npj | computational materials

Article

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

https://doi.org/10.1038/s41524-024-01282-x

Compositional design and phase formation capability of high-entropy rare-earth disilicates from machine learning and decision fusion

Yun Fan[1,2], Yuelei Bai[1] ✉, Qian Li[3], Zhiyao Lu[1], Dong Chen[4], Yuchen Liu[2], Wenxian Li[5] & Bin Liu[2,6] ✉

A key strategy for designing environmental barrier coatings is to incorporate multiple rare-earth (RE) components into β- and γ-$RE_2Si_2O_7$ to achieve multifunctional performance optimization. However, the polymorphic phase presents significant challenges for the design of multicomponent RE disilicates. Here, employing decision fusion, a machine learning (ML) method is crafted to identify multicomponent RE disilicates, showcasing notable accuracy in prediction. The well-trained ML models evaluated the phase formation capability of 117 $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ and $(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$, which are unreported in experiments and validated by first-principles calculations. Utilizing model visualization, essential factors governing the formation of $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ are pinpointed, including the average radius of RE^{3+} and variations in different RE^{3+} combinations. On the other hand, $(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$ must take into account the average mass and the electronegativity deviation of RE^{3+}. This work combines material-oriented ML methods with formation mechanisms of multicomponent RE disilicates, enabling the efficient design of superior materials with exceptional properties for the application of environmental barrier coatings.

Environmental barrier coatings (EBCs) have received significant attention in gas turbine technology, aiming to shield turbine components crafted from SiC_f/SiC ceramic matrix composites (CMCs) against corrosion damage from CaO-MgO-Al_2O_3-SiO_2 (CMAS) and water vapor in high-temperature combustion environments, thus enabling higher inlet temperatures, greater core power, and improving fuel efficiency in turbine engines[1–5]. Rare earth disilicates ($RE_2Si_2O_7$) show great potential for EBC applications, because of their excellent resistance to molten CMAS/water vapor, low thermal conductivity, and compatible thermal expansion coefficients (CTEs) with CMC substrates[6,7]. Particularly, the thermal expansion coefficients ($(4–5)\times10^{-6}K^{-1}$) of the β- and γ-$RE_2Si_2O_7$ are close to that of CMCs ($(4.5–5.5)\times10^{-6}K^{-1}$)[6–10]. However, previous studies have shown that their performance decreases when subjected to the strong coupled attack of thermal steam, molten CMAS, and thermal stresses[8,11,12].

Most $RE_2Si_2O_7$ experience intricate phase transformations (α, β, γ, δ, A, F, and G) at various temperatures. Especially, the β-γ polymorphic transition may result in catastrophic cracking or delamination of the EBCs[13–15]. Multi-component systems often exhibit interesting characteristics, such as high configurational entropy, slow diffusion kinetics, severe lattice distortions, and mixing effects. Through combining the outstanding characteristics of each element, multi-component $RE_2Si_2O_7$ materials allow for exploring and optimizing the properties of single-phase multi-RE principal component $RE_2Si_2O_7$ (namely $(nRE_{xi})_2Si_2O_7$) beyond known single one[16–22]. Currently, β-

[1]National Key Laboratory of Science and Technology on Advanced Composites in Special Environments and Center for Composite Materials and Structure, Harbin Institute of Technology, Harbin 150080, China. [2]School of Materials Science and Engineering, Shanghai University, Shanghai 200444, China. [3]National Engineering Research Center for Magnesium Alloy, Chongqing University, Chongqing 400044, China. [4]School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150080, China. [5]UNSW Materials & Manufacturing Futures Institute, School of Materials Science and Engineering, University of New South Wales, Sydney NSW 2052, Australia. [6]Institute of Coating Technology for Hydrogen Gas Turbines, Liaoning Academy of Materials, Shenyang 110004, China. ✉e-mail: baiyl@hit.edu.cn; binliu@shu.edu.cn

$(n\mathrm{RE}_{xi})_2\mathrm{Si}_2\mathrm{O}_7$ with good phase stability and co-doped solid solutions are commonly used for EBC applications[23]. A key challenge in designing $(n\mathrm{RE}_{xi})_2\mathrm{Si}_2\mathrm{O}_7$ is to determine if a specific composition can eventually form a single-phase $(n\mathrm{RE}_{xi})_2\mathrm{Si}_2\mathrm{O}_7$.

Recently, Machine Learning (ML) methods have achieved great success in many fields, especially in predicting new materials. ML can predict high-performance materials directly from high-dimensional input data (descriptors) through learning, rather than extracting limited information from linear combinations of different descriptors[24–29]. In the field of multicomponent materials, ML models, including Artificial Neural Networks (ANN)[30–32], Random Forest Classification (RFC)[32–35], and Support Vector Machines (SVM)[31–35], have been successfully applied for predicting their single-phase formation ability, with their outstanding predictive capabilities and flexibility in handling new material predictions. For example, Kaufmann et al. developed a regression ML model based on hundreds of properties and CALPHAD features[36], in which the results are consistent with density functional theory (DFT)[37–39] calculations and experiments. In contrast, the effective ML model for designing and identifying the β- and γ-$(n\mathrm{RE}_{xi})_2\mathrm{Si}_2\mathrm{O}_7$ is still rare. Recently, Luo et al. found that the configurational entropy of mixing serves as a dependable descriptor for β-$(n\mathrm{RE}_{xi})_2\mathrm{Si}_2\mathrm{O}_7$ formation, but extensive experiments and first-principles calculations are needed for the configurational entropy of mixing[13]. Therefore, it is essential to develop a visualized ML model that utilizes the characteristics of potential $(n\mathrm{RE}_{xi})_2\mathrm{Si}_2\mathrm{O}_7$ and $\mathrm{RE}_2\mathrm{Si}_2\mathrm{O}_7$ features to predict the phase of $(n\mathrm{RE}_{xi})_2\mathrm{Si}_2\mathrm{O}_7$.

Moreover, the $(n\mathrm{RE}_{xi})_2\mathrm{Si}_2\mathrm{O}_7$ synergistically optimizes its structural stability, mechanical/thermal properties, and corrosion resistance by combining key RE components. The performance of rare earth elements in CMAS corrosion reactions can be categorized into three groups: inert elements (Yb and Lu), active elements (Gd, Tb, and Dy), and neutral elements (Ho, Er, and Tm)[15]. For instance, the Lu element shows the ability to diminish the activity of CMAS, thereby improving the resistance of environmental barrier coatings to CMAS[23,40,41]. Through customization of these elements, $(n\mathrm{RE}_{xi})_2\mathrm{Si}_2\mathrm{O}_7$ may enhance the formation of corrosion-resistant products (apatite) and decrease the activity of CMAS, resulting in a slower corrosion rate[15]. Especially, $\mathrm{Lu}_2\mathrm{Si}_2\mathrm{O}_7$ and $\mathrm{Yb}_2\mathrm{Si}_2\mathrm{O}_7$ exhibit good phase stability and comprehensive performance in EBCs. Therefore, the SVC, ANN, and RFC models are firstly developed to predict the β- and γ-$(\mathrm{RE1}_{0.25}\mathrm{RE2}_{0.25}\mathrm{Yb}_{0.25}\mathrm{Lu}_{0.25})_2\mathrm{Si}_2\mathrm{O}_7$ (RE= La, Ce, Eu, Gd, Tb, Dy, Ho, Er, Tm, and Y) in this work (Fig. 1), where some structural characteristics are extracted from the $(n\mathrm{RE}_{xi})_2\mathrm{Si}_2\mathrm{O}_7$ and $\mathrm{RE}_2\mathrm{Si}_2\mathrm{O}_7$. Moreover, the well-trained SVC, ANN, and RFC models are used to predict the β- and γ-$(\mathrm{RE1}_{0.25}\mathrm{RE2}_{0.25}\mathrm{Yb}_{0.25}\mathrm{Lu}_{0.25})_2\mathrm{Si}_2\mathrm{O}_7$ with a validating by DFT calculations, in which the correlation between the average RE^{3+} radius and deviation with different RE^{3+} combinations are analyzed by model visualization.

Meanwhile, by combining the Bayesian theory and majority voting methods as decision fusion approaches, the decision fusion-based RFC model is optimized and extended for the single-phase $(6\mathrm{RE}_{xi})_2\mathrm{Si}_2\mathrm{O}_7$. Due to the atomic radius of Gd located in the middle of lanthanide RE elements, it plays a dual role in improving both the formation of apatite and facilitating the synthesis of single-phase $(n\mathrm{RE}_{xi})_2\mathrm{Si}_2\mathrm{O}_7$. $(\mathrm{RE1}_{1/6}\mathrm{RE2}_{1/6}\mathrm{RE3}_{1/6}\mathrm{Gd}_{1/6}\mathrm{Yb}_{1/6}\mathrm{Lu}_{1/6})_2\mathrm{Si}_2\mathrm{O}_7$ (RE = Y, La, Ce, Eu, Tb, Dy, Ho, Er, and Tm) are focused on this work to enhance their performance as EBCs, where the formation of regular patterns for single-phase $(\mathrm{RE1}_{1/6}\mathrm{RE2}_{1/6}\mathrm{RE3}_{1/6}\mathrm{Gd}_{1/6}\mathrm{Yb}_{1/6}\mathrm{Lu}_{1/6})_2\mathrm{Si}_2\mathrm{O}_7$ are deduced through model visualization and elemental analysis. Finally, the effectiveness of the decision fusion strategy and robustness of the optimized model are demonstrated by re-predicting the phase formation ability of $(\mathrm{RE1}_{0.25}\mathrm{RE2}_{0.25}\mathrm{Yb}_{0.25}\mathrm{Lu}_{0.25})_2\mathrm{Si}_2\mathrm{O}_7$ using them. It is important that the good consistency between the predictions and reported experiments demonstrates the accuracy of this ML model for the rapid evaluation and design of multicomponent RE disilicates, which will encourage and facilitate both the exploitation and design of innovative single-phase β-$(n\mathrm{RE}_{xi})_2\mathrm{Si}_2\mathrm{O}_7$ and γ-$(n\mathrm{RE}_{xi})_2\mathrm{Si}_2\mathrm{O}_7$.

Results

Feature selection

As shown in Table 1, a list of commonly used descriptors that may influence the phase stability of multicomponent RE disilicates, such as ion radius-related features, is considered for this ML models[22]. It should be noticed that the features which can be obtained directly from the database to replace the input descriptors calculated by DFT is selected. To mitigate the risk of overfitting caused by highly correlated features and enhance the fitting efficiency, the features are eliminated with Pearson coefficients exceeding 0.90[18,42]. The Pearson correlation coefficient (r) is calculated using the

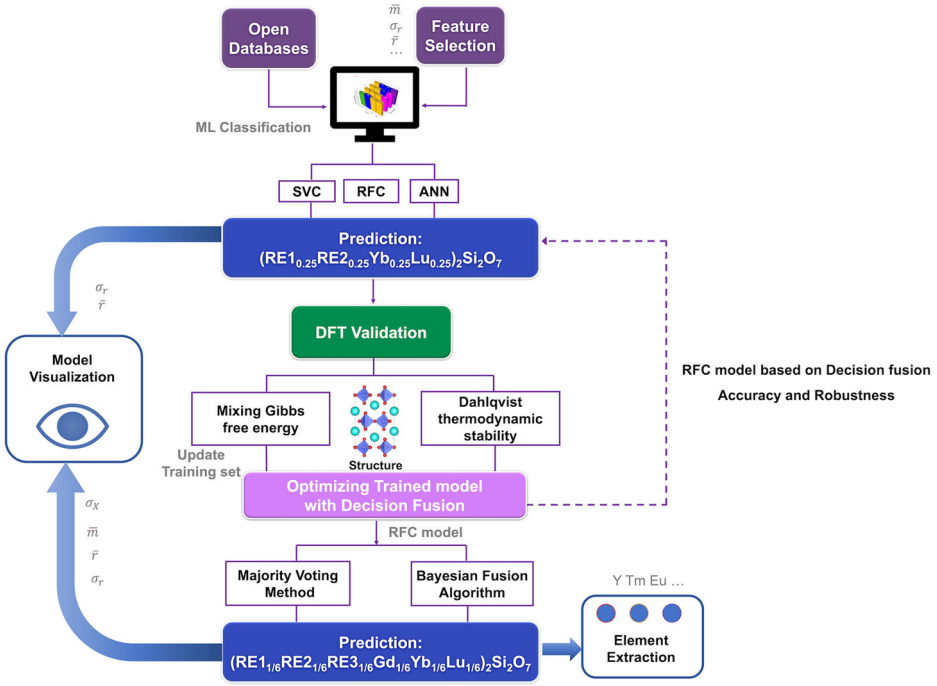

Fig. 1 | Machine Learning framework used in this work. The strategy is composed of machine learning discovery, decision fusion and density functional theory validation modules, in which ANN, RFC and SVC refer to artificial neural network, random forest classification and support vector classification, respectively.

