Machine Learning Prediction of Crystal Structure Stability toward the Design of High-Entropy Oxides

Qiancheng Zhou(周前程)[1], Chunchun Wang(王春春)[1], Liyuan Wang(王丽媛)[1], Zhouzhou Wang(王舟舟)[1], Ming Qiu(邱明)[1*], Mingdong Dong(董明东)[2*], and Ying Yu(余颖)[1*]

[1]*Institution of Nanoscience and Nanotechnology, College of Physical Science and Technology, Central China Normal University, Wuhan 430070, China*
[2]*Interdisciplinary Nanoscience Center, Aahus University, Denmark*

(Received 10 February 2026; accepted manuscript online 30 March 2026)

Energy above the convex hull ($E_{\rm hull}$) is a key thermodynamic criterion for assessing phase stability. However, the enormous computational cost required for phase diagram construction hinders the prediction of $E_{\rm hull}$, underscoring the need for data-driven approaches. Here, a hybrid framework integrating an autoencoder with a random forest classifier was proposed to effectively categorize crystal structures into stable, metastable, and unstable regimes according to $E_{\rm hull}$ thresholds, achieving an overall accuracy above 84%. More importantly, physically interpretable latent features associated with density, symmetry, and lattice were identified for stability prediction. Application to high-entropy oxides (HEOs) further demonstrates the effectiveness of the framework, revealing that structures with high configurational entropies and low cation radius mismatch are overwhelmingly classified as stable or metastable. Beyond confirming the dominant role of density and lattice features in stability prediction, SHAP analysis further suggests that larger disparities in atomic thermal conductivities and the regulation of the magnetic moment by limited magnetic atoms play a critical role in governing the stability of HEO structures. The interpretable and effective AE-RF algorithm developed in this work holds great potential for accelerating the discovery of novel HEOs and multicomponent materials.

DOI: 10.1088/0256-307X/43/5/050801 CSTR: 32039.14.0256-307X.43.5.050801

1. Introduction. In recent decades, material discovery has predominantly relied on trial-and-error experimentation.[1,2] This so-called "Edisonian approach" necessitates substantial research investment from numerous scientists to identify materials with practical application value, leading to prolonged timelines.[3,4] While enhancing laboratory efficiency remains challenging, machine learning (ML) technologies and materials databases have emerged as powerful tools to significantly accelerate the evaluation of unknown materials.[5–7]

As a powerful tool for rapid material property prediction, nearly all ML algorithms operate by fitting statistical models to large datasets.[8] In materials science, the scarcity of well-curated experimental data has led to a reliance on input data derived from density functional theory (DFT) calculations.[9,10] These databases typically contain two types of features. The first is the "direct" features related to the composition and structure, obtained without first-principles calculations. These include atomic types and counts, chemical composition, unit cell parameters, symmetry,[6,9] etc. The "indirect" features are derived from direct features via first-principles calculations, such as band gap, phonon spectra, electron density, and work function.[11] Obtaining indirect features demands substantial computational resources, driving the development of algorithms in materials science to prioritize accurate and rapid prediction of these features using direct features as inputs. While the speed advantage of data science algorithms over human-knowledge-based theoretical methods is inherent,[12–15] limited prediction accuracy—alongside challenges in interpretability, transferability, and data availability—currently prevents ML from replacing computational simulation.[15] Critically, ML predictions do not inherently reflect material reality; experimental validation remains essential to confirm application potential. This underscores the need for close collaboration between ML and experimental validation in material development to minimize blind spots and maximize development efficiency.[16–18] It suggests that material stability, as a fundamental property, thus emerges as a key target for ML prediction, which not only addresses the primary question for experimentalists—whether a material can be synthesized—but also serves as a prerequisite for subsequent experimental characterization.[19–21]

High-entropy oxides (HEOs) are a new class of materials composed of five or more metal elements mixed in near-equimolar ratios, forming single-phase solid solutions stabilized by high configurational entropy. HEOs exhibit complex atomic configurations and strongly disordered lattices, which give rise to intrinsically low thermal conductivity.[22–24] Such characteristics make them promising candidates for thermal management and heat

*Corresponding authors. Email: qium@ccnu.edu.cn; dong@inano.au.dk; yuying01@ccnu.edu.cn
© 2026 Chinese Physical Society and IOP Publishing Ltd. All rights, including for text and data mining, AI training, and similar technologies, are reserved.

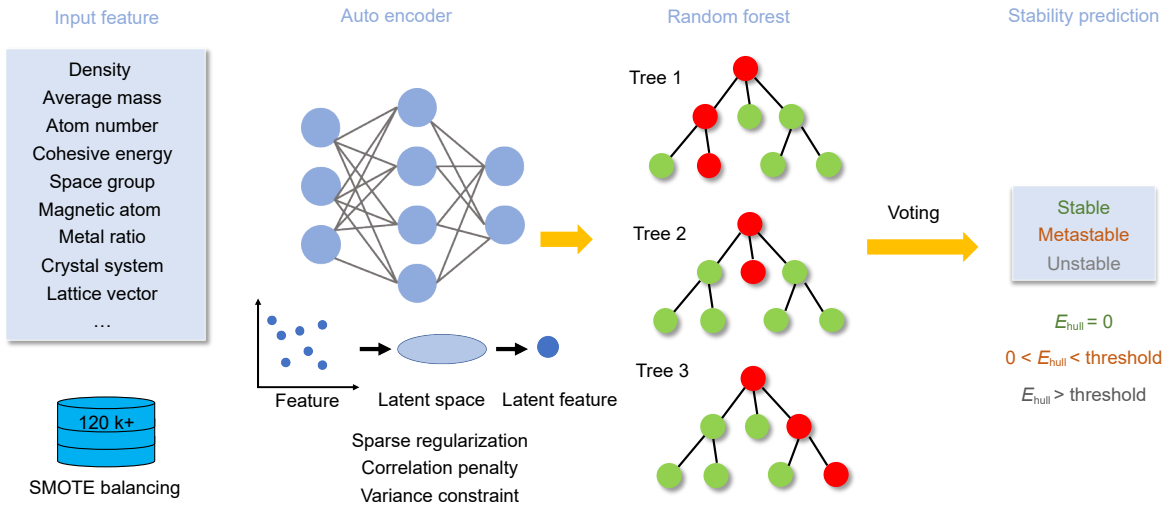

Fig. 1. Workflow of this work.

insulation applications.[23,25] Moreover, understanding the relationship between structural disorder and heat conduction in these systems is essential for guiding the design of next-generation heat conduction materials. However, HEOs possess complex compositions and are difficult to synthesize, which makes predicting their stability using ML methods highly demanded for the development of novel HEOs. Employing neural network models based on atomic coordinates and chemical bonds to predict the stability of HEO supercells is challenging and computationally expensive because they often contain hundreds or even thousands of atoms. Therefore, using simple classification models without deep learning to rapidly assess the stability of HEOs is a more reliable and efficient approach for the development of new HEOs.

To enable ML to predict material stability and synthesizability, a universal stability assessment criterion independent of elemental composition is crucial. Currently, formation energy (E_f) is commonly employed to evaluate structural stability.[26–30] Calculated as the energy difference between a structure and its constituent elements via first-principles methods, E_f provides insights into structural stability but suffers from two significant limitations. First, E_f values are inherently influenced by the elemental composition of the structure.[31–34] Second, while E_f quantifies the tendency of a compound to form from its elements, most materials are not able to be synthesized directly from isolated atoms. In contrast, the energy above the convex hull (E_{hull}) offers a more robust approach. By quantifying the thermodynamic stability of a compound relative to all competing phases, E_{hull} allows for a more rational comparison of synthesis feasibility among structures with the same composition, surpassing the limitations of relying solely on E_f values.[10,21,31] Previous studies have rigorously demonstrated that such compositional structures form a convex hull, the distance from the convex hull to the new crystal structure in the diagram. By definition, determining E_{hull} depends on the E_f values of an extensive set of competing materials, which not only demands comprehensive structure information and physical constants but also incurs substantial computational cost.

