

Machine learning-enabled discovery of refractory high entropy alloys on curated data

Yonggang Yan

Sichuan University

Nan Zhou

Sichuan University

Yalin Liao

University of Delaware

Lijuan Cui

Sichuan University

dahuan Zhu

Chinese Academy of Sciences

Xunxiang Hu

huxx@scu.edu.cn

Sichuan University

Article

Keywords: Refractory high entropy alloys, Phase formation, Machine learning, Most stable oxidation state

Posted Date: August 3rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10354540/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Machine learning-enabled discovery of refractory high entropy alloys on curated data

Yonggang Yan^{1*}, Nan Zhou¹, Yalin Liao², Lijuan Cui¹, Dahuan Zhu^{3*} and Xunxiang Hu^{1*}

¹College of Physics, Sichuan University, Chengdu 610065, China

²Department of Electrical and Computer Engineering, University of Delaware, Evans Hall, 139 The Green, Newark, DE 19716, USA

³Institute of Plasma Physics, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, 230031, China

⁺**Corresponding author.**

E-mail: huxx@scu.edu.cn, yanyonggang24@scu.edu.cn, dhzhu@ipp.ac.cn

Abstract: Refractory high-entropy alloys (RHEAs) exhibit outstanding potential for high-temperature structural applications, yet their rational design remains challenging because of the enormous compositional space and inconsistent literature data. Herein, we develop a machine-learning framework for predicting phase stability using a carefully curated RHEA dataset that combines secondary phase-inspired descriptors, refined synthesis metadata, and reassessed phase labels. The resulting ML model achieves a high predictive performance (F1 score = 0.89) in identifying single-phase solid solution RHEAs. Model interpretation reveals a key finding: RHEAs with a mismatch in the most stable oxidation state among constituent elements (MSO_{dv}) substantially exceeding 15 are prone to forming ordered intermetallic phases, thereby compromising single-phase stability. Theoretical analysis further links MSO to the itinerant electron characteristics of elements, elucidating the electronic origin of its impact on phase formation. Additionally, factors such as annealing temperature and first ionization energy are also shown to govern phase formation by modulating multiphase solid solution formation. Guided by ML predictions, several RHEAs were successfully synthesized and validated experimentally. This study establishes a robust predictive framework and unveils previously unrecognized electronic mechanisms, offering a new paradigm for the rational discovery of high-performance RHEAs.

Keywords: Refractory high entropy alloys, Phase formation, Machine learning, Most stable oxidation state

1. Introduction

The quick-emerging paradigm of metallic alloy development with multiple principal elements has rendered salient advantages over their dilute counterparts in seeking optimal mechanical or functional performances within the seemingly infinite compositional space[1]. Refractory high entropy alloys (RHEAs), a subgroup of these heavily alloyed systems, have stimulated increasing interest since the report of the first example with single-phase body-centered cubic (BCC) structure[2], mostly ascribed to the excellent mechanical properties[3-10]. Subsequently, the field has expanded significantly into other areas, such as electrocatalysis[11], superconductivity[12-14], metallic biomaterials[15], electrocatalysts[16], and energy storage[17-19]. Their phase structures, ranging from simple disordered solid solution[20], to heterogeneous microstructures[21-23], and even to metallic glass[24, 25], directly affect the properties[26]. Therefore, phase prediction is of great significance in the design of RHEAs, as it offers valuable insights into their expected mechanical behavior and other essential properties.

Unfortunately, phase prediction across such a vast compositional space remains extremely challenging due to the absence of efficient design tools. Despite extensive theoretical and experimental efforts, only a limited number of RHEAs have been experimentally confirmed by materials scientists. This can be primarily attributed to the following reasons. Trial-and-error experimental methods are excessively resource-intensive, making it impossible to efficiently explore the vast compositional space. Alternatively, one can design RHEAs using the Calculation of Phase Diagrams (CALPHAD) methods. Nevertheless, the effectiveness of CALPHAD methods is significantly undermined due to the absence of reliable thermodynamic databases[27]. Additionally, enthalpies obtained from density functional theory (DFT) calculations have already been fed into some semi-empirical phase prediction models[28] and CALPHAD[29]. However, DFT calculations for RHEAs require either sufficiently large supercells to capture configurational disorder or statistical approaches such as cluster expansions[30, 31], rendering their computationally prohibitive for high-throughput screening of RHEAs. Moreover, the empirical criteria, such as atomic radius mismatch (R_c), valence electron concentration (VEC) criterion, tabulated mixing enthalpy (ΔH_{mix}), have been developed to rapidly predict the formation of HEAs[32, 33], based on experimentally determined phase information from previously reported alloys. However, the sluggish diffusion kinetics of refractory elements often lead to the formation

of metastable dendritic or lamellar structures in as-cast RHEAs, thereby complicating experimental phase identification. Consequently, the solidness of empirical methods is undermined.

The past decade has witnessed the explosive rise of machine learning (ML) in phase formation investigations, where readily accessible phase data are typically used as labels, while weight-averaged elemental properties, empirical parameters or thermodynamic descriptors - calculated via DFT or CALPHAD - serve as features[34-36]. In addition to the label quality issues arising from experimental characterization discussed above, the inherent theoretical limitations of these empirical parameters also raise concerns about the reliability of the ML models. For instance, the debated values of itinerant electrons in transition metals (TMs) complicate the use of Hume-Rothery electron concentration rule as effective descriptors in ML frameworks[37].

Herein, we proposed a ML framework for the screening of potential single-phase solid solution (SPSS) RHEAs based on literature-reported RHEAs (Fig. 1). To overcome the above-mentioned obstacles, this work begins with the development of a curated dataset featuring three key innovations: (i) to improve the quality of label, phases were carefully re-evaluated based on X-ray diffraction (XRD) and scanning electron microscopy/energy dispersive X-ray spectroscopy (SEM/EDS) data; (ii) robust descriptors were constructed drawing upon the characteristic of the secondary phases presented in RHEAs, which encode critical information about phase formation; and (iii) synthesis-related metadata were carefully considered during dataset construction process. The curated dataset was then used to train advanced ML models, followed by interpretability methods and DFT calculations to uncover the nature of phase stability. Finally, experimental validation showed that the ML model can successfully predict the phase of RHEAs containing cations from groups IV-VI in periodic table of elements (Fig. S1). Our combined approach offers a reliable pathway to accelerate RHEA design and deepen understanding of their phase formation.