Table 1 | Input features and their corresponding descriptions

Feature	Description
σ_r	RE Ionic radius deviation of constituent $RE_2Si_2O_7$
$\bar{r}$	Average RE Ionic radius of constituent $RE_2Si_2O_7$ per formula unit
σ_m	Mass deviation of constituent $RE_2Si_2O_7$
$\bar{m}$	Average mass of constituent $RE_2Si_2O_7$ per formula unit
σ_X	Electronegativity deviation of constituent $RE_2Si_2O_7$
$\bar{X}$	Average electronegativity of constituent $RE_2Si_2O_7$ per formula unit
$\triangle S$	Mixing entropy

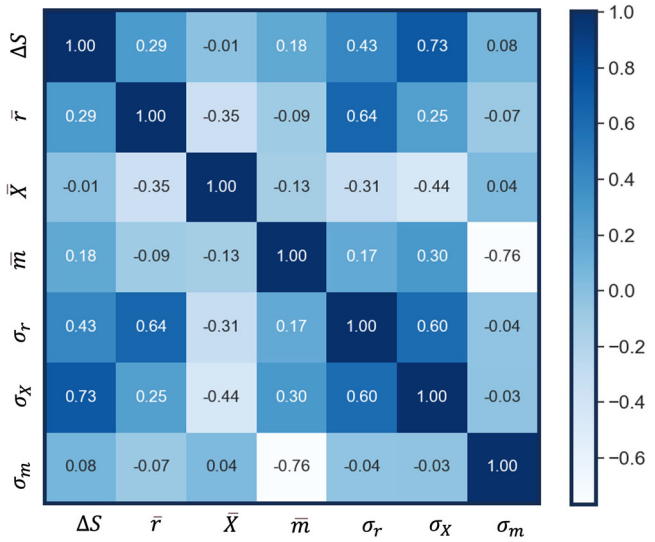

Fig. 2 | Pearson correlation coefficients for all features. The two features with Pearson coefficients close to 1 or −1 indicate a strong correlation between them, whereas those with coefficients close to 0 indicate a weak correlation between them.

following formula:

$$r = \frac{\sum(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{y}_i - \bar{\mathbf{y}})}{\sqrt{\sum(\mathbf{x}_i - \bar{\mathbf{x}})^2(\mathbf{y}_i - \bar{\mathbf{y}})^2}} \quad (1)$$

where, $\mathbf{x}_i$ and $\mathbf{y}_i$ represent the ith values of two different input features, respectively, while $\bar{\mathbf{x}}$ and $\bar{\mathbf{y}}$ represent the expectations of these two input features, respectively. As depicted in Fig. 2, all Pearson correlation coefficients are below 0.8, indicating the rationality of the feature value selection in this work. To expedite the convergence of the ML models, the features are additionally normalized via the formula:

$$\hat{\mathbf{x}}_i = \frac{\mathbf{x}_i - \mathbf{x}_{\text{mean}}}{\mathbf{x}_{\text{std}}} \quad (2)$$

where, $\mathbf{x}_{\text{std}}$ and $\mathbf{x}_{\text{mean}}$ are the variance and mean values, respectively. The processed array shows a mean of 0 and a variance of 1 for each column. $\mathbf{x}_i$ is ith value of the feature ($\mathbf{x}$).

Phase identification of $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ via ML

During the training process, a random splitting technique is employed to divide the data into training and testing sets. Then, the performance of each model is assessed by measuring its validation accuracy at three proportions of training and testing data. Here, the proportions for the test set are 0.15, 0.2, and 0.25, while the corresponding proportions for the training set are 0.85, 0.8, and 0.75, respectively. For the ANN model, the hyper-parameters play a crucial role in the ability of the model to capture intricate patterns from input samples. Few hidden layers and nodes may lead to incomplete feature acquisition. Conversely, excessive hidden layers and nodes can introduce unnecessary noise and lead to overfitting. Therefore, a grid search was conducted on the number of hidden layers and the nodes in each layer. In the case of the SVC model, the kernel type, and the regularization parameter (denoted as C) are optimized. The C parameter determines the tolerance for misclassification errors. For the RFC model, the random state and the number of decision trees are refined to enhance its performance. The random state ensures reproducibility of results, while the number of decision trees affects the model's complexity and predictive power. These three methods, including SVC with hyperplane, ANN with a brain-like structure, and RFC with a tree structure, provide different perspectives. By fine-tuning these parameters and methods, we aim to improve the model's ability to classify data accurately.

As shown in Fig. 3, compared to the testing-to-training ratios of 0.15, 0.20, and 0.25, all three models achieved their respective optimal validation accuracies at a ratio of 0.15. Both the RFC and SVC models achieved a validation accuracy score of 1.000, while the ANN model shows a relatively lower validation accuracy score of 0.857, as depicted in Fig. 3. Meanwhile, the ANN under the PyTorch framework and under the scikit-learn Python package exhibit similar accuracies, as shown in Supplementary Figure 1. To determine the most suitable model for this study, the evaluation metrics of RFC and SVC are further compared, such as the confusion matrices, the receiver operator characteristic (ROC) curve, and AUC (area under the curve). The confusion matrices of RFC and SVC models display similar patterns (Fig. 4a and b), as evidenced by the summation of true positive rate (TPR) and true negative rate (TNR). Additionally, from the AUC-ROC curve (Fig. 4c and d), RFC exhibits higher classification quality with an AUC score of 1.00, in contrast to the AUC score of 0.93 for SVC. Therefore, RFC demonstrates better performance than SVC.

Then, the trained SVC, RFC, and ANN models are used to identify the formation capability of the pure $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$, with their phase composition predicted by the RFC model in Table 2, as well as the ones for SVC and ANN in Supplementary Table 1. Among the 33 predicted samples in the RFC model, 17 samples are identified as pure $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$, in which three compounds are β-$(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ and 14 ones are γ-$(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$. The ML framework offers a high accuracy in prediction but is often perceived as a black-box model. To interpret the prediction process of the RFC model, researchers commonly employ feature importance, which visually reflects the contribution of each feature in enhancing the model's predictive ability[43,44]. Meanwhile, although the Pearson coefficient identifies influential features, it does not provide an explanation of how these features affect the predictions. To address this limitation, the SHAP (Shapley Additive exPlanetions) is introduced to explain the key mechanism of the formation of pure $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$.

The SHAP value and its importance for all features are visualized in Fig. 5. The importance of features is reflected by their average absolute SHAP values (|SHAP|), where the larger the |SHAP| value is, the greater the impact of the feature on the prediction of pure $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ is. Figure 5 (a) illustrates the impact of input features on the prediction by graphing the SHAP values for each feature-sample pair, with each row and dot representing a feature and sample, respectively. The position of the point on the x-axis is determined by its SHAP value, reflecting the impact of features on target attribute prediction. In addition, the point color corresponds to the feature values represented by the color code from low to high, and their specific analysis is described in Supplementary Note 1 and Supplementary Figure 2. The importance of input features that influence the formation capability of pure $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ is ranked in increasing order and displayed in Fig. 5 (b). It is identified that the value of $\bar{r}$ is the most critical factor in predicting the formation capability of pure $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$. In addition, the value of σ_r also holds significance in the prediction results. The specific values of σ_r and $\bar{r}$ for each

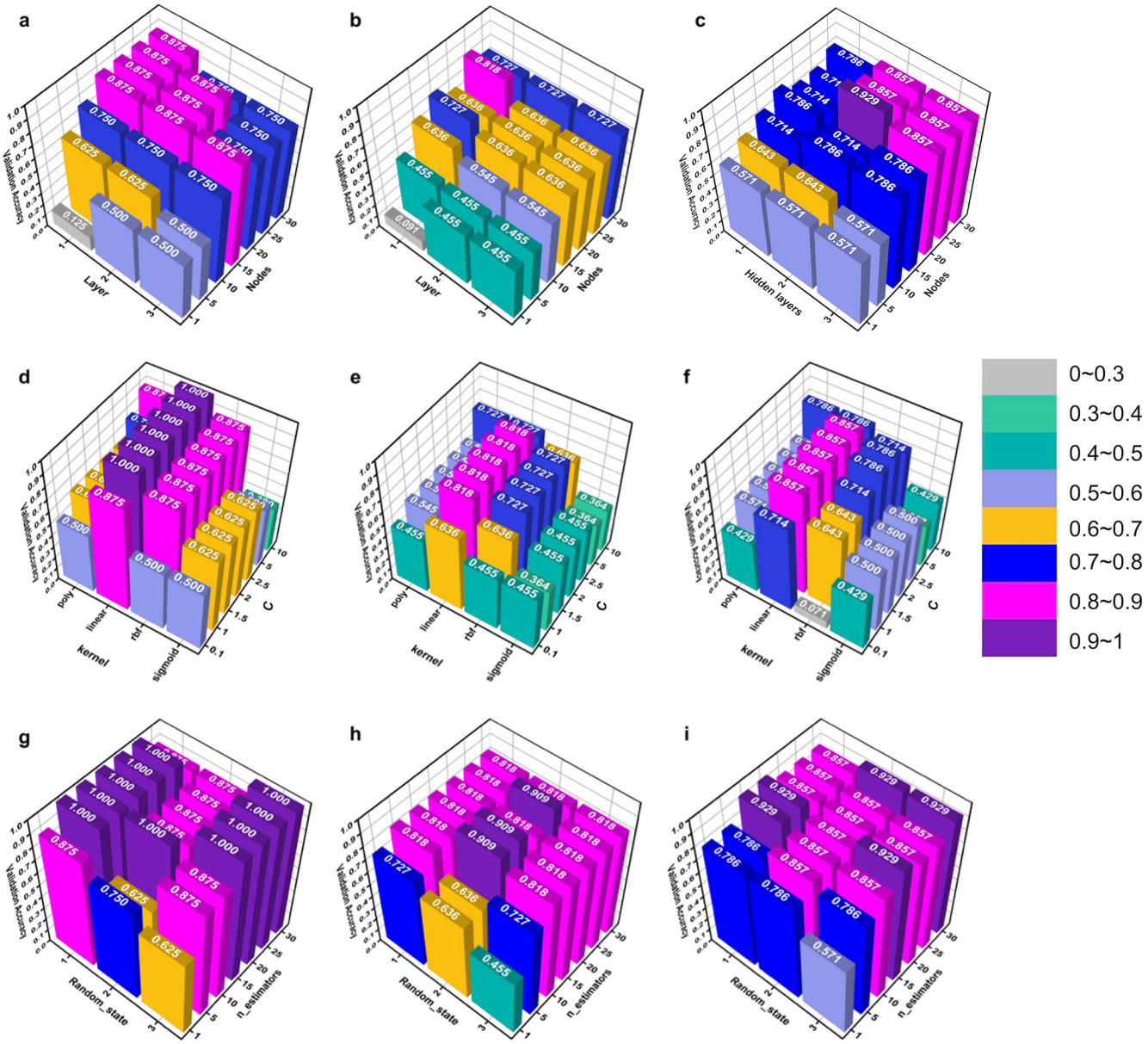

Fig. 3 | Grid search of Machine Learning models. Parameters of (**a**–**c**) ANN, (**d**–**f**) SVC, and (**g**–**i**) RFC models, respectively.The first, second, and third columns are the testing proportions of 0.15, 0.2 and 0.25, respectively.

predicted $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ are summarized in Supplementary Table 2.