A framework combining an autoencoder (AE) and a random forest (RF) classifier for predicting crystal structure stability based on E_{hull} was established, trained on over 120000 compounds from the Materials Project (MP). The AE-RF model achieved an accuracy of 84.3% and remained robust under limited-data conditions, attaining accuracy above 75% with only 40% of the training data. In this work, three physically interpretable hybrid features associated with density, symmetry, and lattice characteristics were also identified as critical descriptors in stability prediction. The proposed framework is further applied to HEOs, where structures with higher configurational entropies were successfully screened. Shapley additive explanations (SHAP) analysis is employed to elucidate the model's decision-making process, consistently highlighting the dominant contributions of density-, symmetry-, and lattice-related hybrid features. Overall, the interpretable AE-RF framework presented here provides an efficient and physically grounded approach for stability assessment, offering a promising route toward accelerated discovery and experimental design of HEOs and other multicomponent materials.

2. Workflow. Figure 1 illustrates the workflow for this study. First, a set of readily accessible physical and structural descriptors, such as atomic density and average atomic mass, is collected for over 120000 material entries. To mitigate the inherent class imbalance in the dataset, the synthetic minority oversampling technique (SMOTE) is applied to generate a balanced training set. These engineered features are then fed into an AE, which compresses the original high-dimensional input into a compact latent representation. The AE extracts the most informative patterns in the data, producing latent features constrained by sparse regularization, correlation penalties, and variance control. The resulting latent features are subsequently used as inputs to a RF classifier composed of multiple decision trees. In this work, to comprehensively characterize the structural and compositional properties of periodic inorganic crystals, a total of 35 representative features were

extracted from the MP database. These features fall into four categories: lattice, local atomic environments, composition, and symmetry. Detailed descriptions are provided in the Methods section. As shown in the heatmap of pairwise correlation coefficients in Fig. S1 of the Supplementary Materials (SM), clear redundancy can be observed among the extracted features, with a mean absolute correlation of 0.21 and some feature pairs exhibiting correlation coefficients as high as 0.90. Such strong correlations indicate that the raw feature set is not directly suitable for supervised classification and requires proper preprocessing to avoid issues such as multicollinearity and overfitting. To address this, a neural network-based AE was introduced to optimize the feature representation and reduce dimensionality before feeding the data into the classification model. To further enhance the interpretability of the latent space, several additional regularization terms were incorporated into the AE training process, including L1 sparsity regularization to encourage a more compressible latent representation, a correlation penalty to reduce interdependence among latent dimensions, and a variance constraint to prevent latent feature collapse. The calculation methods for each component of the AE loss function are provided in the Methods section. The latent features generated by the AE were subsequently fed into the RF algorithm for predicting material stability, and the selection and optimization of RF hyperparameters are described in the Methods section.

In previous studies addressing E_{hull}, materials are typically categorized as stable, unstable, and metastable. A statistical analysis by Sun *et al.* revealed that for all known inorganic crystalline materials in the MP dataset (29902 samples), the DFT-calculated median metastability is 15 ± 0.5 meV/atom, with the 90th percentile at 67 ± 2 meV/atom.[35] Lyngby *et al.* employed generative models to explore the space of unknown 2D materials, noting that many of the predicted 2D materials with an $E_{\text{hull}} < 50$ meV/atom, or even < 100 meV/atom, are likely synthesizable.[36] Most studies adopt 100 meV/atom as the energy-above-hull threshold to include metastable but potentially synthesizable materials, making it a common criterion in high-throughput screening and ML modeling across various material classes, such as MAB phases (M = transition metal, B = boron, A = main-group element), ternary borides, and ABO_3 perovskites.[31,37–39] Based on the aforementioned studies, in this work, inorganic crystalline materials with an E_{hull} value between 0 and 100 meV/atom are considered potentially synthesizable, may exhibit stability under certain conditions, and thus fall into the category of metastable materials. Similarly, in this work, crystal structures are categorized into three classes based on their energy above the convex hull (E_{hull}): stable ($E_{\text{hull}} = 0$), metastable ($0 < E_{\text{hull}} \leq$ threshold), and unstable ($E_{\text{hull}} >$ threshold). For convenience, metastable and unstable structures ($E_{\text{hull}} \geq 0$) are collectively referred to as "above-hull" materials, whereas stable and metastable structures ($0 \leq E_{\text{hull}} \leq$ threshold) are collectively referred to as "low-hull" materials.

3. Result and Discussion. All crystal structures with $E_{\text{hull}} \geq 0$ recorded in the MP database prior to May 2024 were collected for analysis. Figure S2 of the SM presents the distribution of 153215 crystal structures, ranging from one to seven elements, categorized into stable ($E_{\text{hull}} = 0$) and above-hull ($E_{\text{hull}} > 0$) classes. Although the MP dataset includes structures with eight and nine elements, their occurrences are extremely rare; hence, they were excluded from the study. In ternary materials, $\sim$46000 above-hull and $\sim$19000 stable cases are observed, far more than in systems with higher or lower elemental dimensionality, underscoring the central importance of ternaries in the MP dataset. Notably, ternaries show the highest stable-case fraction (29.4%) and account for 56.7% of all stable materials. This indicates that crystal structures composed of three elements are the most common and are more likely to exhibit higher stability compared to those with a different number of elements. On the one hand, ternary systems offer a broader compositional design space than binaries, enabling more flexible local environments and coordination motifs that can lower the formation energy. On the other hand, they remain significantly less complex than quaternary and higher-order systems, where the configurational space expands exponentially and often results in structural frustration or competing metastable states. A comparison between E_{hull} and E_{f} is shown in Fig. S3 of the SM. The results reveal that there is no linear correlation between E_{hull} and E_{f}, except when only a single compound exists within a given chemical space, resulting in identical E_{hull} and E_{f} values. The distributions of E_{hull} and E_{f} values, also depicted in Fig. S3 of the SM, show that E_{f} values are primarily distributed in the range of -5 eV/atom to 3 eV/atom, while E_{hull} values are predominantly confined below 1 eV/atom. These findings suggest that E_{hull} is a more discerning, yet inherently more challenging, metric for predicting crystal stability compared to E_{f}.

Before ML is executed, physical filters on E_{hull}, volume, and atomic density of the crystal structures were applied to ensure that the selected entries are structurally valid and representative of realistic materials. Specifically, E_{hull} is restricted to 0–0.3 eV/atom to exclude highly unstable or non-synthesizable phases. The crystal volume is limited to 10–1000 Å^3 to avoid unphysically small or excessively large cells, and the density is constrained to 0.5–25 g/cm^3 to remove structures with unrealistic atomic packing. After applying these criteria, a total of 118977 structures were retained, ensuring a reliable and physically meaningful dataset for subsequent analysis. The AE used in this work consists of a symmetric encoder–decoder architecture designed to compress the original feature space and reconstruct it with minimal information loss. The encoder maps the input vector of dimension 35 into a compact latent representation through a series of fully connected layers: an input layer followed by a 128-unit hidden layer with rectified linear unit (ReLU) activation, a second hidden layer with 64 units and ReLU activation, and a final linear layer that outputs the latent vector. The

decoder mirrors this structure, expanding the latent vector back to the original 35-dimensional feature space using two hidden layers and an output layer. This architecture enables the model to learn a low-dimensional latent feature set suitable for downstream ML tasks.

For different latent dimensions, the evolution of the training loss over epochs is shown in Fig. S4 of the SM. It is evident that the AE converges after approximately 50 training epochs. Although the training loss decreases rapidly when the latent dimension is increased from small values, the rate of improvement gradually diminishes as the latent dimension increases. Figure S5 of the SM further presents the reconstruction errors obtained after 120 training epochs for latent dimensions ranging from 1 to 48, again demonstrating that the loss monotonically decreases with latent dimension, but with progressively diminishing returns. Notably, once the latent dimension exceeds 16, the reduction in reconstruction error becomes almost negligible, indicating that the AE has already approached its best performance. To determine an appropriate latent dimension from the perspective of stability prediction, latent features of various lengths were fed into an RF classifier to distinguish stable, metastable (those with an E_{hull} between 0 and 100 meV/atom), and unstable materials. As shown in Fig. S6 of the SM, the RF classification accuracy improves steadily as the latent dimension increases. However, when the latent space reaches 16 dimensions, the RF model already achieves an accuracy of 84.3%. Further increases in dimensionality lead to only marginal improvements. For comparison, using the full set of 35 original features yields an accuracy of 85.8%, only 1.5 percentage points higher than that achieved with the 16-dimensional latent features, suggesting that the additional 19 dimensions do not provide discriminative information for the RF classifier. These results indicate that the 16-dimensional latent representation extracted by the AE is sufficient to capture the stability-relevant structural information of materials, enabling the downstream RF model to maintain nearly the same classification performance as when using the full high-dimensional feature set. Therefore, a latent size of 16 represents an optimal balance between model performance and interpretability: it effectively reduces feature redundancy, improves model simplicity and generalization, and retains the key discriminative information.