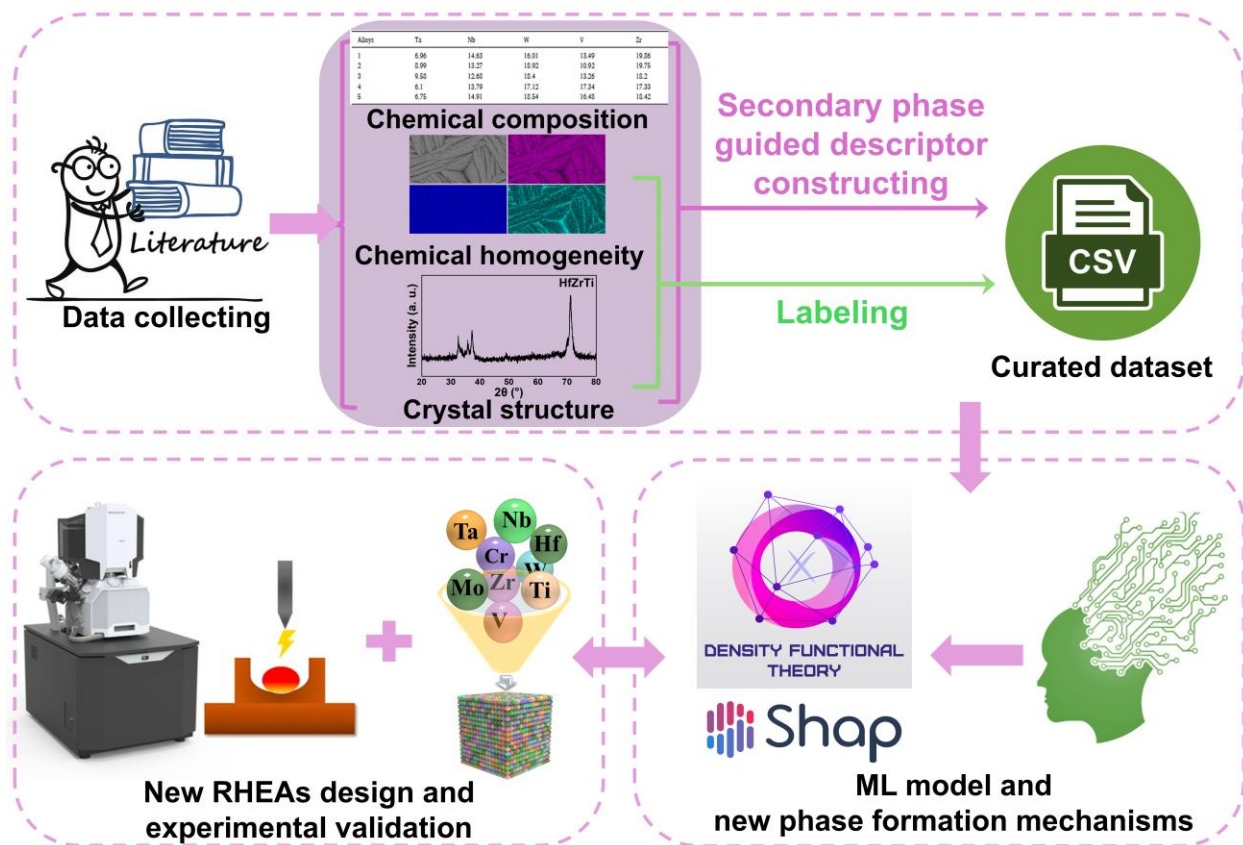


Fig. 1 Workflow for SPSS RHEA discovery. The process begins with literature-based data collection, followed by curated dataset construction with phase labeling and descriptor construction. The ML model is then trained on the curated dataset, and model interpretability methods, together with theoretical simulations, are employed to elucidate the phase formation mechanisms. Finally, the reliability of the model is validated through experimental verification.

2. Results

2.1 Curated dataset

RHEAs only consisting of refractory elements (Ti, V, Cr, Zr, Nb, Mo, Hf, Ta, and W) were collected to construct the dataset. To address data quality challenges, systematic efforts including refinement of alloy metadata, improvement of label accuracy, and the development of secondary phase-informed descriptors, were implemented. To eliminate uncertainties from different fabrication routines, arc-melted RHEAs were selected as candidates. These alloys constitute more than 90% of the literature-reported RHEAs. Furthermore, RHEAs subjected to homogenization or recrystallization were retained. After a rigorous selection process, only heat-treated arc-melted RHEAs were deemed qualified and used for the final dataset construction.

As literature often fails to strictly adhere to the definition of SPSS, directly adopting the reported phase as labels usually compromise model accuracy and the reliability of derived physical insights[38]. Therefore, the phase of each RHEA was re-evaluated using their XRD patterns and SEM-EDS elemental mappings. RHEA is identified as SPSS if it simultaneously meets two key criteria: a simple crystal structure and micro-scale chemical homogeneity. Although the criteria may not fully capture the characteristics of SPSS, they offer a practical strategy given the trade-offs between accuracy and dataset size. For instance, XRD and SEM-EDS lack the resolution to detect minor nanoscale secondary phases[6], so the immiscible nanoparticles or lamellae[39] were inevitably excluded in some cases during phase assessment. Acknowledging these limitations, we prioritized consistency and robustness over excessive precision in phase labeling. In total, 675 RHEAs were screened, including 333 SPSS and 342 multiphase alloys (Fig. 2a). Among the multiphase alloys, 140 featured multiphase SSs, while the remainder consisted of SS matrix with IM. Note that although we provided detailed labeling of multiphase samples here, we only performed binary classification. The labeling strategy adopted in this study provides a robust basis for ensuring the reliability of subsequent model training and mechanism elucidation.

To construct more reliable descriptors, RHEAs were statistically analyzed according to their constituent elements as well as secondary phases. As depicted in Fig. 2b, 77% of Cr-containing alloys exhibit IM phases, suggesting Cr's strong tendency to promote IM formation. By contrast, a high fraction of multiphase SSs is observed in RHEAs that contain Ta and W. An analysis of the secondary phases in the multiphase SSs was also conducted (Fig. 2c). Most multiphase SSs exhibit dual-phase structure, indicating that the refractory elements tend to partition into two subsets. Among these dual-phase SSs, W and Ta tend to form one SS phase, while Ti, Zr, Hf, and V preferentially form another one. Counter to chemical intuition, if a significant difference exists in a specific property between the two groups of elements, it is likely to serve as a key factor governing the multiphase SSs formation. Therefore, elemental properties, such as IP1, cohesive energy (E_c), and melting point (T_m) tend to have strong potential correlations with immiscibility or phase separation, whereas properties like thermal conductivity (TC) and density (ρ) are deemed to be less relevant to phase formation (Fig. 2d). Building on these insights, new descriptors were developed to better characterize the SSs formation of RHEAs (Eq. 2 in Methods section). In addition, the high fraction of IMs formed in Cr-containing RHEAs may be linked to pronounced property differences between Cr and other elements, e.g., MSO and geometric properties, which

could be key contributors to IM formation, as presented in Fig. 2d. Analogous to SSs descriptors, MSO_{dv} was calculated as a descriptor for IM prediction. Since many geometric properties (e.g., molar volume, ionic radius, lattice parameter) are highly correlated based on Pearson correlation analysis, only covalent radius (R_c) was retained in the final dataset. In addition to R_c , electronic properties with known relevance to phase behavior, like VEC and electrons per atom ratio (e/a), were selected to construct descriptors. Thermodynamic parameters (ΔH_{mix} and ΔS_{mix}) were also incorporated. Annealing temperature (T_{ann}) and duration (time) were explicitly incorporated as descriptors. After a correlation analysis, a final descriptor set composed of 12 descriptors (Table S1) were used to train ML models. Their normalized value distribution is shown in Fig. 2e.

Pearson correlation analysis indicates that the selected descriptors exhibit only moderate intercorrelations, suggesting limited multicollinearity. More importantly, Mantel analysis reveals that different descriptors are associated with distinct phase-formation pathways (Fig. 3). MSO_{dv}, ΔH_{mix} and R_c show strong correlations with SS+IM. ΔH_{mix} also exhibits significant associations with SSs. Interestingly, IP1_b is only weakly correlated with SSs, which may result from label uncertainty of multiphase SSs. For example, some SSs alloys were classified as SS+IM due to the presence of IMs, introducing ambiguity in SSs labeling. Additionally, the imbalance between SSs and non-SS samples may hinder the model's ability to accurately identify SSs.