As shown in Fig. 6, β-$(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ corresponds to smaller $\bar{r}$ values, while γ-$(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ tends to form as the $\bar{r}$ increases. There is a potential boundary between the dominant regions of the β and γ phases at "$\bar{r}$ = 8.885 Å". This suggests that to design β-$(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$, the $\bar{r}$ feature for multiple combinations of RE^{3+} should be kept below 0.885 Å, which is consistent with the experimental result inferred by Luo et al. [13]. Furthermore, as indicated by Fig. 6, the formation of pure $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ compound can only occur when the value of σ_r is small enough, while large σ_r values lead to element segregation and phase separation. In the ML prediction results, the σ_r values for pure $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ are all smaller than 0.066, which is consistent with the reported results[45]. In fact, when multiple RE elements reside within the target lattice, a phenomenon known as "imperfect isomorphism" occurs[46]. For example, RE elements apart from Yb and Lu can be incorporated into the β-type lattice. Therefore, it is expected that the size mismatch and property differences among multiple RE^{3+} will introduce extensive disturbances in the lattice, resulting in competition among different phases during the formation of solid solutions. The lattice can tolerate small differences in size between various RE^{3+}, thus forming a thermodynamically favorable pure multicomponent solid solution. However, excessive differences will lead to element segregation and phase separation.

It can be concluded that the insights obtained through SHAP analysis are of great significance for the design of synthetic pure materials. Specifically, the formation ability of the pure $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ depends on the average $\bar{r}$ and the σ_r, following the potential phase formation and possible phase transitions at sufficiently high temperatures in experimental conditions. Since there are no reports on these predicted $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ compounds before, the first-principles calculation is used to validate our prediction via the ML models.

Verification of pure $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$

Based on the predicted results of the RFC model (as shown in Table 2), the structures of 17 pure $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ compounds are

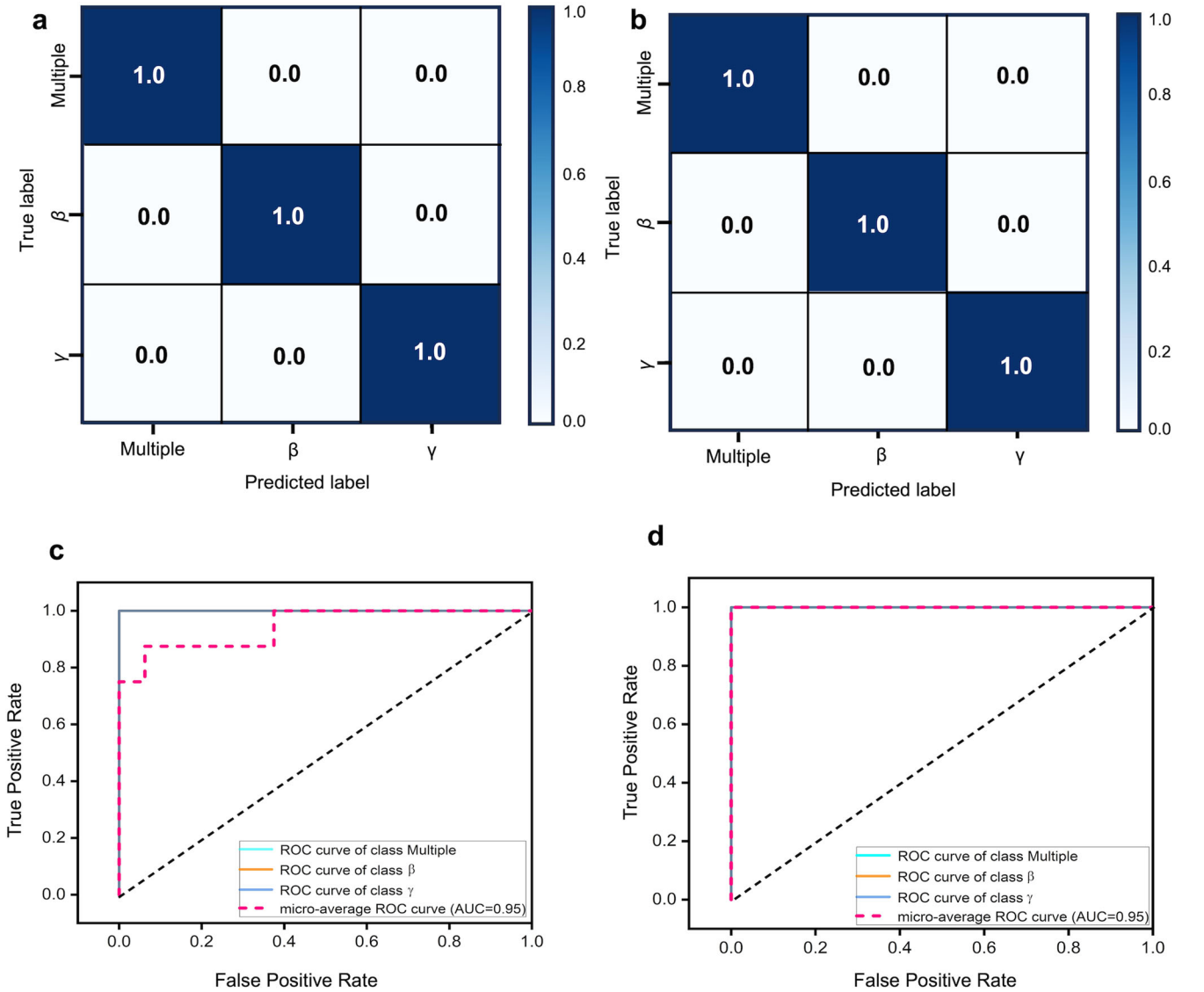

Fig. 4 | The evaluation metrics of (a, c) SVC and (b, d) RFC models. a, **b** Confusion matrix and **c**, **d** ROC-AUC curves.

constructed, including three β phases and fourteen γ ones. As described in Fig. 7, the β-$(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ consists of corner-sharing Si-O tetrahedra that form $[Si_2O_7]$ units, which are stacked along the *y*-axis. The rare earth atoms are coordinated with six oxygen atoms, with a Si-O_{bridge}-Si bond angle of 180°. The bridging oxygen atoms in the $[Si_2O_7]$ unit do not bond with the rare earth atoms. The γ-$(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ is similar to the β phase, while the $[Si_2O_7]$ units are arranged in a staggered pattern, resembling a sinusoidal waveform. The lattice parameters of 17 materials and related $RE_2Si_2O_7$ are listed in Supplementary Table 3, in which the calculated lattice parameters are close to the reported values[8].

The thermodynamic stability of $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ is analyzed through a linear optimization program to verify the stable phases screened by the ML method. Supplementary Table 4 summarizes the calculated formation enthalpies ΔH_{comp} of $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ via a linear optimization procedure. As illustrated in Fig. 8, the ΔH_{comp} values of predicted $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ are all negative, indicating their thermodynamic stability. Among the γ phases, the ΔH_{comp} values follow the order of $(Dy_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7 < (Tb_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7 < (Gd_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7 < (Eu_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ with the increasing of $\bar{r}$ and σ_r (Supplementary Table 2). It means that the thermodynamics stability of $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ synthesis raises with the increasing $\bar{r}$ and σ_r, while the multiple-$(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ will be formed above a certain threshold of $\bar{r}$ and σ_r. This is consistent with reported results[13] and the present ML analysis.

Then, the mixing Gibbs free energy is also introduced to calculate the phase stability of materials against temperatures[47], which can be expressed as:

$$G_{mix} = H_{mix} - T\triangle S \tag{3}$$

where, H_{mix} and ΔS represent the mixing enthalpy and mixing entropy, respectively. T is the temperature. The mixing enthalpy of $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ is defined as the energy relative to four single-component RE disilicates with the same space group, according to the following equation:

$$H_{mix} = \frac{E_{total}\left((RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7\right) - E_{total}\left(RE_2Si_2O_7\right)}{n} \tag{4}$$

where $E_{total}((RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7)$ and $E_{total}(RE_2Si_2O_7)$ are the total energy of multicomponent and single rare-earth disilicates,

Table 2 | Predicted phase composition for some selected $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ via RFC models

Material	Phase	Material	Phase
$(La_{0.25}Eu_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	$(Eu_{0.25}Tb_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ
$(La_{0.25}Gd_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	$(Eu_{0.25}Dy_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ
$(La_{0.25}Tb_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	$(Eu_{0.25}Ho_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ
$(La_{0.25}Dy_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	$(Eu_{0.25}Er_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ
$(La_{0.25}Ho_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	$(Eu_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ
$(La_{0.25}Er_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	$(Gd_{0.25}Tb_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ
$(La_{0.25}Tm_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	$(Gd_{0.25}Dy_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ
$(La_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	$(Gd_{0.25}Er_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ
$(Ce_{0.25}Eu_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	$(Gd_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ
$(Ce_{0.25}Gd_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	$(Tb_{0.25}Dy_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ
$(Ce_{0.25}Tb_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	$(Tb_{0.25}Ho_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ
$(Ce_{0.25}Dy_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	$(Tb_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ
$(Ce_{0.25}Ho_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	$(Dy_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ
$(Ce_{0.25}Er_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	$(Ho_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	β
$(Ce_{0.25}Tm_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	$(Er_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	β
$(Ce_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	$(Tm_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	β
$(Eu_{0.25}Gd_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ		

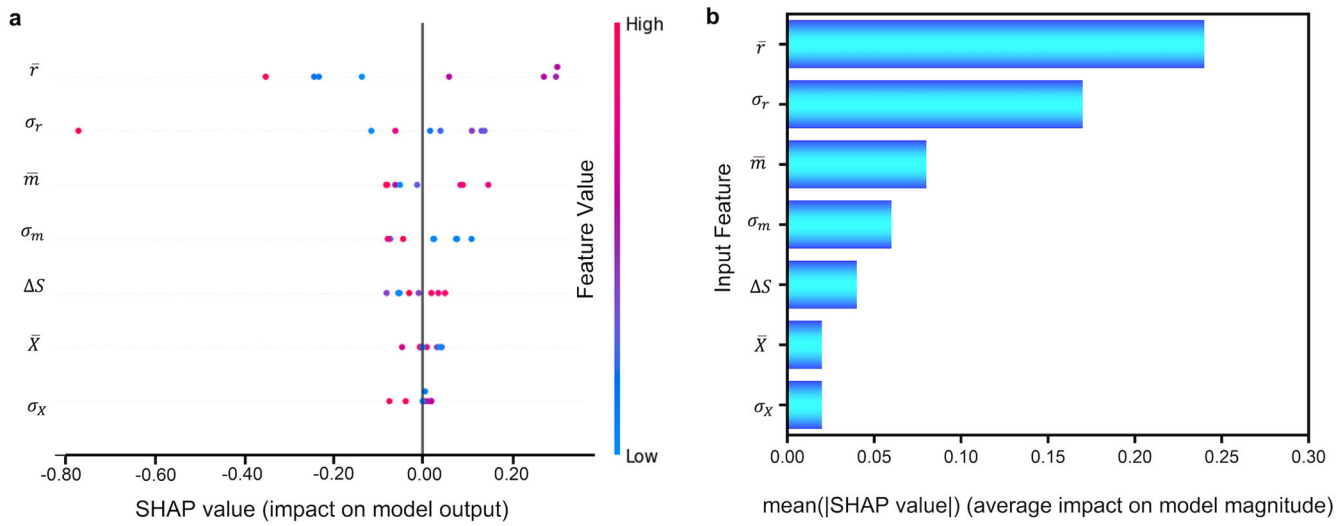

Fig. 5 | The feature visualization via the SHAP value. The SHAP model of (**a**) SHAP values for each feature and (**b**) feature importance for $(RE1_{1/4}RE2_{1/4}Yb_{1/4}Lu_{1/4})_2Si_2O_7$.

respectively. n represents the atom number of the calculated $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$. In addition, the mixing entropy of the supercells $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ is calculated by Eq. (5). As depicted in Fig. 9, the mixing Gibbs free energies are negative, suggesting the stability of $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ above 0 K. Also, Chen et. al successfully synthesized the γ-$(Gd_{1/4}Dy_{/4}Yb_{1/4}Lu_{1/4})_2Si_2O_7$, which verifies our ML prediction of this material[48].