Figure 2(a) shows that the 16-dimensional latent features learned by the AE exhibit significantly reduced correlations. The maximum correlation coefficient is only 0.18, and the mean absolute correlation is merely 0.11, less than half of that observed in the original features. This indicates that the AE not only achieves effective dimensionality reduction but also successfully captures a more orthogonalized representation, making the latent features more independent of each other and substantially reducing feature redundancy. 2D t-distributed stochastic neighbor embedding (t-SNE) analysis was performed on the 16-dimensional features in the dataset, with clustering results shown in Fig. S7a of the SM. After clustering, the 16-dimensional features exhibit a uniform distribution in the 2D space, forming 6 distinct clusters, which is far beyond the typical categorization number of material stability and fails to provide any meaningful reference for assessing material stability. Nonetheless, when clustering results are visualized using three distinct colors for stable, metastable, and unstable samples (see Fig. S7b of the SM), the spatial distributions of the three categories overlap significantly. The clustering result motivates the use of nonlinear classification methods to predict material stability more effectively.

To comprehensively evaluate the performance of different classification algorithms, models based on K-nearest neighbors (KNN), decision tree (DT), gradient boosting tree (GBT), logistic regression (LR), naive Bayes (NB), and RF were trained to categorize materials into distinct stability classes. Among these models, the RF algorithm achieved the highest prediction accuracy of 84.3%, outperforming all other methods in distinguishing material stability categories and thus demonstrating its superior effectiveness for this task [Fig. 2(b)]. The confusion matrices of the prediction results from RF and the other five algorithms are shown in Fig. 2(c) and Fig. S8 of the SM, further validating the superior performance of RF in predicting material stability. Figure 2(b) also shows the 5-fold cross-validation accuracy of models based on the six algorithms as well as the standard deviation of cross-validation. Among all algorithms investigated herein, the mean error from 5-fold cross-validation is consistently lower than the error on the whole dataset, with the cross-validation standard deviation below 0.03. These results indicate that none of the training processes suffer from overfitting.

Figure 2(d) illustrates the variation in model accuracy as a function of training subset size (ranging from 5% to 100% of the filtered dataset with a size of 118977). Model performance was evaluated at three levels: the test set of the subset, the subset itself, and the entire dataset. As the subset size increases, accuracy on the entire dataset increases steadily, indicating that sample size plays a critical role in this ML process and underscoring the importance of building more comprehensive databases. Notably, a model trained with only 40% of the entire dataset achieves over 75% prediction accuracy on the full dataset, demonstrating that the RF model can achieve high predictive performance with reduced datasets and highlighting its excellent performance in ML tasks with limited data availability. Additionally, when different fractions of the dataset (from 5% to 100%) are used while keeping the training, validation, and test set ratios fixed, the error evaluated on the used dataset remains consistently lower than that on the corresponding test sets. This observation suggests that the model maintains stable predictive performance and does not exhibit significant overfitting. Figure 2(e) shows the receiver operating characteristic (ROC) curve, along with the corresponding area under the curve (AUC). The AUC values exhibit a decreasing trend, from unstable (0.92) to stable (0.91) to metastable materials (0.84), indicating that the classifier distinguishes unstable materials most reliably, while the metastable class is inherently more

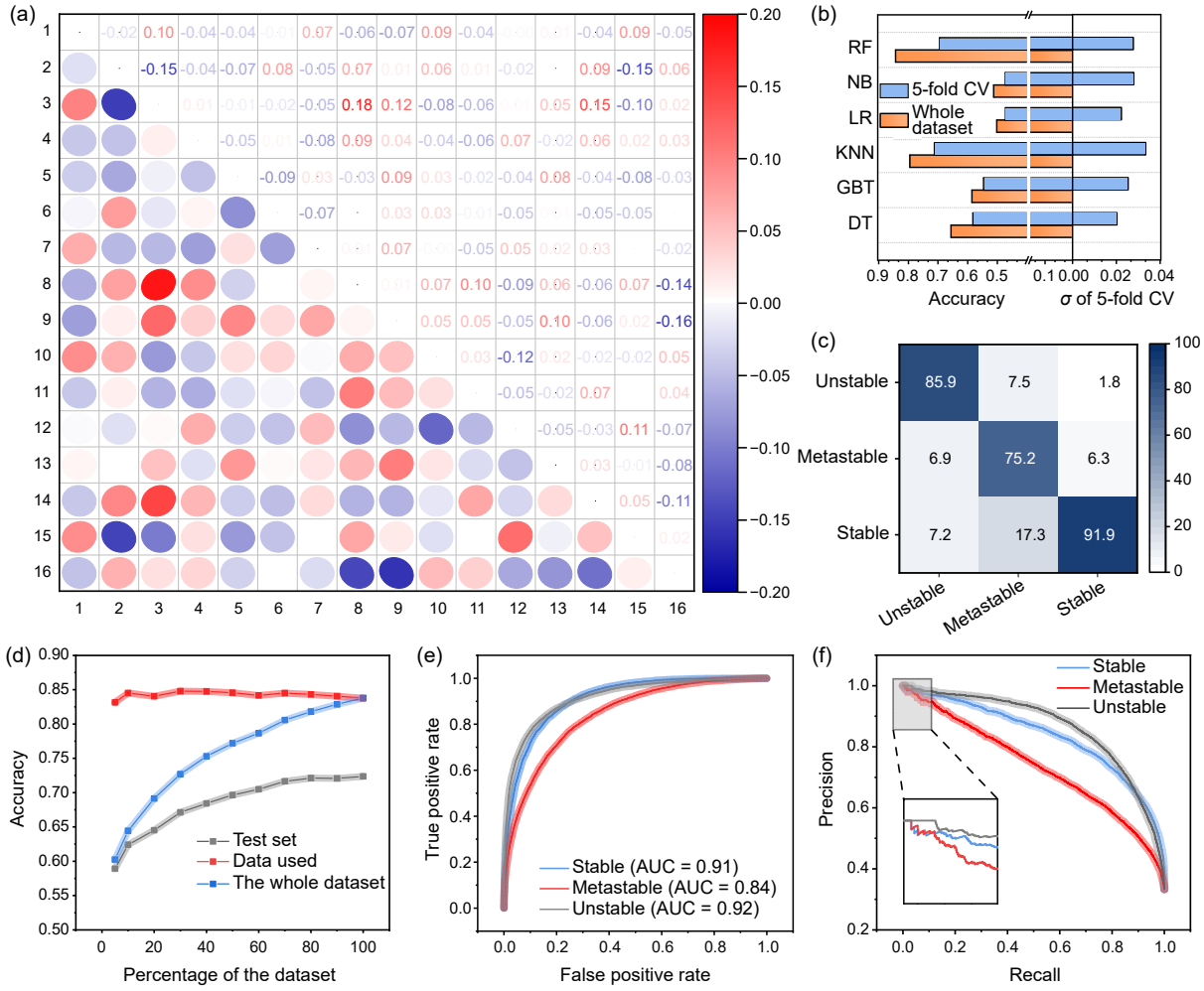

Fig. 2. (a) Correlation coefficients of 16 latent features in heatmap. (b) Prediction accuracy of all algorithms. (c) Confusion matrix of the RF prediction result. (d) Learning curve of RF model using 5% to 100% of dataset. (e) ROC curve of RF classifier. (f) PR curve of RF classifier.

challenging to separate. A similar trend is also observed in the precision-recall (PR) curves [Fig. 2(f)], where the curve corresponding to the metastable class lies noticeably below those of the other two categories. This can be attributed to metastable materials occupying an intermediate zone in the energy landscape, and their structural and chemical features often resembling those of stable phases.