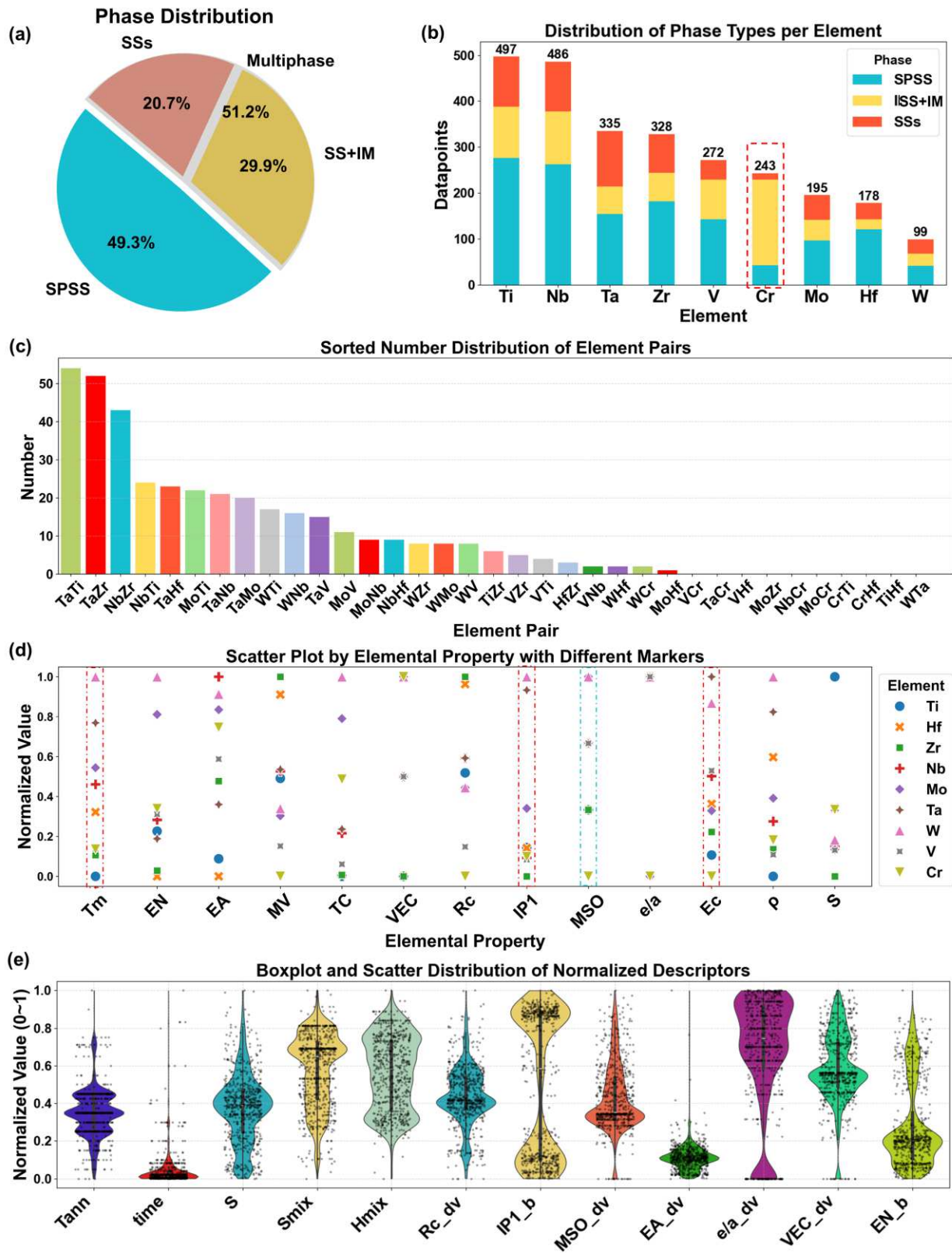


Fig. 2 Statistical analysis of the dataset. (a) Pie chart showing the data distribution categorized by phase type; (b) Histograms displaying phase distributions for different constituent elements; (c)

Dominant elemental pairs in dual-phase SSs, indicating the elements enriched in each corresponding SS phase; (d) Distribution of the properties of different elements; (e) Violin plot demonstrating the distribution of normalized values for each descriptor.

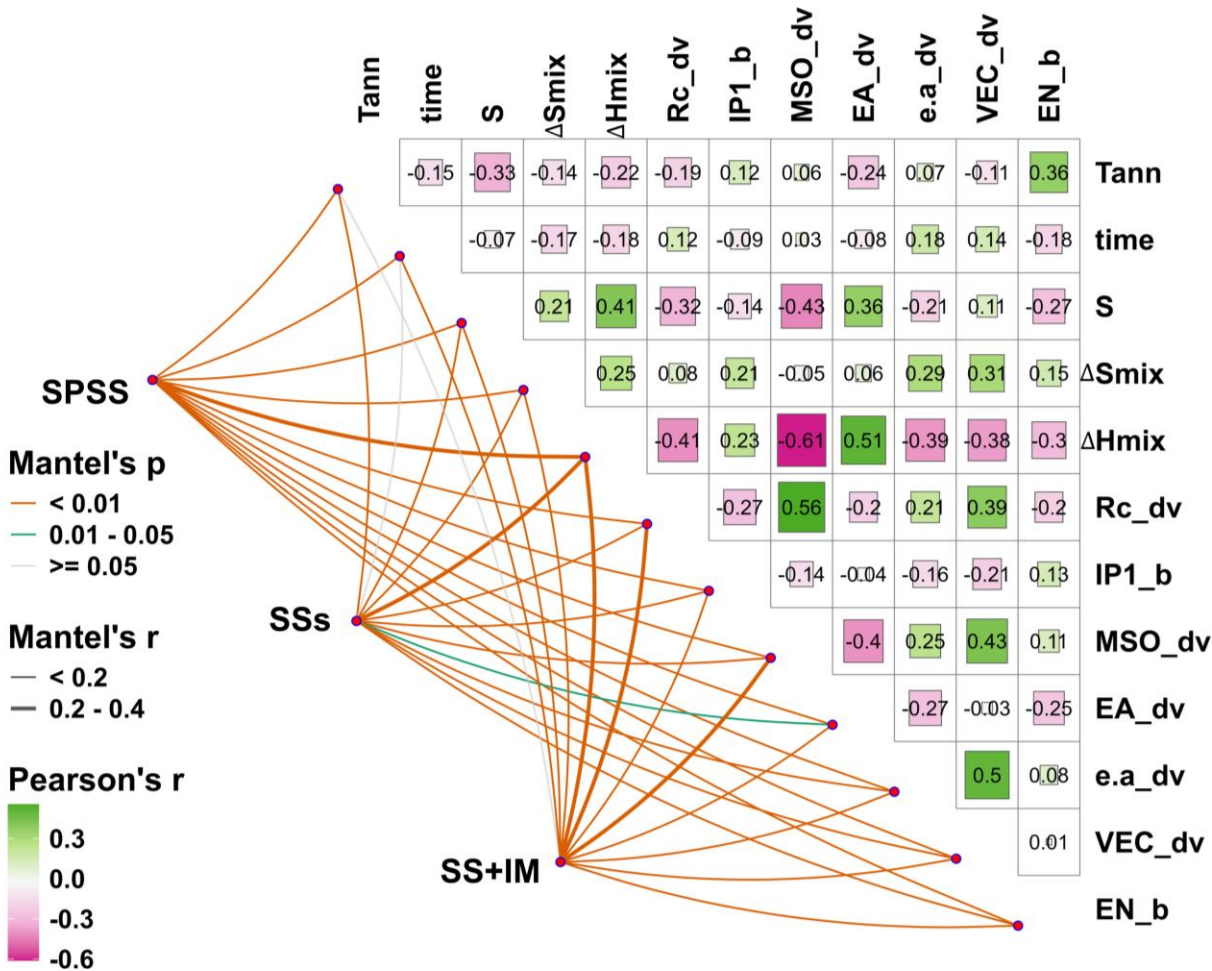


Fig. 3 Pearson correlation and Mantel test analysis between phase and descriptors. The upper triangular matrix displays the correlation between the descriptors, with colors representing the strength and direction of the correlations (red for negative, green for positive). The size of the squares is proportional to the magnitude of the correlation coefficients. Overlaying the correlation matrix, lines represent significant relationships identified by the Mantel test, where line color indicates the p-value category (<0.01, 0.01-0.05, ≥0.05), line thickness represents the Mantel's r value (<0.2, 0.2-0.4), and line type (solid or dashed) corresponds to the p-value significance levels.

