *outer-training* set. Hyperparameters were optimized exclusively within  $\mathcal{I}_{\text{train}}^{(k)}$  via an inner stratified  $K_{\text{inner}} = 5$  cross-validation loop, implemented with GridSearchCV from scikit-learn [44]. The selection criterion was the mean inner-loop accuracy. After identifying the best hyperparameter configuration, the model was refit on the full outer-training set  $\mathcal{I}_{\text{train}}^{(k)}$  and evaluated once on the untouched outer-test set  $\mathcal{I}_{\text{test}}^{(k)}$ .

Crucially, all data-driven operations were restricted to the training portion of each inner split to prevent information leakage. This includes preprocessing steps (e.g., scaling) and any feature processing or selection (e.g., Pearson-correlation filtering). For each inner split, transformations were fit using only the inner-training subset and then applied to the corresponding inner-validation subset using parameters learned solely from inner-training data.

After completing the outer loop, we formed *out-of-fold* (OOF) predictions by concatenating the outer-test predictions across folds, so that each sample is predicted exactly once when it appears in an outer-test fold. Let  $\hat{y}_i^{\text{OOF}}$  denote the OOF-predicted label for sample  $i$ . Global OOF

metrics were computed from the concatenated vectors  $(y_1, \dots, y_N)$  and  $(\hat{y}_1^{\text{OOF}}, \dots, \hat{y}_N^{\text{OOF}})$ .

In addition to OOF-aggregated results, we also report *fold-wise* outer-test metrics summarized as mean  $\pm$  standard deviation across the  $K_{\text{outer}} = 5$  outer folds. Specifically, we report accuracy and macro- and support-weighted precision, recall, and F1 as mean  $\pm$  standard deviation across outer folds.

**Accuracy.** Accuracy was computed from OOF predictions as

$$\text{Accuracy} = \frac{1}{N} \sum_{i=1}^N \mathbb{I}(\hat{y}_i^{\text{OOF}} = y_i), \quad (1)$$

where  $\mathbb{I}(\cdot)$  is the indicator function.

**Class-wise precision, recall, F1, and support.** For class-wise evaluation, we adopted a one-vs-rest scheme. For each class  $c \in \{1, \dots, C\}$ , let  $TP_c$ ,  $FP_c$ , and  $FN_c$  denote the numbers of true positives, false positives, and false negatives, respectively, computed from OOF predictions. Precision, recall, and F1 for class  $c$  are then defined as

$$\begin{aligned} \text{Precision}_c &= \frac{TP_c}{TP_c + FP_c}, \quad \text{Recall}_c = \frac{TP_c}{TP_c + FN_c}, \\ \text{F1}_c &= \frac{2 \text{Precision}_c \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c}. \end{aligned} \quad (2)$$

The *support* of class  $c$  in the full dataset is

$$N_c = \sum_{i=1}^N \mathbb{I}(y_i = c). \quad (3)$$

**Macro- and weighted-average metrics.** To summarize performance under class imbalance, we report macro- and support-weighted averages computed from class-wise values. For a generic class-wise metric  $m_c$  (e.g., precision, recall, or F1 for class  $c$ ),

$$m_{\text{macro}} = \frac{1}{C} \sum_{c=1}^C m_c, \quad m_{\text{weighted}} = \sum_{c=1}^C \frac{N_c}{N} m_c. \quad (4)$$

**Outer-fold mean and standard deviation.** For completeness, we report fold-wise outer-test metrics as mean  $\pm$  standard deviation across the  $K_{\text{outer}} = 5$  outer folds. For a scalar metric  $m$ , let  $m^{(k)}$  be its value computed on outer-test fold  $k$ . The across-fold mean  $\bar{m}$  and sample standard deviation  $s_m$  are

$$\bar{m} = \frac{1}{K_{\text{outer}}} \sum_{k=1}^{K_{\text{outer}}} m^{(k)}, \quad s_m = \sqrt{\frac{1}{K_{\text{outer}} - 1} \sum_{k=1}^{K_{\text{outer}}} (m^{(k)} - \bar{m})^2}, \quad (5)$$

where  $s_m$  denotes the (sample) standard deviation of  $m$  across outer folds. In tables, these results are reported as  $\bar{m} \pm s_m$ .

**Confusion matrices.** We report the *row-normalized* OOF confusion matrix to visualize systematic confusions among phases and to emphasize per-class recall. Let  $M_{ij}$  be the OOF confusion count, defined as the number of samples whose true class is  $i$  and predicted class is  $j$ :

$$M_{ij} = \sum_{n=1}^N \mathbb{I}(y_n = i) \mathbb{I}(\hat{y}_n^{\text{OOF}} = j), \quad (6)$$

for  $i, j \in \{1, \dots, C\}$ . The reported normalized confusion matrix  $\widetilde{M}_{ij}$  is computed by normalizing each row by its total:

$$\widetilde{M}_{ij} = \frac{M_{ij}}{\sum_{j'=1}^C M_{ij'}}, \quad (7)$$

where  $\sum_{j'=1}^C M_{ij'}$  is the total number of samples with true class  $i$  (i.e., the row sum). By construction,  $\sum_{j=1}^C \widetilde{M}_{ij} = 1$  for each  $i$ , and the diagonal entries  $\widetilde{M}_{ii}$  correspond to the class-wise recall.

**ROC curves and AUC (one-vs-rest).** For probabilistic models, we computed one-vs-rest receiver operating characteristic (ROC) curves and the corresponding area under the curve (AUC) using scikit-learn [44]. Let  $p_{i,c} \in [0, 1]$  be the OOF-predicted probability that sample  $i$  belongs to class  $c$ , and let  $y_{i,c} \in \{0, 1\}$  be the binarized ground-truth indicator for class  $c$  (i.e.,  $y_{i,c} = 1$  if  $y_i = c$  and 0 otherwise). Sweeping a threshold  $t \in [0, 1]$  over  $p_{i,c}$  yields the true positive rate (TPR) and false positive rate (FPR):

$$\begin{aligned} \text{TPR}_c(t) &= \frac{\sum_{i=1}^N \mathbb{I}(y_{i,c} = 1) \mathbb{I}(p_{i,c} \geq t)}{\sum_{i=1}^N \mathbb{I}(y_{i,c} = 1)}, \\ \text{FPR}_c(t) &= \frac{\sum_{i=1}^N \mathbb{I}(y_{i,c} = 0) \mathbb{I}(p_{i,c} \geq t)}{\sum_{i=1}^N \mathbb{I}(y_{i,c} = 0)}. \end{aligned} \quad (8)$$

The AUC for class  $c$  is computed via trapezoidal integration over the ROC curve:

$$\text{AUC}_c = \sum_{r=1}^{n-1} \frac{(\text{FPR}_{r+1} - \text{FPR}_r)(\text{TPR}_r + \text{TPR}_{r+1})}{2}, \quad (9)$$

where  $\{(\text{FPR}_r, \text{TPR}_r)\}_{r=1}^n$  are ROC coordinates sorted by increasing FPR and  $n$  is the number of points on the discretized curve. AUC is not reported for the ELM model in our implementation because it outputs hard class labels, whereas AUC requires a continuous score (e.g., a probability or decision function).

## 2.8. Cations combinatorial analysis

The cations combinatorial analysis step is intended to integrate the trained models with experimental synthesis in a practical and informed manner; its goals are to limit the number of inferences generated by the models and to increase the likelihood of successful synthesis.

In a real-world HEO design setting, design constraints naturally restrict the accessible compositional space. For example, when designing a solid electrolyte for a solid oxide fuel cell, high ionic conductivity is required; therefore, candidate inferences can be generated by fixing cerium across the enumerated composition combinations [49].

Regarding increasing the chances of obtaining a single-phase HEO, we highlight two strategies: (1) fixation of cations that previous studies indicate tend to preserve a given structure even when mixed, which increases the likelihood of success [41]; and (2) starting from low- or medium-entropy compositions that have been successfully synthesized as single phase, and using machine-learning models to add the remaining cations such that the composition reaches the high-entropy regime, which can also improve the chances of success [50].

## 2.9. Synthesis method

In this study, we adopted a hydrothermal synthesis route. Other synthesis approaches could also be explored to fabricate compositions suggested by the machine-learning models. In the literature on high-entropy ceramics, it is common to exclude the synthesis method as an input feature for phase-prediction models [5,16–18,27]. This practice is primarily motivated by data scarcity: there are currently too few HEO samples synthesized via a wide range of methods and characterized in a sufficiently consistent manner to encode the synthesis route reliably. In addition, each synthesis technique introduces its own set of process parameters, which further complicates its representation in a unified model. Liu et al. [29] attempted to include synthesis method information, but it had limited influence on the model decisions, likely for the reasons outlined above. Therefore, our models were trained using the data that are presently available. Under these constraints, perfect accuracy is not expected and some degree of error is unavoidable. Nonetheless, the candidates proposed by the models are more likely to succeed in synthesis than compositions selected at random or purely based on intuition.

The starting materials were cerium nitrate  $\text{Ce}(\text{NO}_3)_3 \cdot 6\text{H}_2\text{O}$ , lanthanum nitrate  $\text{La}(\text{NO}_3)_3 \cdot 6\text{H}_2\text{O}$ , neodymium nitrate

$\text{Nd}(\text{NO}_3)_3 \cdot 6\text{H}_2\text{O}$ , magnesium nitrate  $\text{Mg}(\text{NO}_3)_2 \cdot 6\text{H}_2\text{O}$ , and aluminum nitrate  $\text{Al}(\text{NO}_3)_3 \cdot 9\text{H}_2\text{O}$  (analytical grade,  $\geq 99.9\%$ , Sigma-Aldrich), as well as sodium hydroxide NaOH (P.A.-A.C.S., 100%, Synth). All reagents were used as received, without further purification. Deionized water and urea  $(\text{NH}_2)_2\text{CO}$  (P.A.-A.C.S., 100%, Nuclear) were also used.