Given that the metastable class is intrinsically difficult to distinguish, it is reasonable to question whether the commonly used E_{hull} threshold of 0–100 meV/atom provides an optimal definition for metastability in the context of ML classification. The commonly adopted metastability threshold of 100 meV/atom, frequently reported in previous studies, is predominantly based on empirical judgment derived from a comprehensive literature review.[34–36,40,41] A narrower or broader energy window could potentially yield a more separable class distribution and different classifier performance. To investigate this, the E_{hull} threshold was systematically varied for defining metastable materials (from 0 to 300 meV/atom), and by examining how model performance evolves with different metastability criteria, it is possible to assess the robustness of the classifier and identify an energy cutoff that yields more reliable predictions. The resulting changes in classification accuracy on the whole dataset and test set are summarized in Figs. 3(a) and 3(b). Notably, the model retained relatively high accuracy (above 86%) when the metastability threshold is higher than 120 meV/atom. This behavior can be attributed to the simplification of the three-class classification task into a binary classification when the metastable threshold is large enough to encompass most structures. More importantly, an unexpected increase in classification difficulty occurs when the metastability threshold is set to 60 meV/atom, resulting in a classification accuracy of 82.0%. Prediction accuracy decreases below 60 meV/atom and increases above it monotonically. To investigate the underlying cause of this phenomenon, the average and standard deviation of all 35 features for structures with E_{hull} lower than 120 meV/atom are shown in Fig. S9 of the SM. The lower E_{hull} threshold (<120 meV/atom) dataset exhibits smaller standard deviations for 28 of the 35 feature dimensions compared with the whole dataset. This indicates that structures within this energy range occupy a narrower region in the feature space, suggesting that their structural and compositional characteristics are more homogeneous compared with those of the entire dataset.

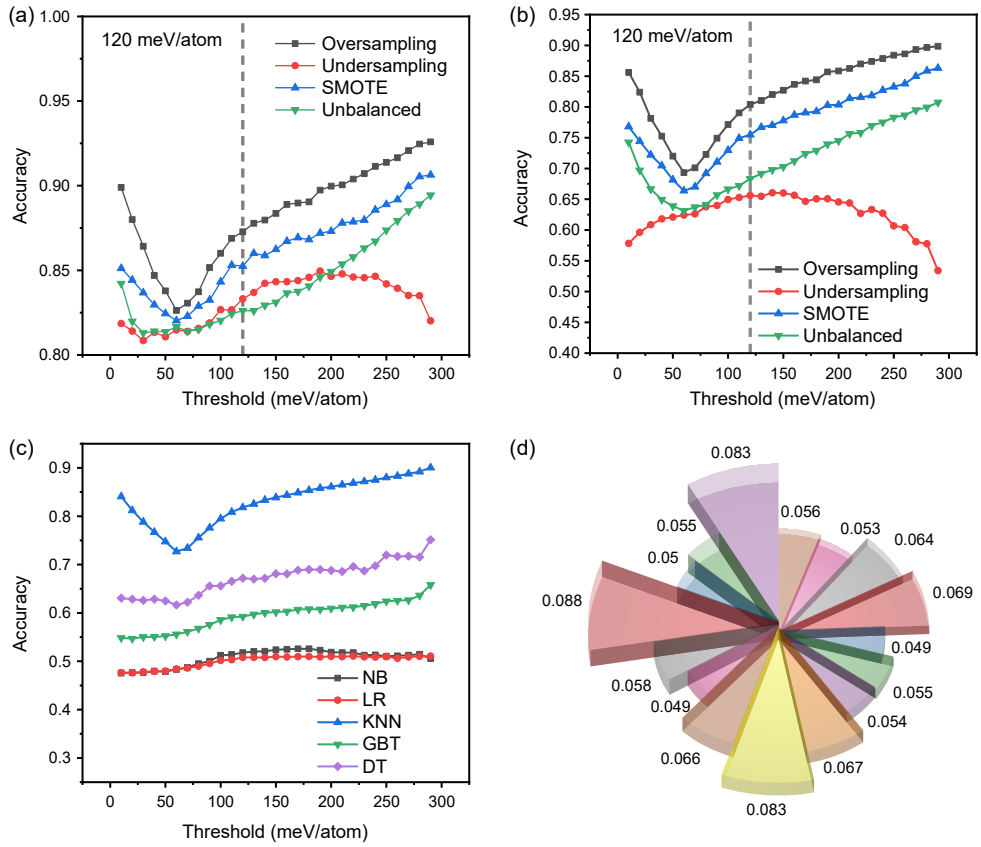

Fig. 3. Accuracy as a function of classification threshold on the (a) whole dataset and (b) test set, with different resampling strategies, including oversampling, undersampling, SMOTE, and unbalanced datasets. (c) Accuracy changes of different ML algorithms (NB, LR, KNN, GBT, DT) on varying thresholds. (d) Normalized feature importance.

Feature value distribution of structures with E_{hull} less than 120 meV/atom and higher than 120 meV/atom is also shown in Fig. S10 of the SM. The largest difference between structures with E_{hull} below 120 meV/atom and above 120 meV/atom is observed for feature 21, corresponding to the number of elements, with lower E_{hull} structures containing fewer elements. This suggests that fewer elements are associated with lower E_{hull} values. When the metastability threshold is below 120 meV/atom, the reduced predictive performance can be attributed to the fact that structures with fewer elements exhibit near-zero values in standard-deviation-related features, such as the standard deviation of atomic electronegativity and atomic radius. As a result, multiple feature dimensions of these structures become very small, leading to overly sparse representations that weaken the model's ability to capture structural characteristics and cause performance decline. To further assess whether the observed trend persists under different data distributions, model performance was evaluated using the original imbalanced dataset, as well as datasets balanced through random oversampling and undersampling. After applying oversampling, undersampling, and SMOTE methods, the classification accuracy of RF on the whole dataset with different E_{hull} thresholds is also shown in Figs. 3(a) and 3(b). It is evident that the model's predictive performance exhibits a trend of initial decline followed by subsequent improvement within the range of 0–300 meV/atom for the unbalanced, oversampled, and SMOTE datasets. However, this non-monotonic behavior was not observed when the model was trained on datasets balanced through undersampling. This discrepancy can be attributed to the inherent characteristics of the undersampling strategy. By randomly removing a large portion of majority-class samples, undersampling substantially reduces the diversity and structural heterogeneity of the training data. As a result, the remaining subset becomes overly simplified and no longer preserves the fine structure of the latent feature space. Nonetheless, this exception does not alter the overall conclusion that these non-monotonic accuracy trends are intrinsic to the underlying data structure rather than an artifact of the balancing method.

To examine whether this phenomenon is algorithm-dependent, the performance of five additional models across varying E_{hull} thresholds was further evaluated [Fig. 3(c)]. A similar non-monotonic trend was observed only in the KNN model. This behavior can be attributed to the intrinsic characteristics shared by RF and KNN: both methods rely on local decision patterns rather than global functional approximations. The latent features transformed by the AE form locally clustered manifolds in the original feature space, where materials with similar structures aggregate. When the metastability threshold falls within the ambiguous energy range, these local

Table 1. Top 5 correlated physical features related to latent features with high importance.

Index	Description	Top 5 related physic features (correlation coefficients)
13	Density	Atomic density (−0.774), Average coordination number (0.679), Average atomic mass(−0.628), Average atomic radius (−0.567), Space group(−0.553)
16	Symmetry-bonding	Space group (0.474), Average melting point (0.465), Crystal system (−0.454), Kurtosis of the interatomic distance (−0.437), Average cohesive energy (0.416)
9	Lattice	Number of atoms (0.752), Lattice volume (0.725), Lattice a (0.680), Lattice b (0.667), Standard deviation of the coordination number (0.555)

clusters become compositionally mixed, making the boundary between metastable and unstable materials ambiguous. KNN is particularly sensitive to such local clusters, and its decision boundary evolves dramatically when the metastable threshold changes. Therefore, its performance exhibits the same non-monotonic fluctuation as observed with RF. In contrast, the remaining algorithms (LR, NB, DT, and GBT) depend primarily on global decision boundaries. These models approximate the boundary through either global linear decision functions (LR), parametric likelihood assumptions (NB), a single hierarchical partition (DT), or gradient-boosted global refinements (GBT). Such approaches are less sensitive to the local structure of the dataset and produce comparatively smoother changes in their classification performance as the metastability threshold varies, whereas the other algorithms maintain relatively stable predictive accuracy as the metastable threshold varies. This conclusion is further supported by the significant fluctuations of metastable samples proportion in the predictions when the metastable threshold changes in Fig. S11 of the SM. Nevertheless, RF consistently outperforms all other algorithms across the entire range of thresholds, and—even when accounting for these performance fluctuations—remains the optimal choice for stability classification.