2.2 Machine learning models

Fig. 4a presents the average F1 score obtained from 40 repetitions of 5-fold cross-validation (F1_CV) of the final model across different models. For relatively small datasets, model performance is more reliably evaluated using the average F1 score obtained from repeated cross-validation rather than the performance on a single train-test split. Overall, several tree-based

ensemble models (RF, GB, XGB) achieved statistically comparable performances, consistently outperform the remaining algorithms, such as LR, KNN and SVM. Notably, the GB model demonstrated both strong performance and robustness, achieving an F1_CV of 0.894 ± 0.024 . The hyperparameters of the GB model are listed in Table S2. Moreover, the low standard deviation (<0.03) of F1_CV indicates that the GB model's hyperparameters are well-tuned (Fig. 4a). The fluctuations in F1_CV are mainly attributed to the limited dataset size, which restricts the representation of the full data distribution and increases sensitivity to individual samples. Therefore, the GB model retrained on the entire dataset was selected for subsequent phase prediction and elucidation of the underlying physical mechanisms.

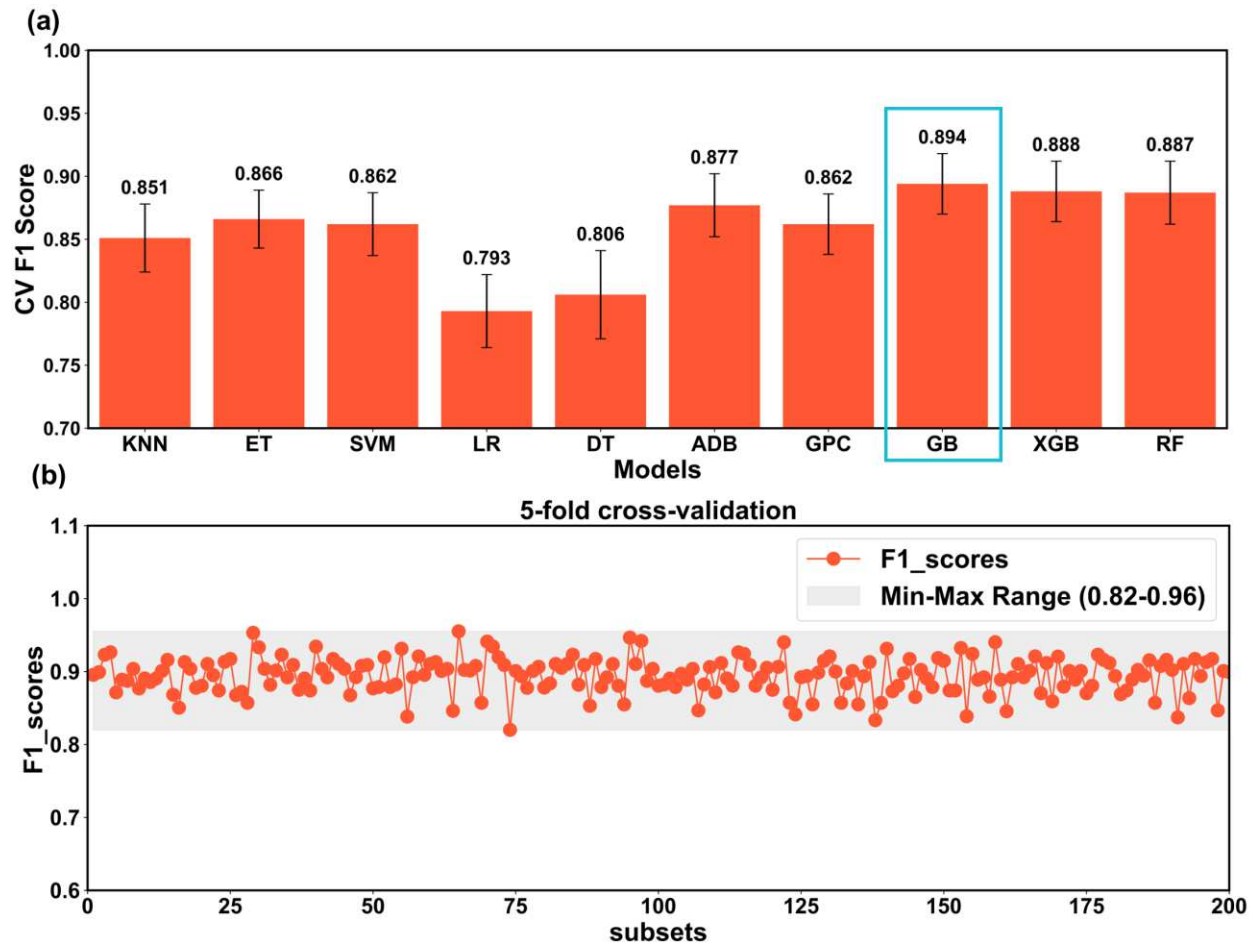


Fig. 4 Performance of models for SPSS prediction. (a) Comparison of average CV-F1 scores among different ML models based on 40 repeated five-fold cross-validation runs; (b) F1_scores of the GB models among 40 repetitions of 5-fold CV.

2.3 Phase formation mechanisms

SHAP analysis of the GB model for SPSS prediction reveals the following order of descriptor importance: MSO_{dv} , ΔH_{mix} , S , T_{ann} , IP1_{b} , $R_{\text{c_dv}}$, EA_{dv} , VEC_{dv} , EN_{b} , time , e/a_{dv} and ΔS_{mix} (Fig. 5a-b). In addition to the primary SPSS prediction model, two auxiliary binary classification models for SS+IM and multiphase SS formation were independently developed to facilitate the mechanistic interpretation of descriptor effects. The predictive performance and feature importance rankings of these auxiliary models are presented in Figs. S4-5. The details of each descriptor are listed in Table S1.