High-entropy oxide powders  $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{Al}_{0.2}\text{O}_{2-\delta}$  (equimolar, 20% each cation) were synthesized via a hydrothermal process, following a procedure similar to that reported elsewhere [51]. Briefly, 0.20 M solutions of each precursor nitrate were prepared separately, and urea was added to a final concentration of 0.01 M. The precursor solutions were then added to 30 mL of 7 M NaOH and stirred for 30 min. The resulting mixture was transferred to a 100 mL autoclave and heated at 70 °C for 24 h. After cooling, the product was washed five times with ultrapure water and three times with ethanol, and then dried at 70 °C for 12 h. Finally, the dried powder was calcined at 800 °C for 1 h using a heating rate of 5 °C min<sup>-1</sup>.

X-ray powder diffraction measurements were performed on a Bruker AXS D8 Focus diffractometer equipped with a Ni filter and a Lynxeye detector, using Cu K $\alpha$  radiation ( $\lambda = 1.5418 \text{ \AA}$ ). Patterns were collected at room temperature over  $2\theta = 20\text{--}90^\circ$  with a step size of  $0.05^\circ$  and a scan rate of  $1^\circ \text{ min}^{-1}$ . Rietveld refinement was performed using reference data from the Inorganic Crystal Structure Database (ICSD), and the refinements were carried out with TOPAS Academic 7 software. Instrumental peak broadening was corrected using a fundamental-parameters refinement approach. Additional analyses to validate the effective phase were performed by TEM using a Talos F200X G2 (Thermo Scientific), including energy-dispersive X-ray spectroscopy (EDS) elemental mapping and selected-area electron diffraction (SAED). TEM micrographs and diffraction patterns were processed and analyzed using ImageJ (v1.54g) [52].

## 3. Results and discussion

### 3.1. Nested stratified cross-validation (evaluation protocol)

Pearson correlation quantifies the strength of linear dependence between pairs of features and provides a simple criterion for removing redundant descriptors. In our workflow, we applied correlation-based filtering by discarding features whose absolute pairwise correlation exceeded 0.6. This reduces multicollinearity and limits dimensionality, while keeping a compact feature set consistent with the target dimensionality discussed in Section 2.5. In addition, we removed clearly uninformative descriptors (e.g., constant-valued features) prior to model training.

Table 3 summarizes the most frequently retained descriptors selected by the nested stratified cross-validation pipeline under the Pearson-based feature selection procedure (including a tie for the tenth position). For each descriptor, we report the number of occurrences, defined as the number of times it was retained across all outer-fold runs and models, together with a brief description. The Supplementary Information includes Table S1, which provides an analogous summary for the permutation feature importance selection strategy. Overall, the Pearson-based procedure was substantially more stable: the top nine Pearson-selected features were identical across all outer folds and models, whereas under permutation feature importance no single feature was retained consistently across all outer folds and models.

An additional observation is that the descriptors retained by both selection strategies (Pearson filtering and PFI) share closely related physical meaning, spanning multiple electronegativity scales, electron affinity, atomic-size measures, and enthalpy-related descriptors. These quantities are frequently implicated in discussions of stable single-phase HEO formation [3,5,10,13,35]. A few selected descriptors do not have an equally direct or intuitive connection to phase formation and may instead reflect indirect correlations with the phenomenon (e.g., `atomic_spacegroupnum_max`). This agreement supports the

**Table 3**

Most frequent features selected across the outer folds (Pearson approach). Occurrences count how many times each feature was retained across all outer-fold runs and models (maximum = 5 folds  $\times$  4 models = 20).

| Feature                  | Occurrences | Description                                                                       |
|--------------------------|-------------|-----------------------------------------------------------------------------------|
| atomic_ea_min            | 20          | Minimum electron affinity among constituent elements.                             |
| atomic_ea_sum            | 20          | Sum of electron affinity over all constituent elements.                           |
| atomic_en_allen_std      | 20          | Standard deviation of Allen electronegativity across constituent elements.        |
| atomic_en_allen_sum      | 20          | Sum of Allen electronegativity over constituent elements.                         |
| atomic_en_allredroch_min | 20          | Minimum Allred-Rochow electronegativity among constituent elements.               |
| atomic_hatm_min          | 20          | Minimum enthalpy of atomization ( $H_{\text{atom}}$ ) among constituent elements. |
| electrical_resist_max    | 20          | Maximum elemental electrical resistivity among constituent elements.              |
| mineral_hardness_max     | 20          | Maximum elemental mineral hardness among constituent elements.                    |
| van_der_waals_rad_min    | 20          | Minimum van der Waals radius among constituent elements.                          |
| atomic_ea_max            | 16          | Maximum electron affinity among constituent elements.                             |
| atomic_spacegroupnum_max | 16          | Maximum space-group number of the elemental solid among constituent elements.     |

**Table 4**

OOF-aggregated performance metrics from the nested stratified cross-validation evaluation. Acc = accuracy;  $P_{\text{macro}}$ ,  $R_{\text{macro}}$ ,  $F1_{\text{macro}}$  = macro-averaged precision, recall, and F1;  $P_w$ ,  $R_w$ ,  $F1_w$  = support-weighted (frequency-weighted) precision, recall, and F1.

| Model             | Acc   | $P_{\text{macro}}$ | $R_{\text{macro}}$ | $F1_{\text{macro}}$ | $P_w$ | $R_w$ | $F1_w$ |
|-------------------|-------|--------------------|--------------------|---------------------|-------|-------|--------|
| Random Forest     | 0.781 | 0.645              | 0.677              | 0.656               | 0.732 | 0.781 | 0.751  |
| Gradient Boosting | 0.759 | 0.663              | 0.666              | 0.663               | 0.746 | 0.759 | 0.751  |
| MLP               | 0.788 | 0.688              | 0.693              | 0.683               | 0.760 | 0.788 | 0.769  |
| ELM               | 0.737 | 0.636              | 0.638              | 0.633               | 0.710 | 0.737 | 0.720  |
| Ensemble          | 0.788 | 0.686              | 0.695              | 0.687               | 0.764 | 0.788 | 0.773  |

suitability of our feature generation and selection procedure and suggests that the learned patterns reflect physically meaningful structure-composition relationships rather than purely statistical associations. Notably, all retained descriptors in Table 3 originate from the NOMAD Lab atomic features package, which is widely used in materials-informatics workflows [53].

Given the similarity in feature type and the substantially lower computational cost of correlation filtering, the Pearson-based approach is preferable for the present application, particularly because we train four base learners and an ensemble model. Importantly, predictive performance was statistically equivalent for the two selection strategies: the ensemble achieved an accuracy of  $0.79 \pm 0.05$  with Pearson selection and  $0.80 \pm 0.04$  with permutation feature importance. Because these differences are not significant and Pearson selection provides clear stability and efficiency advantages, we adopt the Pearson-selected feature set for the more detailed evaluation and for training the final ensemble used in subsequent inference.

In addition to dimensionality reduction to mitigate overfitting, careful definition of the hyperparameter search space is essential for robust model selection. We therefore bounded each hyperparameter grid to reflect the limited dataset size. The resulting search spaces for the four trained models are reported in Tables S2-S5 in the Supplementary Information.

Table 4 reports the out-of-fold aggregated performance metrics. Under this metric set, the ensemble achieves the highest values for all evaluated criteria except macro precision, where it ranks second by a margin of 0.002. Overall, the ensemble consistently outperforms the individual base learners, consistent with the expected behavior of ensemble methods, which are typically more robust than single models [54]. Among the individual models, the multilayer perceptron yields the strongest results. Nevertheless, all standalone learners and the ensemble exhibit broadly similar performance; for example, the accuracy standard deviations are on the order of 2%.

Macro-averaged metrics assess model behavior under the assumption that all classes are equally important. In class-imbalanced settings, however, macro metrics can appear lower because they weight minority classes equally with majority classes. In the present study, the Mixed class is the smallest class and is also intrinsically difficult to learn because it represents the coexistence of two or more phases.

**Table 5**

Class-wise OOF-aggregated performance for the ensemble model. P = precision, R = recall, F1 = F1-score (computed in a one-vs-rest manner for each class), and Support = number of samples in the class.

| Class      | Support | P     | R     | F1    |
|------------|---------|-------|-------|-------|
| Fluorite   | 25      | 0.767 | 0.920 | 0.836 |
| Mixed      | 11      | 0.167 | 0.091 | 0.118 |
| Perovskite | 21      | 0.800 | 0.762 | 0.780 |
| Pyrochlore | 16      | 0.600 | 0.562 | 0.581 |
| RockSalt   | 38      | 0.900 | 0.947 | 0.923 |
| Spinel     | 26      | 0.885 | 0.885 | 0.885 |

Moreover, even minor secondary phases lead to the Mixed label under our protocol, consistent with the definition adopted by Leong et al. [36]. Consequently, the trend in Table 4-macro metrics being lower than weighted metrics-is expected. Despite these challenges, the ensemble achieves a macro F1-score of 69%, which is reasonable for a six-class problem (uniform random guessing yields an expected accuracy of  $\sim 17\%$ ), particularly given the limited dataset size and the minority-class difficulty. In this internal CV evaluation, the models are intentionally left uncalibrated. In the present study, calibration is reserved for the final inference framework and is used as a post-processing step to improve the reliability and interpretability of the predicted class probabilities, i.e., to make the `predict_proba` outputs more closely aligned with empirical confidence levels [44,55]. This is particularly important because these probabilities are used downstream to support the prioritization and selection of candidate compositions for experimental synthesis. Accordingly, calibration is not included in the model-development evaluation protocol used to estimate the internal metrics reported here, which instead quantify the discrimination and classification performance of the native outputs. The ensemble prediction is computed as the average of the four model probability vectors, where the ELM contributes a one-hot vector with entries 0 or 1.