Using datasets labeled with different E_{hull} thresholds, multiple models were trained. These models and datasets were then used for cross-prediction. The relative accuracy changes across all model–dataset combinations, relative to their original performance, are visualized as a heatmap in Fig. S12 of the SM. Besides SMOTE, two additional approaches were employed to balance the relative proportions of stable, metastable, and unstable materials: oversampling and undersampling. Notably, nearly all models achieve their best performance on datasets labeled with consistent thresholds. In contrast, for imbalanced datasets, optimal predictive performance is not consistently attained when the training and testing datasets are labeled with identical E_{hull} thresholds; in fact, ten models exhibit higher accuracy when evaluated on datasets labeled with other thresholds. Together, these results suggest that overfitting is negligible on balanced datasets and highlight the critical role of data balancing strategies in achieving robust stability prediction. Overall, these results demonstrate that the decreased classification performance observed within the intermediate E_{hull} range is a robust phenomenon that persists across algorithms based on local decision and remains largely insensitive to the choice of data balancing strategy. This consistent trend highlights that the difficulty in identifying metastable materials is fundamentally rooted in the intrinsic nature of their structural and energetic characteristics. Metastable materials inherently occupy a transitional region in the energy landscape and share overlapping structural features with both stable and unstable phases, making their classification intrinsically ambiguous. It also emphasizes the need for more descriptors or adaptive classification strategies when dealing with metastable materials.

To further understand how the AE-RF framework performs stability classification and to enable its interpretable application to HEOs, a detailed feature-importance analysis was conducted on the 16-dimensional latent representation learned by the AE. Since stability prediction in HEOs often depends on subtle combinations of structural complexity, local disorder, and composition-dependent bonding environments, identifying which latent variables capture the most physically meaningful information is essential for model interpretability and transferability. Based on feature importance ranking in Fig. 3(d), the latent features can be grouped into three categories: high importance (3 features), medium importance (10 features), and low importance (3 features). To further understand the physical meaning of the latent representations learned by the AE, the correlations between the three most important latent features [indexed as 13, 16, and 9 in Fig. 2(a)] and the original 35 features were analyzed. For each latent feature, the top five most strongly correlated original features were identified, as summarized in Table 1. By categorizing the physical properties represented by these original features, the physical essence encoded in the learned latent space can be elucidated.

Latent 13 shows strong correlations with atomic density and average coordination number, with correlation coefficients of −0.774 and 0.679, respectively. This result implies that it captures the characteristic of densely packed light atoms. Therefore, this latent feature can be interpreted as a descriptor of atomic packing density and is hereafter referred to as the density feature. In addition, its moderate negative correlations with atomic mass (−0.628), atomic radius (−0.567), and space group (−0.553) suggest that the density feature also weakly encodes atomic mass, size, and structural symmetry information. The density feature is learned by the AE because crystals with compact packing generally tend to have lower energies and higher structural stability. Such densely packed structures favor stronger interatomic bonding and reduced lattice distortions, thereby stabilizing the crystal framework. Latent 16 does not exhibit a strong correlation with any individual original feature; instead, it shows moderate negative correlations with symmetry descriptors (such as space group and crystal system) and bond strength descriptors (such as melting point, kurtosis of the interatomic distance distribution, and average cohesive energy). It can thus be interpreted as a symmetry–bonding feature. Given the critical role of local disorder and bonding strength in thermal stability, the symmetry–bonding feature naturally exhibits high importance in stability estimation. Latent 9 exhibits strong correlations with the number of atoms, lattice volume, and lattice constants a and b, indicating that it reflects lattice dimensions and their anisotropy. Therefore, it can be interpreted as a lattice feature. It also shows a weak correlation with coordination number dispersion, partially capturing the presence of locally high-coordination environments. Materials with excessively expanded or anisotropic lattices often exhibit higher strain energy, while variations in local coordination environments may signal structural distortions or motifs. Thus, the strong response of the lattice feature to these attributes highlights its role as a stability descriptor learned by the AE. SHAP analysis was performed, and this interpretability framework enabled a comparative assessment of feature importance in distinguishing between stability categories, as presented in the SHAP summary plot (Fig. S13 of the SM). Notably, the resulting feature importance ranking is highly consistent with that obtained from Gini importance analysis based on tree models, further reinforcing the robustness and reliability of the interpretability results. Collectively, these three latent features—density, symmetry–bonding, and lattice features—capture complementary structural and energetic factors governing crystal stability, including atomic packing compactness, bonding strength, local disorder, and lattice geometry. Their clear physical interpretations demonstrate that the AE learns stability-relevant representations rather than purely statistical patterns, highlighting the strong physical interpretability and practical guidance of the proposed framework for stability assessment and materials design.

In recent years, HEOs have found broad applications in energy storage, conversion, and heat conduction.[42] However, HEO structures exhibit high structural complexity and diversity owing to their definition: oxides composed of at least five different metal elements. Significant challenges for the rational design and synthesis of novel HEOs can be overcome by the AE-RF algorithm presented in this work, because it enables rapid evaluation of the structural stability of multicomponent crystalline materials, making it particularly advantageous for the design of new HEOs. Specifically, four types of oxides were constructed, including perovskite, fluorite, rock-salt, and spinel types, with the A-site occupied by Sr, Ba, Ca, Nd, Na, K, Rb, Cs, and La, and the B-site occupied by Fe, Ni, Mn, Ti, V, Cr, and Co. Besides, lattice distortion was applied by randomly scaling the three lattice vectors within a specified distortion amplitude. In addition, atomic positions were displaced by small random offsets proportional to the distortion level. This procedure is intended to mimic the lattice strain and local atomic perturbations that arise when cations are substituted in high-entropy oxide structures. Furthermore, a random 5% concentration of oxygen vacancies was introduced to reflect their frequent occurrence in experimentally synthesized HEOs (detailed construction procedures are provided in the Methods section). To verify the capability of the AE-RF algorithm in screening low-hull structures, new oxide structures were initially generated with both element types and their proportions randomly assigned. Figure 4(a) presents the numbers of stable, metastable, and unstable oxide structures predicted by the AE-RF framework for 1000 perovskite-, fluorite-, rock-salt-, and spinel-type oxides. Clear structure-dependent stability trends are observed across the four lattice types. Among them, the rock-salt and fluorite lattices exhibit a higher number of stable or metastable cases than the perovskite and spinel lattices, which can be attributed to their simpler and more symmetric structures. Their strong tolerance to ionic radius mismatch, charge variation, and oxygen vacancy formation makes them particularly favorable for forming low-hull metal oxides. Figure 4(b) shows the distribution of configurational entropy (in units of R, where R is the gas constant, $8.314\,\mathrm{J{\cdot}mol^{-1}{\cdot}K^{-1}}$) across the dataset for the four lattice types. For most structures, the configurational entropy exceeds $1.5R$, which falls within the range of high-entropy materials, according to previous studies. This observation suggests that the AE-RF model successfully identifies structures with intrinsically favorable thermodynamic characteristics. However, three prominent peaks are observed around $0R$, $0.7R$, and $1.0R$, corresponding to monometallic oxides, binary and ternary compositions with metals in equal molar ratios, respectively. This can be attributed to the fact that equimolar compositions exhibit higher chemical uniformity and lattice symmetry compared with compositions having uneven elemental ratios, which reduces internal strain and stabilizes the crystal structure. Consequently, even though their overall configurational entropy is relatively low, the AE-RF model recognizes them as thermodynamically favorable due to the energetically advantageous arrangement of atoms in these proportionally balanced systems. In