The mechanisms by which the key descriptors influence phase formation were elucidated. MSO_{dv} unexpectedly exhibits the highest impact on SPSS prediction, indicating the significant role of MSO in phase formation of SPSS. SHAP analysis further reveals a negative correlation between MSO_{dv} and SPSS, where greater MSO disparity among elements hinders SPSS formation (Fig. 5b). Despite the lack of a well-defined theoretical framework linking oxidation states to phase formation, oxidation states likely capture electron concentration of the constituent elements that influence phase formation[40, 41]. MSO, the typical number of electrons donated by a metal to form its most stable oxide, is presumed to reflect the itinerant electrons to chemical bonding[42]. Therefore, the strong effect of MSO for phase formation can be surprising at first but is rationalized by its correlation with electron concentration. The bonding of ordered IM involves a complex interplay of metallic, covalent, and ionic interactions, largely influenced by the itinerant electrons of constituent elements[43-45]. Therefore, we propose that MSO tends to influence SPSS formation by modulating IM formation. This hypothesis is strongly supported by both IM (or SS+IM) and SSs models (Figs. S4-5). Specifically, interpretability analysis reveals that MSO_{dv} strongly influences IM prediction, while exerting minimal impact on multiphase SSs prediction (Fig. S5a). Further, SHAP analysis indicates a positive correlation between MSO_{dv} and IM prediction (Fig. S5b). Additionally, PDP and the corresponding ICE plots visualize that the probability of IM prediction rises sharply when MSO_{dv} exceeds around (Fig. 5c2). Electron concentration is commonly represented by VEC or e/a in previous studies. VEC includes all d and s electrons without distinguishing their itinerancy, whereas e/a considers only s electrons[46]. However, d -electrons of TMs exist in a quantum superposition of localized and itinerant states[47, 48]. In addition, d orbitals dominate the valence behavior of TMs[49]. Consequently, their ability to represent itinerant electrons in TM-based alloys remains debated[50], which could be indirectly proved by the uncertainty in RHEAs phase investigations[45, 51]. This aligns with our finding that

212 VEC_{dv} and e/a_{dv} play minor roles in phase formation. The markedly higher importance of
213 MSO_{dv} over both VEC_{dv} and e/a_{dv} supports the hypothesis that MSO reflects the number of
214 itinerant electrons more accurately (Fig. 5a-b and Fig. S5a-b), thereby better governing IM
215 formation trends. This hypothesis is exemplified by Cr-containing RHEAs, which often form IMs
216 of the type XCr₂, such as HfCr₂, NbCr₂, TaCr₂, TiCr₂, and ZrCr₂. Notably, all refractory elements
217 in this work (except Cr) share identical MSO and VEC values (Table S3). Cr's distinctive
218 electronic configuration, with its half-filled 3*d* subshell, spin-density wave behavior, and
219 localized-itinerant duality, complicates the determination of its itinerant electrons. MSO (+3) of
220 Cr reflects a propensity to donate three bonding electrons, in contrast to VEC (6), which does not
221 account for itinerancy. Hence, MSO offers a more realistic measure of Cr's contribution to IM
222 formation.

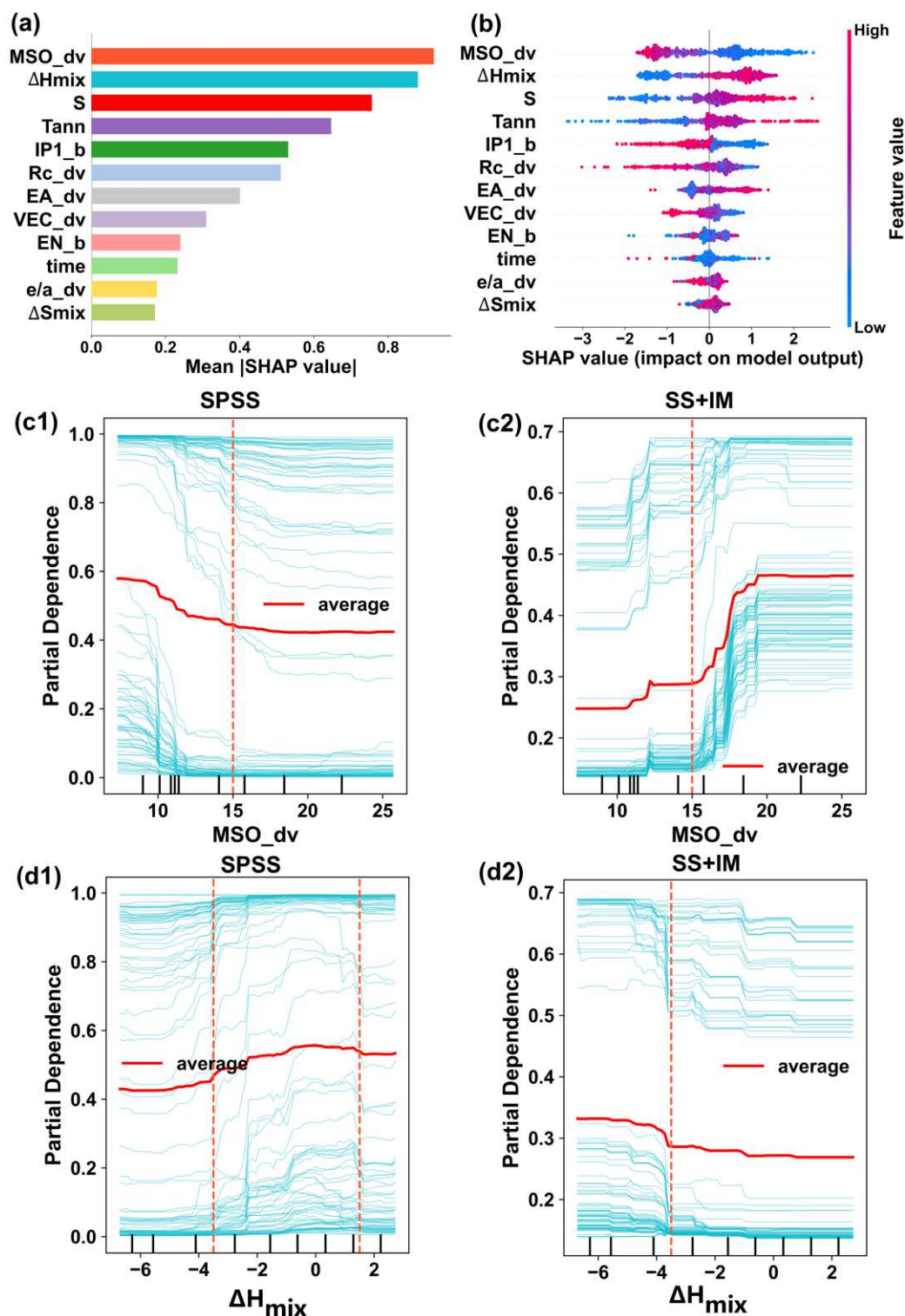


Fig. 5 Interpretation of the GB model for SPSS prediction. (a) Relative importance of the descriptors ranked by the mean absolute SHAP values. (b) SHAP summary plot, where each point represents one sample and the color indicates the original feature value. The horizontal position represents the contribution of each descriptor to the model output. (c-d) Descriptor-dependent phase-transition thresholds revealed by partial dependence plots (PDPs) and individual conditional

expectation (ICE) plots. The light-blue curves represent ICE plots, the red curves denote the corresponding PDPs, and the orange dashed lines indicate characteristic thresholds. (c1-c2) Effect of MSO_{dv} on the prediction probabilities of SPSS and SS+IM, respectively. Increasing MSO_{dv} markedly suppresses SPSS formation while promoting SS+IM formation. (d1-d2) Effect of ΔH_{mix} on phase prediction. SPSS exhibits the highest probability within a ΔH_{mix} window of approximately -3.5 to 1.5 kJ/mol, whereas increasingly negative ΔH_{mix} promotes SS+IM formation.

