

J Adv Ceram 2026, **15**: 9221385

<https://doi.org/10.26599/JAC.2026.9221385>

Research Article

Machine learning empowered compositional design of multiple rare-earth principal component disilicates

Yixiu Luo¹, Xinyu Gao¹, Ziyu Wang^{1,2}, Cui Zhou³, Jiemin Wang¹, Tiefeng Du¹,
Luchao Sun^{1,*}, Jingyang Wang^{3,*}, Ying Xiong⁴

¹*Shenyang National Laboratory for Materials Science, Institute of Metal Research,
Chinese Academy of Sciences, Shenyang 110016, China*

²*School of Materials Science and Engineering, University of Science and Technology
of China, Shenyang 110016, China*

³*Institute of Coating Technology for Hydrogen Gas Turbines, Liaoning Academy of
Materials, Shenyang 110167, China*

⁴*AECC Shenyang Liming AERO-ENGINE Science and Technology Co., Ltd.,
Shenyang 110043, China*

*Corresponding author.

Email: L. Sun, lcsun@imr.ac.cn;

J. Wang, jywang@imr.ac.cn

Abstract

The targeted design of multi-RE-principal-component $\text{RE}_2\text{Si}_2\text{O}_7$ disilicates ($(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$) for environmental barrier coatings (EBCs) applications requires customizing the multi-RE compositions to achieve maximal optimization of the target properties. A critical prerequisite is the retention of a stable β - or γ -polymorphic phase under high-temperature service conditions. This, however, is challenged by their rich polymorphic phases, which varies with the elemental properties of the RE cationic sites. In this study, a random forest (RF) model with high accuracy is developed to classify the four types of phase composition – single- β , single- γ , single- δ /mixed $\delta+\gamma$, and separate phase – identifying the average RE^{3+} cationic radius ($\bar{r}$) and the deviation of RE^{3+} cationic radius (σ_r) as the most influential factors. The well-trained model is validated by predicting the phase compositions of $(\text{Gd}_{x1}\text{Ho}_{x2}\text{Yb}_{x3}\text{Lu}_{x4})_2\text{Si}_2\text{O}_7$ and $(\text{Nd}_{x1}\text{Ho}_{x2}\text{Yb}_{x3}\text{Lu}_{x4})_2\text{Si}_2\text{O}_7$ systems, supported by experimental characterization of representative compositions. High-throughput DFT calculations reveal that the formation of their phases correlates with the low energy costs to accommodate configurational randomness into the multicomponent system, characterized by rapid convergence of the configurational entropy of mixing with increased excitation energy.

The quantitative design criteria for single-phase β -($n\text{RE}_{xi}$)₂Si₂O₇ and γ -($n\text{RE}_{xi}$)₂Si₂O₇ disilicates are established: (i) $\bar{r} < 0.885 \text{ \AA}$ for β -polymorphs and $0.885 \text{ \AA} < \bar{r} < 0.900 \text{ \AA}$ for the γ -polymorphs; and (ii) sufficiently small σ_r , whose upper bound increases monotonically with $\bar{r}$, reaching $\sigma_r \sim 0.04$ at the vicinity of $\bar{r} \sim 0.885$ and $0.900 \text{ \AA}$. This work provides an investigation paradigm enabling the targeted design of ($n\text{RE}_{xi}$)₂Si₂O₇ EBCs candidates.

Keywords

High-entropy rare-earth disilicates, Phase composition, Machine learning, Environmental barrier coatings

1. Introduction

In the gas turbine technology, the development of environmental barrier coatings (EBCs) is a critical research branch – specifically aimed at protecting the SiC_f/SiC ceramic matrix composites (CMCs) engine components from high-temperature high-velocity steam-induced oxidation and recession during service [1,2]. Rare-earth disilicates (RE₂Si₂O₇) show great promise for the EBCs applications due to their good CTE compatibility (specially for the β - and γ -polymorphs) with SiC_f/SiC CMCs substrates [3-7]. They have been integrated as a critical intermediate layer in the state-of-the-art multilayer EBCs to effectively mitigate the CTE mismatch between the SiC_f/SiC CMCs substrates and the top coats, which minimizes the inter-layer thermal stress during thermal cycling, thereby preventing delamination and coating failure [8]. Constrained by up to seven polymorphic types and the complex polymorphic phase transformations amongst [9], the commercialized research and development of

RE₂Si₂O₇ disilicates for EBCs applications are mostly concentrated on β -RE₂Si₂O₇ (RE = Sc, Yb, and Lu) disilicates mainly due to their good phase stability up to melting point [6]. Meanwhile, γ -RE₂Si₂O₇ disilicates are also being developed as acceptable EBCs materials, because the high-temperature γ phase can be kinetically stabilized down to lower temperatures; plus, the volume change associated with $\beta \rightarrow \gamma$ polymorphic transformations is extremely small, due to the crystallographic similarity between these two polymorphic structures as well as their identical CTE values [10], making the product phases relatively benign during coating service. However, the performance of traditional single-RE-principal-component β -RE₂Si₂O₇ and γ -RE₂Si₂O₇ EBCs under high-temperature corrosive environment cannot meet the requirements of the next-generation gas turbine engine [11], and therefore need to be comprehensively optimized.

A promising strategy to overcome these challenges is to design multi-RE-principal-component RE₂Si₂O₇ disilicates (abbreviated as $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$). The concept of high-entropy or multicomponent materials was originally proposed in metallic alloys in 2004 [12], subsequently extended to ceramics in 2015 [13]; and has since undergone rapid development in a wide range of material systems, with a particular burst in high-temperature protective coating materials [14-21]. Recently, significant progress has been achieved in the development of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ materials, wherein a collection of properties, including superior high-temperature stability, low or even amorphous-like thermal conductivity, matched CTE with SiC_f/SiC CMCs, excellent resistance against water vapor and CMAS corrosion, etc. have been reported, showing remarkable optimizations as compared to the constituent single-RE-principal-component RE₂Si₂O₇

materials [22-25]. Particularly, dedicated efforts have been focused on elucidating the critical role of individual RE constituents as well as their synergistic effects in tailoring the specific properties of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ materials. For instance, Sun et al. reported that RE elements with larger cationic radius (such as Er and Tm) take effects in suppressing the corrosion activity of CMAS melts, whereas those with smaller cationic radius (such as Yb and Lu) helps retaining the structural stability of disilicates [23]; and thus, the overall CMAS corrosion resistance could be optimized by adjusting the relative proportion of the two types of RE elements [26]. Wang et al. reported that the incorporation of Sc into the multi-RE (RE = lanthanides) composition enables electrons transfer from Si-O bond to Sc, facilitated by the local charge disorder, and thereby yields adjustable CTE values [27]. Thanks to these efforts, the cutting-edge development of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ materials is progressively transitioning from a conventional trial-and-error paradigm to a rational, target-driven design framework. Within this framework, the multi-RE combinations can be exquisitely customized by leveraging the distinct functional roles of individual RE elements, and then the $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ compositions can be fine-tuned to retain a desired polymorphic structure, thereby harvesting the maximum optimization on target properties.

The prime challenge in precisely designing $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ EBCs candidates is the fast identification of available multi-RE combinations – from all the permutations and combinations of seventeen RE elements – that can form into a single-phase β - or γ -polymorph with good high-temperature stability. Many pioneering works have revealed that the stabilized phase composition of high-entropy or multicomponent ceramics,

such as oxides, carbides, borides, silicates, etc., should be governed by numerous factors, including lattice constants, bulk modulus, Fermi level, band gap, atomic radius, molar fraction, valence electron concentration, mixing enthalpy, and the configurational entropy, just to name a few [28-31]. However, such knowledge may not be directly transferred to the $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ systems, in that the prediction of their stabilized phase composition is a non-independent multi-objective problem, due to their complex crystal structures and various polymorphic types. To circumvent this obstacle, one may leverage the recent prosperity of machine learning (ML) method, which is particularly well-suited for modeling problems involving non-independent variables and capable of establishing a relatively straightforward physics-informed relationship with the target property. Up till now, ML method has been innovatively utilized to investigate the phase formation capability, mechanical, thermal and corrosion properties of single-RE-principal-component or multi-RE-principal-component silicates [31-38]. Particularly, Meng et al.'s work has established a multi-class classification ML model to predict the polymorphic phases of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ disilicates, identifying the average ionic radius as a single criterion in distinguishing the phases [39]. And, Fan et al.'s work has developed a ML model combined with decision fusion strategies to predict the phase formation capability of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ disilicates, revealing that phase stability is governed by descriptors such as average ionic radius and ionic radius deviation, average mass, and electronegativity deviation, with their relative importance varying with the compositional complexity [31]. These studies demonstrate the capability of ML in decoding the composition-phase relationships within the $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ systems; and

together with our prior work [40], consistently identified the ionic radius – average and deviation – as a critical descriptor governing the stabilized phases. In this respect, further efforts to outline the quantitative boundaries of the ionic-radius-based descriptors for stabilizing specific polymorphs would be of significant value, as it could provide direct design guidelines for rapidly screening viable multi-RE combinations and, more importantly, for quickly identifying the most practically useful β - and γ -phase candidates for EBCs applications.

In this study, the phase composition of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ is investigated by combining machine-learning, high-throughput calculation and experimental validation. Firstly, the phase composition of 114 $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ disilicates is quantitatively investigated by ML algorithms, with the aim of developing a most reliable ML model as well as elucidating the key factors determining the phase compositions of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$. Subsequently, the optimal random forest model (RF) is utilized to predict the phase composition of $(\text{Gd}_{x1}\text{Ho}_{x2}\text{Yb}_{x3}\text{Lu}_{x4})_2\text{Si}_2\text{O}_7$ and $(\text{Nd}_{x1}\text{Ho}_{x2}\text{Yb}_{x3}\text{Lu}_{x4})_2\text{Si}_2\text{O}_7$ systems, which validates the robustness of the RF model. Among these, six representative samples are synthesized, and their phase composition is examined through X-ray diffraction (XRD) measurements, Rietveld refinement analysis, scanning electron microscope (SEM) observation, and transmission electron microscope (TEM) characterizations. The underlying mechanisms governing the formation of their phases are revealed by evaluating the configurational entropy of mixing through high-throughput density functional theory (DFT) calculations coupled with statistical mechanics theory. Finally, the phase compositions of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ materials are projected onto the dual-

descriptor space of average RE^{3+} cationic radius ($\bar{r}$) and the deviation of RE^{3+} cationic radius (σ_r). The critical thresholds of $\bar{r}$ and σ_r parameter as well as their interdependence are established, which set the criteria of designing single-phase β - $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ and γ - $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ disilicates as potential candidates for advanced EBCs applications.

2. Methods

2.1 Machine learning model

The training dataset contains the phase composition of 114 $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ materials collected from the published papers as well as our own experiments, including 40 materials in single- β phase, 36 materials in single- γ phase, 8 materials in single- δ phase or mixed $\delta+\gamma$ phase due to the reversible $\gamma \rightleftharpoons \delta$ phase transformation, and 30 materials showing separate phase (listed in Supplementary Table S1). These compositions are widely collected to sample over various RE elements (lanthanides and Y), while considering different total number of multi-RE components, in equimolar and non-equimolar ratios. And, the phase composition information was manually checked to ensure that only the polymorphic phases retained after high-temperature processing were designated for each compound, whereas any low-temperature phase appeared intermediately during materials synthesis were excluded. The machine learning model was constructed based on Random Forest (RF), Support Vector Machine (SVM), and Artificial Neural Network (ANN) algorithms using the scikit-learn Python package [41,42]. The hyperparameters of the ML models, such as the number of decision trees and the maximum depth in RF model, the kernel type and regularization parameters in

SVM model, the number of hidden layers and the number of neurons in each layer in ANN model, etc., were optimized via a grid-search technique combined with fivefold cross validation to enhance the model generalization and avoid overfitting.