In summary, all 17 predicted $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ selected by the RFC model exhibit stability through ΔH_{comp} and mixing Gibbs free energy. This verifies the reliability of the present ML model.

Extended phase prediction of $(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$ based on decision fusion

As the component of RE elements increases, the difficulty of synthesizing multi-component rare earth silicates escalates. Therefore, based on the validated data of quaternary rare earth disilicates, we retrained the RFC model in ML to predict the single-phase formation capability of $(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$. To replace computationally expensive DFT verification calculations, two decision fusion theories based on cross-validation, i.e., relative voting and Bayesian theory, are incorporated into the retrained model to enhance its reliability in this study. Compared with the proportional division approach used in $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$, the k-fold strategy is considered to take full advantage of decision fusion and improves the results by accentuating decision fusion based on the validated effective algorithms. After the parameter testing, the training model with high precision (0.875), namely $k = 80$ and $n = 200$ (Fig. 10) are selected. The retrained RFC model exhibits a favorable performance, which indicates its generalization ability for high-throughput predictions.

To enhance the accuracy of phase stability prediction, decision fusion is applied to integrate and leverage multiple well-trained models. We mainly evaluate two distinct fusion strategies: a voting-based approach and a Bayesian fusion-based method. The former utilizes the common majority voting scheme, treating the classification result of each model as one vote and integrating the results. The latter effectively combines experimental results and prior information and provides specific prediction probabilities for each

class[49–51]. Table 3 presents the results obtained from the above two decision fusion methods, revealing that the outputs from both methods are consistent. Among them, P_{Mul}, P_β, and P_γ represent the probabilities of forming multiple, β, and γ classes after Bayesian fusion, respectively, while V_{Mul}, V_β, and V_γ denote the number of votes obtained by multiple, β, and γ classes after majority voting, respectively. Among the studied 84 $(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$ materials, 35 ones are predicted to be single phases, including 7 β and 28 γ phases. In experiments, Sun et al. have successfully synthesized γ-$(Tb_{1/6}Dy_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$ that is predicted as a single γ phase in this work[40], validating our ML prediction results.

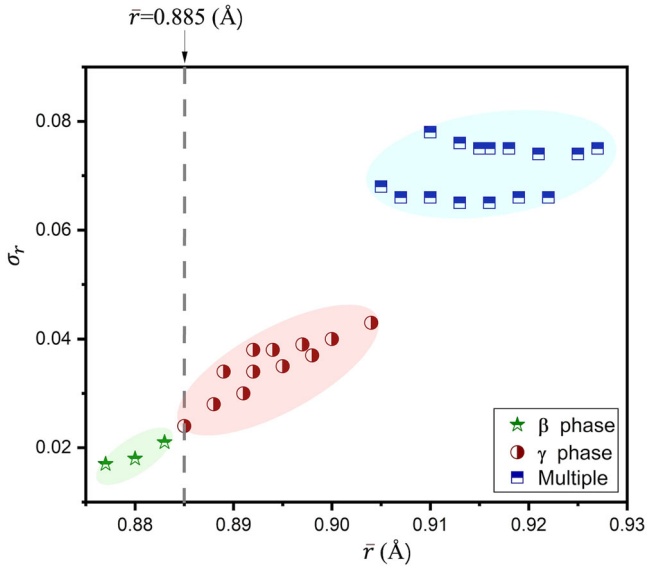

Fig. 6 | The relationship between the feature values related to the ionic radius (σ_r and $\bar{r}$) and the predicted phases.

To further analyze the theoretical implications of the ML prediction results, we selected one of the cross-validation models with high validation accuracy to analyze its SHAP values. As shown in Fig. 11a, there is a negative correlation between σ_r and the formation ability of pure $(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$. Additionally, Fig. 11b shows that being similar to the SHAP model of $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$, the features related to ionic radii still play a crucial role, with the greatly increased importance of σ_X.

The specific values of σ_r, $\bar{r}$, $\bar{m}$ and σ_X for each predicted $(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$ are plotted in Supplementary Table 5. It can be found that σ_r plays a decisive role in the formation of pure $(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$, as the predicted σ_r values for the 35 pure materials are all below 4.0% (as displayed in Fig. 12a), which is consistent with the reported literature[47]. The reported key transition point for the β- and γ-$(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ is $\bar{r}$ = 0.885 Å[13]. According to the SHAP model, $\bar{m}$ and σ_X also play a crucial role in $(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$ located at the boundary for $\bar{r}$ = 0.885 Å. As plotted in Fig. 12b, the $\bar{r}$ values of β-$(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$ are all smaller than 0.900 Å (around 0.885 Å), and the σ_X values of them are less than 0.046. Furthermore, the σ_X values of all the materials predicted as β phase are less than 478.5 g·mol^{-1}. This also explains why the materials satisfied the σ_X, σ_r and $\bar{r}$ conditions (as listed in Supplementary Table 5). For example, $(Dy_{1/6}Er_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$ and $(Dy_{1/6}Ho_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$, are predicted as γ phases. Figure 13 presents the frequency of each RE element in predicted β-$(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$. The Y element shows the highest frequency, indicating its positive contribution to the formation of β-$(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$, while the frequency of Tb, La, and Ce elements is 0, suggesting that these elements hinder the formation of β phases.

In a word, the formation of pure $(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$ cannot be determined solely based on $\bar{r}$, with increase of

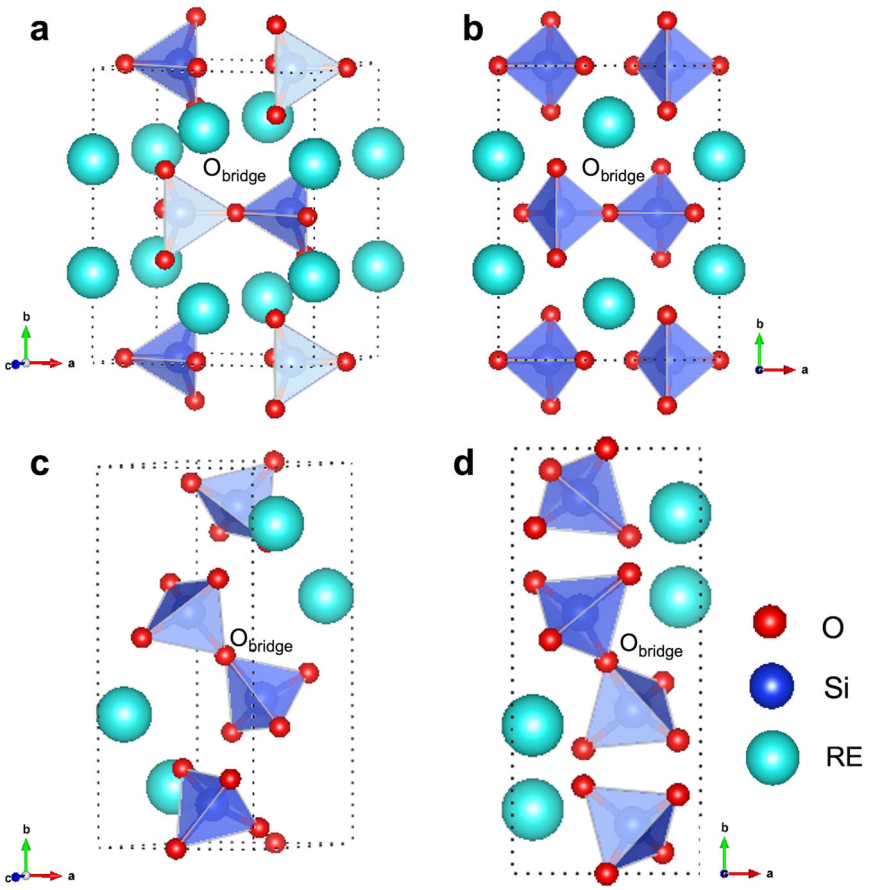

Fig. 7 | The crystal structure of $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$. a, c Crystal structure of β- and γ-$(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$, respectively; **b, d** projection of β- and γ-$(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ along the c axis, respectively.

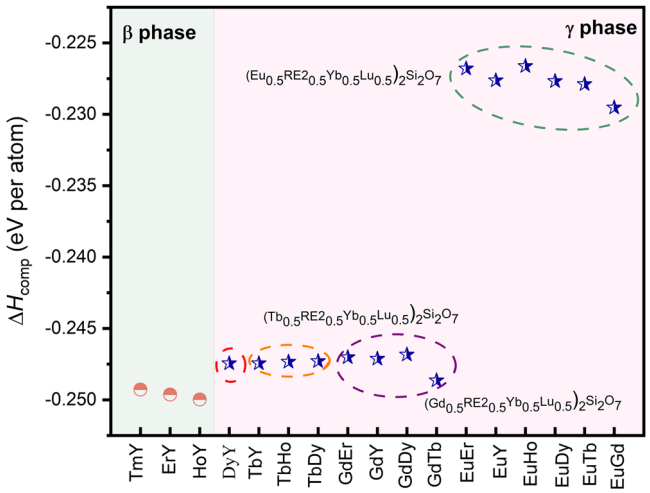

Fig. 8 | Calculated formation enthalpies (eV per atom) of $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$. The red, orange, purple, and green dashed circles point out the $(Dy_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$, $(Tb_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$, $(Gd_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ and $(Eu_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ classification, respectively.

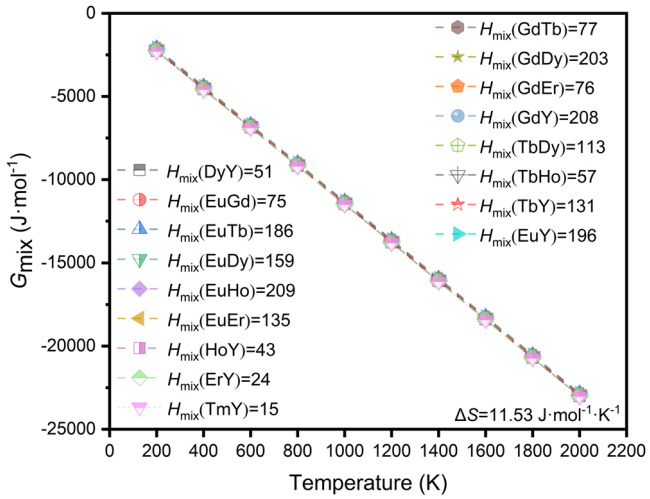

Fig. 9 | The mixing Gibbs free energy of $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ at different temperatures. The calculated mixing enthalpies and mixing entropy of various $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ are all listed in the figure, respectively.