To elucidate the atomic-scale mechanism underlying Cr's tendency to form IMs in RHEAs, we analyzed the electronic structure via ELF and DOS, using equal-molar TaNbVX (X = 0, W, Mo, Cr) as model systems. TaNbV, TaNbVW, and TaNbVMo were known as SS phases, whereas TaNbVCr exhibits the formation of IM phase (Table S4). ELF analysis reveals that W and Mo share similar electron distributions, whereas Cr shows more delocalized electrons (Fig. 6b). In particular, pronounced charge localization appears between Cr and V atoms (Fig. 6b2), indicating stronger electron sharing and a tendency toward covalent-like bonding. DOS results demonstrate that *d*-electrons dominate metallic interactions. W and Mo show similar *d*-electron features (Fig. 5c1-2), while Cr displays significantly lower *d*-electron density below the Fermi level (Fig. 6c3). This observation indicates that Cr retains fewer delocalized electrons for metallic bonding than its group 6 congeners W and Mo. By integrating *d*- and *s*-electron densities below the Fermi level, Cr, Mo, and W count 3.5, 4.63, and 4.44 electrons with per atom, respectively. Though W, Ta, and Cr all have a VEC of 6, their MSO differs: +6 for W and Ta, +3 for Cr. The MSO trend aligns with the total number of *d*- and *s*-electrons below the Fermi level, suggesting MSO accurately reflects the itinerant electrons participating in metallic bonding. Additionally, Bader charge analysis demonstrates that the addition of Cr to TaNbV results in minimal changes in overall charge transfer, in contrast to the notable effects observed with Mo and W (Table S4). This suggests that the interaction between Cr and the base alloy atoms exhibits quasi-covalent characteristics, thereby promoting its tendency to form IM phases. Intriguingly, Mo gains nearly twice the amount of charge as Cr, while W acquires almost three times as much - despite Mo and W having comparable MSO. This quantitative disparity may stem from the combined influence of the MSO and electronegativity: W possesses a significantly higher electronegativity than Mo, enhancing its capacity to attract electrons during charge transfer. Therefore, MSO serves as a proxy for itinerant electron characteristics governing IM formation compared to VEC and *e/a*.

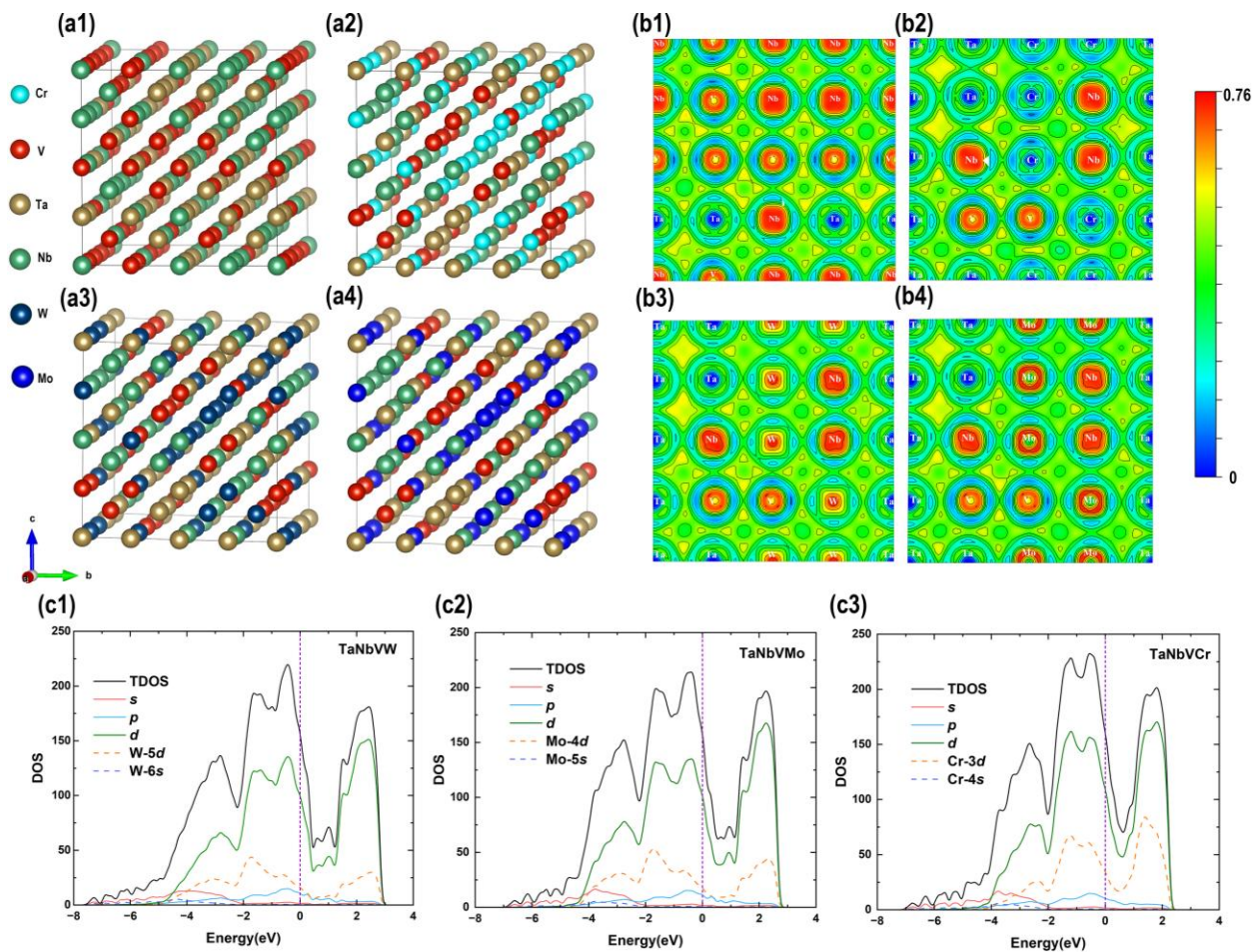


Fig. 6 DFT calculations showing the electronic characteristics of RHEAs. (a) The generated SQS structures for RHEAs; (b) The 2D-electron localization function (ELF) of different RHEAs; (c) Total and partial densities of states (DOS) of RHEAs.

As illustrated in Fig. 5a-b, Miedema-based ΔH_{mix} ranks just behind MSO_{dv} , highlighting its effectiveness for SPSS prediction. SHAP reveals an overall positive contribution, whereas PDP further shows that the relationship is highly nonlinear, with an optimal ΔH_{mix} window of -3.5 to 1.5 kJ/mol (Fig. 5d1). When ΔH_{mix} values below -3.5 kJ/mol, the alloy exhibits a stronger tendency toward IM formation (Fig. 6e and Fig. 5d2). Within the Miedema framework, ΔH_{mix} is determined by binary mixing enthalpies ($\Delta H_{\text{mix}}^{\text{bin}}$) of constituent elemental pairs. Therefore, increasingly negative ΔH_{mix} values indicate the presence of strongly attractive binary interactions, providing a larger thermodynamic driving force for ordered compound formation than for random atomic mixing. This behavior is exemplified by Cr-containing RHEAs, in which highly negative $\Delta H_{\text{mix}}^{\text{bin}}$ of Cr-X frequently promote Laves phases such as NbCr_2 , TaCr_2 , TiCr_2 , and HfCr_2 (Fig. 2b). Nevertheless, ΔH_{mix} alone cannot fully determine phase stability, as W-Mo-Ta-Nb systems often

retain SPSS despite exhibiting negative ΔH_{mix} [52]. Conversely, when ΔH_{mix} values above 1.5 kJ/mol, increasingly repulsive binary interactions promote phase separation, leading to multiphase SSs. Although SHAP suggests an overall positive contribution of ΔH_{mix} to SPSS prediction, this apparent discrepancy mainly reflects entropy stabilization at elevated temperatures, the persistence of metastable solid solutions under limited annealing conditions, and the limited detectability of micro- or nanoscale phase separation by conventional XRD or SEM-EDS characterization.