The descriptors utilized in this study include the fundamental physical and chemical properties of RE elements (ionic radius, atomic mass and electronegativity), taken as their weighed average and deviation:

$$\bar{p} = \sum_{i=1}^n x_i \cdot p_i \quad (1)$$

$$\sigma_p = \sqrt{\sum_{i=1}^n x_i \cdot (p_i - \bar{p})^2} \quad (2)$$

Herein, $\bar{p}$ and σ_p are the weighed average and standard deviation of a considered elemental parameter p , which could be specified as the RE³⁺ cationic radius (r), atomic mass (m), and electronegativity (X); n is the total number of RE components in a $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ composition; p_i and x_i are the elemental parameter and molar fraction of the i^{th} RE component, wherein $\sum_{i=1}^n x_i = 1$. Another descriptor taken into consideration is the theoretical configurational entropy of mixing ($S_{\text{config}}^{\infty}$) for an ideal $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ solid solution:

$$S_{\text{config}}^{\infty} = -R \left[2 \sum_{i=1}^{N^{\text{RE}}} x_i \ln(x_i) + 2 \sum_{j=1}^{N^{\text{Si}}} x_j \ln(x_j) + 7 \sum_{k=1}^{N^{\text{O}}} x_k \ln(x_k) \right] \quad (3)$$

wherein, $R=8.31 \text{ J} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$ is the ideal gas constant; N^{RE} , N^{Si} and N^{O} are the number of components in the sublattice of RE, Si and O, and x_i , x_j and x_k are their molar fraction.

These descriptors have been proved to be effective in predicting the formation capability of high-entropy alloys and ceramics [30,43]. In total, seven descriptors ($\bar{r}$, σ_r , $\bar{m}$, σ_m , $\bar{X}$, σ_X , $S_{\text{config}}^{\infty}$) for all samples constituting a 114×7 matrix was randomly partitioned into nonoverlapping training subset (80%) versus testing subset (20%).

Feature scaling was performed using a standard scaler fitted exclusively on the training data, with the same parameters subsequently applied to transform the test set, thereby preventing data leakage from the test set into the preprocessing stage. The rationale of constructing the training dataset and the descriptor set is presented in the Supplementary Note 1.

To address the class imbalance issue in the training dataset – particularly the underrepresentation of the single- δ /mixed $\delta+\gamma$ category (only 8 samples), the Synthetic Minority Over-sampling Technique (SMOTE) was applied to balance the four phase-composition categories prior to model training, with three nearest neighbors used to define the neighborhood during synthetic sample generation. Herein, the SMOTE was applied exclusively to the training subset, so that the synthetic samples never enter the test subset. This prevents the optimistic bias that would arise from evaluating the model on synthetic data generated from the same minority-class instances used for training. A detailed discussion of the SMOTE implementation and its influence on model generalization is provided in Supplementary Note 2.

2.2 Configurational entropy of mixing

The temperature-dependent configurational entropy of mixing for the $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ materials is calculated based on the equilibrium statistical mechanics. An ensemble constituted by large amounts of metastable configurations is constructed, wherein the multiple RE^{3+} cations are randomly distributed on the sublattice sites. The partition function is used to describe how the probabilities are partitioned among different metastable configurations according to their individual configurational

energies [28,44]:

$$Z = \sum_{l=1}^m \exp\left(-\frac{E_l}{k_B T}\right) \quad (4)$$

Herein, m is the total number of the metastable configurations in the ensemble, which is factorially related to the number of doped RE elements and the number of available RE sites within the lattice; k_B is the Boltzmann constant; T is the absolute temperature.

E_l is the mixing energy for the l^{th} metastable configuration:

$$E_l = E_{\text{mix}} - \sum_{i=1}^n x_i E_{\text{RE}^i} \quad (5)$$

where E_{mix} is the total energy of the $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ unit cell; E_{RE^i} is the total energy of the i^{th} single-RE-principal-component $\text{RE}_2\text{Si}_2\text{O}_7$, calculated using the same polymorphic structure with that of the $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ system; and n is the total number of RE components in the composition. Then, the configurational energy and the free energy (due to configurational disorders) could be written as:

$$\langle E \rangle = \frac{1}{Z} \sum_{l=1}^m E_l \exp\left(-\frac{E_l}{k_B T}\right) \quad (6)$$

$$F = -k_B T \ln Z \quad (7)$$

The ensemble average is performed over a representative subset of the whole population, restricted by:

$$\frac{Z}{m} = \frac{Z'}{m'} \quad (8)$$

Herein, m' is the total number of the sampled metastable configurations; and Z' is their partition function. Therefore, the free energy can be reformulated into:

$$F = -k_B T \ln m - k_B T \ln \frac{\sum_{i=1}^{m'} \exp\left(-\frac{E_i}{k_B T}\right)}{m'} \quad (9)$$

And, the configurational entropy of mixing is derived from:

$$S_{\text{config}} = (\langle E \rangle - F)/T \quad (10)$$

In this study, the β -type (γ -type, or δ -type) configurational ensembles are constructed to include 119 (114, or 31) unique configurations, each configuration containing 110 (110, or 220) lattice sites. The metastable configurations (supercells) in each ensemble are generated using the special quasirandom structures (SQS) generation code implemented in the Alloy Theoretic Automated Toolkit (ATAT) package [45]. The sizes of supercells are carefully chosen to ensure reliable reproduction of the compositional stoichiometry without causing unbearable computational expense. Herein, each configurational ensemble contains over 100 unique configurations (except for the δ -polymorph ensemble), which, based on the convergence test in our previous work, is expected to yield reliable simulation of the energy distribution among the metastable configurations within the ensemble [40]. Specially, the δ -type configurational ensemble is constructed using larger but fewer supercells, primarily due to the highly complex crystal structure of the δ -RE₂Si₂O₇ disilicate (44 atoms in the primitive unit cell with two types of crystallographically inequivalent RE sites). This means that larger “minimum cell” is required to reproduce the compositional stoichiometry, and consequently, the computational expense for the Density-functional-theory energy calculations is excessively high.

2.3 Details of DFT calculations

Density functional theory (DFT) calculations are conducted using the Vienna Ab Initio Simulation Package (VASP) [46]. The projector augmented wave (PAW) method [47] is used to describe the interactions between core electrons and valence electrons. The Perdew-Burke-Ernzerhof (PBE) [48] form of the generalized gradient

approximation (GGA) is used to describe the electronic exchange and correlation interactions. The energy cut-off for plane-wave basis is set as 600 eV. The Brillouin zone (BZ) is sampled using the Γ -centered Monkhorst-Pack [49] k -mesh in the resolution of $2\pi \times 0.03 \text{ \AA}^{-1}$. The convergence criteria for geometric optimization are set as 1×10^{-7} eV for the total energy and 5×10^{-3} eV/ $\text{\AA}$ for the ionic Hellmann-Feynman forces, respectively. The valence electron configuration is $5s^2 5p^6 5d^1 6s^2$ for Nd, and $5p^6 5d^1 6s^2$ for Gd, Ho, Yb and Lu.

2.4 Experimental

Commercially available powders of Nd_2O_3 , Gd_2O_3 , Ho_2O_3 , Yb_2O_3 , and Lu_2O_3 (Rear-Chem. Hi-Tech. Co. Ltd., Huizhou, China), and SiO_2 (Sinopharm Chemical Reagent Co. Ltd., Shanghai, China) were used as starting materials to fabricate the $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ samples. For each composition, RE_2O_3 and SiO_2 powders were weighed according to the desired stoichiometry, and then mixed with ethanol in a Si_3N_4 mill jar for 24 h. The obtained slurry was dried at 60°C for 24 h and then passed through a 120-mesh sieve. Pressure-less synthesis method was adopted to synthesize the $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ powders at 1550°C for 6 h. And, in order to sufficiently promote the mass transfer in the solid-state synthesis reaction, the as-synthesized $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ powders were hot-pressed in a BN-coated graphite mold at 1800°C under the pressure of 30 MPa for 2 h, in a flowing argon atmosphere. The phase compositions of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ samples were characterized by X-ray diffractometer (D/max-2400, Rigaku, Tokyo, Japan). Rietveld refinement method was used for quantitative phase analysis using the General Structural Analysis System (GSAS) program [50]. Phase identification and rare-earth

elements distribution of bulk samples were observed with a scanning electron microscope (SUPRA 55 SAPPHERE, Zeiss, Oberkochen, Germany) equipped with an energy dispersive spectrometer (X-MAX, Oxford Instruments, Oxford, UK), and the point analysis for each sample was repeated for at least 10 times. The atomic distributions of rare earth elements in $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ samples were characterized by a double aberration-corrected scanning transmission electron microscope (STEM) (Titan Themis Z, Thermo Fisher Scientific, Waltham, USA) equipped with high angle annular dark field (HAADF) detector.

3. Results and discussion

3.1 Machine learning prediction on phase composition

A correlation analysis on the input features is performed prior to ML modeling training, wherein highly correlated (i.e. correlation coefficient exceeding 0.8) and redundant descriptors should be identified and eliminated prior to model training. As shown in Fig. 1(a), the Spearman correlation coefficients (ρ) of the seven selected features are all below 0.8, indicating no significant correlation among them. The relatively higher negative correlation ($\rho = -0.79$) between $\bar{r}$ and $\bar{m}$ is a reasonable consequence of the systematically decreased RE^{3+} cationic radius and increased atomic mass as the RE elements switch from La to Lu; with the exception of Y element, whose r value fits in the series from La to Lu but its m value is significantly lower than that of the lanthanides. Therefore, all the seven selected features are retained as effective descriptors to train the ML models.

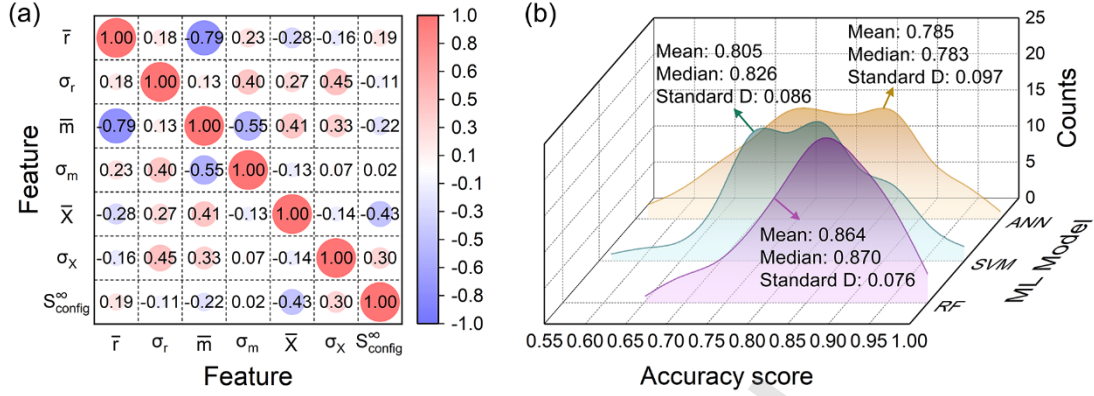


Figure 1. Feature selection and model selection. (a) Spearman correlogram of all features. (b) The spectrum of accuracy score for RF, SVM and ANN models, collected from 100 independent runs using randomly selected training-testing subsets for each model. The mean and median values as well as the standard deviation of accuracy score for each model are marked in the figure.