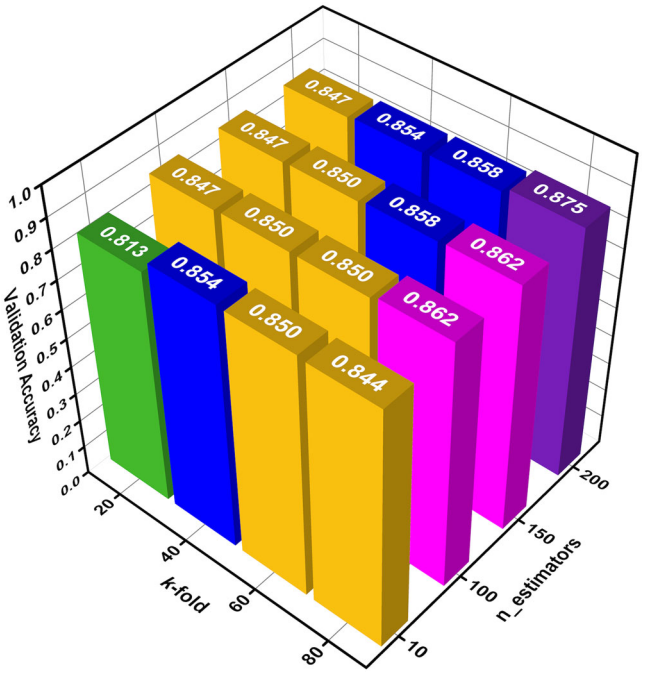

Fig. 10 | Grid search of optimized RFC Machine Learning models. The green, yellow, blue, red, and purple bars represent validation accuracy below 0.840, 0.840~0.850, 0.850~0.860, 0.860-0.870, and above 0.870, respectively.

the RE component. It requires considering various factors, such as $\bar{m}$ and σ_X of the materials. Moreover, by examining the predicted probabilities of different phases via Bayesian theory in Table 3, there is small difference in the predicted probabilities of β- and γ-$(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$ by the decision fusion-based RFC model for these β-$(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$. This indicates that the prediction results show a certain degree of tolerance for the prediction errors.

Discussion

To verify the accuracy and robustness of the RFC model based on decision fusion for the prediction of multicomponent rare earth disilicates, the results of $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ phase stability are re-predicted by an optimized RFC model. After parameter testing, the results of the model with the highest accuracy are listed in Table 4. Specifically, the model achieved the optimal performance when the $k = 20$ is selected.

Combined with Table 2, the inclusion of decision fusion does not change the results of $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ with true labels. This indicates that the RFC model based on decision fusion exhibits high accuracy and generalization in predicting multicomponent disilicates. Additionally, it has been reported that the results obtained through the model based on decision fusion are equal to and/or better than those of the individual models[49,50]. Therefore, the RFC model based on decision fusion demonstrates theoretical and experimental feasibility in the prediction of multicomponent disilicates. In addition, although ML methods are applied to predict the mechanical and thermal properties of different material systems[19–21,24], research on mechanical and thermal properties of high-entropy rare earth disilicates via ML methods is still rare. The foundation for exploring the mechanical and thermal properties of high-entropy rare earth disilicates lies in the investigation of their compositional design and phase formation capability (i.e. phase stability). Thus, this work is expected to lay the foundation for further studying the mechanical and thermal properties of stable high-entropy rare earth silicates.

This work successfully establishes a series of models based on Artificial Neural Network, Random Forest Classification, and Support Vector Machine with a high accuracy for predicting the single-phase formation ability of $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ compounds. The results indicate that σ_r and $\bar{r}$ are the most significant impact on the formation of pure $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$. Specifically, β phases can be formed and further stabilized at high temperatures when satisfying the criteria of (1) $\bar{r} < 0.885$ Å and (2) sufficiently small σ_r. Then, the predictions validated via DFT calculations show excellent agreement with experimental results. Furthermore, by combining two decision fusion approaches (Bayesian theory and the majority voting method), we further optimize the Random Forest Classification model and successfully predict the single-phase formation ability of 84 un-synthesized $(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$ compounds. Among them, 7 β and 28 γ phases are predicted. The results reveal that, unlike $(RE1_{1/4}RE2_{1/4}Yb_{1/4}Lu_{1/4})_2Si_2O_7$, $(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$ also require additional attention to the factors of σ_X and $\bar{m}$. The β phases can be formed when satisfying the criteria of (1) $\bar{r} < 0.900$ Å, (2) sufficiently small σ_r, (3) $\sigma_X < 0.046$, and (4) $\bar{m} < 478$ g·mol^{-1}. Finally, by re-predicting the phase formation ability of $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ via the Random Forest Classification model based on decision fusion, the effectiveness of the decision fusion strategy and the

Table 3 | Predicted phase composition for some selected $(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$ via RFC models based on decision fusion

Material	Bayesian fusion				Majority voting			
	Phase	$P_{Mul.}$	P_{β}	P_{γ}	Phase	$V_{Mul.}$	V_{β}	V_{γ}
$(La_{1/6}Ce_{1/6}Eu_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.854	0.043	0.103	Multiple	80	0	0
$(La_{1/6}Ce_{1/6}Tb_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.897	0.033	0.070	Multiple	80	0	0
$(La_{1/6}Ce_{1/6}Dy_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.869	0.039	0.092	Multiple	80	0	0
$(La_{1/6}Ce_{1/6}Ho_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.896	0.035	0.069	Multiple	80	0	0
$(La_{1/6}Ce_{1/6}Er_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.954	0.037	0.009	Multiple	80	0	0
$(La_{1/6}Ce_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.955	0.036	0.009	Multiple	80	0	0
$(La_{1/6}Ce_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.856	0.047	0.097	Multiple	80	0	0
$(La_{1/6}Eu_{1/6}Tb_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.834	0.048	0.118	Multiple	80	0	0
$(La_{1/6}Eu_{1/6}Dy_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.784	0.044	0.173	Multiple	80	0	0
$(La_{1/6}Eu_{1/6}Ho_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.801	0.043	0.155	Multiple	80	0	0
$(La_{1/6}Eu_{1/6}Er_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.799	0.051	0.150	Multiple	80	0	0
$(La_{1/6}Eu_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.832	0.048	0.119	Multiple	80	0	0
$(La_{1/6}Eu_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.783	0.060	0.158	Multiple	80	0	0
$(La_{1/6}Tb_{1/6}Dy_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.879	0.050	0.070	Multiple	80	0	0
$(La_{1/6}Tb_{1/6}Ho_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.880	0.050	0.069	Multiple	80	0	0
$(La_{1/6}Tb_{1/6}Er_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.875	0.057	0.068	Multiple	80	0	0
$(La_{1/6}Tb_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.950	0.039	0.011	Multiple	80	0	0
$(La_{1/6}Tb_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.867	0.060	0.073	Multiple	80	0	0
$(La_{1/6}Dy_{1/6}Ho_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.783	0.056	0.161	Multiple	80	0	0
$(La_{1/6}Dy_{1/6}Er_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.775	0.057	0.167	Multiple	80	0	0
$(La_{1/6}Dy_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.840	0.041	0.119	Multiple	80	0	0
$(La_{1/6}Dy_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.806	0.055	0.139	Multiple	80	0	0
$(La_{1/6}Ho_{1/6}Er_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.766	0.056	0.178	Multiple	80	0	0
$(La_{1/6}Ho_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.844	0.046	0.110	Multiple	80	0	0
$(La_{1/6}Ho_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.807	0.064	0.129	Multiple	80	0	0
$(La_{1/6}Er_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.901	0.050	0.049	Multiple	80	0	0
$(La_{1/6}Er_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.795	0.065	0.140	Multiple	80	0	0
$(La_{1/6}Tm_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.853	0.058	0.089	Multiple	80	0	0
$(Ce_{1/6}Eu_{1/6}Tb_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.789	0.050	0.162	Multiple	80	0	0
$(Ce_{1/6}Eu_{1/6}Dy_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.756	0.051	0.193	Multiple	80	0	0
$(Ce_{1/6}Eu_{1/6}Ho_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.740	0.053	0.207	Multiple	80	0	0
$(Ce_{1/6}Eu_{1/6}Er_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.740	0.047	0.213	Multiple	80	0	0
$(Ce_{1/6}Eu_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.739	0.047	0.214	Multiple	80	0	0
$(Ce_{1/6}Eu_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.747	0.078	0.175	Multiple	80	0	0
$(Ce_{1/6}Tb_{1/6}Dy_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.691	0.048	0.261	Multiple	79	0	1
$(Ce_{1/6}Tb_{1/6}Ho_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.749	0.048	0.203	Multiple	80	0	0
$(Ce_{1/6}Tb_{1/6}Er_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.826	0.050	0.123	Multiple	80	0	0
$(Ce_{1/6}Tb_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.879	0.050	0.071	Multiple	80	0	0
$(Ce_{1/6}Tb_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.805	0.054	0.142	Multiple	80	0	0
$(Ce_{1/6}Dy_{1/6}Ho_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.650	0.047	0.303	Multiple	79	0	1
$(Ce_{1/6}Dy_{1/6}Er_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.662	0.047	0.291	Multiple	80	0	0
$(Ce_{1/6}Dy_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.699	0.048	0.253	Multiple	80	0	0
$(Ce_{1/6}Dy_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.729	0.071	0.200	Multiple	80	0	0
$(Ce_{1/6}Ho_{1/6}Er_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.674	0.049	0.277	Multiple	80	0	0
$(Ce_{1/6}Ho_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.708	0.056	0.236	Multiple	80	0	0
$(Ce_{1/6}Ho_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.739	0.061	0.200	Multiple	80	0	0
$(Ce_{1/6}Er_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.750	0.066	0.183	Multiple	80	0	0
$(Ce_{1/6}Er_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.733	0.068	0.198	Multiple	80	0	0
$(Ce_{1/6}Tm_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	Multiple	0.763	0.067	0.170	Multiple	80	0	0

Table 3 (continued) | Predicted phase composition for some selected (RE11/6RE21/6RE31/6Gd1/6Yb1/6Lu1/6)2Si2O7 via RFC models based on decision fusion

Material	Bayesian fusion				Majority voting			
	Phase	$P_{Mul.}$	P_{β}	P_{γ}	Phase	$V_{Mul.}$	V_{β}	V_{γ}
$(Eu_{1/6}Tb_{1/6}Dy_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.358	0.051	0.592	γ	0	0	80
$(Eu_{1/6}Tb_{1/6}Ho_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.362	0.049	0.589	γ	0	0	80
$(Eu_{1/6}Tb_{1/6}Er_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.340	0.076	0.584	γ	0	0	80
$(Eu_{1/6}Tb_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.226	0.079	0.695	γ	0	0	80
$(Eu_{1/6}Tb_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.418	0.104	0.479	γ	11	0	69
$(Eu_{1/6}Dy_{1/6}Ho_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.288	0.093	0.618	γ	0	0	80
$(Eu_{1/6}Dy_{1/6}Er_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.144	0.284	0.572	γ	0	0	80
$(Eu_{1/6}Dy_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.075	0.401	0.524	γ	0	0	80
$(Eu_{1/6}Dy_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.283	0.168	0.549	γ	0	0	80
$(Eu_{1/6}Ho_{1/6}Er_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.075	0.377	0.549	γ	0	0	80
$(Eu_{1/6}Ho_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.034	0.459	0.507	γ	0	14	66
$(Eu_{1/6}Ho_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.269	0.180	0.551	γ	0	0	80
$(Eu_{1/6}Er_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.019	0.464	0.517	γ	0	12	68
$(Eu_{1/6}Er_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.156	0.409	0.435	γ	0	23	57
$(Eu_{1/6}Tm_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	β	0.075	0.524	0.401	β	0	78	2
$(Tb_{1/6}Dy_{1/6}Ho_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.112	0.057	0.831	γ	0	0	80
$(Tb_{1/6}Dy_{1/6}Er_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.027	0.056	0.917	γ	0	0	80
$(Tb_{1/6}Dy_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.016	0.046	0.938	γ	0	0	80
$(Tb_{1/6}Dy_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.244	0.134	0.622	γ	0	0	80
$(Tb_{1/6}Ho_{1/6}Er_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.022	0.05	0.928	γ	0	0	80
$(Tb_{1/6}Ho_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.010	0.056	0.934	γ	0	0	80
$(Tb_{1/6}Ho_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.109	0.173	0.718	γ	0	0	80
$(Tb_{1/6}Er_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.020	0.075	0.905	γ	0	0	80
$(Tb_{1/6}Er_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.095	0.176	0.729	γ	0	0	80
$(Tb_{1/6}Tm_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.096	0.173	0.731	γ	0	0	80
$(Dy_{1/6}Ho_{1/6}Er_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.013	0.359	0.628	γ	0	0	80
$(Dy_{1/6}Ho_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.011	0.366	0.622	γ	0	0	80
$(Dy_{1/6}Ho_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	β	0.042	0.516	0.442	β	0	74	6
$(Dy_{1/6}Er_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.010	0.369	0.621	γ	0	0	80
$(Dy_{1/6}Er_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	β	0.039	0.559	0.402	β	0	78	2
$(Dy_{1/6}Tm_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	β	0.034	0.582	0.384	β	0	79	1
$(Ho_{1/6}Er_{1/6}Tm_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	γ	0.010	0.378	0.612	γ	0	0	80
$(Ho_{1/6}Er_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	β	0.033	0.583	0.383	β	0	79	1
$(Ho_{1/6}Tm_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	β	0.031	0.592	0.377	β	0	79	1
$(Er_{1/6}Tm_{1/6}Y_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$	β	0.040	0.593	0.367	β	0	79	1