As shown in Fig. 5a-b, the solubility parameter (S) is identified as a key descriptor for SPSS prediction, with SHAP analysis indicating that higher S values favor SPSS formation. S represents the average maximum equilibrium solubility of the constituent elements, reflecting their intrinsic chemical compatibility in alloy formation[53]. IP1_b exhibits relatively weak influence on SPSS prediction (Fig. 5a), while it ranks as the top descriptor in SSs model (Fig. S5c). Note that SSs model was trained on 473 alloys after excluding all IM-containing entries. The role of IP1_b in SPSS prediction is best understood via its influence on multiphase SSs formation. IP1 reflects the binding strength of first outer electron from an atom and relates to metallic bonding behavior[54]. Small IP1_b promotes uniform electron delocalization and bonding, favoring SPSS. Conversely, larger IP1_b can induce phase segregation, thus IP1_b affects SPSS indirectly via phase separation tendencies.

Beyond the intrinsic compositional descriptors, T_{ann} also significantly influences SPSS prediction (Fig. 5a). Increasing T_{ann} enhances atomic diffusion, promoting chemical homogenization and the dissolution of secondary phases, thereby favoring SPSS formation. However, prolonged annealing may induce phase decomposition because many RHEAs are thermodynamically metastable at intermediate temperatures. In contrast, T_{ann} ranks among the least important descriptors in the IM model (Fig. S5c), reflecting the exceptional thermal stability of refractory intermetallic compounds, which remain resistant to dissolution even at elevated temperatures. Although annealing time also affects phase evolution through diffusion kinetics, its contribution to the present ML model is limited because of the relatively small dataset and the insufficient coverage of annealing-time conditions.

Collectively, the descriptors regulate phase formation through complementary physical mechanisms. Electronic descriptors (MSO_dv and IP1_b) primarily control the tendency toward

IM or multiphase formation, ΔH_{mix} provides the thermodynamic driving force, whereas T_{ann} determines whether these thermodynamic preferences can be realized through diffusion kinetics.

2.4 Experimental validation

To validate the reliability of the ML model, the phases of over 500 RHEAs (not in dataset) were predicted using our GB model. Most were previously reported as-cast state, primarily produced by arc melting, and the rest were newly designed in present study. These compositions exhibit promising potential across diverse fields such as nuclear energy[55], metallic biomaterials[56], superconductivity[57], and other fields[58]. Among these RHEAs, 40% were predicted to stabilize as SPSS after annealing at 1400 °C for 4 h (Fig. 7a). Ten alloys were randomly selected for experimental validation (Table S5), six of which are shown in Fig. 7b-g. $\text{W}_{0.33}\text{Mo}_{0.33}\text{Ta}_{0.33}$, $\text{Nb}_{0.26}\text{Ta}_{0.24}\text{Ti}_{0.21}\text{V}_{0.24}\text{W}_{0.05}$ and $\text{W}_{0.5}\text{Ta}_{0.5}\text{Mo}_1\text{Nb}_1$ - all exhibiting a negative ΔH_{mix} and small MSO_{dv} - demonstrate excellent chemical homogeneity at the microscale and adopt a simple BCC crystal structure. In addition, $\text{Hf}_1\text{Ti}_1\text{Nb}_1\text{Ta}_{0.2}$ (minor Ta oxides particle separated in SS matrix) and $\text{Hf}_1\text{Ti}_1\text{Nb}_1\text{Zr}_{0.2}$ with a positive ΔH_{mix} and small MSO_{dv} exhibits quasi-chemical homogeneity. Moreover, $\text{W}_1\text{Mo}_1\text{Ta}_1\text{V}_1\text{Cr}_1$ and $\text{W}_1\text{Mo}_1\text{Ta}_1\text{V}_{0.25}\text{Cr}_{0.25}$ were identified as multiphase SSs due to their obvious elemental segregations. $\text{W}_1\text{Mo}_1\text{Ta}_1\text{V}_{0.25}\text{Cr}_{0.25}$ demonstrates an obvious heterogeneous morphology (Fig. 7g). The EDS mappings confirm the secondary phases are MoVCr rich SS and WTa rich SS phases (Fig. S10b). Furthermore, as-cast $\text{Hf}_{0.33}\text{Zr}_{0.33}\text{Ti}_{0.34}$ shows perfect chemical homogenous and simple hexagonal close-packed (HCP) crystal structure (Fig. S10), while phase decomposition was observed after annealing at 900 °C and 1500 °C for 4 h, respectively. This is further confirmed by the XRD pattern of annealed $\text{Hf}_{0.33}\text{Zr}_{0.33}\text{Ti}_{0.34}$ (Fig. 7f). $\Delta H_{\text{mix}}^{\text{bin}}$ of $\text{Hf}_{0.33}\text{Zr}_{0.33}\text{Ti}_{0.34}$ being equal to zero suggests weak atomic interactions, potentially resulting in phase decomposition at elevated temperatures. Overall, the experimental results show strong agreement with model predictions, validating the reliability of the ML model.

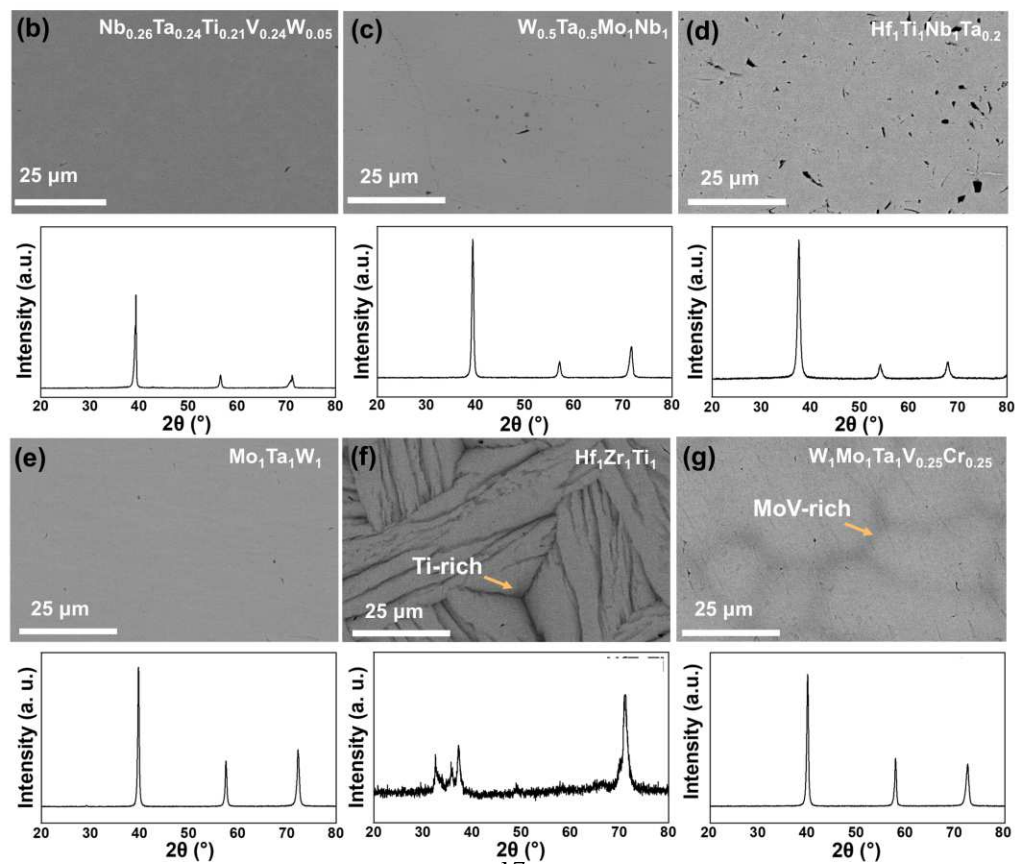
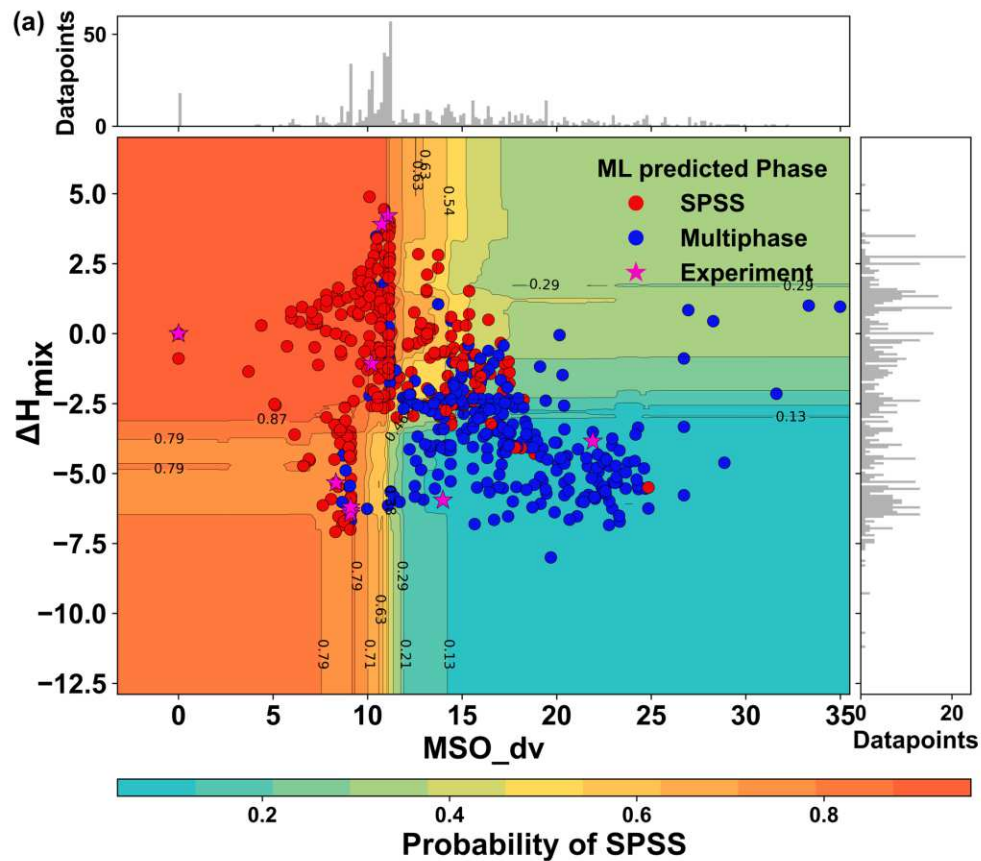