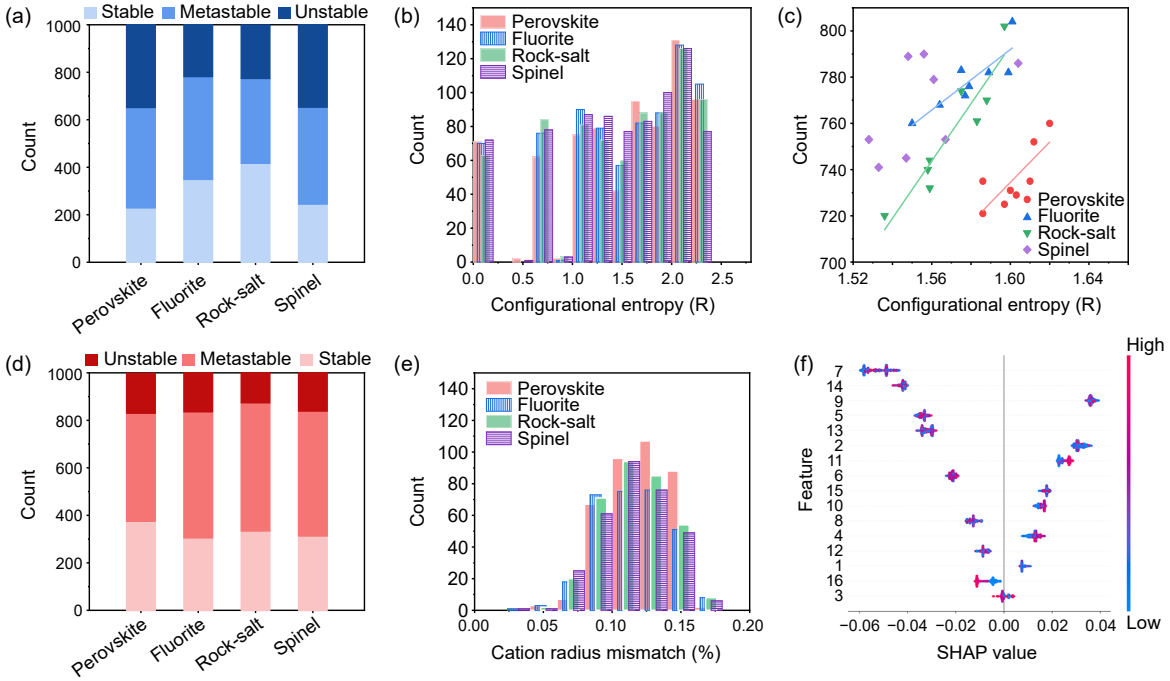

Fig. 4. (a) Numbers of stable, metastable, and unstable cases predicted by the AE-RF framework for oxide structures with perovskite, fluorite, rock-salt, and spinel lattices. (b) Histogram of the configurational entropy of oxides. (c) Correlation between the number of low-hull structures predicted by the AE-RF framework and the average configurational entropy across the four oxide lattice-type datasets. (d) Numbers of stable, metastable, and unstable structures predicted by the AE-RF framework for HEO structures. (e) Histogram of the configurational entropy of HEOs. (f) SHAP summary plot for stable HEOs in the AE-RF algorithm.

Fig. 4(c), a larger number of novel oxides were explored within the current workflow. The relationship between configurational entropy and the number of low-hull oxide structures was illustrated, with each lattice type represented by eight independent datasets containing 1000 structures each. A clear positive linear correlation is observed for perovskite, fluorite, and rock-salt lattices, where the number of low-hull structures increases steadily with increasing configurational entropy. Configurational entropy reflects the degree of atomic disorder in a material, and higher entropy generally enhances structural stability by lowering the free energy, thereby stabilizing complex solid solutions. The consistency between the predicted trends and established thermodynamic principles indicates that the AE-RF framework successfully identifies relatively stable oxide structures with high reliability. However, the spinel lattice exhibits a much weaker correlation, with the low-hull structure count showing only a limited dependence on configurational entropy. This is because spinel structures (AB_2O_4) contain two nonequivalent cation sites: tetrahedrally coordinated sites (A sites) and octahedrally coordinated sites (B sites). These two types of sites differ significantly in coordination environment, bond lengths, and local symmetry. This leads to the system's free energy being governed more strongly by local interactions rather than entropy effects.

To more rigorously demonstrate the screening capability of the AE-RF model for HEOs, stricter generation towards HEOs were applied. Specifically, the generated structures were required to exhibit high-entropy characteristics, with at least one of the A or B sites composed of five different metal elements in approximately equal proportions. According to previous studies, HEOs with five metals at the A site and one metal at the B site were generated, and the low-hull cases were selected using the AE-RF algorithm. To visually assess the reliability of the predictions, the numbers of stable, metastable, and unstable HEO structures predicted by the AE-RF framework for 1000 perovskite-, fluorite-, rock-salt-, and spinel-type oxides are shown in Fig. 4(d). It can be clearly observed that the proportion of low-hull structures is higher compared to the oxide dataset without strict entropy requirements. Since both the number and the ratio of components are strictly defined during the generation of HEOs, cation radius mismatch serves as a better descriptor of HEO stability than configurational entropy. Figure 4(e) shows the distribution of cation radius mismatch ratios for materials predicted to be stable or metastable across the four lattice types. It can be seen that the mismatch ratios are all below 0.15, indicating that the screened novel HEOs indeed possess excellent thermal stability, consistent with the cation radius mismatch theory for HEOs. Interestingly, elements that are less commonly used in practical applications, such as La and Nd at the A site, and Zr and Hf at the B site, also appear in the candidate HEOs. This demonstrates the superior predictive capability of the model: although it was trained on common crystal structures where these rare elements are underrepresented, it still accurately identifies stable structures containing these rarely studied metals. In a dataset consisting of 13608 structures containing La, Nd,

Table 2. Top 5 correlated physical features related to latent features with highest importance in HEOs stability prediction, as determined by SHAP values.

Index	Description	Top 5 related physic features
7	Density feature	Kurtosis of the interatomic distance (−0.410), Standard deviation of the thermal conductivity (−0.371), Lattice b (0.291), Number of magnetic sites (0.275), Number of atoms (−0.259)
14	Lattice feature	Lattice angel γ (0.337), Lattice b (−0.316), Lattice angel α (−0.289), Average atomic mass (−0.269), Standard deviation of the thermal conductivity (−0.268)

Hf, and Zr downloaded from MP, the performance of the AE-RF framework was evaluated and is shown in Fig. S14 of the SM. The prediction accuracy for this rare-element dataset is 78.5%, which is only 4% lower than that obtained for a dataset of the same size containing common elements. This indicates that the AE-RF model still exhibits high transferability to elements that are relatively underrepresented in the training set. In addition, the SHAP summary plot for the rare-element dataset (Fig. S15 of the SM) shows a similar pattern to that of the common-element dataset, with almost no change in the ranking of latent feature importance. This further suggests that the model does not fail when handling structures containing La, Nd, Hf, and Zr. These elements possess large atomic masses and strong lattice distortion tendencies, both of which are known to intensify phonon scattering and thermal conductivity.[43,44]

To achieve a data-driven assessment of feature importance in distinguishing low-hull HEOs, SHAP analysis was performed, and the summary plot is shown in Fig. 4(f) (stable cases) and Fig. S16 of the SM (metastable and unstable cases). Table 2 lists the top five physical features most strongly correlated with the latent features identified as most important for HEO stability prediction: Latent 7 for stable and metastable cases, and Latent 14 for unstable cases. Compared with the important latent features identified based on Gini importance in Table 1, the top two latent features for HEO stability prediction have constituent physical features almost identical to those of the density feature and the lattice feature, except for the newly appearing standard deviation of thermal conductivity and the number of magnetic sites. This observation is particularly intriguing, as it suggests that thermal transport properties and magnetic characteristics of constituent elements can play a significant role in determining the stability of HEOs. The number of magnetic sites is positively correlated with Latent 7, with a correlation coefficient of 0.275. This relatively low correlation coefficient suggests that an increase in the number of magnetic atoms in the lattice slightly increases the value of Latent 7. However, the red and blue points largely overlap and are concentrated in the negative and positive SHAP regions in Fig. 4(f) and Fig. S16a of the SM, indicating that both high and low values of this feature tend to decrease the predicted probability of the stable class and increase the predicted probability of the metastable class. This indicates that the importance of the latent feature in the RF model is not simply linearly separable; that is, the magnitude of the latent feature cannot be used alone to distinguish whether a crystal structure belongs to the stable or metastable class. Since there is no explicit functional relationship between the prediction results and the latent vectors, we can only conclude that the number of magnetic atoms plays a certain role in distinguishing stable from metastable structures; however, this effect is relatively weak and cannot be described by a simple function. Besides the number of magnetic sites, the standard deviation of thermal conductivity is negatively correlated with Latent 7 and Latent 14, with correlation coefficients of −0.371 and −0.268, respectively. This implies that larger differences in the thermal conductivity of atoms in the lattice lead to lower values of Latent 7 and Latent 14. Compared with the number of magnetic atoms, this property has a greater influence on the prediction results and contributes to distinguishing among stable, metastable structures and unstable structures. Physically, variations in thermal conductivity among constituent cations may influence lattice vibrations and phonon scattering, thereby affecting overall lattice stability. Similarly, the presence of magnetic sites could modify local magnetic ordering, which in turn may stabilize configurations through reorganization of magnetic moments. The absence of the symmetry feature in Table 2 suggests that symmetry is no longer a dominant predictor for HEO stability, because once the primitive cell of the constituent elements is determined, the symmetry is effectively fixed. Nevertheless, the lattice feature still plays a significant role in the prediction, indicating that HEOs retain the same sensitivity to lattice variations as conventional crystals. These results highlight that, beyond conventional structural descriptors, intrinsic electronic and thermal properties of the components are critical factors in HEO stability, offering valuable guidance for the rational design of new HEOs with tailored thermal and magnetic functionalities. By the way, using AE-RF, for 1000 newly generated HEO structures, structural descriptors were extracted, compressed into latent features using the trained AE, and subsequently classified by the RF model in only 671 seconds by using CPU