robustness of the optimized model is demonstrated. In summary, this paper explores a materials-oriented machine learning approach that incorporates the mechanisms behind the phase formation of multicomponent rare-earth disilicates into materials design. It also opens avenues for rational design of multicomponent rare-earth disilicates within the embedded phase space, thereby allowing for effective tuning of their properties.

Methods

Machine Learning

The training dataset was derived from various published sources[7,8,13,23,40,41,52–56]. In the process of ML model training, samples representing the mix phase, β phase, and γ phase were marked as "0", "1", and "2" respectively. These samples were characterized by 7 input features (Table 1), which captured the properties of potential multicomponent $RE_2Si_2O_7$. Notably, these input features were selected based on their proven effectiveness in predicting other multicomponent materials[57,58]. In the calculations, the $RE_2Si_2O_7$ was all composed of equiatomic ratios of the constituent elements, whose mixing entropy was estimated as:

$$\Delta S = -R\sum_i \left(c_i \mathrm{In} c_i\right) \tag{5}$$

where R represented the molar gas constant, and c_i was the molecular concentration of the i-th $RE_2Si_2O_7$. The mean value of a specific property ($\overline{\mathrm{prop}}$) was expressed by:

$$\overline{\mathrm{prop}} = \sum_1^n c_i \mathrm{prop}_i \tag{6}$$

where prop_i was considered as the values of each constituent $RE_2Si_2O_7$. To account for the variation in the constituent $RE_2Si_2O_7$, the deviation (σ_{prop})

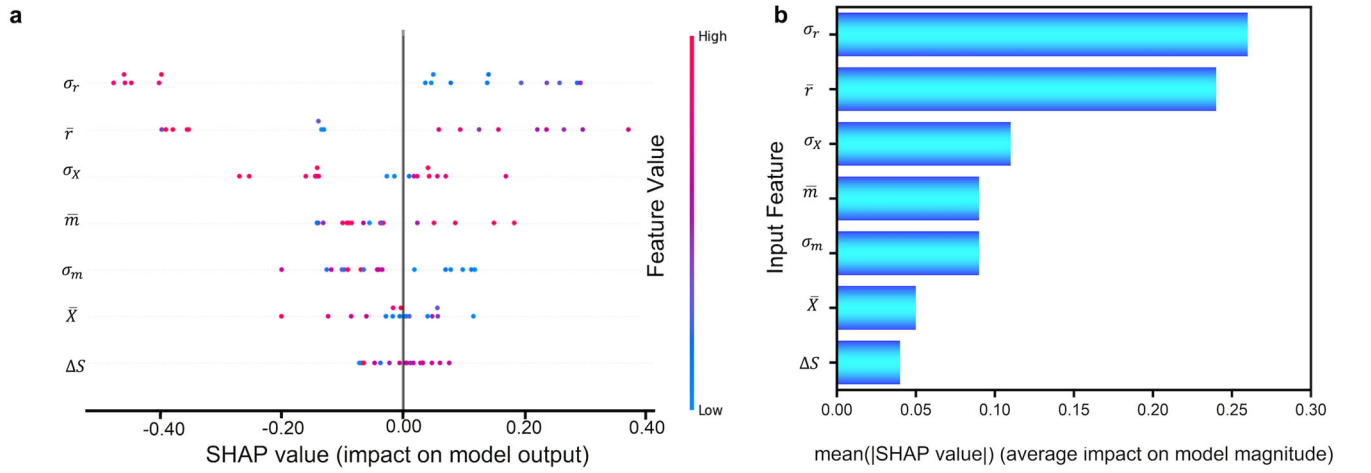

Fig. 11 | The feature visualization via the SHAP value. The SHAP model of (**a**) SHAP values for each feature and (**b**) feature importance for $(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$.

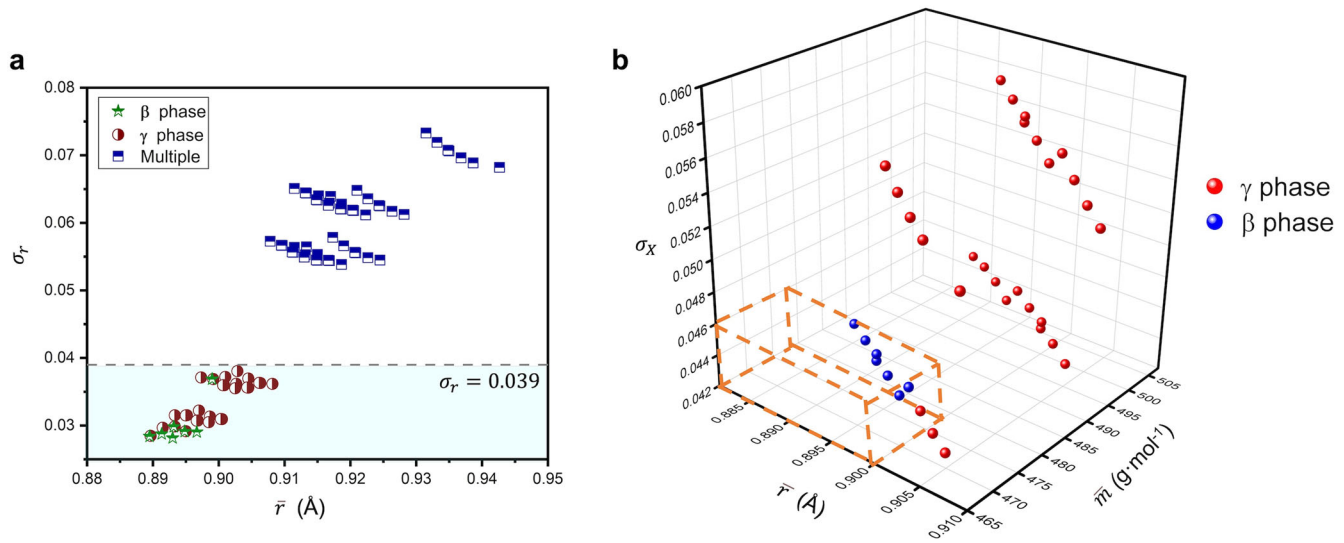

Fig. 12 | The relevant features and phase formation of $(RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$. a The relationship between the σ_X and $\bar{r}$ for the predicted phases, in which the blue area highlights the β and γ phases. **b** The relationship among the σ_X, $\bar{m}$ and $\bar{r}$ values for the predicted phases, in which the dashed box points out the β phases.

of the considered properties was calculated as following:

$$\sigma_{\text{prop}} = \sqrt{\sum_1^n c_i \left(1 - \frac{\text{prop}_i}{\text{prop}}\right)^2} \quad (7)$$

As shown in Fig. 1, the collected samples were used to extract features, which were then transformed into vectors. These input vectors were randomly shuffled and supplied as input to the ANN, SVC, and RFC models. These trained ML models were employed to predict the phase formation ability of multicomponent RE disilicates, with the subsequent validation using first-principles calculations for selected quaternary multicomponent RE disilicates. The validation results could be leveraged to improve the robustness of the ML models and extended the prediction capability to six-RE-principal-component disilicates by including the ML models based on decision fusion in the future training dataset.

The SVC and RFC models were employed utilizing the Scikit-learn (sklearn) Python package[59,60]. The former used a linear kernel with a C parameter of 1.5, which enabled it to effectively find a hyperplane for separating the input samples by mapping them to a higher-dimensional space, while the latter consisted of 200 decision trees, with the predictions by combining the results from all the trees. The ANN models were implemented using two approaches: one utilizing the PyTorch framework and the other employing the sklearn Python package, both with three densely connected hidden layers. The calculation process for obtaining the output vectors in each layer was described as follows:

$$\mathbf{X}_{l+1} = \sigma(\mathbf{W}\mathbf{X}_l + \mathbf{b}) \quad (8)$$

where $\mathbf{X}_{l+1}$ and $\mathbf{X}_l$ corresponded to the feature representations in the $(l+1)$-th and lth layers, respectively. When l was 0, $\mathbf{X}_l$ represented the input feature vector. The weight matrix and bias vector were denoted by $\mathbf{W}$ and $\mathbf{b}$, respectively. The softmax activation function ($\sigma(\mathbf{z})$) was applied to map the perceptron outputs into a non-linear space. During the training process, the PyTorch-based ANN model employed the Adam optimizer to iteratively minimize the cross-entropy loss function. The loss ($H(p,q)$) between the labels and predictions was calculated as follows:

$$H(p,q) = -\sum_x [p(x)\log q(x)] \quad (9)$$

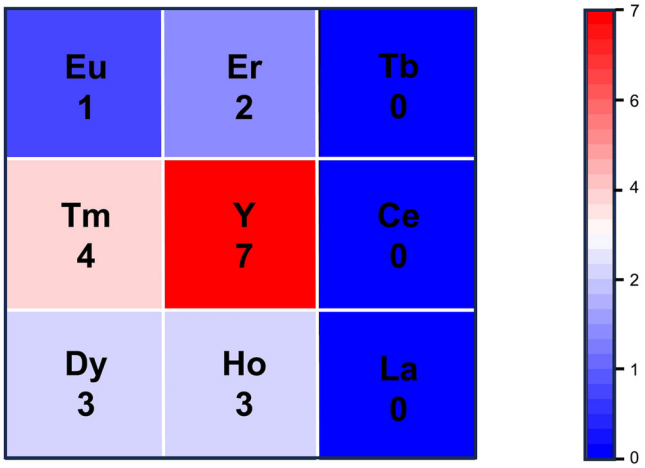

Fig. 13 | Frequencies of various RE elements appeared in the β-($RE1_{1/6}RE2_{1/6}RE3_{1/6}Gd_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$. The numbers in this figure represent the frequencies of the elements, in which the red and blue square areas denote the highest and lowest elemental frequencies, respectively.

where $p(x)$ represented the one-hot encoding of real label, while $q(x)$ was the predict distribution.

The paper employed two training strategies, namely k-fold cross-validation and random sampling[61], to make optimal use of the limited available data. In the former, the number of folds was set to 20 for the selected quaternary multicomponent RE disilicates and 80 for six-RE-principal-component disilicates, based on training accuracy. Multiple models obtained from these strategies could be further combined for the subsequent decision fusion to achieve more accurate results. In the latter, a fixed seed was set to ensure that different models can be compared under the same data partitioning conditions.