For high-entropy oxide design, the objective is often to obtain a targeted single phase for a specific application; therefore, class-wise performance is particularly informative. As shown in Table 5, the ensemble attains high F1-scores for Spinel, Fluorite, RockSalt, and Perovskite, a moderate F1-score for Pyrochlore (a low-support class), and a low F1-score for Mixed (the smallest and most heterogeneous class). Thus, across six classes, performance is strong for four classes, intermediate for one class, and weak for the minority Mixed class. The confusion matrix (Fig. 3) provides a graphical summary of this outcome and shows that samples with true label Mixed are misclassified into multiple classes. The dominant source of confusion is Pyrochlore, which likely reflects, at least in part, its low support. Figure S1 in the Supplementary Information presents the confusion matrices for the individual models and shows behavior broadly consistent with the ensemble.

Fig. 4 shows the ROC curves for the individual models and for the ensemble. ROC curves are not reported for the ELM because it does

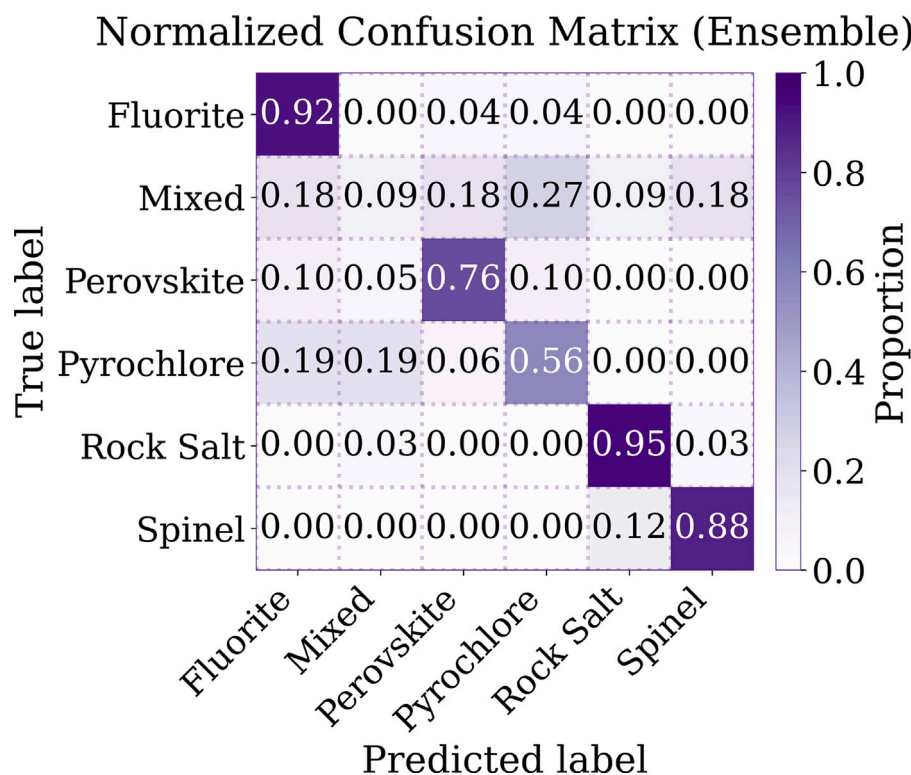


Fig. 3. Row-normalized confusion matrix computed from out-of-fold (OOF) predictions for the ensemble model.

**Table 6**

Cross-validation performance reported as mean  $\pm$  standard deviation across the  $K_{\text{outer}} = 5$  outer folds. Acc = accuracy;  $P_{\text{macro}}$ ,  $R_{\text{macro}}$ ,  $F1_{\text{macro}}$  = macro-averaged precision, recall, and F1;  $P_w$ ,  $R_w$ ,  $F1_w$  = support-weighted (frequency-weighted) precision, recall, and F1.

| Model | Acc             | $P_{\text{macro}}$ | $R_{\text{macro}}$ | $F1_{\text{macro}}$ | $P_w$           | $R_w$           | $F1_w$          |
|-------|-----------------|--------------------|--------------------|---------------------|-----------------|-----------------|-----------------|
| RF    | $0.78 \pm 0.08$ | $0.65 \pm 0.08$    | $0.69 \pm 0.10$    | $0.65 \pm 0.10$     | $0.74 \pm 0.08$ | $0.78 \pm 0.08$ | $0.75 \pm 0.09$ |
| GB    | $0.76 \pm 0.04$ | $0.68 \pm 0.09$    | $0.68 \pm 0.06$    | $0.66 \pm 0.06$     | $0.76 \pm 0.09$ | $0.76 \pm 0.04$ | $0.74 \pm 0.05$ |
| MLP   | $0.79 \pm 0.08$ | $0.68 \pm 0.08$    | $0.70 \pm 0.09$    | $0.68 \pm 0.08$     | $0.76 \pm 0.08$ | $0.79 \pm 0.08$ | $0.76 \pm 0.08$ |
| ELM   | $0.74 \pm 0.08$ | $0.63 \pm 0.08$    | $0.65 \pm 0.09$    | $0.62 \pm 0.08$     | $0.72 \pm 0.08$ | $0.74 \pm 0.08$ | $0.71 \pm 0.08$ |
| Ens.  | $0.79 \pm 0.05$ | $0.69 \pm 0.04$    | $0.70 \pm 0.07$    | $0.68 \pm 0.06$     | $0.77 \pm 0.05$ | $0.79 \pm 0.05$ | $0.77 \pm 0.05$ |

not produce probabilistic outputs and provides only hard predictions. Overall, the random forest achieves the highest class-wise AUC values, whereas gradient boosting yields the highest AUC for the minority Mixed class. This complementarity – where different models capture different class-specific patterns – supports the use of ensemble learning. For the ensemble, the mean AUC across classes is 90% (macro) and 93% (weighted), with all non-minority classes exhibiting AUC values close to 1. In contrast, for the minority Mixed class, the ensemble achieves an AUC of 57%, indicating only a slight improvement over random discrimination. Accordingly, the class-wise AUC values suggest good discrimination for all classes except Mixed, for which performance remains only marginally above chance; the class-wise metrics in Table 5 corroborate this conclusion.

Table 6 reports cross-validation performance as mean  $\pm$  standard deviation across the  $K_{\text{outer}} = 5$  outer folds. The results are broadly consistent with Table 4, while additionally quantifying split-to-split variability. The observed standard deviations (which reach 0.10 for the random forest  $F1_{\text{macro}}$ , for example) indicate substantial split sensitivity and reinforce the suitability of the adopted nested stratified cross-validation protocol for this dataset; in contrast, a single train-test split could yield a misleading estimate of generalization due to sampling variability. A further noteworthy observation is that the ensemble generally exhibits smaller standard deviations than the individual models, supporting the increased robustness typically associated with ensemble learning.

### 3.2. Final ensemble (inference framework)

For the final inference framework, we refit the four base learners on the full dataset using the hyperparameter settings that were selected most frequently as optimal across the outer folds of the nested stratified cross-validation procedure. The final feature set was obtained by applying the Pearson correlation filter to the full dataset, yielding a fixed set of non-redundant descriptors for deployment. The resulting hyperparameter configurations and the final feature list are reported in the Supplementary Information (Table S6 and Fig. S3).

To improve the reliability of the predicted probabilities, the random forest, gradient boosting, and multilayer perceptron models were calibrated using `CalibratedClassifierCV` from scikit-learn [44] with sigmoid calibration (Platt scaling) and internal  $K$ -fold cross-validation ( $K = 5$ ). The final ensemble prediction was then obtained by uniformly averaging the class-probability vectors from the four constituent models, such that each model contributed an equal weight of 1/4. This approach corresponds to a soft-voting ensemble classifier [48].

#### 3.2.1. External validation

Internal validation strategies, such as cross-validation, are widely used to estimate predictive performance while reducing the bias associated with a single hold-out split. Nevertheless, because these evaluations are still performed on data drawn from the same population used

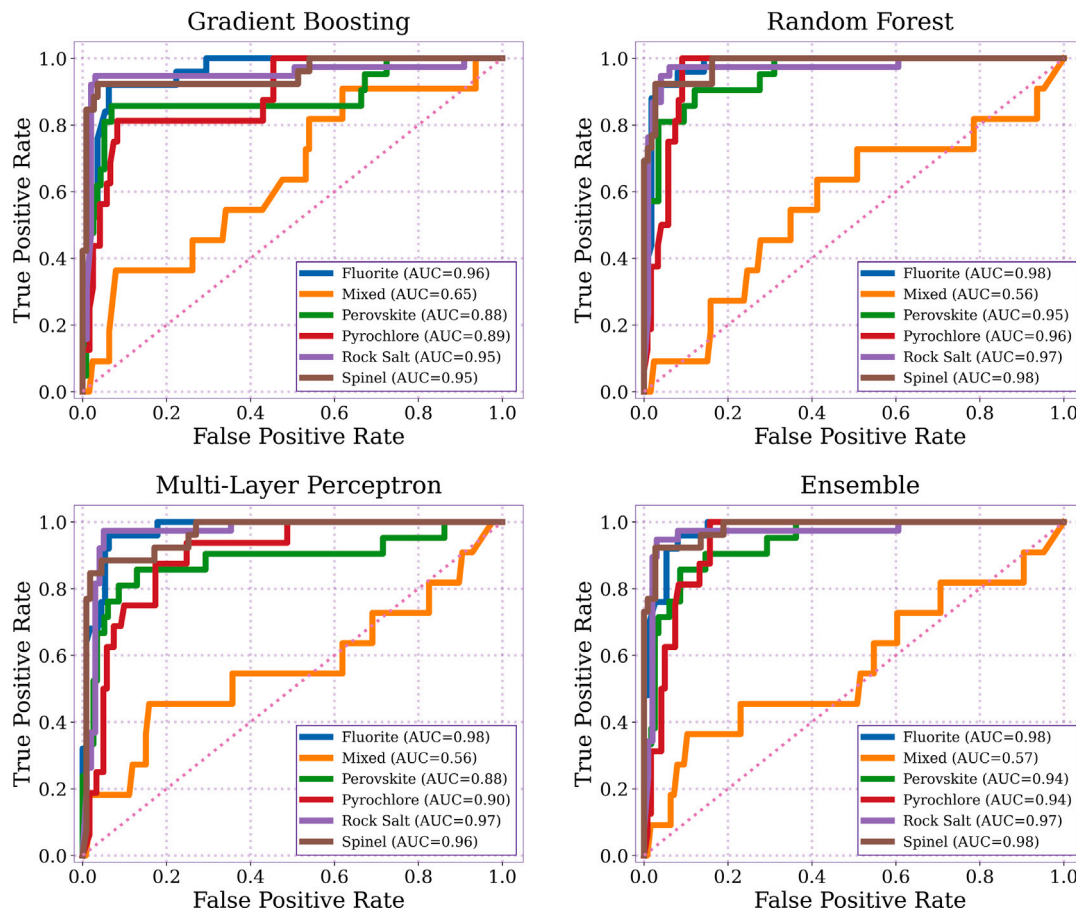


Fig. 4. One-vs-rest ROC curves (and corresponding AUC) computed from OOF predicted probabilities for the individual models and the ensemble.

for model development, the resulting metrics can remain optimistic and may not fully reflect real-world generalization [56].