Three ML classification models, namely, RF, SVM and ANN are trained to classify the phase composition of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$. A randomly selected 80% versus 20% training-testing subsets of the training dataset is used for statistical learning and evaluating the performance of the trained model on previously unseen data. The hyperparameters in ML models are refined through a grid search technique for each ML model, and the performance of the ML models is evaluated by measuring the validation and prediction accuracy through fivefold cross validation. Herein, we carefully checked that the cross-validation accuracy for the optimal hyperparameters is only marginally higher – but not significantly superior to – as compared with that of the suboptimal combinations, which renders the model more robust against overfitting concerns. In order to assess the variability of the model predictions resulting from the random selection of training and testing subsets, each ML model is independently run for 100 times, wherein the training-testing subsets in each run are randomly selected and re-split, and the statistical distributions of prediction accuracy for each model are analyzed in Fig. 1(b). It is shown

that, the RF model demonstrates the highest mean and median values of the prediction accuracy score, followed by the SVM model, and the ANN model has the lowest overall accuracy scores. Meanwhile, the prediction accuracy score of RF model variates from 0.65 to 1.0, in obviously smaller variation range than the SVM and ANN (from 0.57 to 1.0) models; and the standard deviation for the RF model (0.076) is obviously lower than the other two models (0.086 for SVM and 0.097 for ANN). These results indicate that the RF model exhibits higher overall prediction accuracy and better reproducibility associated with randomly selected training-testing subsets. This advantage of the RF algorithm over SVM and ANN could be rationalized in terms of the specific data characteristics and the complexity of the feature-label relationship of this study. The phase prediction of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ in this study involves coupled, non-linear interactions among multiple descriptors, together with limited size of training dataset. In such case, the fundamental principles of RF algorithm – fitting multiple decision trees on various sub-samples of the dataset and averaging their predictions – inherently mitigates overfitting and delivers stable performance on small-to-moderate datasets [51]. And, its hierarchical tree splitting naturally captures non-linear interactions and higher-order feature combinations without assuming a predefined functional form, which is particularly advantageous for polymorphic phase prediction where the decision boundaries are irregular and interdependent. Therefore, the RF model stands out as the optimal model to address the classification problem in this study, and will be the focus of following discussions.

The confusion matrix and the area under the curve (AUC) of the receiver operating

characteristic (ROC) graph are utilized to evaluate the performance of the RF classification model. As shown in Fig. 2(a), all of the four types of phase compositions are precisely classified. More importantly, the relative proportion of the four categories of phase compositions in the testing subsets (7:7:2:7 for the categories of single- β , single- γ , single- δ or mixed $\delta+\gamma$, and separate phases) is close to that in the original training dataset (40:36:8:30), which means that the randomly selected testing subsets properly capture the data distribution of the original dataset. This is a critical advantage for classification models, as it prevents misleading evaluation of the model performance due to mismatched data distributions. Furthermore, the ROC curve visualizes the model's ability to distinguish between different categories, by plotting the prediction rate for each category (correctly predicted against incorrectly predicted) across the entire spectrum of classification thresholds. As shown in Fig. 2(b), the four categories of phase compositions can be accurately classified with AUC score of 1.0, indicative of high classification quality of the RF model. Herein, the class imbalance in the original dataset is effectively mitigated by applying SMOTE prior to model training, i.e., the RF model trained on the balanced dataset achieves a prediction accuracy of 100% on the test subset, outperforming the conventional RF model (accuracy of 95.65%) by 4.35%.

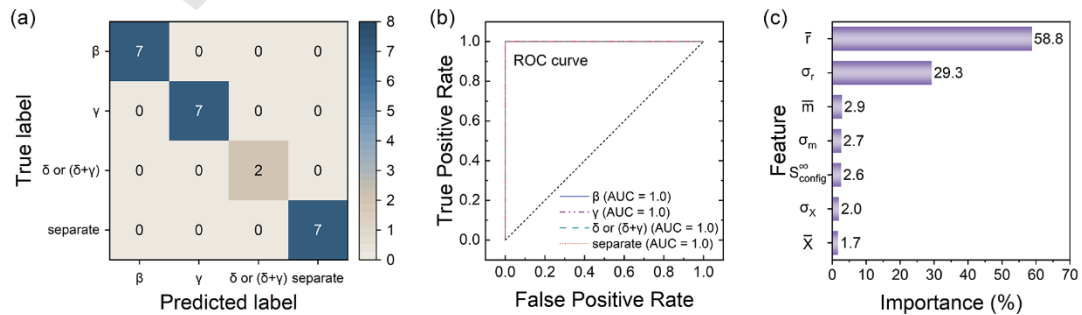


Figure 2. The evaluation metrics of the RF model. The (a) confusion matrix, (b) ROC-AUC curves, and (c) feature importance of the RF model.

The importance of input features that influence the RF model prediction is analyzed in Fig. 2(c). Herein, the average RE³⁺ cationic radius ($\bar{r}$) and the deviation of RE³⁺ cationic radius (σ_r) are the two overwhelmingly influential parameters, with their importance factors added up to over 88%. The $\bar{r}$ parameter fundamentally indicates the ionic field strength around the RE sites, which primarily determines the unit cell dimensions and the overall lattice packing (such as the RE-O bond length and the coordination geometries) of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$, and thus directly dictate which polymorphic framework is favored. Meanwhile, the σ_r parameter measures the mismatch of ionic sizes among the multiple RE constituents. That is to say, while $\bar{r}$ determines which lattice framework is adopted, σ_r governs whether the framework can remain stable as a single-phase solid solution. A larger σ_r introduces greater local strain fields, and ultimately drives towards phase separation when exceeding the critical value. This demonstrates that the RE³⁺ cationic radii serve as sufficiently good indication of the phase composition of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ materials, which lays the foundation for the classical perception that employs the $\bar{r}$ and σ_r parameters to describe the phase formation capability of rare-earth silicates [31,39,40]. In comparison, the features associated with the atomic mass and electronegativity ($\bar{m}$, σ_m , $\bar{X}$, and σ_X) have minor significance. This diminished significance can be attributed, in part, to their indirect relationship with ionic field strength – and consequently with atomic bonding and phase structure; and, in part, to the inherent dependence of machine learning models on the characteristics of the training dataset when assessing the relevance of lower-ranked descriptors. Indeed, this latter factor most likely accounts for the discrepancy between

the descriptor rankings observed in the present study and those reported in prior works [31,39]. Interestingly, the theoretical configurational entropy of mixing ($S_{\text{config}}^{\infty}$) shows marginal influence in describing the phase compositions of $(n\text{RE}_{x_i})_2\text{Si}_2\text{O}_7$ materials. This could be attributed to the fact that the $S_{\text{config}}^{\infty}$ parameter is derived by assuming homogeneous and uniform solid solution with infinitely large lattice and completely random occupations on the doping sites, and thus is depend only on the number and overall molar fractions of the principal RE components, but irrelevant to the specific combination of RE species, their interchanged relative proportions, or the crystallographic polymorph. Detailed discussions on the influence of $S_{\text{config}}^{\infty}$ are presented in the Supplementary Note 3.

The generalization ability of the trained RF model is then evaluated by making interpolative prediction on the phase composition of new $(n\text{RE}_{x_i})_2\text{Si}_2\text{O}_7$ materials. This is a critical procedure for not only measuring how accurate the model is, but also testing whether the model reasonably captures the fundamental rules governing the phase composition of $(n\text{RE}_{x_i})_2\text{Si}_2\text{O}_7$ materials. The $(\text{Gd}_{x_1}\text{Ho}_{x_2}\text{Yb}_{x_3}\text{Lu}_{x_4})_2\text{Si}_2\text{O}_7$ and $(\text{Nd}_{x_1}\text{Ho}_{x_2}\text{Yb}_{x_3}\text{Lu}_{x_4})_2\text{Si}_2\text{O}_7$ ($x_1+x_2+x_3+x_4=1$; x_i variates from 0.1 to 0.9, in interval of 0.05), which are not contained in the training dataset, are used as representative compositions. The rationale for selecting the $(\text{Gd}_{x_1}\text{Ho}_{x_2}\text{Yb}_{x_3}\text{Lu}_{x_4})_2\text{Si}_2\text{O}_7$ and $(\text{Nd}_{x_1}\text{Ho}_{x_2}\text{Yb}_{x_3}\text{Lu}_{x_4})_2\text{Si}_2\text{O}_7$ as validation compositions are explained in the Supplementary Note 4. The RF model predicts 124 single- β phases, 235 single- γ phases, 93 single- δ or mixed $\gamma+\delta$ phases, and 3 separate phases for $(\text{Gd}_{x_1}\text{Ho}_{x_2}\text{Yb}_{x_3}\text{Lu}_{x_4})_2\text{Si}_2\text{O}_7$; while 34 single- β phases, 19 single- γ phases, 3 single- δ or mixed $\gamma+\delta$ phases, and 399

separate phases for $(\text{Nd}_{x1}\text{Ho}_{x2}\text{Yb}_{x3}\text{Lu}_{x4})_2\text{Si}_2\text{O}_7$. Among these, six representative compositions, namely, $(\text{Gd}_{0.15}\text{Ho}_{0.15}\text{Yb}_{0.35}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$, $(\text{Gd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.35}\text{Lu}_{0.15})_2\text{Si}_2\text{O}_7$, $(\text{Gd}_{0.35}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.15})_2\text{Si}_2\text{O}_7$, $(\text{Nd}_{0.10}\text{Ho}_{0.20}\text{Yb}_{0.30}\text{Lu}_{0.40})_2\text{Si}_2\text{O}_7$, $(\text{Nd}_{0.10}\text{Ho}_{0.50}\text{Yb}_{0.30}\text{Lu}_{0.10})_2\text{Si}_2\text{O}_7$, and $(\text{Nd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$, are chosen for further experimental validations in the following section. Their phase composition predicted by the RF model are summarized in Table 1, covering the four types of phase compositions.

Before proceeding to the experimental validation, we performed a critical assessment of whether the six experimental-validation compositions are adequately distinct from the training dataset. Firstly, we verified that none of the six experimental-validation compositions is present in the training dataset. And, the closest training-set neighbors for each experimental-validation sample are recognized by using the Euclidean distance in the two-dimensional $\bar{r}$ - σ_r descriptor space; with $\bar{r}$ and σ_r being the two most influential factors in determining the phases of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$, according to the trained RF model. As listed in Supplementary Table S2, the nearest training neighbors exhibit Euclidean distances ranging from 0.04 to 0.23, with matched polymorphic phases, confirming that these experimental-validation samples are not memorized from identical training instances. Furthermore, we performed leave-one-combination-out (LOCO) cross-validation on a held-out subset, which is consisted of (i) the closest training compositions for the six experimental-validation samples (listed in the second column of Supplementary Table S2), and (ii) all Gd-Ho-Yb-Lu quaternary samples present in the training dataset, added up to be seven samples in total. Here we

note that no Nd-Ho-Yb-Lu quaternary samples are present in the original training dataset, and thus no corresponding held-out composition is added. This design enables testing the model's performance on the most similar but non-identical training samples as well as the ability to generalize to an entirely withheld elemental combination. Thereafter, we systematically excluded all the contained samples in the held-out subset from the original training dataset, retrained the RF model, and tested its performance on the withheld system. The LOCO validation accuracy for the held-out systems remains high (100%), confirming that the RF model generalizes beyond the specific compositions in the training dataset and captures the intrinsic physical descriptors governing the phase compositions, rather than merely memorizing compositional patterns.