The ROC curve, AUC, and confusion matrix were commonly utilized for classifier performance evaluation. The former plotted the true positive rate (TPR) against the false positive rate (FPR) at different discrimination thresholds, while a higher AUC value of the ROC curve, particularly in the top-left region, generally signified better statistical performance of the classifier[62].

Decision fusion is a process of forming unified decisions by integrating the outputs of multiple sensors, systems, or algorithms. The decision fusion used in this work is implemented by synthesizing outcomes from diverse trained models through majority voting and Bayesian fusion. It can maximize the information contained in the current data, provide a comprehensive and accurate decision-making, and construct an appropriate material-oriented modeling approach. For the majority voting method, each classifier casted a vote for a class label, which with the highest number of votes was selected as the final output[51]. The ensemble's output class label was represented as follows:

$$H(x) = c_{\arg\max_j \sum_{i=1}^{T} h_i^j(x)} \tag{10}$$

where $h_i^j(x)$ represented predictive output for classifier h_i on category marker c_j. T was the number of a classifier.

For the Bayesian fusion algorithm, the validation accuracy obtained under the k-fold cross-validation method was served as prior information in this paper, while the predicted probabilities was used as the experimental information[50]. The probability fusion formula was as follows and the specific derivation process is shown in Supplementary Note 2:

$$P\left(B_1, B_2, \ldots, B_n | A_j\right) = \frac{\sum_{i=1}^{n} P\left(A_j | B_i\right) P\left(B_i\right)}{\sum_{j=1}^{c} \sum_{i=1}^{n} \left(A_j | B_i\right) P\left(B_i\right)} \tag{11}$$

where A_j and B_i represented the jth category and the ith model, respectively. n was consistent with the number of cross-validation, which was the number of models. During the implementation of the Bayesian fusion algorithm, the validation accuracy of models was taken as the weights for each model, with the adjusted results to calculate the final predicted probabilities.

As a post-hoc model interpretation method, the SHAP model analysis is a component of interpretable machine learning[63,64]. It establishes a connection between the model's output and input, offering insights into "black box models" from both overall and local situations. Regarded as a paramount method for visual analysis and model interpretation, the SHAP was approached for interpreting predictions generated by the RFC and SVC models to shed light on complex models, with assigning an important value to each feature of a sample in the dataset. These SHAP values acted as sensitivity coefficients, providing insights into the most significant features of the ML model and illustrating how each feature influenced the predictions[64]. Notably, a positive SHAP value indicated that a feature increased the predicted value and had a positive effect, and vice versa. In this way, these SHAP values allowed for a delicately understanding of feature contributions and their impact on the predictions.

DFT calculation

All calculations in this study were conducted utilizing the Vienna Ab Initio Simulation Package (VASP), which implemented the density functional theory (DFT). The projector augmented wave (PAW) potentials were employed for an accurate description of the electron-ion interactions, with an energy cutoff of 520 eV that converged the energy uncertainty below 1 meV per atom. Valence electrons were $5p^6 5d^1 6s^2$, $3s^2 3p^2$, and $2s^2 2p^4$ for RE, Si, and O atoms, respectively. The Perdew-Burke-Ernzerhof (PBE) functional was employed to describe the exchange and correlation interactions within the generalized gradient approximation (GGA)[37–39]. The Brillouin zone (BZ) was sampled using a Γ- centered 2×3×3 and 3×2×2 Monkhorst-Pack k- mesh for β and γ phase multicomponent rare earth disilicates and $RE_2Si_2O_7$. During the geometric optimization process, electronic self-consistency was obtained when the energy difference fell below 10^{-6} eV. As for ionic optimization, it was terminated when the forces acting on atoms reached a value less than 0.01 eV $\cdot$Å^{-1}. During the geometrical optimization, the atomic positions, super-cell volume, and supercell shapes were fully relaxed.

To construct special quasirandom structures (SQS) of multi-component RE disilicates, Monte-Carlo simulations were employed[65,66]. The random swap process was confined to the cation sublattice and excluded swaps between cations and anions. To ensure the convergence of lattice constants and energies concerning the local environments, three independent SQS (Special Quasirandom Structure) structures were computed for each rare earth (RE) disilicate.

The phase stability was analyzed by examining their thermodynamics, in which the targeted phases should show a lower energy than all the other competing ones and their combinations. In the study, a linear optimization procedure introduced by Dahlqvist et al. [67]. was employed:

$$\triangle H_{\text{comp}} = E - E_{\text{comp}}\left(b^{\text{RE1}}, b^{\text{RE2}}, b^{\text{Yb}}, b^{\text{Lu}}, b^{\text{Si}}, b^{\text{O}}\right) = E - \min\left\{\sum\nolimits_i^n x_i E_i \right. \tag{12}$$

where E and E_i represented the total energy of predicted and competing phases, respectively. b^{RE1}, b^{RE2}, b^{Yb}, b^{Lu}, b^{Si} and b^{O} correspond to the elemental fraction of the RE1, RE2, Yb, Lu, Si, and O elements, respectively.

Table 4 | Results of predicted $(RE1_{0.25}RE2_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ via RFC models based on decision fusion

Material	Bayesian fusion				Majority voting			
	Phase	$P_{Mul.}$	P_{β}	P_{γ}	Phase	$V_{Mul.}$	V_{β}	V_{γ}
$(La_{0.25}Eu_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	0.815	0.083	0.102	Multiple	20	0	0
$(La_{0.25}Gd_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	0.815	0.083	0.102	Multiple	20	0	0
$(La_{0.25}Tb_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	0.816	0.087	0.097	Multiple	20	0	0
$(La_{0.25}Dy_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	0.780	0.096	0.124	Multiple	20	0	0
$(La_{0.25}Ho_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	0.768	0.105	0.126	Multiple	20	0	0
$(La_{0.25}Er_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	0.707	0.124	0.169	Multiple	20	0	0
$(La_{0.25}Tm_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	0.660	0.168	0.172	Multiple	20	0	0
$(La_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	0.702	0.127	0.171	Multiple	20	0	0
$(Ce_{0.25}Eu_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	0.746	0.109	0.145	Multiple	20	0	0
$(Ce_{0.25}Gd_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	0.723	0.116	0.161	Multiple	20	0	0
$(Ce_{0.25}Tb_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	0.774	0.108	0.118	Multiple	20	0	0
$(Ce_{0.25}Dy_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	0.701	0.127	0.172	Multiple	20	0	0
$(Ce_{0.25}Ho_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	0.655	0.145	0.200	Multiple	20	0	0
$(Ce_{0.25}Er_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	0.657	0.147	0.197	Multiple	20	0	0
$(Ce_{0.25}Tm_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	0.636	0.161	0.202	Multiple	20	0	0
$(Ce_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	Multiple	0.607	0.156	0.237	Multiple	20	0	0
$(Eu_{0.25}Gd_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ	0.368	0.245	0.387	γ	9	0	11
$(Eu_{0.25}Tb_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ	0.307	0.271	0.422	γ	1	0	19
$(Eu_{0.25}Dy_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ	0.067	0.308	0.626	γ	0	0	20
$(Eu_{0.25}Ho_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ	0.012	0.287	0.701	γ	0	0	20
$(Eu_{0.25}Er_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ	0.012	0.332	0.656	γ	0	0	20
$(Eu_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ	0.062	0.262	0.676	γ	0	0	20
$(Gd_{0.25}Tb_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ	0.080	0.156	0.764	γ	0	0	20
$(Gd_{0.25}Dy_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ	0.007	0.036	0.958	γ	0	0	20
$(Gd_{0.25}Er_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ	0.004	0.030	0.966	γ	0	0	20
$(Gd_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ	0.022	0.166	0.812	γ	0	0	20
$(Tb_{0.25}Dy_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ	0.053	0.054	0.893	γ	0	0	20
$(Tb_{0.25}Ho_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ	0.048	0.062	0.890	γ	0	0	20
$(Tb_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ	0.109	0.206	0.685	γ	0	0	20
$(Dy_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	γ	0.020	0.212	0.768	γ	0	0	20
$(Ho_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	β	0.011	0.695	0.294	β	0	20	0
$(Er_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	β	0.010	0.746	0.244	β	0	20	0
$(Tm_{0.25}Y_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$	β	0.009	0.765	0.226	β	0	20	0

E_{comp} should be minimized subject to the constraints:

$$x_i \geq 0, \sum_i^n x_i = 1 \tag{13}$$

$$\begin{cases} \sum_i^n x_i b_i^{RE1} = b_{RE1}, \sum_i^n x_i b_i^{RE2} = b_{RE2}, \sum_i^n x_i b_i^{Yb} = b_{Yb} \\ \sum_i^n x_i b_i^{Lu} = b_{Lu}, \sum_i^n x_i b_i^{Si} = b_{Si}, \sum_i^n x_i b_i^{O} = b_O \end{cases} \tag{14}$$

$$b_{RE1} + b_{RE2} + b_{Yb} + b_{Lu} + b_{Si} + b_O = 1 \tag{15}$$

where, b_i^{RE1} was the proportion of RE1 atomic number in compound i, etc.

Data availability

All data used in this work are publicly available. Original datasets could be found in corresponding literature[7,8,13,23,40,41,52–56]. Besides, the original and processed datasets used in this work are also available at https://github.com/Yun-Fann/ML-HEC.

Code availability

The codes developed for this work are available at https://github.com/Yun-Fann/ML-HEC.

Received: 4 January 2024; Accepted: 19 April 2024;
Published online: 07 May 2024

References

1. Turcer, L. R. & Padture, N. P. Towards multifunctional thermal environmental barrier coatings (TEBCs) based on rare-earth pyrosilicate solid-solution ceramics. *Scr. Mater.* **154**, 111–117 (2018).
2. Liu, B. et al. Advances on strategies for searching for next generation thermal barrier coating materials. *J. Mater. Sci. Technol.* **35**, 833–851 (2019).
3. Liu, B. et al. Application of high-throughput first-principles calculations in ceramic innovation. *J. Mater. Sci. Technol.* **88**, 143–157 (2021).