only. For comparison, a typical DFT relaxation and total-energy evaluation of a $LaCoO_3$-derived oxide on a standard HPC compute node (e.g., 32 cores, 2.6–3.0 GHz) usually requires 3–6 hours per structure, even with well-converged settings. Screening 1000 structures would therefore take approximately 3000–6000 core-hours. Consequently, our ML pipeline offers a computational acceleration exceeding 2×10^4, enabling large-scale exploration of complex composition space that would be prohibitively time-consuming using first-principles methods.

4. Conclusion. In this study, a comprehensive ML study on the prediction of crystal structure stability based on E_{hull} was conducted by systematically analyzing over 120000 compounds from MP. Among the tested algorithms, the AE-RF model achieved the best overall performance, with an accuracy of 84.3% in distinguishing between stable, metastable, and unstable materials. Beyond its superior classification performance, the model maintained an accuracy exceeding 75% even when only 40% of the data were used for training. A systematic feature importance analysis further identified density, symmetry–bonding, and lattice latent features as crucial contributors to the model's performance. Interestingly, our results highlight that the classification of metastable materials with E_{hull} less than 120 meV/atom presents inherent challenges that persist even after the sample size was balanced.

By constructing and evaluating four types of HEO structures, the capability of the AE-RF framework to identify novel HEOs efficiently was further validated. In particular, the AE-RF model effectively identifies low-hull HEO candidates, selecting structures with intrinsically favorable thermodynamic characteristics such as high configurational entropy ($> 1.5R$) and low cation radius mismatch (< 0.15). More importantly, the standard deviation of atomic thermal conductivity in the lattice and the number of magnetic atoms were, for the first time, found to play important roles in the stability of HEOs. In this work, the AE successfully identified that a uniform distribution of atomic thermal conductivities and the regulation of magnetic moments throughout the crystal can effectively enhance the stability of HEOs.

In addition to classifying materials based on their E_{hull} values and training models for this classification task, ensuring the absence of imaginary phonon modes is essential for confirming stability. However, first-principles phonon calculations based on DFT are computationally expensive, particularly for large or high-entropy systems.[45] Recent advances in ML offer promising alternatives, primarily through predicting the phonon density of states (DOS) using deep learning models. Representative algorithms, including cluster-based graph models,[46] symmetry-aware neural networks,[47] and attention-based graph neural networks,[48] have demonstrated the ability to capture the relationship between structure and phonon spectrum, reducing computational cost while maintaining reasonable accuracy. Despite these progresses, predicting phonon DOS remains challenging due to its high-dimensional, continuous, and strongly correlated nature, as well as limited data availability and complex chemical environments.[23] Future efforts should focus on integrating physical priors, improving spectral representations, and developing data transfer strategies to enhance model robustness and transferability. These advances are expected to facilitate reliable and efficient prediction of dynamic stability, thereby accelerating the discovery of novel materials with desirable phononic and thermal properties.

Overall, this study establishes a reliable and interpretable ML framework for the prediction of crystal structure stability and demonstrates its utility in accelerating the discovery of energetically accessible HEOs. This approach also paves the way for data-driven strategies in the rational design of multicomponent materials.

5. Methods.

5.1. Algorithm and Hyperparameters. To explore the influence of training-validation-test set splits on training performance, the sizes of the validation and test sets were fixed, while the training set was varied to account for 60%, 70%, 80%, and 90% of the total samples. RF is an ensemble learning algorithm that builds multiple decision trees during training and aggregates their predictions to improve accuracy and control overfitting. In this work, the following parameters were optimized to achieve optimal predictive performance: the number of decision trees, the maximum depth for each decision tree, the minimum number of samples required to split an internal node, the minimum number of samples required to be at a leaf node, the number of features evaluated when determining the optimal split at a node, and whether bootstrap is used in sampling. The principle of GridSearchCV involves exhaustively exploring all combinations of hyperparameters within a predefined search space. For each combination, performance is evaluated via cross-validation, and the set of hyperparameters yielding the highest performance on the validation set is selected. As a data-driven tool, it is widely utilized for automated hyperparameter tuning. However, its computational complexity escalates significantly when the hyperparameter search space is large. For the RF algorithm employed in this study, the aforementioned parameters were optimized, with the optimal values summarized in Table S2 of the SM and the ranges of parameter selection provided in Table S3 of the SM.

5.2. Feature Selection and Description. Based on the structure of the MP database, 35 features related to composition and structure that do not require first-principles calculations were considered in this work, including the mean and standard deviation of elemental atomic mass, atomic radius, melting point, thermal conductivity, and electronegativity of all elements in the structure; structural and lattice descriptors consisting of lattice parameters (a, b, c, α, β, γ), lattice volume, lattice anisotropy, and the maximum singular value (MSV) of the lattice matrix. MSV refers to the largest scaling factor of the crystal lattice matrix under an orthogonal transformation, indicating the maximum deformation or elongation of the lattice. A lattice matrix is a 3×3 matrix describing the vectors that define the unit cell of the crystal in 3D space. When such a

matrix A is decomposed into three matrices: $A = U\Sigma V^{\mathrm{T}}$, where U is an orthogonal matrix (3×3), Σ is a diagonal matrix containing the singular values (three values for a 3×3 matrix), and V^{T} is the transpose of another orthogonal matrix (3×3). The singular values are stored in Σ, which represent the stretching factors of the lattice along the principal axes of the matrix. They are always non-negative and sorted in descending order. Besides, the mean and standard deviation of coordination numbers, average coordination, coordination uniformity, the mean and standard deviation of interatomic distances,skewness and kurtosis of interatomic distances were used to quantify the distribution of local bonding environments. Global structural attributes, such as the number of atoms per unit cell, number of constituent elements, atomic density, and the presence of magnetic sites, are also included. Chemical and energetic features, including average cohesive energy and the metal–nonmetal ratio, further capture bonding strength. Finally, crystallographic symmetry is encoded through the crystal system and the space group index. Together, these descriptors provide a comprehensive representation of periodic multicomponent structures.

5.3. Construction of HEO. In this work, HEO structures were systematically constructed on the basis of the parent perovskite $LaCoO_3$ (mp-1185232, $E_{\mathrm{hull}} = 47\,\mathrm{meV/atom}$), rock-salt MgO (mp-1265, $E_{\mathrm{hull}} = 0$), fluorite CeO_2 (mp-20194, $E_{\mathrm{hull}} = 0$), and spinel $MgAl_2O_4$ (mp-3536, $E_{\mathrm{hull}} = 0$). Specifically, the A-site (La, Mg, Ce) and B-site (Co, Mg, Ce, and Al) atoms were partially substituted with candidate elements. For the A-site or B-site, all possible combinations containing 5 distinct elements were enumerated; this procedure ensured that each configuration consisted of at least five different elements distributed across the A- and B-sites. Once a given combination was selected, the corresponding elements were evenly allocated to the available crystallographic sites. Each resulting configuration was exported as a CIF file for subsequent structural optimization and property evaluation.