External validation addresses this limitation by evaluating a model trained on one dataset (study A) using an independent dataset collected in a different study (study B). This setting provides a more stringent test of generalization and is commonly regarded as stronger evidence of out-of-sample performance [57].

To obtain a rigorous estimate of generalization for our final calibrated soft-voting ensemble (constructed from the four base learners), we evaluated it on the dataset compiled by Akrami et al. [40], which was also used by Liu et al. [29]. Notably, the Akrami dataset contains only three crystal-structure classes (fluorite, rock salt, and spinel), whereas our models were trained to discriminate among six classes.

To ensure a *strictly* external evaluation, we first removed from the Akrami dataset all compositions that also appear in the Leong dataset [36]. This de-duplication step yields an external test set composed exclusively of previously unseen compositions. After removing the overlaps, the external set comprised 29 samples with representation across the three available phases. On this fully independent set, the ensemble achieved an overall accuracy of 66%. This performance indicates a reasonable level of generalization to new compositions (for reference, uniform random guessing over six classes would yield an expected accuracy of ~17%). Although the external set spans only a subset of the phase space considered during training, the results suggest that the ensemble learned discriminative patterns that transfer to unseen compositions.

Table 7 reports the overall and class-wise metrics, with F1-scores of 70% or higher for each of the three classes in the external set. Fig. 5 shows the one-vs-rest ROC curves for the same classes, for which the minimum AUC is 0.84. Taken together, these class-wise F1-scores

and AUC values under *strict* external validation provide evidence that the ensemble retains robust predictive performance beyond the training distribution.

More generally, performance estimates can be inflated when evaluation data overlap with, or are processed jointly with, the data used for model fitting. For this reason, strict external validation typically provides a more realistic estimate of real-world performance. Liu et al. [29], for instance, trained and evaluated their models on the full Akrami dataset [40] and reported AUC values of 1.00, 0.98, and 0.98 for fluorite, rock salt, and spinel, respectively, along with an overall F1-score of 98%. Although these values are numerically higher than those obtained here, direct comparison is limited by differences in both task definition and evaluation protocol. In particular, the Akrami benchmark involves three classes, whereas our framework is trained to discriminate among six phases, which is a more challenging classification setting and can lead to lower aggregate metrics even when the underlying discriminative patterns are similar. In addition, differences in evaluation protocol can substantially inflate performance estimates in small-sample settings. For example, the protocol described by Liu et al. [29] includes applying Adaptive Synthetic Sampling (ADASYN) to the full dataset prior to defining an independent test split. This procedure may inflate reported performance in small-sample settings because synthetic samples are generated using nearest-neighbor relationships derived from the overall data distribution and may indirectly introduce information from the eventual test set into the training stage. Furthermore, the absence of results on a separate, de-duplicated external test set limits the conclusions that can be drawn about generalization to truly unseen compositions.

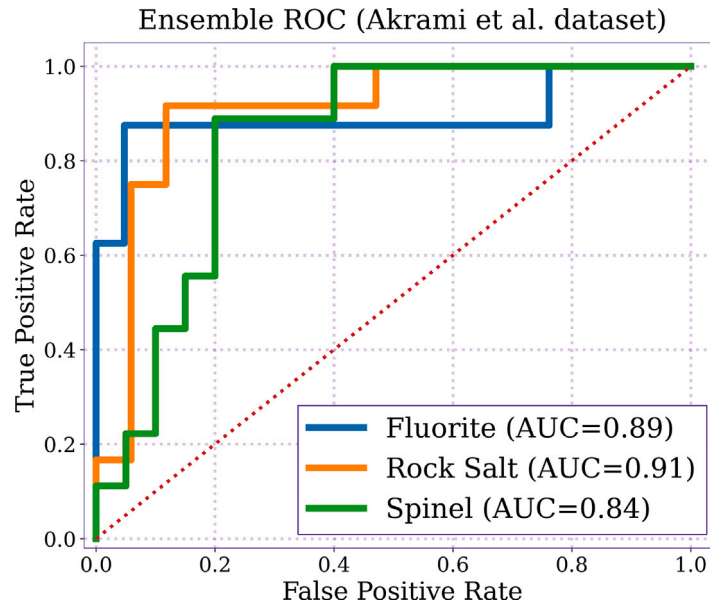


Fig. 5. One-vs-rest ROC curves for the final ensemble on the strictly external, de-duplicated Akrami test set [40].

Table 7

Overall and class-wise performance metrics of the final ensemble on a strictly external, de-duplicated test set derived from Akrami et al. [40]. P, R, and F1 denote precision, recall, and F1-score, respectively. Support indicates the number of samples.

| Item       | P     | R     | F1    | Support |
|------------|-------|-------|-------|---------|
| Overall    |       |       |       |         |
| Weighted   | 0.789 | 0.655 | 0.704 | 29      |
| Class-wise |       |       |       |         |
| Fluorite   | 0.833 | 0.625 | 0.714 | 8       |
| Rock salt  | 0.875 | 0.583 | 0.700 | 12      |
| Spinel     | 0.636 | 0.778 | 0.700 | 9       |

Table 8

Final ensemble predictions for compositions classified as fluorite, including the fluorite class probability and the second-ranked phase (with its probability).

| Composition                                                                                      | Inference | $P(\text{Fluorite})$ | Second phase | $P(\text{second})$ |
|--------------------------------------------------------------------------------------------------|-----------|----------------------|--------------|--------------------|
| $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{Zr}_{0.2}\text{O}_{2-\delta}$ | Fluorite  | 0.559                | Rock salt    | 0.106              |
| $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{Zn}_{0.2}\text{O}_{2-\delta}$ | Fluorite  | 0.540                | Rock salt    | 0.165              |
| $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{Ti}_{0.2}\text{O}_{2-\delta}$ | Fluorite  | 0.520                | Perovskite   | 0.149              |
| $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{Co}_{0.2}\text{O}_{2-\delta}$ | Fluorite  | 0.510                | Rock salt    | 0.119              |
| $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{Pd}_{0.2}\text{O}_{2-\delta}$ | Fluorite  | 0.493                | Rock salt    | 0.134              |
| $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{Fe}_{0.2}\text{O}_{2-\delta}$ | Fluorite  | 0.455                | Perovskite   | 0.119              |
| $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{Ba}_{0.2}\text{O}_{2-\delta}$ | Fluorite  | 0.439                | Perovskite   | 0.174              |
| $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{Ga}_{0.2}\text{O}_{2-\delta}$ | Fluorite  | 0.402                | Perovskite   | 0.195              |
| $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{Ca}_{0.2}\text{O}_{2-\delta}$ | Fluorite  | 0.380                | Perovskite   | 0.245              |
| $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{Ge}_{0.2}\text{O}_{2-\delta}$ | Fluorite  | 0.379                | Perovskite   | 0.214              |
| $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{In}_{0.2}\text{O}_{2-\delta}$ | Fluorite  | 0.377                | Perovskite   | 0.253              |
| $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{Al}_{0.2}\text{O}_{2-\delta}$ | Fluorite  | 0.367                | Perovskite   | 0.199              |
| $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{Li}_{0.2}\text{O}_{2-\delta}$ | Fluorite  | 0.314                | Perovskite   | 0.300              |

### 3.3. Cation combinatorial analysis, synthesis, and characterization

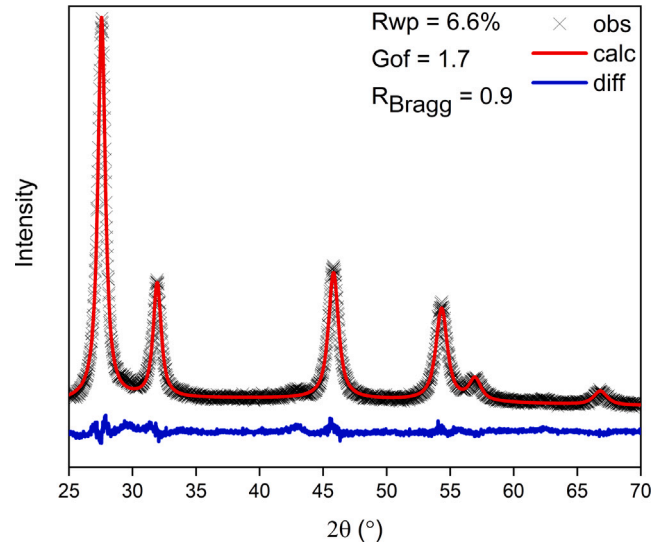
Cerium-based oxides are widely investigated for renewable-energy technologies, particularly as functional components in solid oxide fuel cells operating at intermediate temperatures [49]. Motivated by this relevance, we constructed a constrained combinatorial design space in which the cerium cation was fixed. To keep the inference space tractable and to satisfy the minimum compositional requirement for classification as a high-entropy oxide, we restricted the search to equimolar compositions containing exactly five cations. In addition, to avoid trivially inheriting fluorite behavior from the  $\text{CeO}_2$  end member and to pose a more stringent test for phase prediction, we fixed three additional cations whose binary oxides adopt non-fluorite structures. Importantly, these three fixed cations (besides Ce) were selected at random from a pool of ~40 experimentally accessible oxides. Specifically,  $\text{La}_2\text{O}_3$  and  $\text{Nd}_2\text{O}_3$  (hexagonal) and  $\text{MgO}$  (rocksalt) were included as fixed constituents, while the fifth cation was selected from the same pool of ~40 experimentally accessible oxides.