4. Dang, X. L. et al. Oxidation behaviors of carbon fiber reinforced multilayer SiC-Si_3N_4 matrix composites. *J. Adv. Ceram.* **11**, 354–364 (2022).
5. Dong, L. et al. Pressure infiltration of molten aluminum for densification of environmental barrier coatings. *J. Adv. Ceram.* **11**, 145–157 (2022).
6. Fernandez-Carrion, A. J., Allix, M. & Becerro, A. I. Thermal expansion of rare-earth pyrosilicates. *J. Am. Ceram. Soc.* **96**, 2298–2305 (2013).
7. Xu, Y., Hu, X. X., Xu, F. F. & Li, K. W. Rare earth silicate environmental barrier coatings: Present status and prospective. *Ceram. Int.* **43**, 5847–5855 (2017).
8. Luo, Y. X. et al. Material-genome perspective towards tunable thermal expansion of rare-earth di-silicates. *J. Eur. Ceram. Soc.* **38**, 3547–3554 (2018).
9. Lv, X. R. et al. Rare earth monosilicates as oxidation resistant interphase for SiC_f/SiC CMC: Investigation of SiC_f/Yb_2SiO_5 model composites. *J. Adv. Ceram.* **11**, 702–711 (2022).
10. Yang, L. W. et al. Dynamic oxidation mechanism of carbon fiber reinforced SiC matrix composite in high-enthalpy and high-speed plasmas. *J. Adv. Ceram.* **11**, 365–377 (2022).
11. Poerschke, D. L., Van Sluytman, J. S., Wong, K. B. & Levi, C. G. Thermochemical compatibility of ytterbia-(hafnia/silica) multilayers for environmental barrier coatings. *Acta Mater.* **61**, 6743–6755 (2013).
12. Richards, B. T. et al. Response of ytterbium disilicate-silicon environmental barrier coatings to thermal cycling in water vapor. *Acta Mater.* **106**, 1–14 (2016).
13. Luo, Y. X. et al. Phase formation capability and compositional design of β-phase multiple rare-earth principal component disilicates. *Nat. Commun.* **14**, 1275 (2023).
14. Soetebier, F. & Urland, W. Crystal structure of lutetium disilicate, $Lu_2Si_2O_7$. *Z. Krist.-N. Cryst. St.* **217**, 22 (2002). 22.
15. Poerschke, D. L., Jackson, R. W. & Levi, C. G. Silicate deposit degradation of engineered coatings in gas turbines: progress toward models and materials solutions. *Annu. Rev. Mater. Res.* **47**, 297–330 (2017).
16. Yeh, J. W. et al. Nanostructured high-entropy alloys with multiple principal elements: Novel alloy design concepts and outcomes. *Adv. Eng. Mater.* **6**, 299–303 (2004).
17. Rost, C. M. et al. Entropy-stabilized oxides. *Nat. Commun.* **6**, 8485 (2015).
18. Gild, J. et al. High-entropy metal diborides: a new class of high-entropy materials and a new type of ultrahigh temperature ceramics. *Sci. Rep.* **6**, 37946 (2016).
19. Sarker, P. et al. High-entropy high-hardness metal carbides discovered by entropy descriptors. *Nat. Commun.* **9**, 4980 (2018).
20. Castle, E., Csanadi, T., Grasso, S., Dusza, J. & Reece, M. Processing and properties of high-entropy ultra-high temperature carbides. *Sci. Rep.* **8**, 8609 (2018).
21. Harrington, T. J. et al. Phase stability and mechanical properties of novel high entropy transition metal carbides. *Acta Mater.* **166**, 271–280 (2019).
22. Zhang, J. et al. Design high-entropy carbide ceramics from machine learning. *Npj Comput. Mater.* **8**, 5 (2022).
23. Sun, L. C. et al. High temperature corrosion of $(Er_{0.25}Tm_{0.25}Yb_{0.25}Lu_{0.25})_2Si_2O_7$ environmental barrier coating material subjected to water vapor and molten calcium-magnesium-aluminosilicate (CMAS). *Corros. Sci.* **175**, 108881 (2020).
24. Schmidt, J., Marques, M. R. G., Botti, S. & Marques, M. A. L. Recent advances and applications of machine learning in solid-state materials science. *Npj Comput. Mater.* **5**, 83 (2019).
25. Jablonka, K. M., Ongari, D., Moosavi, S. M. & Smit, B. Big-data science in porous materials: materials genomics and machine learning. *Chem. Rev.* **120**, 8066–8129 (2020).
26. Chen, C. et al. A critical review of machine learning of energy materials. *Adv. Energy Mater.* **10**, 1903242 (2020).
27. Liu, H., Fu, Z. P., Yang, K., Xu, X. Y. & Bauchy, M. Machine learning for glass science and engineering: A review. *J. Non-Cryst. Solids* **557**, 119419 (2021).
28. Hart, G. L. W., Mueller, T., Toher, C. & Curtarolo, S. Machine learning for alloys. *Nat. Rev. Mater.* **6**, 730–755 (2021).
29. Guo, Y. N. et al. Cracking behavior of newly-developed high strength eutectic high entropy alloy matrix composites manufactured by laser powder b e d fusion. *J. Mater. Sci. Technol.* **163**, 81–91 (2023).
30. Huang, W. J., Martin, P. & Zhuang, H. L. L. Machine-learning phase prediction of high-entropy alloys. *Acta Mater.* **169**, 225–236 (2019).
31. Islam, N., Huang, W. J. & Zhuang, H. L. L. Machine learning for phase selection in multi-principal element alloys. *Comp. Mater. Sci.* **150**, 230–235 (2018).
32. Zhou, Z. Q. et al. Machine learning guided appraisal and exploration of phase design for high entropy alloys. *Npj Comput. Mater.* **5**, 128 (2019).
33. Kaufmann, K. & Vecchio, K. S. Searching for high entropy alloys: A machine learning approach. *Acta Mater.* **198**, 178–222 (2020).
34. Zhang, L. et al. Machine learning reveals the importance of the formation enthalpy and atom-size difference in forming phases of high entropy alloys. *Mater. Des.* **193**, 108835 (2020).
35. Zhang, Y. et al. Phase prediction in high entropy alloys with a rational selection of materials descriptors and machine learning models. *Acta Mater.* **185**, 528–539 (2020).
36. Kaufmann, K. et al. Discovery of high-entropy ceramics via machine learning. *Npj Comput. Mater.* **6**, 42 (2020).
37. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B.* **59**, 1758–1775 (1999).
38. Perdew, J. P. et al. Restoring the density-gradient expansion for exchange in solids and surfaces. *Phys. Rev. Lett.* **100**, 136406 (2008).
39. Zhao, J. L. et al. Native point defects and oxygen migration of rare earth zirconate and stannate pyrochlores. *J. Mater. Sci. Technol.* **73**, 23–30 (2021).
40. Sun, L. C. et al. A multicomponent γ-type $(Gd_{1/6}Tb_{1/6}Dy_{1/6}Tm_{1/6}Yb_{1/6}Lu_{1/6})_2Si_2O_7$ disilicate with outstanding thermal stability. *Mater. Res. Lett.* **8**, 424–430 (2020).
41. Sun, L. C. et al. High entropy engineering: new strategy for the critical property optimizations of rare earth silicates. *J. Inorg. Mater.* **36**, 339–346 (2021).
42. Jung, H. W., Sauerland, L., Stocker, S., Reuter, K. & Margraf, J. T. Machine-learning driven global optimization of surface adsorbate geometries. *Npj Comput. Mater.* **9**, 114 (2023).
43. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
44. Zhu, X. Z. et al. Machine learning exploration of the critical factors for CO_2 adsorption capacity on porous carbon materials at different pressures. *J. Clean. Prod.* **273**, 122915 (2020).
45. Yang, X. & Zhang, Y. Prediction of high-entropy stabilized solid-solution in multi-component alloys. *Mater. Chem. Phys.* **132**, 233–238 (2012).
46. Bondar, I. A. Rare-earth silicates. *Ceram. Int.* **8**, 83–89 (1982).
47. Wang, J. et al. High-entropy ferroelastic rare-earth tantalite ceramic: $(Y_{0.2}Ce_{0.2}Sm_{0.2}Gd_{0.2}Dy_{0.2})TaO_4$. *J. Am. Ceram. Soc.* **104**, 5873–5882 (2021).
48. Chen, Z. Y. et al. Mechanism of enhanced corrosion resistance against molten CMAS for pyrosilicates by high-entropy design. *J. Am. Ceram. Soc.* **106**, 6000–6013 (2023).
49. Chen, B. & Varshney, P. K. A Bayesian sampling approach to decision fusion using hierarchical models. *IEEE T. Signal Proces.* **50**, 1809–1818 (2002).
50. He, J. P., Tu, Y. Y. & Shi, Y. Q. Fusion model of multi monitoring points on dam based on Bayes theory. *Procedia Eng.* **15**, 2133–2138 (2011).
51. Kittler, J. & Alkoot, F. M. Sum versus vote fusion in multiple classifier systems. *IEEE T. Pattern Anal.* **25**, 110–115 (2003).

52. Wang, X. et al. Preparation and corrosion resistance of high-entropy disilicate $(Y_{0.25}Yb_{0.25}Er_{0.25}Sc_{0.25})Si_2O_7$ ceramics. *Corros. Sci.* **192**, 109786 (2021).
53. Fujii, S., Ioki, A., Yokoi, T. & Yoshiya, M. Role of phonons on phase stabilization of $RE_2Si_2O_7$ over wide temperature range (*RE* = Yb, Gd). *J. Eur. Ceram. Soc.* **40**, 780–788 (2020).
54. Guo, X. T. et al. High-entropy rare-earth disilicate $(Lu_{0.2}Yb_{0.2}Er_{0.2}Tm_{0.2}Sc_{0.2})Si_2O_7$: A potential environmental barrier coating material. *J. Eur. Ceram. Soc.* **42**, 3570–3578 (2022).
55. Stokes, J. L., Harder, B. J., Wiesner, V. L. & Wolfe, D. E. Effects of crystal structure and cation size on molten silicate reactivity with environmental barrier coating materials. *J. Am. Ceram. Soc.* **103**, 622–634 (2020).
56. Salanova, A., Brummel, I. A., Yakovenko, A. A., Opila, E. J. & Ihlefeld, J. F. Phase stability and tensorial thermal expansion properties of single to high-entropy rare-earth disilicates. *J. Am. Ceram. Soc.* **106**, 3228–3238 (2023).
57. Troparevsky, M. C., Morris, J. R., Kent, P. R. C., Lupini, A. R. & Stocks, G. M. Criteria for predicting the formation of single-phase high-entropy alloys. *Phys. Rev. X.* **5**, 011041 (2015).
58. Li, Y. & Guo, W. L. Machine-learning model for predicting phase formations of high-entropy alloys. *Phys. Rev. Mater.* **3**, 095005 (2019).
59. Pedregosa, F. et al. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).
60. Hao, J. G. & Ho, T. K. Machine learning made easy: a review of Scikit-learn Package in Python programming language. *J. Educ. Behav. Stat.* **44**, 348–361 (2019).
61. Oyedele, O. Determining the optimal number of folds to use in a K-fold cross-validation: A neural network classification experiment. *Res. Math.* **10**, 2201015 (2023).
62. Marzban, C. The ROC curve and the area under it as performance measures. *Weather Forecast* **19**, 1106–1114 (2004).
63. Mangalathu, S., Hwang, S. H. & Jeon, J. S. Failure mode and effects analysis of RC members based on machine-learning-based SHapley Additive exPlanations (SHAP) approach. *Eng. Struct.* **219**, 110927 (2020).
64. Rodriguez-Perez, R. & Bajorath, J. Interpretation of compound activity predictions from complex machine learning models using local approximations and shapley values. *J. Med. Chem.* **63**, 8761–8777 (2020).
65. Zhang, J., Ma, S. H., Xiong, Y. X., Xu, B. A. & Zhao, S. J. Elemental partitions and deformation mechanisms of $L1_2$-type multicomponent intermetallics. *Acta Mater.* **219**, 117238 (2021).
66. Zhao, S. J., Stocks, G. M. & Zhang, Y. W. Stacking fault energies of face-centered cubic concentrated solid solution alloys. *Acta Mater.* **134**, 334–345 (2017).
67. Dahlqvist, M., Alling, B. & Rosen, J. Stability trends of MAX phases from first principles. *Phys. Rev. B.* **81**, 220102 (2010).

Acknowledgements

This work is supported by the National Natural Science Foundation of China (No. U21A2063, 52172071 and 51972080). Bin Liu acknowledges research support from Shanghai Technical Service Center for Advanced Ceramics Structure Design and Precision Manufacturing (No. 20DZ2294000).

Author contributions

Yun Fan: Conceptualization, Investigation, Formal analysis, Data Curation, Mading and training the ML model, Writing-Original draft preparation. Yuelei Bai: Writing-Reviewing and Editing, Supervision, Funding acquisition. Qian Li: Writing-Reviewing and Editing. Zhiyao Lu: Visualization, Data Curation. Dong Chen: Mading and training the ML model. Yuchen Liu: Investigation, Formal analysis. Wenxian Li: Writing-Reviewing and Editing, Visualization. Bin Liu: Investigation, Resources, Supervision, Funding acquisition, Writing-Reviewing and Editing.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41524-024-01282-x.

Correspondence and requests for materials should be addressed to Yuelei Bai or Bin Liu.