3.2 Experiment validation on representative materials

Six representative $(\text{Gd}_{x1}\text{Ho}_{x2}\text{Yb}_{x3}\text{Lu}_{x4})_2\text{Si}_2\text{O}_7$ and $(\text{Nd}_{x1}\text{Ho}_{x2}\text{Yb}_{x3}\text{Lu}_{x4})_2\text{Si}_2\text{O}_7$ samples are experimentally synthesized. Their phase compositions are characterized by XRD measurements, and results are presented in Fig. 3. Herein, the XRD patterns are collected for both the as-synthesized powders obtained from solid-state reaction at 1550 °C as well as the bulk materials obtained after hot-pressed at 1800 °C, in order to ensure that the phase compositions retained after high-temperature processing is correctly recognized for each sample. Meanwhile, scanning electron microscope (SEM) energy dispersive spectrometer (EDS) point analysis is performed on the hot-pressed bulk samples to quantitatively illustrate the distribution of RE elements, and results are presented in Table 2. Herein, the SEM-EDS point analysis for each sample is repeated

at least ten times, and then averaged to obtain the relative ratio of constituent RE elements. As shown in Fig. 3(a), all the detected peaks for as-synthesized powders and the corresponding hot-pressed bulk materials of $(\text{Gd}_{0.15}\text{Ho}_{0.15}\text{Yb}_{0.35}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$ could be indexed to $\beta\text{-Yb}_2\text{Si}_2\text{O}_7$ (PDF#25-1345), indicative of its being stabilized as single- β phase after high-temperature processing. The homogeneous distribution of the RE elements is confirmed by the SEM-EDS analysis (Table 2), wherein the atomic ratio of Gd:Ho:Yb:Lu is measured to be approximately 0.15:0.16:0.35:0.34, close to the designed ratio of 0.15:0.15:0.35:0.35. As shown in Fig. 3(b), the $(\text{Gd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.35}\text{Lu}_{0.15})_2\text{Si}_2\text{O}_7$ powders synthesized at 1550 °C contain γ -phase (indexed to $\gamma\text{-Y}_2\text{Si}_2\text{O}_7$, PDF#42-0167) as the major phase, together with a few β -phase as the secondary phase; and the β -phase is completely eliminated after hot-pressing the sample at 1800 °C, due to the $\beta \rightarrow \gamma$ polymorphic transformation. And thus, the $(\text{Gd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.35}\text{Lu}_{0.15})_2\text{Si}_2\text{O}_7$ sample are ultimately stabilized as single- γ phase. SEM-EDS analysis on the hot-pressed bulk sample shows the atomic ratio of Gd:Ho:Yb:Lu being approximately 0.15:0.35:0.36:0.14, close to the designed ratio of 0.15:0.35:0.35:0.15. To further characterize the phase composition and RE elemental distribution at atomic scale, aberration-corrected scanning transmission electron microscopy (STEM) observations are performed on the bulk samples of $(\text{Gd}_{0.15}\text{Ho}_{0.15}\text{Yb}_{0.35}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$ and $(\text{Gd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.35}\text{Lu}_{0.15})_2\text{Si}_2\text{O}_7$. Fig. 4(a) presents the STEM high-angle annular dark-field (HAADF) image of a chosen area in the $(\text{Gd}_{0.15}\text{Ho}_{0.15}\text{Yb}_{0.35}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$ sample, and the fast Fourier transform (FFT) pattern (Fig. 4(b)) is resolved to be $\beta\text{-(Gd}_{0.15}\text{Ho}_{0.15}\text{Yb}_{0.35}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$, taken in the $[\bar{1}12]$ zone

axis. Fig. 4(d) presents the HAADF-STEM-image of $(\text{Gd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.35}\text{Lu}_{0.15})_2\text{Si}_2\text{O}_7$ sample, and the corresponding FFT pattern (Fig. 4(e)) is resolved to be the $[110]$ zone axis of γ - $(\text{Gd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.35}\text{Lu}_{0.15})_2\text{Si}_2\text{O}_7$. Fig. 4(c) and 4(f) present the selected compositional maps from EDS for both materials, which directly visualizes the compositional homogeneity of RE elements (Gd, Ho, Yb and Lu) on the lattice of β - $(\text{Gd}_{0.15}\text{Ho}_{0.15}\text{Yb}_{0.35}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$ and γ - $(\text{Gd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.35}\text{Lu}_{0.15})_2\text{Si}_2\text{O}_7$ samples, respectively. These observations demonstrate that the $(\text{Gd}_{0.15}\text{Ho}_{0.15}\text{Yb}_{0.35}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$ and $(\text{Gd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.35}\text{Lu}_{0.15})_2\text{Si}_2\text{O}_7$ samples are stabilized as pure single- β and single- γ phases, respectively, with homogeneous distribution of the RE elements.

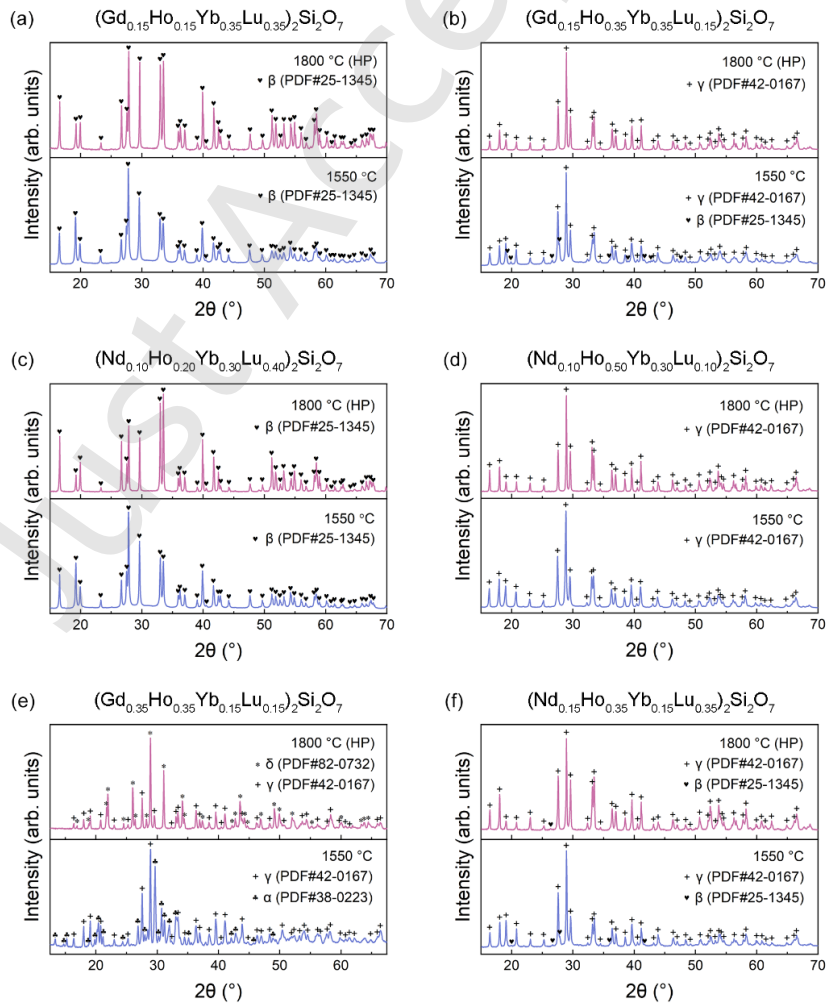


Figure 3. XRD patterns of the synthesized ($n\text{RE}_{xi}$) $_2\text{Si}_2\text{O}_7$ samples. XRD patterns for the (a) $(\text{Gd}_{0.15}\text{Ho}_{0.15}\text{Yb}_{0.35}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$, (b) $(\text{Gd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.35}\text{Lu}_{0.15})_2\text{Si}_2\text{O}_7$, (c) $(\text{Nd}_{0.10}\text{Ho}_{0.20}\text{Yb}_{0.30}\text{Lu}_{0.40})_2\text{Si}_2\text{O}_7$, (d) $(\text{Nd}_{0.10}\text{Ho}_{0.50}\text{Yb}_{0.30}\text{Lu}_{0.10})_2\text{Si}_2\text{O}_7$, (e) $(\text{Gd}_{0.35}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.15})_2\text{Si}_2\text{O}_7$, and (f) $(\text{Nd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$ samples prepared at 1550°C (as-synthesized powders) and 1800°C (hot-pressed bulk samples). “HP” is the abbreviation for “hot-pressed”.

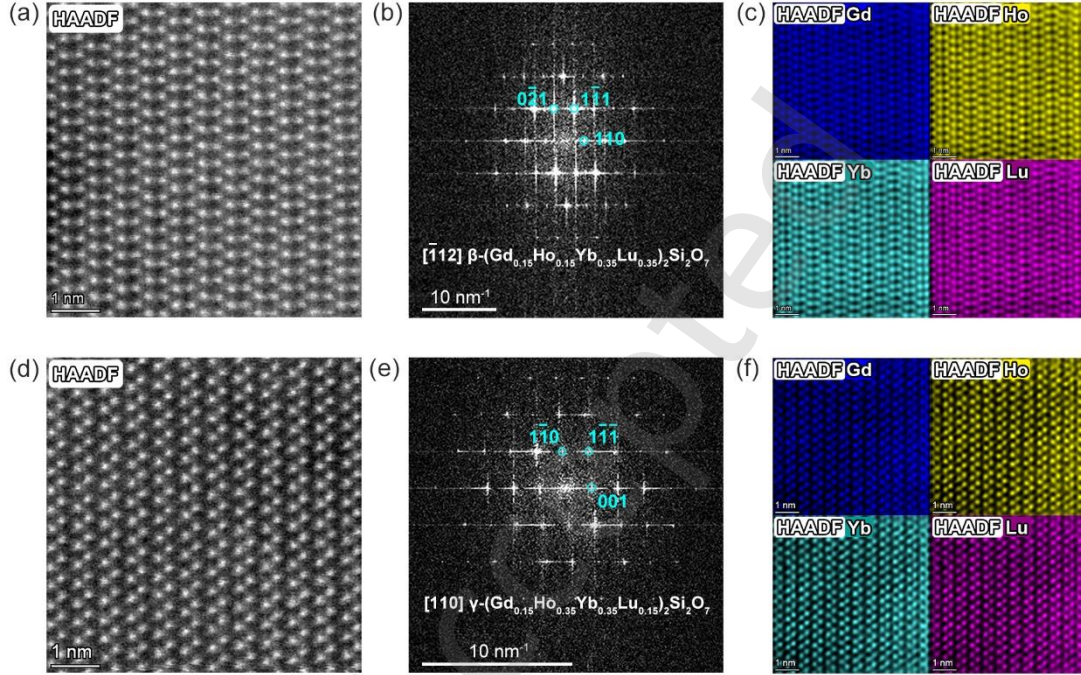


Figure 4. TEM characterization results of the $(\text{Gd}_{0.15}\text{Ho}_{0.15}\text{Yb}_{0.35}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$ and $(\text{Gd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.35}\text{Lu}_{0.15})_2\text{Si}_2\text{O}_7$ bulk samples. (a) The scanning transmission electron microscopic high-angle annular dark-field (HAADF-STEM) images for $(\text{Gd}_{0.15}\text{Ho}_{0.15}\text{Yb}_{0.35}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$; (b) the Fast Fourier transform (FFT) of the HAADF-STEM images of the crystal taken in the $[\bar{1}12]$ zone axis; and (c) the corresponding atomic-resolution EDS maps for Gd, Ho, Yb and Lu elements. (d) The HAADF-STEM images for $(\text{Gd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.35}\text{Lu}_{0.15})_2\text{Si}_2\text{O}_7$; (e) the FFT of HAADF-STEM images the crystal taken in the $[110]$ zone axis; and (f) the corresponding atomic-resolution EDS maps for Gd, Ho, Yb and Lu elements. The cyan circles and numbers in (b) and (e) mark the Miller indices of corresponding patterns.

As shown in Fig. 3(c) and 3(d), the $(\text{Nd}_{0.1}\text{Ho}_{0.2}\text{Yb}_{0.3}\text{Lu}_{0.4})_2\text{Si}_2\text{O}_7$ and $(\text{Nd}_{0.1}\text{Ho}_{0.5}\text{Yb}_{0.3}\text{Lu}_{0.1})_2\text{Si}_2\text{O}_7$ samples are found to be stabilized in single- β and single- γ phase, respectively. Herein, their phase compositions remain unchanged for both powder samples synthesized at 1550 °C as well as bulk samples obtained after hot-pressing at 1800 °C. According to the SEM-EDS results in Table 2, the atomic ratio of

Nd:Ho:Yb:Lu in the hot-pressed β -(Nd_{0.1}Ho_{0.2}Yb_{0.3}Lu_{0.4})₂Si₂O₇ sample is approximated to be 0.09:0.21:0.31:0.39, and that for the γ -(Nd_{0.1}Ho_{0.5}Yb_{0.3}Lu_{0.1})₂Si₂O₇ sample is 0.08:0.52:0.31:0.09, both in consistent with their designed RE atomic ratios. These results demonstrate the successful formation of single-phase β -(Nd_{0.1}Ho_{0.2}Yb_{0.3}Lu_{0.4})₂Si₂O₇ and γ -(Nd_{0.1}Ho_{0.5}Yb_{0.3}Lu_{0.1})₂Si₂O₇ samples exhibiting homogeneous RE elemental distributions.