Using the combinatorial space defined above, the soft-voting Ensemble classifier described in Section 3.2 was applied to generate phase predictions for all hypothetical compositions. Because fluorite-type structures are frequently targeted in oxide systems (e.g., in solid-electrolyte

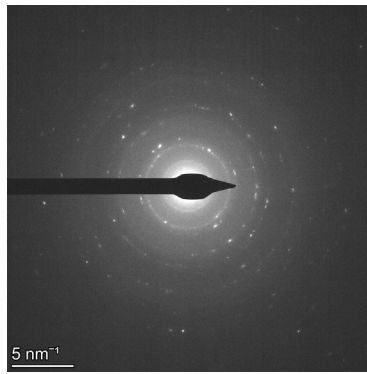
and electrostrictive contexts), we focus on candidates inferred as fluorite [49,58]. Table 8 summarizes the compositions selected for experimental consideration, reporting the Ensemble probability assigned to the fluorite class as well as the second most probable phase and its associated probability.

For practical reasons, we report here the synthesis of  $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{Al}_{0.2}\text{O}_{2-\delta}$ . This composition was initially selected because an earlier version of the Ensemble model ranked it as the highest-probability fluorite candidate at the time the experimental work was initiated. In addition to its predicted fluorite likelihood, the choice of an Al-containing composition was also guided by practical considerations, as Al precursors are typically more cost-effective than several alternatives listed in Table 8. While data-driven screening cannot guarantee phase formation with 100% success, the Ensemble provides an efficient, low-cost strategy expected to outperform unguided trial-and-error exploration, while avoiding the prohibitive computational expense of large-scale *ab initio* screening across extensive compositional spaces.

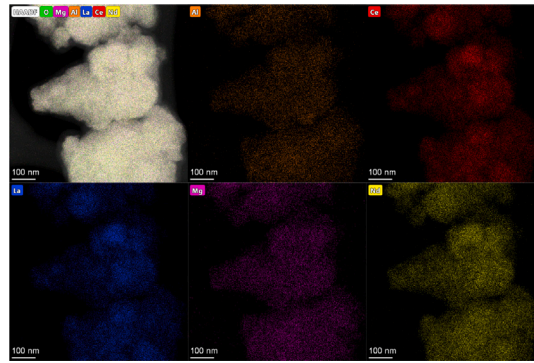
To determine the most plausible crystal structure for the synthesized sample, we first performed a Search-Match routine using the EXPO program [59], which suggested a defective-fluorite prototype.



(a) XRD pattern and Rietveld refinement fit.



(b) SAED pattern.



(c) TEM-EDS elemental maps and overlay.

**Fig. 6.** Structural and microstructural characterization of  $\text{Ce}_{0.2}\text{La}_{0.2}\text{Nd}_{0.2}\text{Mg}_{0.2}\text{Al}_{0.2}\text{O}_{2-\delta}$ : (A) XRD pattern with Rietveld refinement, (B) SAED pattern, and (C) TEM-EDS elemental mapping.

Subsequently, Rietveld refinement was carried out using TOPAS-7 Academic. Fig. 6a shows the observed and refined XRD profiles along with the refinement indicators  $R_{\text{wp}}$  (weighted profile  $R$ -factor),  $\text{GoF}$  (goodness of fit), and  $R_{\text{Bragg}}$  (Bragg  $R$ -factor). The obtained values fall within commonly adopted quality criteria ( $R_{\text{wp}} < 20\%$ ,  $\text{GoF} < 5$ , and  $R_{\text{Bragg}} < 3$ ) [60], supporting a reliable assignment to a defective-fluorite structure. The refinement further indicates substantial lattice distortion. Both oxygen-vacancy defects (as suggested by the Search-Match result) and pronounced lattice distortion are consistent with expectations for high-entropy oxides [61], arising from the coexistence of cations with disparate ionic sizes and charges. In particular, size mismatch promotes lattice strain, while charge mismatch is accommodated through oxygen vacancies to maintain electroneutrality.

In Fig. 6c, the TEM-EDS elemental maps (Al: orange, Ce: red, La: blue, Mg: magenta, and Nd: yellow) show a uniform spatial distribution of all constituent cations, with no evidence of segregation at the probed length scale; the composite overlay further emphasizes this homogeneity. Selected-area electron diffraction (SAED) (Fig. 6b) independently corroborates the fluorite-type assignment: the pattern exhibits Debye-Scherrer rings typical of polycrystalline materials, and the extracted interplanar spacings ( $d$ -spacings) are reproducible and can be indexed to fluorite reflections. As summarized in Table 9, the measured  $d$ -spacings closely match the reference values for fluorite  $\text{CeO}_2$  obtained from the ICSD entry provided as Crystallographic Information File (CIF) 24887, supporting the phase identification. In addition, the presence

**Table 9**

Experimental  $1/R$  values and derived interplanar spacings ( $d_{hkl}$ ), compared with reference  $d_{hkl}$  values from CIF 24887.

| Crystallographic plane | $1/R$ ( $\text{nm}^{-1}$ ) | $d_{hkl}$ ( $\text{\AA}$ ) | $d_{hkl}$ CIF 24887 ( $\text{\AA}$ ) |
|------------------------|----------------------------|----------------------------|--------------------------------------|
| (111)                  | 3.35975                    | 2.97641                    | 3.127968                             |
| (200)                  | 3.86636                    | 2.58641                    | 2.70890                              |
| (220)                  | 5.49441                    | 1.82003                    | 1.915482                             |
| (311)                  | 6.43359                    | 1.55434                    | 1.633528                             |

of parallel lattice fringes further indicates the high crystallinity of the synthesized powder.

### 3.4. Insights into the phase-formation phenomenon in HEOs

To gain mechanistic insight into the ensemble classifier and quantify how each descriptor contributes to phase discrimination, we employed SHapley Additive exPlanations, a model-agnostic interpretability framework [62]. Fig. 7 summarizes the five most influential descriptors, showing (a) the global mean absolute SHAP values (average feature importance across the dataset) and (b) a layered violin (summary) plot for the fluorite class, which relates feature magnitude (low vs. high values) to both the direction and the strength of the model output. Precise definitions of the descriptors are provided in Table 3.

As shown in Fig. 7a, electronegativity-related descriptors are among the dominant drivers of phase discrimination, consistent with prior reports [29,36,63]. Whereas these earlier studies often rely on aggregation schemes based on electronegativity differences, our results indicate that alternative composition-to-feature constructions can also yield strong discriminative performance. In addition, although Pauling and Mulliken electronegativities are frequently adopted in high-entropy materials informatics, we find that electron affinity and the Allen and Allred-Rochow electronegativity scales also provide substantial predictive signal across multiple phase classes. Although different electronegativity scales may exhibit similar qualitative trends, their predictive capabilities can differ substantially [64]. Notably, the minimum Allred-Rochow electronegativity exhibits strong ability to discriminate among all six phases, suggesting that the lower tail of the constituent electronegativity distribution may encode chemically meaningful constraints on bonding preferences and phase stability. From a physical standpoint, electronegativity is closely related to bond character and charge-transfer propensity and is therefore expected to correlate with the relative stability of competing crystal structures in compositionally complex oxides [63].

Fig. 7a also highlights the minimum atomization enthalpy as an influential descriptor, with particular relevance for separating the rock salt class from the remaining phases. Atomization enthalpy reflects the energetic cost of disrupting atomic bonding and thus correlates with both the number and the strength of bonds that an element can form. Its prominence is consistent with the expectation that cohesion and bond energetics influence which oxide frameworks are stabilized under high configurational disorder.

The layered violin summary in Fig. 7b further indicates that, for fluorite classification, one of the strongest drivers is the sum of electron affinity. This trend aligns with the conclusions of Syarif et al. [65], who identified electron affinity as a key variable governing phase formation in high-entropy alloys using a machine-learning-based analysis. In addition, the relevance of an electronegativity-dispersion descriptor (e.g., the standard deviation of electronegativity) for fluorite is broadly consistent with the findings of Liu et al. [29], where electronegativity differences were reported as central descriptors for fluorite classification.

The importance of electron-affinity- and electronegativity-related descriptors for fluorite can be rationalized by their close connection to polarization and charge-transfer tendencies. In ionic ceramics, variations in polarization and local charge distribution can strongly affect geometric preferences and, consequently, structural stability [29]. Moreover, HEOs often deviate from simple average-site pictures; instead, local environments and the chemistry of participating ions can play an outsized role in determining the adopted structure [66]. This perspective is particularly relevant when interpreting the fluorite phase, which contrasts with more compositionally flexible oxide frameworks such as perovskites [67]. In such structures – and likewise in spinel and pyrochlore – multiple crystallographically distinct cation sublattices are available, which can accommodate chemical complexity through site partitioning and thus broaden the accessible composition space. In fluorite-type HEOs, by contrast, random cation mixing occurs on a single cation site, and stabilization is favored when cations can randomly share a single site while avoiding short-range cation-anion repulsion and anion-anion repulsion (double repulsion), i.e., by balancing ionic sizes and local electrostatics [68,69]. The maximum elemental space-group number in the composition (`atomic_spacegroupnum_max`) also ranks among the most relevant descriptors for fluorite. Although this feature is an empirical proxy derived from the stable elemental solids, it may capture whether at least one constituent tends to adopt high-symmetry packing and coordination environments, which can be associated with the stabilization of fluorite-type structures.