5.4. Other Methods. t-SNE: A dimensionality reduction technique primarily used for visualizing high-dimensional data in lower-dimensional space. In this work, to ensure the speed of clustering, 10000 randomly selected data points from the total dataset were used, with the perplexity set to 50. The number of clusters was automatically chosen within the range of 2 to 20, aiming to maintain a high silhouette score. The silhouette score is a metric used to evaluate the quality of clustering in unsupervised learning. It measures how well each data point fits within its assigned cluster compared to other clusters. The score is calculated for each data point and ranges from -1 to 1. For a data point i,

$$s(i) = \frac{b(i) - a(i)}{\max[a(i), b(i)]},$$

where $a(i)$ is the average distance of i to all other points within its own cluster (intra-cluster distance), and $b(i)$ is the average distance of i to all points in the nearest neighboring cluster (inter-cluster distance). A score close to $+1$ indicates that the point is well-clustered and well-separated from other clusters.

Data and Code Available. This work was implemented using Python, and all the programs can be found on the following webpage: github.com/zhouqc2022/stability_convex_hull.

Acknowledgements. This work was supported by the National Key Research and Development Program of China (Grant No. 2026YFE0199500), the National Natural Science Foundation of China (Grant No. 52472205), the Fundamental Research Funds for the Central Universities (Grant Nos. CCNU25ZH006 and JC2026TS-006), and the Hubei Provincial Natural Science Foundation of China (Grant No. 2025EHA032).

References

[1] Moosavi S M, Jablonka KM, and Smit B 2020 *J. Am. Chem. Soc.* **142** 20273
[2] Golmohammadi M and Aryanpour M 2023 *Mater. Today Commun.* **35** 105494
[3] Utela B, Storti D, Anderson R, and Ganter M 2008 *J. Manuf. Process.* **10** 96
[4] Valero A, Valero A, Calvo G, and Ortego A 2018 *Renew. Sustain. Energy Rev.* **93** 178
[5] Ramprasad R, Batra R, Pilania G, MannodiK A, and Kim C 2017 *npj Comput. Mater.* **3** 54
[6] Zhu L G, Zhou J, and Sun Z M 2022 *J. Phys. Chem. Lett.* **13** 3965
[7] Ward L, Agrawal A, Choudhary A, and Wolverton C 2016 *npj Comput. Mater.* **2** 16028
[8] Xu P C, Ji X B, Li M J, and Lu W C 2023 *npj Comput. Mater.* **9** 42
[9] Kirklin S, Saal J E, Meredig B, Thompson A, Doak J W, Aykol M, Rühl S, and Wolverton C 2015 *npj Comput. Mater.* **1** 15010
[10] Carbogno C, Thygesen K S, Bieniek B *et al.* 2022 *npj Comput. Mater.* **8** 69
[11] Kondakci H E, Abouraddy A F, and Saleh B E A 2015 *Nat. Phys.* **11** 930
[12] Thakur A and Anasori B 2024 *Science* **383** 1182
[13] Pyzer-Knapp E O, Pitera J W, Staar P W J, Takeda S, Laino T, Sanders D P, Sexton J, Smith J R, and Curioni A 2022 *npj Comput. Mater.* **8** 84
[14] Tabor D P, Roch L M, Saikin S K *et al.* 2018 *Nat. Rev. Mater.* **3** 5
[15] Juan Y F, Dai Y B, Yang Y, and Zhang J 2021 *J. Mater. Sci. Technol.* **79** 178
[16] Liu Y Y, Mu Z L, Hong P C, Yang Y, and Lin C X 2025 *Nanoscale* **17** 2200
[17] Wang Z L, Funada T, Onda T, and Chen Z C 2023 *Mater. Today Phys.* **31** 100971
[18] Dinic F, Singh K, Dong T, Rezazadeh M, Wang Z B, Khosrozadeh A, Yuan T G, and Voznyy O 2021 *Adv. Funct. Mater.* **31** 2104195
[19] Singh A K, Zhou L, Shinde A, Suram S K, Montoya J H, Winston D, Gregoire J M, and Persson K A 2017 *Chem. Mater.* **29** 10159
[20] Aykol M, Dwaraknath S S, Sun W H, and Persson K A 2018 *Sci. Adv.* **4** 0148
[21] Therrien F, Jones E B, and Stevanović V 2021 *Appl. Phys. Rev.* **8** 031310
[22] Liang J Q, Liu J X, Jing Y F, Yang H C, Zhou S M, Xia Y P, Xu F, Sun L X, and Huang P R 2025 *Mater. Des.* **255** 114157
[23] Pan D K, Li T H, Wan X, Zong Z C, Qin Y J, and Yang N 2025 *Chin. Phys. Lett.* **42** 070801

[24] Wang T T, Li X Z, Ju Z, Zhang G, Ma D K, Li W, and Zhang L F 2025 *Chin. Phys. Lett.* **42** 070802
[25] Chung K M, Adapa S R, Pei Y, Yeerella R H, Chen L, Shivakumar S, Huang W, Liu Z W, Cai S Q, Luo J, and Chen R K 2025 *Adv. Mater.* **37** 2406732
[26] Olsson P, Abrikosov I A, Vitos L, and Wallenius J 2003 *J. Nucl. Mater.* **321** 84
[27] Emery A A and Wolverton C 2017 *Sci. Data* **4** 170153
[28] Korzhavyi P A, Abrikosov I A, Johansson B, Ruban A V, and Skriver H L 1999 *Phys. Rev. B* **59** 11693
[29] Wuliji H, Zhao K P, Cai X M, Jing H R, Wang Y W, Huang H R, Wei T R, Zhu H, and Shi X 2023 *Mater. Today Phys.* **35** 101129
[30] Xie Y L, Li X, Wang Y, Li B W, Yang L, Zhao N, Liu M F, Wang X Z, Yu Y, and Liu J M 2020 *Appl. Surf. Sci.* **499** 143964
[31] Bartel C J, Trewartha A, Wang Q, Dunn A, Jain A, and Ceder G 2020 *npj Comput. Mater.* **6** 97
[32] Sun W H, Holder A, Orvañanos B, Arca E, Zakutayev A, Lany S, and Ceder G 2017 *Chem. Mater.* **29** 6936
[33] Liu M, Rong Z Q, Malik R, Canepa P, Jain A, Ceder G, and Persson K A 2015 *Energy Environ. Sci.* **8** 964
[34] Li W, Jacobs R, and Morgan D 2018 *Comput. Mater. Sci.* **150** 454
[35] Sun W H, Dacek S T, Ong S P, Hautier G, Jain A, Richards W D, Gamst A C, Persson K A, and Ceder G 2016 *Sci. Adv.* **2** e1600225
[36] Lyngby P and Thygesen K S 2022 *npj Comput. Mater.* **8** 232
[37] Young S D, Chen J D, Sun W H, Goldsmith B R, and Pilania G 2023 *Chem. Mater.* **35** 5975
[38] Shen C, Gao Q, Fortunato N M, Singh H K, Opahle I, Gutfleisch O, and Zhang H B 2021 *J. Mater. Chem. A* **9** 8805
[39] Twyman N M, Walsh A, and Buonassisi T 2022 *Chem. Mater.* **34** 2545
[40] Anelli A, Engel E A, Pickard C J, and Ceriotti M 2018 *Phys. Rev. Mater.* **2** 103804
[41] Takahiro I and Takashi M 2020 *Phys. Rev. B* **101** 214106
[42] Sørensen S S, Pedersen E J, Paulsen F K, Adamsen I H, Laursen J L, Christensen S, Johra H, Jensen L R, and Smedskjaer M M 2020 *Appl. Phys. Lett.* **117** 31901
[43] Zhang H S, Tong Y P, Yang X F, Sang W W, Zhang H M, and Zhao Y T 2020 *Ceram. Int.* **46** 2862
[44] Yu X X, Ma D K, Deng C C, Wan X, An M, Meng H, Li X B, Huang X M, and Yang N 2021 *Chin. Phys. Lett.* **38** 014401
[45] Zong Z C, Chen X D, Yan B, Xie Y L, Pang J, Li G Y, Hu J Q, Wu Z P, Li B, Fang H S, and Yang N 2024 *Chin. Phys. Lett.* **41** 106301
[46] Liang C, Rouzhahong Y, Yao S W, Liang J H, Yu C L, Wang B, and Li H 2024 *Adv. Sci.* **11** 2406183
[47] Chen Z T, Andrejevic N, Smidt T, Ding Z W, Xu Q, Chi Y T, Nguyen Q T, Alatas A, Kong J, and Li M D 2021 *Adv. Sci.* **8** 2004214
[48] Al-Fahdi M, Lin C P, Shen C, Zhang H B, and Hu M 2025 *Mater. Today Phys.* **50** 101632