As for the designed composition of (Gd_{0.35}Ho_{0.35}Yb_{0.15}Lu_{0.15})₂Si₂O₇, XRD results (Fig. 3(e)) show that the as-synthesized powder is the mixture of γ -phase and α -phase (indexed to α -Y₂Si₂O₇, PDF#38-0223) disilicates. Quantitative phase analysis based on the Rietveld refinement of XRD patterns (Fig. 5(a)) shows that, the γ -phase accounts for 45.13 wt% in the mixture, whereas the α -phase accounts for 54.87 wt%. The hot-pressed bulk sample consists of γ -phase and δ -phase (indexed to δ -Y₂Si₂O₇, PDF#82-0732) disilicates, in 43.69 wt% and 56.31 wt%, respectively (shown in Fig. 5(b)). SEM-EDS analysis (Table 2) on the hot-pressed bulk sample reveals slight elemental segregation in both phases. Specifically, the ratio of Gd:Ho:Yb:Lu in one phase is estimated to be 0.27:0.32:0.22:0.19, wherein the concentration of Gd and Ho (or, Yb and Lu) elements are slightly lower (or higher) than the designed composition of 0.35:0.35:0.15:0.15. And the situation is reversed in the other phase, wherein the ratio of Gd:Ho:Yb:Lu is approximated to be 0.42:0.36:0.11:0.11, and thus the concentration of Gd and Ho (or, Yb and Lu) is slightly increased (or decreased) as compared with the designed composition. Therefore, the (Gd_{0.35}Ho_{0.35}Yb_{0.15}Lu_{0.15})₂Si₂O₇ sample exhibit phase separation – mixed δ + γ phase – with slight elemental segregation.

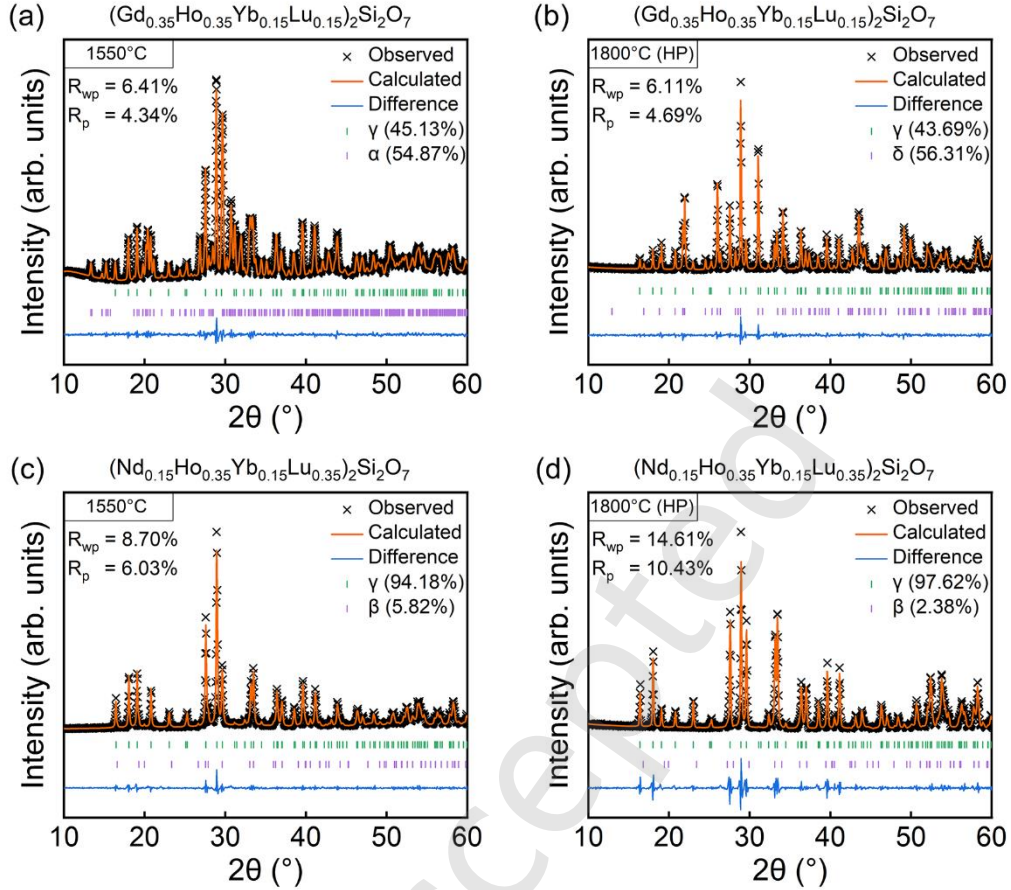


Figure 5. Rietveld refinement of room-temperature XRD patterns for the studied materials. (a) and (b) are the results for the as-synthesized powders of $(\text{Gd}_{0.35}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.15})_2\text{Si}_2\text{O}_7$ prepared at 1550 °C and the bulk sample fabricated via hot-pressing sintering method at 1800 °C. (c) and (d) are the results for the as-synthesized powders of $(\text{Nd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$ prepared at 1550 °C and the bulk sample fabricated via hot-pressing sintering method at 1800 °C. The reliability factors R_p and R_{wp} are also presented in the figures. The short vertical lines at the bottom of each panel denote the positions of XRD peaks of the constituent phases.

As show in Fig. 3(f), the as-synthesized powders of $(\text{Nd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$ contain γ -phase as the major phase, together with small amount of β -phase as the secondary phase. Rietveld refinement analysis (Fig. 5(c)) shows that, the γ -phase accounts for 94.18 wt% in the mixture, whereas the β -phase accounts for 5.82 wt%. And, fewer β -phase (2.38 wt% shown in Fig. 5(d)) can be detected in the hot-pressed bulk sample. SEM-EDS analysis shows that the two phases exhibit Nd:Ho:Yb:Lu of 0.39:0.32:0.09:0.20 and 0.11:0.36:0.16:0.37, respectively.

Interestingly, the ratios of Ho:Yb:Lu in the latter phase are quite close to the designed ratio of 0.35:0.15:0.35; nevertheless, the Nd element is seriously enriched in one phase and depleted in the other. Such considerable deviation of the RE ratio from the designed composition demonstrate tremendous elemental segregation in this sample. As summarized in Table 1, the experimentally observed phase compositions for the six representative samples are in consistent with the results predicted from the RF model, which demonstrates the prediction reliability of the RF model in this study.

Apart from the above discussed six representative materials, we further evaluate the robustness of the trained RF model across different compositional types, by compiling an additional independent experimental-validation dataset (Supplementary Table S3) consisting of 67 equimolar and 20 non-equimolar $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ compositions collected from literatures as well as our own experiments, none of which had been involved in the original model training or hyperparameter optimization processes. The well-trained RF model was applied to predict the phase compositions of these new samples, achieving accuracies of 97.01% (65 out of 67) for the equimolar group and 95.00% (19 out of 20) for the non-equimolar group. The consistently high accuracies across both groups, with only a marginal difference between the two groups, indicate that the model is not biased toward a specific compositional type and exhibits robust generalization capability over diverse compositional spaces. Overall, the high accuracy of 96.55% on this unseen validation dataset further confirms its reliable predictive performance in practical scenarios. And, the discrepancy between the RF-predicted and experimentally measured phase

compositions of the three validation samples may arise from (i) the kinetics effects on the entropy-stabilized phase induced by the different synthesis conditions, such as synthesis temperature, synthesis time, heating or cooling rates [39], (ii) the sluggish polymorphic transformations between neighboring phases [40], or (iii) the inherent prediction errors associated with the present ML model.

3.3 Insights from configurational entropy of mixing

In addition to ML predictions as well as experimental validations, the formation of the phases for $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ disilicates can be rationalized from the perspective of statistical mechanics, by evaluating how much the configurational disorder arising from the random occupation of multiple RE atoms at the RE cationic sites can be tolerated by a specific type of polymorphic lattice. To address this issue, the configurational ensembles of the six representative $(\text{Gd}_{x1}\text{Ho}_{x2}\text{Yb}_{x3}\text{Lu}_{x4})_2\text{Si}_2\text{O}_7$ and $(\text{Nd}_{x1}\text{Ho}_{x2}\text{Yb}_{x3}\text{Lu}_{x4})_2\text{Si}_2\text{O}_7$ compositions are constructed, wherein the polymorphic type for the metastable configurations in each ensemble is constructed according to the experimentally observed phase compositions after high-temperature processing for each sample. In total, eight configurational ensembles, namely, β -($\text{Gd}_{0.15}\text{Ho}_{0.15}\text{Yb}_{0.35}\text{Lu}_{0.35}$) $_2\text{Si}_2\text{O}_7$, γ -($\text{Gd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.35}\text{Lu}_{0.15}$) $_2\text{Si}_2\text{O}_7$, γ -($\text{Gd}_{0.35}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.15}$) $_2\text{Si}_2\text{O}_7$, δ -($\text{Gd}_{0.35}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.15}$) $_2\text{Si}_2\text{O}_7$, β -($\text{Nd}_{0.10}\text{Ho}_{0.20}\text{Yb}_{0.30}\text{Lu}_{0.40}$) $_2\text{Si}_2\text{O}_7$, γ -($\text{Nd}_{0.10}\text{Ho}_{0.50}\text{Yb}_{0.30}\text{Lu}_{0.10}$) $_2\text{Si}_2\text{O}_7$, β -($\text{Nd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.35}$) $_2\text{Si}_2\text{O}_7$ and γ -($\text{Nd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.35}$) $_2\text{Si}_2\text{O}_7$ are constructed, each containing more than 100 unique metastable configurations; except for the δ -polymorph ensemble, which is constructed in smaller size due to the high

computational cost for the DFT ground-state energy calculation rising from its complex crystal structure.

The configurational entropy of mixing (S_{config}) for the eight ensembles is calculated as a function of thermal-energy excitation ($k_B T$) and temperature (T). The convergence behavior of the $S_{\text{config}} \sim k_B T$ curve measures the number of metastable configurations that are excited with increased thermal-energy excitation above the ground state, and therefore is a direct indicator of the energy cost to achieve configurational randomness. As shown in Fig. 6(a), the $S_{\text{config}} \sim k_B T$ curve for the β -(Nd_{0.15}Ho_{0.35}Yb_{0.15}Lu_{0.35})₂Si₂O₇ and γ -(Nd_{0.15}Ho_{0.35}Yb_{0.15}Lu_{0.35})₂Si₂O₇ ensembles cannot reach convergence at sufficiently high energy excitations, as compared with the good convergence of the other six ensembles. Such different trend is inherited from the dispersive features of mixing energy distributions (discussed in Supplementary Note 5), and is fundamentally related to the level of strain-field mismatch arising from the occupation of the RE³⁺ sublattice by multiple RE³⁺ cations – each possessing a distinct cationic radius. From the perspective of statistical mechanics, for the β -(Gd_{0.15}Ho_{0.15}Yb_{0.35}Lu_{0.35})₂Si₂O₇, γ -(Gd_{0.15}Ho_{0.35}Yb_{0.35}Lu_{0.15})₂Si₂O₇, β -(Nd_{0.10}Ho_{0.20}Yb_{0.30}Lu_{0.40})₂Si₂O₇, γ -(Nd_{0.10}Ho_{0.50}Yb_{0.30}Lu_{0.10})₂Si₂O₇, γ -(Gd_{0.35}Ho_{0.35}Yb_{0.15}Lu_{0.15})₂Si₂O₇, and δ -(Gd_{0.35}Ho_{0.35}Yb_{0.15}Lu_{0.15})₂Si₂O₇ ensembles, sufficient numbers of metastable configurations are energetically accessible under modest energy excitation, which indicates a relatively low energy barrier to achieve substantial configurational randomness, thereby promoting the formation of a homogeneous solid solution. By contrast, achieving similar level of configurational randomness in the β -

$(\text{Nd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$ and γ -($\text{Nd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$ ensembles requires extremely high energy excitation. Therefore, it is highly impossible to form $(\text{Nd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$ solid solutions – neither in β nor in γ polymorphic types – with homogeneous RE distribution; rather, the system tends to results in elemental segregation or phase separation.