Finally, we emphasize that the summary visualization in Fig. 7b does not explicitly represent feature-feature interactions. Such interactions are captured by the underlying machine-learning model, whereas

SHAP summary plots primarily report associations between descriptor magnitudes and model output. Therefore, although SHAP is valuable for generating physically motivated hypotheses regarding phase-formation trends, these results should not be interpreted as direct causal rules that fully explain the phenomenon.

#### 4. Conclusion

The results of this study support the premise that machine-learning models, by extracting latent relationships from patterns in experimental datasets, can enable efficient exploration of the vast compositional space of high-entropy oxides. From a modeling perspective, the proposed soft-voting ensemble achieves solid performance on a six-class phase-classification task, as evaluated using a leakage-safe nested stratified cross-validation protocol and an external validation dataset (F1-scores of 77% and 70%, respectively). In addition, the SHAP analysis suggests that the learned decision rules are physically meaningful. Notably, the minimum Allred-Rochow electronegativity emerges as a key descriptor, exerting a strong influence on ensemble predictions across all six phase classes.

From an experimental perspective, we successfully synthesized a fluorite HEO composition predicted by the framework and confirmed its phase by X-ray diffraction with Rietveld refinement and transmission electron microscopy. Overall, these combined theoretical and experimental results indicate that machine-learning-assisted design can increase the likelihood of obtaining targeted HEO phases and can provide interpretable insights into structure-composition relationships through feature-importance analysis.

Despite the encouraging overall results, some limitations should be noted. In particular, performance for the Mixed class remained comparatively weaker than for the other classes, indicating that this category is still more challenging for the proposed framework. This may be partly explained by both the way the labels were constructed and the fact that Mixed is the minority class in the dataset. In addition, the external test set was relatively small, which limits the strength of generalizations regarding real-world performance.

#### CRedit authorship contribution statement

**Arthur da Silva Sousa Santos:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **James Moraes de Almeida:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Fabiane de Jesus Trindade:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Marissol Rodrigues Felez:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Methodology, Investigation, Formal analysis, Data curation. **Larissa da Silva Marques:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation. **Daniel Zanetti de Florio:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

#### Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used an AI-based language tool to improve clarity and readability. After using this tool, the author(s) reviewed and edited the content as needed and take full responsibility for the content of the publication.

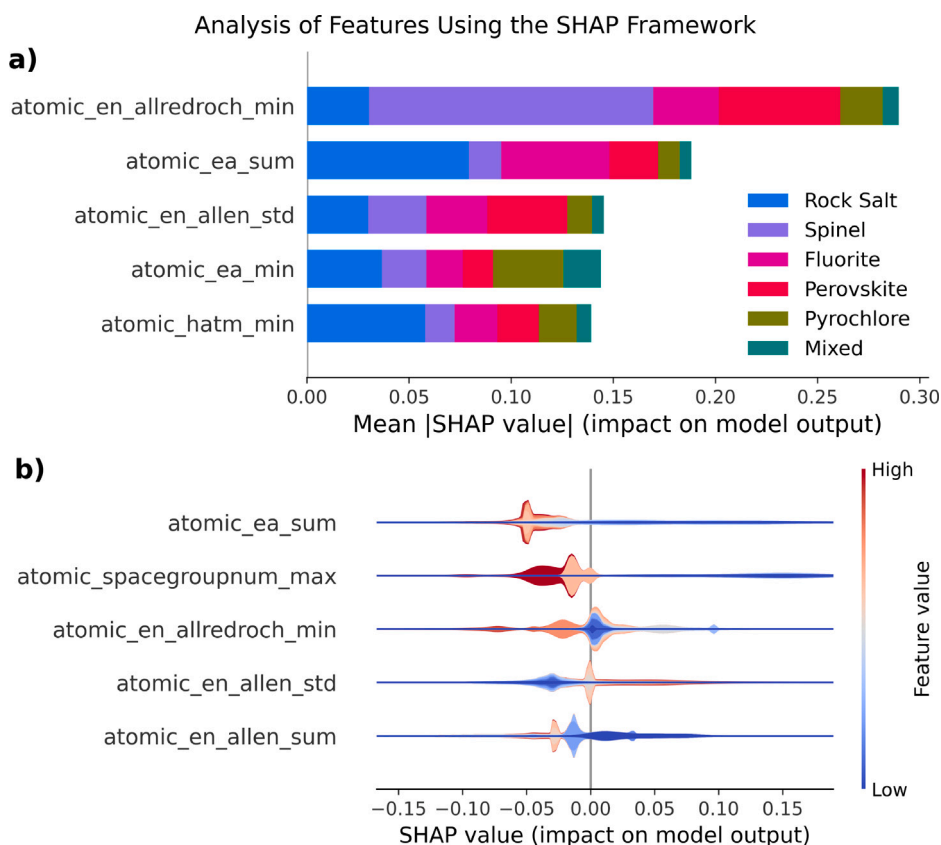


**Fig. 7.** SHAP analysis of the five most influential descriptors. (a) Global mean absolute SHAP values. (b) Layered violin summary plot for the fluorite class, relating descriptor values (low to high) to the direction and magnitude of the model output.

### Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgments

The authors acknowledge financial support from the São Paulo Research Foundation (FAPESP), the Coordination for the Improvement of Higher Education Personnel (CAPES), the National Council for Scientific and Technological Development (CNPq), and the National Institute of Science and Technology in Materials Informatics (INCT Materials Informatics). The authors also thank the Multiuser Computational Center (CCM-UFABC), Federal University of ABC (UFABC), for providing computational resources. A.S.S.S. acknowledges a FAPESP scholarship (Grant No. 2025/06024-6). J.M.A. acknowledges support from FAPESP research grants (Grant Nos. 2023/09820-2 and 2025/09430-5). D.Z.F. and M.R.F. acknowledge support from M-ERA-NET/FAPESP (Grant No. 2022/09589-6). D.Z.F. also acknowledges support from the National Institute of Science and Technology in Rare-Earth-Based Advanced Materials: Innovations and Applications (INCT MATERIA) and from a FAPESP research grant (Grant No. 2023/14931-8). L.S.M. acknowledges a FAPESP scholarship (Grant No. 2024/13678-0).

### Appendix A. Supplementary data

The following files are available free of charge.

- Supplementary.pdf: Supplementary tables and figures providing detailed information on the machine-learning models trained in this study.

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.oceram.2026.100982>.

### Data availability

To support reproducibility and future extensions of this work, the Python scripts, datasets, and trained models are publicly available on GitHub under the open-source MIT License: <https://github.com/Arthurs16/Machine-Learning-Assisted-Prediction-and-Synthesis-of-New-Single-Phase-High-Entropy-Oxides>. The composition-phase pairs used to construct the dataset were sourced from Leong *et al.* [36] (DOI: <https://doi.org/10.3390/jcs5120311>).

### References

- [1] H. Xiang, Y. Xing, F.-z. Dai, H. Wang, L. Su, L. Miao, G. Zhang, Y. Wang, X. Qi, L. Yao, H. Wang, B. Zhao, J. Li, Y. Zhou, High-entropy ceramics: Present status, challenges, and a look forward, *J. Adv. Ceram.* 10 (3) (2021) 385–441, <http://dx.doi.org/10.1007/s40145-021-0477-y>.
- [2] Y. Jiao, J. Dai, Z. Fan, J. Cheng, G. Zheng, L. Grema, J. Zhong, H.-F. Li, D. Wang, Overview of high-entropy oxide ceramics, *Mater. Today* 77 (2024) 92–117, <http://dx.doi.org/10.1016/j.mattod.2024.06.005>.
- [3] G.M. Tomboc, X. Zhang, S. Choi, D. Kim, L.Y.S. Lee, K. Lee, Stabilization, characterization, and electrochemical applications of high-entropy oxides: Critical assessment of crystal phase-properties relationship, *Adv. Funct. Mater.* 32 (43) (2022) <http://dx.doi.org/10.1002/adfm.202205142>.
- [4] S.S. Aamlid, M. Oudah, J. Rottler, A.M. Hallas, Understanding the role of entropy in high entropy oxides, *J. Am. Chem. Soc.* 145 (11) (2023) 5991–6006, <http://dx.doi.org/10.1021/jacs.2c11608>.
- [5] J. Zhang, X. Xiang, B. Xu, S. Huang, Y. Xiong, S. Ma, H. Fu, Y. Ma, H. Chen, Z. Wu, S. Zhao, Rational design of high-entropy ceramics based on machine learning – a critical review, *Curr. Opin. Solid State Mater. Sci.* 27 (2023) 101057, <http://dx.doi.org/10.1016/j.COSSMS.2023.101057>.