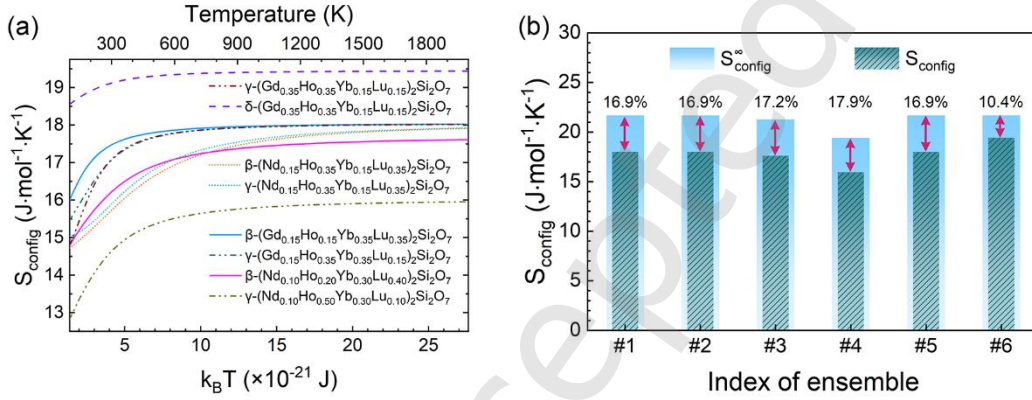


Figure 6. The calculated configurational entropy of mixing (S_{config}) for each configurational ensemble. (a) The dependence of S_{config} on thermal-energy excitation ($k_B T$) and temperature (T). (b) The converged S_{config} values at higher thermal-energy excitation (i.e., extracted at $2000k_B$) and the ideal configuration entropy of mixing ($S_{\text{config}}^{\infty}$) for each ensemble. The differences between the converged S_{config} and $S_{\text{config}}^{\infty}$ (in %) are marked in pink arrows and the values are presented on the top. The indices correspond to the ensemble of β -($\text{Gd}_{0.15}\text{Ho}_{0.15}\text{Yb}_{0.35}\text{Lu}_{0.35}$) $_2\text{Si}_2\text{O}_7$ (#1), γ -($\text{Gd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.35}\text{Lu}_{0.15}$) $_2\text{Si}_2\text{O}_7$ (#2), β -($\text{Nd}_{0.10}\text{Ho}_{0.20}\text{Yb}_{0.30}\text{Lu}_{0.40}$) $_2\text{Si}_2\text{O}_7$ (#3), γ -($\text{Nd}_{0.10}\text{Ho}_{0.50}\text{Yb}_{0.30}\text{Lu}_{0.10}$) $_2\text{Si}_2\text{O}_7$ (#4), γ -($\text{Gd}_{0.35}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.15}$) $_2\text{Si}_2\text{O}_7$ (#5), and δ -($\text{Gd}_{0.35}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.15}$) $_2\text{Si}_2\text{O}_7$ (#6), respectively.

Besides, it is worth noting that the ensembles of γ - and δ -($\text{Gd}_{0.35}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.15}$) $_2\text{Si}_2\text{O}_7$ exhibit small fluctuation in mixing energy as well as the fast convergence of the $S_{\text{config}} \sim k_B T$ curve. Both features indicate a high propensity for forming homogeneous solid solutions, either in the lower-temperature phase (γ) or the higher-temperature phase (δ); nevertheless, experimental observations have revealed coexistence of γ and δ phases with minor elemental segregation –

Gd:Ho:Yb:Lu deviating slightly from 0.35:0.35:0.15:0.15 – in both phases. This discrepancy could be ascribed to the reversible polymorphic transitions of $\gamma \rightleftharpoons \delta$ during solid-state reactions [52,53]. In this scenario, the phase formation is not only governed by the configurational entropy of mixing; but more intricately, influenced by the different diffusion rates of multiple RE elements accompanied by the reversible $\gamma \rightleftharpoons \delta$ polymorphic transitions, resulting in insufficient RE elemental diffusion in either phase and thus heterogeneous distribution of RE elements in the final phase compositions. The latter effect is beyond the scope of statistical mechanisms addressed in this study, and requires further investigations in future works, specifically into the thermodynamics and kinetics of the $\gamma \rightleftharpoons \delta$ reversible polymorphic transitions.

Another point worthy of mentioning is the systematic deviation of the converged S_{config} values – obtained at higher thermal-energy excitation (e.g. extracted at $2000k_B$) – from their respective ideal configuration entropy of mixing ($S_{\text{config}}^{\infty}$), as illustrated in the Fig. 6(b). This could be understood from the framework of statistical mechanics, that the $S_{\text{config}}^{\infty}$ parameter defines the ideal configurational entropy, or rather, the theoretical upper bound of S_{config} for perfectly disordered $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ solid solution with infinitely large lattice. Specifically, the β -(Gd_{0.15}Ho_{0.15}Yb_{0.35}Lu_{0.35})₂Si₂O₇, γ -(Gd_{0.15}Ho_{0.35}Yb_{0.35}Lu_{0.15})₂Si₂O₇, γ -(Gd_{0.35}Ho_{0.35}Yb_{0.15}Lu_{0.15})₂Si₂O₇ and δ -(Gd_{0.35}Ho_{0.35}Yb_{0.15}Lu_{0.15})₂Si₂O₇ ensembles have the same values of $S_{\text{config}}^{\infty}=21.68$ J·mol⁻¹·K⁻¹, whereas the $S_{\text{config}}^{\infty}$ values for the β -(Nd_{0.10}Ho_{0.20}Yb_{0.30}Lu_{0.40})₂Si₂O₇ and γ -(Nd_{0.10}Ho_{0.50}Yb_{0.30}Lu_{0.10})₂Si₂O₇ ensembles are 21.28 J·mol⁻¹·K⁻¹ and 19.43 J·mol⁻¹·K⁻¹, respectively. Herein the converged S_{config} values for the β -type and γ -type

ensembles show systematic deviation (around 17%) from the respective $S_{\text{config}}^{\infty}$ values; on the other hand, the gap is narrowed down to 10.4% for the δ -type ensemble. Such deviations are originated from the insufficient sampling of the configurational space due to the limited size of the supercell used for simulation, namely, 110, 110 and 220 lattice sites in the supercells of β -type, γ -type and δ -type ensembles, respectively. The larger supercell used in simulating the metastable configurations in an ensemble, the closer the converged S_{config} values at high energy excitations approach the theoretical limit of $S_{\text{config}}^{\infty}$.

At last, we clarify that the primary objective of the high-throughput DFT calculations (performed at 0 K) coupled with statistical-mechanics theories in this section is to rationalize the relative tendency of different polymorphic lattices to accommodate configurational disorder arising from multi-RE mixing. The derived $S_{\text{config}}-T$ dependency reflects the energy cost to achieve configurational randomness, and thus fundamentally quantifies the relative propensity of a given polymorphic lattice to tolerate configurational disorder, without implying any kinetic effects governing phase stabilization at experimentally relevant temperatures. Notably, the vibrational contributions to the free energy are not explicitly considered in this framework. A full thermodynamic description at the typical operating temperatures of EBCs (>1200 °C) would indeed require inclusion of lattice vibrational entropy; however, the involved phonon calculations on $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ solid solutions are computationally challenging due to their complex crystal structures, multicomponent compositions, and low crystal symmetry. In this regard, incorporating vibrational entropy would be valuable for

accurately predicting polymorphic phase-transition temperatures of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ and represents a promising direction for future work.

3.4 Design criterion for β - and γ -($n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$

Previous discussions on feature importance of the RF model have demonstrated that the average RE^{3+} cationic radius ($\bar{r}$) and deviation of RE^{3+} cationic radius (σ_r) are the two most important parameters that influence the phase composition of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ materials. Herein, the phase composition of all the 120 $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ samples, used in the training dataset [22,40,54-60] as well as fabricated in this study, is projected onto the dual-descriptor space of $\bar{r}$ and σ_r , and the results are depicted in Fig. 7. In this two-dimensional space, the four types of phase composition are substantially segregated. Here we note that the screening of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ materials for potential EBCs applications mainly focus on those having single- β or single- γ phases, due to their matched CTE with the SiC_f/SiC substrates, as well as showing good phase stability during thermal cycling. As shown in Fig. 7, the boundary that defines the stabilization of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ as single- β and single- γ phases is clearly delineated by the $\bar{r}$ and σ_r parameters. For one thing, the region corresponding to single- β phase falls within $\bar{r} < 0.885 \text{ \AA}$, which is consistent with the findings from our previous research; while the region for single- γ phase falls within $0.885 \text{ \AA} < \bar{r} < 0.900 \text{ \AA}$; and at $\bar{r} > 0.900 \text{ \AA}$, δ phase begins to appear as the primary or secondary phase. For another thing, single phases of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ could be formed when the σ_r values are sufficiently small, whereas larger σ_r values lead to phase separation. Most interestingly, the boundary defined by the σ_r is clearly dependent on $\bar{r}$ values for both single- β and single- γ phases

regions, and exhibits discontinuity at the boundary between the two regions. Specifically, the upper limit of σ_r , below which the $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ samples are stabilized as single- β phase, initially increases with increased $\bar{r}$ and then saturates as $\bar{r}$ approach the demarcation line between single- β phase and single- γ phase ($\bar{r} \sim 0.885 \text{ \AA}$); similarly, for the region of single- γ phase, the boundary curve of σ_r increases with $\bar{r}$ and then levels off at the vicinity of the demarcation line against the region of δ or mixed $\gamma+\delta$ phase ($\bar{r} \sim 0.900 \text{ \AA}$); and in both regions, the maximum σ_r value could reach up to approximately $0.04 \text{ \AA}$. Such dependence of σ_r -boundary on $\bar{r}$ values could be understood by digging into the mechanisms of phase formation and stabilization of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ materials. For a $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ material to be stabilized in single- β or single- γ phase, the mixed multiple RE elements need to construct a spatial framework of a thermodynamically favored polymorphic lattice, wherein the RE sites could be viewed as being occupied by a pseudo atom having the average bonding characteristics of the mixing RE atoms (represented by $\bar{r}$) [40]. Meanwhile, the perturbation caused by size mismatch and chemical difference among multiple RE^{3+} cations (represented by σ_r) must be accommodated by the lattice, in order to prevent competition from any other semi-stable phases that may lead to phase separation. In this scenario, while maintaining a certain type of polymorphic lattice – governed by $\bar{r}$ criterion, as the spatial framework expands (corresponding to an increase in $\bar{r}$), it can accommodate a higher degree of perturbation (corresponding to higher σ_r) within the lattice; until the σ_r reaches a certain threshold, surpassing which the extensive size and chemical mismatch of multiple RE elements will lead to collapse of the lattice. In addition, this

criterion is testified to apply to a broader family of multi-RE-component $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ systems, and is compatible to the phase formation and transformation diagram of single-RE-principal-component $\text{RE}_2\text{Si}_2\text{O}_7$ materials. Detailed results and discussions are presented in Supplementary Note 6 and Supplementary Fig. S3.