- [6] Z. Yang, X. Xiang, J. Yang, Z.-Y. Zhao, High-entropy oxides as energy materials: from complexity to rational design, *Mater. Futur.* 3 (4) (2024) 042103, <http://dx.doi.org/10.1088/2752-5724/ad8463>.
- [7] C. Liu, S. Li, Y. Zheng, M. Xu, H. Su, X. Miao, Y. Liu, Z. Zhou, J. Qi, B. Yang, D. Chen, C.-W. Nan, Y.-H. Lin, Advances in high entropy oxides: synthesis, structure, properties and beyond, *Prog. Mater. Sci.* 148 (2025) 101385, <http://dx.doi.org/10.1016/j.pmatsci.2024.101385>.
- [8] S. Dong, R. Ren, J. Zhang, X. Bao, X. Liu, Q. Shi, Z. Chen, H. Shao, High-entropy oxides: Emergent materials for electrochemical energy storage and conversion, *J. Mater. Sci. Technol.* 227 (2025) 192–215, <http://dx.doi.org/10.1016/j.jmst.2024.11.065>.
- [9] A. Sarkar, L. Velasco, D. Wang, Q. Wang, G. Talasila, L. de Biasi, C. Kübel, T. Brezesinski, S.S. Bhattacharya, H. Hahn, B. Breitung, High entropy oxides for reversible energy storage, *Nat. Commun.* 2018 9:1 9 (2018) 1–9, <http://dx.doi.org/10.1038/s41467-018-05774-5>, URL <https://www.nature.com/articles/s41467-018-05774-5>.
- [10] A. Sarkar, Q. Wang, A. Schiele, M.R. Chellali, S.S. Bhattacharya, D. Wang, T. Brezesinski, H. Hahn, L. Velasco, B. Breitung, High-entropy oxides: Fundamental aspects and electrochemical properties, *Adv. Mater.* 31 (2019) 1970189, <http://dx.doi.org/10.1002/ADMA.201970189>, URL <https://onlinelibrary.wiley.com/doi/full/10.1002/adma.201970189>.
- [11] M. Anandkumar, E. Trofimov, Synthesis, properties, and applications of high-entropy oxide ceramics: Current progress and future perspectives, *J. Alloys Compd.* 960 (2023) 170690, <http://dx.doi.org/10.1016/j.jallcom.2023.170690>.
- [12] C. Oses, C. Toher, S. Curtarolo, High-entropy ceramics, *Nat. Rev. Mater.* 5 (4) (2020) 295–309, <http://dx.doi.org/10.1038/s41578-019-0170-8>.
- [13] R.Z. Zhang, M.J. Reece, Review of high entropy ceramics: design, synthesis, structure and properties, *J. Mater. Chem. A* 7 (2019) 22148–22162, <http://dx.doi.org/10.1039/C9TA05698J>, URL <https://pubs.rsc.org/en/content/articlehtml/2019/ta/c9ta05698j> <https://pubs.rsc.org/en/content/articlelanding/2019/ta/c9ta05698j>.
- [14] S. Divya, S. Panda, S. Hajira, R. Jeyaraj, A. Paul, S.H. Park, H.J. Kim, T.H. Oh, Smart data processing for energy harvesting systems using artificial intelligence, *Nano Energy* 106 (2023) 108084, <http://dx.doi.org/10.1016/J.NANOEN.2022.108084>.
- [15] L.E.B. Lucchetti, J.M. de Almeida, S. Siahrostami, Revolutionizing ORR catalyst design through computational methodologies and materials informatics, *EES Catal.* (Accepted Manuscript) (2024).
- [16] K. Kaufmann, D. Maryanovsky, W.M. Mellor, C. Zhu, A.S. Rosengarten, T.J. Harrington, C. Oses, C. Toher, S. Curtarolo, K.S. Vecchio, Discovery of high-entropy ceramics via machine learning, *Npj Comput. Mater.* 2020 6:1 6 (2020) 1–9, <http://dx.doi.org/10.1038/s41524-020-0317-6>, URL <https://www.nature.com/articles/s41524-020-0317-6>.
- [17] J. Zhang, B. Xu, Y. Xiong, S. Ma, Z. Wang, Z. Wu, S. Zhao, Design high-entropy carbide ceramics from machine learning, *Npj Comput. Mater.* 2022 8:1 8 (2022) 1–12, <http://dx.doi.org/10.1038/s41524-021-00678-3>, URL <https://www.nature.com/articles/s41524-021-00678-3>.
- [18] R. Mitra, A. Bajpai, K. Biswas, Machine learning based approach for phase prediction in high entropy borides, *Ceram. Int.* 48 (2022) 16695–16706, <http://dx.doi.org/10.1016/J.CERAMINT.2022.02.218>.
- [19] C.J. Leverant, J.A. Harvey, Accelerating the discovery of new, single phase high entropy ceramics via active learning, *Chem. Mater.* 36 (22) (2024) 10994–11003, <http://dx.doi.org/10.1021/acs.chemmater.4c00303>.
- [20] W. Guo, L. Wang, L. Zhu, Y. Ye, Z. Zhang, B. Yang, S. Zhang, S. Bai, Accelerated discovery and formation mechanism of high-entropy carbide ceramics using machine learning based on low-cost descriptors, *J. Alloys Compd.* 1004 (2024) 175929, <http://dx.doi.org/10.1016/j.jallcom.2024.175929>.
- [21] R. Mitra, A. Gupta, K. Biswas, A simple recipe for designing multicomponent ultra-high temperature ceramic classes by using structure maps coupled with machine learning, *J. Appl. Phys.* 136 (4) (2024) <http://dx.doi.org/10.1063/5.0200666>.
- [22] Y. Yan, Z. Pei, M.C. Gao, S. Mixture, K. Wang, Data-driven discovery of a formation prediction rule on high-entropy ceramics, *Acta Mater.* 253 (2023) 118955, <http://dx.doi.org/10.1016/j.actamat.2023.118955>.
- [23] R. Mitra, A. Bajpai, K. Biswas, ADASYN-assisted machine learning for phase prediction of high entropy carbides, *Comput. Mater. Sci.* 223 (2023) 112142, <http://dx.doi.org/10.1016/j.commatsci.2023.112142>.
- [24] Q. Zhou, F. Xu, C. Gao, D. Zhang, X. Shi, M.-F. Yuen, D. Zuo, Machine learning-assisted mechanical property prediction and descriptor-property correlation analysis of high-entropy ceramics, *Ceram. Int.* 49 (4) (2023) 5760–5769, <http://dx.doi.org/10.1016/j.ceramint.2022.10.105>.
- [25] W. Wen, X. Yan, X. Pei, Q. Li, T. Jiang, J. Cao, W. Wang, M. Zhang, Multi-dimensional anisotropic feature interaction with machine learning to predict the thermal conductivity of A2B2O7-type high-entropy ceramics, *Ceram. Int.* 51 (13) (2025) 17860–17869, <http://dx.doi.org/10.1016/j.ceramint.2025.01.556>.
- [26] V.A. Mints, K.L. Svane, J. Rossmeisl, M. Arenz, Exploring the high-entropy oxide composition space: Insights through comparing experimental with theoretical models for the oxygen evolution reaction, *ACS Catal.* 14 (9) (2024) 6936–6944, <http://dx.doi.org/10.1021/acscatal.3c05915>.
- [27] S. Nie, Y. Xiang, L. Wu, G. Lin, Q. Liu, S. Chu, X. Wang, Active learning guided discovery of high entropy oxides featuring high H<sub>2</sub>-production, *J. Am. Chem. Soc.* 146 (43) (2024) 29325–29334, <http://dx.doi.org/10.1021/jacs.4c06272>.
- [28] C.-C. Lin, C.-W. Chang, C.-C. Kaun, Y.-H. Su, Stepwise evolution of photocatalytic spinel-structured (Co, Cr, Fe, Mn, Ni)<sub>3</sub>O<sub>4</sub> high entropy oxides from first-principles calculations to machine learning, *Crystals* 11 (9) (2021) 1035, <http://dx.doi.org/10.3390/cryst11091035>.
- [29] J. Liu, A. Wang, P. Gao, R. Bai, J. Liu, B. Du, C. Fang, Machine learning-based crystal structure prediction for high-entropy oxide ceramics, *J. Am. Ceram. Soc.* 107 (2) (2023) 1361–1371, <http://dx.doi.org/10.1111/jace.19518>.
- [30] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, Machine learning in materials informatics: recent applications and prospects, *Npj Comput. Mater.* 2017 3:1 3 (2017) 1–13, <http://dx.doi.org/10.1038/s41524-017-0056-5>, URL <https://www.nature.com/articles/s41524-017-0056-5>.
- [31] D. Morgan, R. Jacobs, Opportunities and challenges for machine learning in materials science, *Annu. Rev. Mater. Res.* 50 (1) (2020) 71–103, <http://dx.doi.org/10.1146/annurev-matsci-070218-010015>.
- [32] M. Scheidgen, L. Himanen, A.N. Lladines, D. Sikter, M. Nakhaee, Á. Fekete, T. Chang, A. Golparvar, J.A. Márquez, S. Brockhauser, S. Brückner, L.M. Ghiringhelli, F. Dietrich, D. Lehmberg, T. Denell, A. Albino, H. Näsström, S. Shabih, F. Dobener, M. Kühbach, R. Mozumder, J.F. Rudzinski, N. Daelman, J.M. Pizarro, M. Kuban, C. Salazar, P. Ondračka, H.-J. Bungartz, C. Draxl, NOMAD: A distributed web-based platform for managing materials science research data, *J. Open Source Softw.* 8 (2023) 5388, <http://dx.doi.org/10.21105/JOSS.05388>, URL <https://joss.theoj.org/papers/10.21105/joss.05388>.
- [33] C.W. Bale, E. Bélsisle, P. Chartrand, S.A. Decterov, G. Eriksson, A.E. Gheribi, K. Hack, I.H. Jung, Y.B. Kang, J. Melançon, A.D. Pelton, S. Petersen, C. Robelin, J. Sangster, P. Spencer, M.A.V. Ende, FactSage thermochemical software and databases, 2010–2016, *Calphad* 54 (2016) 35–53, <http://dx.doi.org/10.1016/J.CALPHAD.2016.05.002>.
- [34] L. Mentel, mendeleeev – a python resource for properties of chemical elements, ions and isotopes, 2014–, URL <https://github.com/lmmmentel/mendeleeev>.
- [35] O. Isayev, C. Oses, C. Toher, E. Gossett, S. Curtarolo, A. Tropsha, Universal fragment descriptors for predicting properties of inorganic crystals, 2017, <http://dx.doi.org/10.1038/ncomms15679>.
- [36] Z. Leong, P. Desai, N. Morley, Can empirical biplots predict high entropy oxide phases? *J. Compos. Sci.* 5 (12) (2021) 311, <http://dx.doi.org/10.3390/jcs5120311>.
- [37] C. Deng, Z. Han, P. Wu, H. Zhu, R. Liu, F. Liu, H. Ji, Y. She, W. Zhu, C. Xu, Enthalpy-driven synthesis of magnetic high entropy oxide catalyst for efficient oxidative desulfurization and high-value product recovery, *Sep. Purif. Technol.* 330 (2024) 125403, <http://dx.doi.org/10.1016/j.seppur.2023.125403>.
- [38] N. Qu, Y. Liu, Y. Zhang, D. Yang, T. Han, M. Liao, Z. Lai, J. Zhu, L. Zhang, Machine learning guided phase formation prediction of high entropy alloys, *Mater. Today Commun.* 32 (2022) 104146, <http://dx.doi.org/10.1016/j.jmtcomm.2022.104146>.
- [39] A. Takeuchi, A. Inoue, Mixing enthalpy of liquid phase calculated by miedema's scheme and approximated with sub-regular solution model for assessing forming ability of amorphous and glassy alloys, *Intermetallics* 18 (9) (2010) 1779–1789, <http://dx.doi.org/10.1016/j.intermet.2010.06.003>.
- [40] S. Akrami, P. Edalati, M. Fuji, K. Edalati, High-entropy ceramics: Review of principles, production and applications, *Mater. Sci. Eng.: R* 146 (2021) 100644, <http://dx.doi.org/10.1016/j.mser.2021.100644>.
- [41] A. Sarkar, C. Loh, L. Velasco, T. Thomas, S.S. Bhattacharya, H. Hahn, R. Djennadi, Multicomponent equiatomic rare earth oxides with a narrow band gap and associated praseodymium multivalency, *Dalton Trans.* 46 (2017) 12167–12176, <http://dx.doi.org/10.1039/C7DT02077E>, URL <https://pubs.rsc.org/en/content/articlehtml/2017/dt/c7dt02077e>.
- [42] A.J. Wright, C. Huang, M.J. Walock, A. Ghoshal, M. Murugan, J. Luo, Sand corrosion, thermal expansion, and ablation of medium- and high-entropy compositionally complex fluorite oxides, *J. Am. Ceram. Soc.* 104 (1) (2020) 448–462, <http://dx.doi.org/10.1111/jace.17448>.
- [43] J. Wang, Y. Cui, Q. Wang, K. Wang, X. Huang, D. Stenzel, A. Sarkar, R. Azmi, T. Bergfeldt, S.S. Bhattacharya, R. Kruk, H. Hahn, S. Schweidler, T. Brezesinski, B. Breitung, Lithium containing layered high entropy oxide structures, *Sci. Rep.* 10 (1) (2020) <http://dx.doi.org/10.1038/s41598-020-75134-1>.
- [44] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine learning in python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [45] L. Xudong, GitHub - 5663015/elm: Extreme Learning Machine(ELM): Python code, 2018, <https://github.com/5663015/elm>. (Accessed 13 September 2024).
- [46] A. Blanco-Justicia, J. Domingo-Ferrer, Machine learning explainability through comprehensible decision trees, in: *Machine Learning and Knowledge Extraction*, Springer International Publishing, 2019, pp. 15–26, [http://dx.doi.org/10.1007/978-3-030-29726-8\\_2](http://dx.doi.org/10.1007/978-3-030-29726-8_2).
- [47] U.A. Umoh, I.J. Eyoh, V.S. Murugesan, E.E. Nyoho, Fuzzy-machine learning models for the prediction of fire outbreaks: A comparative analysis, in: *Artificial Intelligence and Machine Learning for EDGE Computing*, Elsevier, 2022, pp. 207–233, <http://dx.doi.org/10.1016/b978-0-12-824054-0.00025-3>.

- [48] D. Gautam, A. Bhattarai, R. Rupakhety, Machine learning and soft voting ensemble classification for earthquake induced damage to bridges, *Eng. Struct.* 303 (2024) 117534, <http://dx.doi.org/10.1016/j.engstruct.2024.117534>.
- [49] P. Arunkumar, M. Meena, K.S. Babu, A review on cerium oxide-based electrolytes for ITSOFC, *Nanomater. Energy* 1 (5) (2012) 288–305, <http://dx.doi.org/10.1680/nme.12.00015>.
- [50] L. Velasco, J.S. Castillo, M.V. Kante, J.J. Olaya, P. Friederich, H. Hahn, Phase-property diagrams for multicomponent oxide systems toward materials libraries, *Adv. Mater.* 33 (2021) 2102301, <http://dx.doi.org/10.1002/ADMA.202102301>, URL <https://onlinelibrary.wiley.com/doi/full/10.1002/adma.202102301>.
- [51] F.J. Trindade, S. Damasceno, L. Otubo, M.R. Felez, D.Z. de Florio, F.C. Fonseca, A.S. Ferlauto, Tuning of shape, defects, and disorder in lanthanum-doped ceria nanoparticles: Implications for high-temperature catalysis, *ACS Appl. Nano Mater.* 5 (7) (2022) 8859–8867, <http://dx.doi.org/10.1021/acsanm.2c00942>.
- [52] C.A. Schneider, W.S. Rasband, K.W. Eliceiri, NIH image to ImageJ: 25 years of image analysis, *Nature Methods* 9 (7) (2012) 671–675, <http://dx.doi.org/10.1038/nmeth.2089>.
- [53] L. Sbailo, A. Fekete, L.M. Ghiringhelli, M. Scheffler, The NOMAD artificial-intelligence toolkit: turning materials-science data into knowledge and understanding, *Npj Comput. Mater.* 8 (1) (2022) <http://dx.doi.org/10.1038/s41524-022-00935-z>.
- [54] I.D. Mienye, Y. Sun, A survey of ensemble learning: Concepts, algorithms, applications, and prospects, *IEEE Access* 10 (2022) 99129–99149, <http://dx.doi.org/10.1109/access.2022.3207287>.
- [55] C. Guo, G. Pleiss, Y. Sun, K.Q. Weinberger, On calibration of modern neural networks, in: D. Precup, Y.W. Teh (Eds.), *Proceedings of the 34th International Conference on Machine Learning*, in: *Proceedings of Machine Learning Research*, 70, PMLR, 2017, pp. 1321–1330, <https://proceedings.mlr.press/v70/guo17a.html>.
- [56] G. Gallitto, R. Englert, B. Kincses, R. Kotikalapudi, J. Li, K. Hoffschlag, U. Bingel, T. Spisak, External validation of machine learning models—registered models and adaptive sample splitting, *GigaScience* 14 (2025) <http://dx.doi.org/10.1093/gigascience/giaf036>.
- [57] S.Y. Ho, K. Phua, L. Wong, W.W. Bin Goh, Extensions of the external validation for checking learned model interpretability and generalizability, *Patterns* 1 (8) (2020) 100129, <http://dx.doi.org/10.1016/j.patter.2020.100129>.
- [58] V.B. Tinti, J.K. Han, V. Frederiksen, H. Chen, J. Wallentin, I. Kantor, A. Lyksborg-Andersen, T.W. Hansen, G. Bae, W. Song, E. Stamate, D.Z. de Florio, H. Bruus, V. Esposito, Oxygen-defective electrostrictors for soft electromechanics, *Sci. Adv.* 10 (35) (2024) <http://dx.doi.org/10.1126/sciadv.adq3444>.
- [59] A. Altomare, N. Corriero, C. Cuocci, A. Falcicchio, A. Moliterni, R. Rizzi, EXPO software for solving crystal structures by powder diffraction data: methods and application, *Cryst. Res. Technol.* 50 (9–10) (2015) 737–742, <http://dx.doi.org/10.1002/crat.201500024>.
- [60] M.R. Felez, J.C.B. Monteiro, D.Z. de Florio, F.C. Gandra, S. Gama, Effects of Fe substitution on Mn<sub>2</sub>Sn alloy on its structural, magnetic and magnetocaloric properties, *J. Magn. Magn. Mater.* 505 (2020) 166676, <http://dx.doi.org/10.1016/j.jmmm.2020.166676>.
- [61] D. Hong, Y.-H. Choi, High-entropy oxide (Mg<sub>0.2</sub>Fe<sub>0.2</sub>Co<sub>0.2</sub>Cu<sub>0.2</sub>Zn<sub>0.2</sub>)O with rocksalt-to-spinel transformation and its electrocatalytic activity for the oxygen evolution reaction, *J. Alloys Compd.* 985 (2024) 174029, <http://dx.doi.org/10.1016/j.jallcom.2024.174029>.
- [62] S. Lundberg, S.-I. Lee, A unified approach to interpreting model predictions, 2017, <http://dx.doi.org/10.48550/ARXIV.1705.07874>, URL <https://arxiv.org/abs/1705.07874>.
- [63] Y. Li, H. Yan, S. Wang, X. Luo, L. Kurpaska, F. Fang, J. Jiang, H.S. Kim, W. Huo, Toward predictable phase structures in high-entropy oxides: A strategy for screening multicomponent compositions, *Mater. Des.* 248 (2024) 113497, <http://dx.doi.org/10.1016/j.matdes.2024.113497>.
- [64] G.D. Sproul, Evaluation of electronegativity scales, *ACS Omega* 5 (20) (2020) 11585–11594, <http://dx.doi.org/10.1021/acsomega.0c00831>.
- [65] J. Syarif, M.B. Elbeltagy, A.B. Nassif, A machine learning framework for discovering high entropy alloys phase formation drivers, *Heliyon* 9 (1) (2023) e12859, <http://dx.doi.org/10.1016/j.heliyon.2023.e12859>.
- [66] B. Jiang, C.A. Bridges, R.R. Unocic, K.C. Pitike, V.R. Cooper, Y. Zhang, D.-Y. Lin, K. Page, Probing the local site disorder and distortion in pyrochlore high-entropy oxides, *J. Am. Chem. Soc.* 143 (11) (2020) 4193–4204, <http://dx.doi.org/10.1021/jacs.0c10739>.
- [67] M. Johnsson, P. Lemmens, Crystallography and chemistry of perovskites, 2007, <http://dx.doi.org/10.1002/9780470022184.hmm411>.
- [68] E.C. O'Quinn, D.L. Drey, M.K. Lang, Defining the structural stability field of disordered fluorite oxides, *Front. Chem.* 9 (2021) <http://dx.doi.org/10.3389/fchem.2021.733718>.
- [69] A.F. Manchón-Gordón, P. Panadero-Medianero, J.S. Blázquez, Descriptors for predicting single- and multi-phase formation in high-entropy oxides: A unified framework approach, *Materials* 18 (16) (2025) 3862, <http://dx.doi.org/10.3390/ma18163862>.