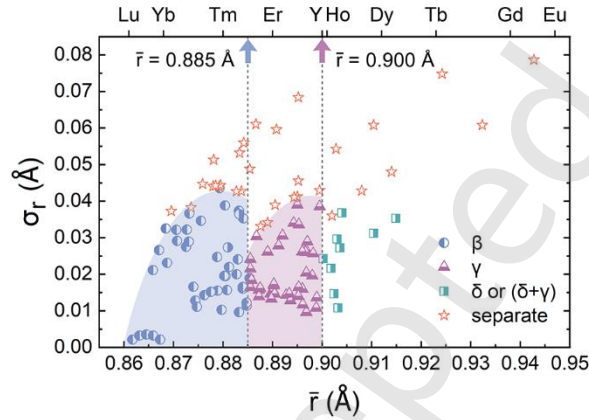


Figure 7. The diagram of average RE^{3+} radius ($\bar{r}$) and the deviations (σ_r) for the $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ (RE = lanthanides and Y) materials. The blue and pink shaded areas outline the region dominated by the $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ that are stabilized in single- β and single- γ phases at high temperatures, and the corresponding boundaries at $\bar{r} \sim 0.885$ Å and $\bar{r} \sim 0.900$ Å are marked in dash lines and arrows. And, the RE^{3+} radius is marked on the upper x-axis as a guide for the eye.

At last, it is worthy to clarify that this work is oriented towards the targeted design of single-phase β -($n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ and γ -($n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ for potential EBCs applications, which offer favorable CTE matching with SiC_f/SiC CMCs. In this context, rather than a broader polymorph classification, the present scheme is organized around the specific design decision of EBCs candidates screening, i.e. retaining single-phase β - or γ -($n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ while avoiding δ -phase formation and phase separation. Compared with recent works focusing on the phase prediction of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ disilicates empowered by machine-learning techniques [31,39] (key ML modeling strategies are compared in Supplementary Table S4), we herein establish a coupled $\bar{r}$ - σ_r design criterion with

quantitative interdependence. This is enabled by the inclusion of deliberately designed diversity of compositions in the training dataset, which covers the number of RE components ranging from two to six, equimolar versus non-equimolar, single-phases versus phase-separated compositions. For one thing, the proposed design criteria – $\bar{r} < 0.885 \text{ \AA}$ for single- β phase and the $0.885 \text{ \AA} < \bar{r} < 0.900 \text{ \AA}$ for single- γ phase – are generally consistent with the previously reported phase-composition rules [31,39,40]. For another, the inclusion of compositions approaching phase separation in our training dataset enables systematic exploration of the upper limit of σ_r , revealing for the first time a quantitative interdependence wherein the σ_r boundary rises monotonically with increased $\bar{r}$ and exhibits discontinuous jumps across different phase regions. In this respect, the dual-descriptor $\bar{r}$ - σ_r design criterion with coupled boundaries could be used as a concise and direct template to guide the compositional design of single-phase β - and γ -($n\text{RE}_{xi}$)₂Si₂O₇ (RE = lanthanides and Y) materials. In the meantime, we emphasize that the compositional design of single- β and single- γ ($n\text{RE}_{xi}$)₂Si₂O₇ constitutes the initial step; and, comprehensive investigations into EBCs-relevant properties, including the coefficient of thermal expansion, thermal conductivity, and resistance to water vapor and CMAS corrosion, etc., are critically prioritized to advance their practical EBCs applications.

4. Conclusions

In summary, the phase composition of ($n\text{RE}_{xi}$)₂Si₂O₇ disilicates is investigated via machine-learning (ML), high-throughput DFT calculation and experiments. A random forest (RF) model is developed to reliably predict the phase composition of

$(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ disilicates, which exhibits high prediction accuracy and robust generalization capability over diverse compositional spaces. The reliability of the RF model is established by successfully classifying the phase compositions of $(\text{Gd}_{x1}\text{Ho}_{x2}\text{Yb}_{x3}\text{Lu}_{x4})_2\text{Si}_2\text{O}_7$ and $(\text{Nd}_{x1}\text{Ho}_{x2}\text{Yb}_{x3}\text{Lu}_{x4})_2\text{Si}_2\text{O}_7$ materials into four types – namely, single- β phase, single- γ phase, single- δ or mixed $\delta+\gamma$ phase, and separate phase – followed by experimental validation on representative materials. Further investigations through high-throughput DFT calculations reveal that the formation of single-phase $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ disilicates correlates with the low energy cost associated with accessing metastable configurations in a configurational ensemble, featured by fast convergence of the $S_{\text{config}} \sim k_B T$ relationship; whereas high energy cost will lead to diminished ability to form homogeneous solid solutions; nevertheless, such descriptor is not applicable to cases involving the δ phase, due to insufficient RE elemental diffusion accompanied by reversible $\gamma \rightleftharpoons \delta$ polymorphic transitions. In addition, the average RE^{3+} cationic radius ($\bar{r}$) and the deviation of RE^{3+} cationic radius (σ_r) are found to be the most influential factors that determines the phase composition of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ disilicates, based on which the criteria of designing single-phase β - $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ and γ - $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ are proposed: (i) $\bar{r} < 0.885 \text{ \AA}$ for β -polymorphs and $0.885 \text{ \AA} < \bar{r} < 0.900 \text{ \AA}$ for the γ -polymorphs, respectively; and (ii) sufficiently small σ_r values, with its upper bound increasing monotonically with $\bar{r}$ values and reaching $\sigma_r \sim 0.04$ as $\bar{r}$ approaches 0.885 and 0.900 $\text{\AA}$. This work is expected to provide an investigation paradigm that accelerates the targeted design of $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ disilicates for their potential applications in the advanced EBCs system.

Funding

This work was supported by National Natural Science Foundation of China (U21A2063, 92360304), LiaoNing Revitalization Talents Program (XLYC2203090), and International Partnership Program of the Chinese Academy of Sciences (172GJHZ2022094FN).

Author contributions

Yixiu Luo: Conceptualization, Investigation, Data curation, Methodology, Writing – original draft

Xinyu Gao: Data curation, Formal analysis

Ziyu Wang: Formal analysis, Visualization

Cui Zhou: Visualization

Jiemin Wang: Methodology

Tiefeng Du: Data curation

Luchao Sun: Funding acquisition, Writing – review & editing

Jingyang Wang: Conceptualization, Funding acquisition, Supervision

Ying Xiong: Validation

Availability of data and codes

The raw data generated in this study have been deposited in the Materials Cloud, which

can be accessed from <https://doi.org/10.24435/materialscloud:p8-ge>. The custom codes have been deposited in Zenodo, which can be accessed from <https://doi.org/10.5281/zenodo.21974722>. All data and codes are available from the corresponding author upon reasonable request.

Competing interests

The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Padture NP. Environmental degradation of high-temperature protective coatings for ceramic-matrix composites in gas-turbine engines. *npj Materials Degradation* 2019, **3**: 1-6.
2. Lee KN. Current status of environmental barrier coatings for Si-based ceramics. *Surface and Coatings Technology* 2000, **133-134**: 1-7.
3. Fernández-Carrión AJ, Allix M, Becerro AI, *et al.* Thermal Expansion of Rare-Earth Pyrosilicates. *J. Am. Ceram. Soc.* 2013, **96**: 2298-2305.
4. Lee KN, Fox DS, Bansal NP. Rare earth silicate environmental barrier coatings for SiC/SiC composites and Si₃N₄ ceramics. *J. Eur. Ceram. Soc.* 2005, **25**: 1705-1715.
5. Zhu D. Advanced Environmental Barrier Coatings for SiC/SiC Ceramic Matrix Composite Turbine Components. In *Engineered Ceramics: Current Status and Future Prospects*, ed. T Ohji, M Singh. Cleveland, OH, USA: John Wiley & Sons, Inc., 2016.
6. Turcer LR, Padture NP. Towards multifunctional thermal environmental barrier coatings (TEBCs) based on rare-earth pyrosilicate solid-solution ceramics. *Scripta Mater.* 2018, **154**: 111-117.
7. Xu Y, Hu X, Xu F, *et al.* Rare earth silicate environmental barrier coatings: Present status and prospective. *Ceram. Int.* 2017, **43**: 5847-5855.
8. Poerschke DL, Van Sluytman JS, Wong KB, *et al.* Thermochemical compatibility of ytterbia–(hafnia/silica) multilayers for environmental barrier coatings. *Acta Mater.* 2013, **61**: 6743-6755.
9. Bondar IA. Rare-earth silicates. *Ceram. Int.* 1982, **8**: 83-89.
10. Luo Y, Sun L, Wang J, *et al.* Material-genome perspective towards tunable thermal expansion of rare-earth di-silicates. *J. Eur. Ceram. Soc.* 2018, **38**: 3547-3554.
11. Richards BT, Young KA, de Francqueville F, *et al.* Response of ytterbium disilicate–silicon environmental barrier coatings to thermal cycling in water vapor. *Acta Mater.* 2016, **106**: 1-14.
12. Yeh J-W, Shen S-K, Lin S-J, *et al.* Nanostructured High-Entropy Alloys with Multiple Principal

- Elements: Novel Alloy Design Concepts and Outcomes. *Adv. Eng. Mater.* 2004, **6**: 299-303.
13. Rost CM, Sachet E, Borman T, *et al.* Entropy-stabilized oxides. *Nat. Commun.* 2015, **6**: 8485.
 14. Oses C, Toher C, Curtarolo S. High-entropy ceramics. *Nat. Rev. Mater.* 2020, **5**: 295-309.
 15. Wang X, Meng M, Xu F, *et al.* $(\text{Lu}_{1/7}\text{Yb}_{1/7}\text{Sc}_{1/7}\text{Er}_{1/7}\text{Y}_{1/7}\text{Ho}_{1/7}\text{Dy}_{1/7})_2\text{Si}_2\text{O}_7$ high entropy rare-earth disilicate with low thermal conductivity and excellent resistance to CMAS corrosion. *J. Adv. Ceram.* 2024, **13**: 549-560.
 16. Bryce K, Majee BP, Huang L, *et al.* A systematic study of thermomechanical properties and calcium-magnesium-aluminosilicate (CMAS) corrosion of multicomponent rare-earth phosphates. *J. Adv. Ceram.* 2024, **13**: 1807-1822.
 17. Ni D, Cheng Y, Zhang J, *et al.* Advances in ultra-high temperature ceramics, composites, and coatings. *J. Adv. Ceram.* 2021, **11**: 1-56.
 18. Zhao Z, Chen H, Xiang H, *et al.* High entropy defective fluorite structured rare-earth niobates and tantalates for thermal barrier applications. *J. Adv. Ceram.* 2020, **9**: 303-311.
 19. Xiang H, Xing Y, Dai F-z, *et al.* High-entropy ceramics: Present status, challenges, and a look forward. *J. Adv. Ceram.* 2021, **10**: 385-441.
 20. Chen L, Li B, Guo J, *et al.* High-entropy perovskite RETa_3O_9 ceramics for high-temperature environmental/thermal barrier coatings. *J. Adv. Ceram.* 2022, **11**: 556-569.
 21. Zhang G, Wang H, Li X, *et al.* Microstructure and thermal properties of hafnium-doped high-entropy oxide $(\text{Y}_{0.25}\text{Ho}_{0.25}\text{Er}_{0.25}\text{Yb}_{0.25})_2\text{O}_3$ coatings. *J. Adv. Ceram.* 2026, **15**:
 22. Sun L, Luo Y, Ren X, *et al.* A multicomponent γ -type $(\text{Gd}_{1/6}\text{Tb}_{1/6}\text{Dy}_{1/6}\text{Tm}_{1/6}\text{Yb}_{1/6}\text{Lu}_{1/6})_2\text{Si}_2\text{O}_7$ disilicate with outstanding thermal stability. *Mater. Res. Lett.* 2020, **8**: 424-430.
 23. Sun L, Ren X, Luo Y, *et al.* Exploration of the mechanism of enhanced CMAS corrosion resistance at 1500 °C for multicomponent $(\text{Er}_{0.25}\text{Tm}_{0.25}\text{Yb}_{0.25}\text{Lu}_{0.25})_2\text{Si}_2\text{O}_7$ disilicate. *Corros. Sci.* 2022, **203**: 110343.
 24. Liu D, Jia X, Shi B, *et al.* $(\text{Sm}_{0.2}\text{Eu}_{0.2}\text{Tb}_{0.2}\text{Dy}_{0.2}\text{Lu}_{0.2})_2\text{Si}_2\text{O}_7$: A novel high-entropy rare earth disilicate porous ceramics with high porosity and low thermal conductivity. *Mater. Chem. Phys.* 2022, **286**: 126181.
 25. Chen G, Zhang Y, Guo X, *et al.* Microstructure and water corrosion behavior of $(\text{Lu}_{0.2}\text{Yb}_{0.2}\text{Er}_{0.2}\text{Tm}_{0.2}\text{Sc}_{0.2})_2\text{Si}_2\text{O}_7$ high-entropy rare-earth disilicate coating for SiC coated C/C composites. *J. Eur. Ceram. Soc.* 2023, **43**: 3647-3657.
 26. Gao X, Wang Z, Luo Y, *et al.* Multicomponent $(\text{Gd}_{x1}\text{Ho}_{x2}\text{Yb}_{x3}\text{Lu}_{x4})_2\text{Si}_2\text{O}_7$ Disilicates: Compositional Modulation and CMAS Corrosion Resistance at 1300°C. *J. Am. Ceram. Soc.* 2026, **109**: e70533.
 27. Wang Y, Zhu J, Li M, *et al.* Thermal properties of high-entropy RE-disilicates controlled by high throughput composition design and optimization. *Mater. Design* 2023, **236**: 112485.
 28. Anand G, Wynn AP, Handley CM, *et al.* Phase stability and distortion in high-entropy oxides. *Acta Mater.* 2018, **146**: 119-125.
 29. Sarker P, Harrington T, Toher C, *et al.* High-entropy high-hardness metal carbides discovered by entropy descriptors. *Nat. Commun.* 2018, **9**: 4980.
 30. Meng H, Yu R, Tang Z, *et al.* Formation ability descriptors for high-entropy diborides established through high-throughput experiments and machine learning. *Acta Mater.* 2023, **256**: 119132.
 31. Fan Y, Bai Y, Li Q, *et al.* Compositional design and phase formation capability of high-entropy rare-earth disilicates from machine learning and decision fusion. *npj Comput. Mater.* 2024, **10**:

- 95.
32. Ayyasamy MV, Deijkers JA, Wadley HNG, *et al.* Density functional theory and machine learning guided search for RE₂Si₂O₇ with targeted coefficient of thermal expansion. *J. Am. Ceram. Soc.* 2020, **103**: 4489-4497.
33. Fan Y, Bai Y, Lu Z, *et al.* Exploring high-performance environmental barrier coatings for rare earth silicates: A combined approach of first principles calculations and machine learning. *J. Materiomics* 2025, **11**: 100921.
34. Wang Y, Liu D, Li Z, *et al.* Data-driven design of high-entropy silicate ceramics with low thermal conductivity. *Materials & Design* 2025, **256**: 114277.
35. Qian B, Liang W, Tu Y, *et al.* Data-driven accelerated design of non-equimolar rare-earth silicates with phase stability and high CMAS resistance at 1550 °C. *J. Alloys and Compd.* 2025, **1042**: 184100.
36. Fan Y, Bai Y, Lu Z, *et al.* Unlocking compositional design and mechanism for mechanical/thermal properties of high-entropy rare-earth disilicates. *J. Am. Ceram. Soc.* 2025, **109**: e70388.
37. Meng H, Wei P, Tang Z, *et al.* Data-driven discovery of formation ability descriptors for high-entropy rare-earth monosilicates. *J. Materiomics* 2024, **10**: 738-747.
38. Fan Y, Bai Y, He X, *et al.* Revealing corrosion mechanisms and enabling predictive lifetime assessment of high-entropy rare-earth disilicates with superior CMAS corrosion resistance. *J. Adv. Ceram.* 2026, **15**: 9221251.
39. Meng H, Bai H, Wei P, *et al.* Data-driven phase prediction for polymorphic high-entropy rare-earth disilicates. *J. Am. Ceram. Soc.* 2025, **109**: e70248.
40. Luo Y, Sun L, Wang J, *et al.* Phase formation capability and compositional design of β -phase multiple rare-earth principal component disilicates. *Nat. Commun.* 2023, **14**: 1275.
41. Pedregosa F, Varoquaux G, Gramfort A, *et al.* Scikit-learn: machine learning in python. *J. Mach. Learn. Res.* 2011, **12**: 2825-2830.
42. Hao J, Ho TK. Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language. *J. Educ. Behav. Stat.* 2019, **44**: 348-361.
43. He Z, Zhang H, Cheng H, *et al.* Machine learning guided BCC or FCC phase prediction in high entropy alloys. *J. Mater. Res. Technol.* 2024, **29**: 3477-3486.
44. Glazer M, Wark J. *Statistical Mechanics: A Survival Guide*. Oxford: Oxford University Press, 2002.
45. van de Walle A, Tiwary P, de Jong M, *et al.* Efficient stochastic generation of special quasirandom structures. *Calphad* 2013, **42**: 13-18.
46. Kresse G, Furthmüller J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* 1996, **54**: 11169-11186.
47. Blochl PE. Projector augmented-wave method. *Phys. Rev. B* 1994, **50**: 17953-17979.
48. Perdew JP, Burke K, Ernzerhof M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* 1996, **77**: 3865-3868.
49. Monkhorst HJ, Pack JD. Special points for Brillouin-zone integrations. *Phys. Rev. B* 1976, **13**: 5188-5192.
50. Toby BH. EXPGUI, a graphical user interface for GSAS. *J. Appl. Crystallogr.* 2001, **34**: 210-230.
51. Breiman L. Random Forest. *Mach. Learn* 2001, **45**: 5-32.

52. Ito J, Johnson H. Synthesis and study of yttrialite. *American Mineralogist* 1968, **53**: 1940-1952.
53. Sun Z, Li M, Zhou Y. Recent progress on synthesis, multi-scale structure, and properties of Y-Si-O oxides. *Inter. Mater. Rev.* 2014, **59**: 357-383.
54. Becerro AI, Escudero A. Polymorphism in the $\text{Lu}_{2-x}\text{Y}_x\text{Si}_2\text{O}_7$ system at high temperatures. *J. Eur. Ceram. Soc.* 2006, **26**: 2293-2299.
55. Becerro AI, Escudero A. Phase transitions in Lu-doped $\text{Y}_2\text{Si}_2\text{O}_7$ at high temperatures. *Chem. Mater.* 2005, **17**: 112-117.
56. Fernández-Carrión AJ, Alba MD, Escudero A, *et al.* Solid solubility of $\text{Yb}_2\text{Si}_2\text{O}_7$ in β -, γ - and δ - $\text{Y}_2\text{Si}_2\text{O}_7$. *J. Solid State Chem.* 2011, **184**: 1882-1889.
57. Chen Z, Lin C, Zheng W, *et al.* A high-entropy $(\text{Yb}_{0.2}\text{Y}_{0.2}\text{Lu}_{0.2}\text{Ho}_{0.2}\text{Er}_{0.2})_2\text{Si}_2\text{O}_7$ environmental barrier coating prepared by atmospheric plasma-spray. *Ceram. Int.* 2023, **49**: 11323-11333.
58. Abrar S, Ma Z, Liu L, *et al.* Excellent CMAS resistance of a newly developed equiatomic high entropy $(\text{Dy}_{1/4}\text{Ho}_{1/4}\text{Tm}_{1/4}\text{Yb}_{1/4})_2\text{Si}_2\text{O}_7$ ceramic pyrosilicate. *Ceram. Int.* 2023, **49**: 19840-19850.
59. Lv X, Cui J, Zhang J, *et al.* Phase composition and property evolution of $(\text{Yb}_{1-x}\text{Ho}_x)_2\text{Si}_2\text{O}_7$ solid solution as environmental/thermal barrier coating candidates. *J. Eur. Ceram. Soc.* 2022, **42**: 4377-4387.
60. Sun L, Luo Y, Tian Z, *et al.* High temperature corrosion of $(\text{Er}_{0.25}\text{Tm}_{0.25}\text{Yb}_{0.25}\text{Lu}_{0.25})_2\text{Si}_2\text{O}_7$ environmental barrier coating material subjected to water vapor and molten calcium-magnesium-aluminosilicate (CMAS). *Corros. Sci.* 2020, **175**: 108881.

Tables

Table 1. Phase composition of representative $(n\text{RE}_{xi})_2\text{Si}_2\text{O}_7$ materials predicted using the RF model, in comparison with the experimental results.

Composition	$\bar{r}$ (Å)	σ_r	ML-predicted (RF model)	Experimental
$(\text{Gd}_{0.15}\text{Ho}_{0.15}\text{Yb}_{0.35}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$	0.881	0.02732	single- β	single- β
$(\text{Gd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.35}\text{Lu}_{0.15})_2\text{Si}_2\text{O}_7$	0.889	0.02612	single- γ	single- γ
$(\text{Gd}_{0.35}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.15})_2\text{Si}_2\text{O}_7$	0.903	0.02964	single- δ or mixed $\gamma+\delta$	mixed $\gamma+\delta$
$(\text{Nd}_{0.10}\text{Ho}_{0.20}\text{Yb}_{0.30}\text{Lu}_{0.40})_2\text{Si}_2\text{O}_7$	0.8833	0.03641	single- β	single- β
$(\text{Nd}_{0.10}\text{Ho}_{0.50}\text{Yb}_{0.30}\text{Lu}_{0.10})_2\text{Si}_2\text{O}_7$	0.8953	0.03357	single- γ	single- γ
$(\text{Nd}_{0.15}\text{Ho}_{0.35}\text{Yb}_{0.15}\text{Lu}_{0.35})_2\text{Si}_2\text{O}_7$	0.89435	0.04108	separate	separate

Table 2. The atomic percentage of RE elements (%) in hot-pressed ($n\text{RE}_{xi}$)₂Si₂O₇ samples, analyzed based on SEM-EDS point analysis.

Composition	Phase	Nd	Gd	Ho	Yb	Lu	Ratio of RE elements
(Gd _{0.15} Ho _{0.15} Yb _{0.35} Lu _{0.35}) ₂ Si ₂ O ₇	β	/	4.92±0.17	5.24±0.52	11.56±0.29	11.29±0.34	Gd _{0.15} Ho _{0.16} Yb _{0.35} Lu _{0.34}
(Gd _{0.15} Ho _{0.35} Yb _{0.35} Lu _{0.15}) ₂ Si ₂ O ₇	γ	/	4.66±0.11	10.87±0.25	10.89±0.17	4.26±0.15	Gd _{0.15} Ho _{0.35} Yb _{0.36} Lu _{0.14}
(Gd _{0.35} Ho _{0.35} Yb _{0.15} Lu _{0.15}) ₂ Si ₂ O ₇	/	/	12.88±0.54	15.69±1.15	10.85±1.55	9.01±1.53	Gd _{0.27} Ho _{0.32} Yb _{0.22} Lu _{0.19}
	/	/	20.58±0.58	17.46±0.81	5.54±0.53	5.25±0.57	Gd _{0.42} Ho _{0.36} Yb _{0.11} Lu _{0.11}
(Nd _{0.10} Ho _{0.20} Yb _{0.30} Lu _{0.40}) ₂ Si ₂ O ₇	β	2.78±0.06	/	6.64±0.21	9.91±0.13	12.62±0.18	Nd _{0.09} Ho _{0.21} Yb _{0.31} Lu _{0.39}
(Nd _{0.10} Ho _{0.50} Yb _{0.30} Lu _{0.10}) ₂ Si ₂ O ₇	γ	2.59±0.09	/	16.17±0.23	9.86±0.18	2.79±0.17	Nd _{0.08} Ho _{0.52} Yb _{0.31} Lu _{0.09}
(Nd _{0.15} Ho _{0.35} Yb _{0.15} Lu _{0.35}) ₂ Si ₂ O ₇	/	19.76±1.06	/	16.12±0.36	4.93±0.34	10.18±0.32	Nd _{0.39} Ho _{0.32} Yb _{0.09} Lu _{0.20}
	/	5.17±0.32	/	17.20±0.29	7.49±0.26	17.47±0.44	Nd _{0.11} Ho _{0.36} Yb _{0.16} Lu _{0.37}

Just Accepted