Fig. 7. Two-way PDPs showing the joint influence of top two descriptors on phase prediction outcomes. (a) Probability of SPSS formation, the accompanying histograms show the distribution of sample counts along the MSO_{dv} and ΔH_{mix} axes. Solid circles represent as-cast RHEAs, while red stars indicate the alloys developed in this study. BSE-SEM images and XRD pattern of the typical 6 new RHEAs: (b) Nb_{0.26}Ta_{0.24}Ti_{0.21}V_{0.24}W_{0.05}; (c) W_{0.5}Ta_{0.5}Mo₁Nb₁; (d) Hf₁Ti₁Nb₁Ta_{0.2}; (e) W_{0.33}Mo_{0.33}Ta_{0.33}; (f) Hf_{0.33}Zr_{0.33}Ti_{0.34} exhibits a dual-phase microstructure, comprising Hf rich and Ti rich SS phases; (g) W₁Mo₁Ta₁V_{0.25}Cr_{0.25} composed of MoV-rich and W-rich SS phases.

3. Discussion

Herein, we presented an ML framework for high-throughput identification of potential SPSS RHEAs, based on a curated dataset from readily accessible experimental data. Notably, the ML model enhancement is achieved by the curated dataset, which integrates alloy processing routines, secondary phase-guided descriptors, and refined phase label. The model enables large-scale compositional screening at low computational cost. The ML model was applied to a wide chemical space of RHEAs constructed from nine refractory elements, the majority of which are reported in the literature in their as-cast form. Among 500 RHEAs, 40% were predicted to stabilize as SPSS after short-term high temperature annealing. The successful synthesis of the ten RHEAs further validates the ML model's predictive power.

The physical mechanisms underlying the formation of RHEAs were unveiled through a combination of model interpretability and theoretical simulations. MSO_{dv} captures the electronic characteristics associated with IM formation, challenging the conventional view that such processes are primarily controlled by δ or VEC. This stems from the fact that MSO more accurately reflects the number of itinerant electrons participating in metallic bonding, thereby serving as a physically meaningful descriptor. Specifically, the complex nature of Cr's 3d electrons complicates itinerant electron counting. As a result, VEC, which does not differentiate *d*- and *s*-orbital itinerancy, falls short in phase prediction of RHEAs. Furthermore, MSO may extend beyond phase prediction and serve as a physically meaningful proxy for the electronic characteristics that govern a wide range of material properties, including mechanical strength and irradiation resistance. In addition, the relationship between Miedema-based ΔH_{mix} and SPSS formation was visualized, offering physical insight into the role of IM formation and immiscibility/phase separation - both of which hinder the formation of SPSS. Interestingly, by tuning positive ΔH_{mix} , microstructures can be engineered through controlled immiscibility or phase separation, whereas negative ΔH_{mix} , particularly under aging conditions, may facilitate local ordering. Both mechanisms offer feasible

strategies for designing high-performance RHEAs. Meanwhile, first ionization potential, atomic size, and cohesive energy are key drivers of multiphase SSs formation.

Although annealing temperature and time are known to influence phase evolution through diffusion kinetics, their importance in the present ML model is relatively limited. This result suggests that the intrinsic thermodynamic stability determined by alloy composition dominates the final phase constitution, whereas annealing conditions mainly regulate the kinetics of approaching equilibrium. Moreover, the literature-derived dataset spans a relatively narrow range of annealing conditions compared with the vast compositional space, reducing the additional predictive information provided by processing parameters. Consequently, composition-derived descriptors account for most of the model's predictive capability, while annealing parameters serve as secondary kinetic modifiers.

4. Methods

4.1 Dataset construction

During dataset construction, the chemical compositions, fabrication methods, heat treatment conditions, and phase characterization data of reported RHEAs were meticulously collected from the published literature.

Reliable supervised learning critically depends on accurate phase labels. However, phase annotations reported in the literature are often inconsistent because different characterization techniques, synthesis conditions, and interpretation criteria have been adopted. Therefore, rather than directly using the reported phase labels, all RHEAs were systematically re-evaluated based on the original experimental characterization data. Phase characterization data mainly contains XRD patterns, SEM-EDS mappings, transmission electron microscope (TEM) images, and others. To avoid inconsistencies in phase determination reported in different literature sources, the phases of RHEAs were rigorously re-evaluated based on the experimental characterization data. Technically, synchrotron X-ray techniques, electron back-scatter diffraction (EBSD), or powder neutron scattering (PNS) are the most suitable methods for an accurate phase characterization of polycrystalline RHEAs. While these rare experimental resources are straightforward only in the rare cases of studies on RHEAs. To optimize the trade-off between data quality and size, we mainly leveraged the available XRD and SEM-EDS data to identify phases in current investigation. The

phases were classified into SPSS, multiphase SSs and SS+IMs, which were subsequently used as labels in ML framework.

With reliable phase labels established, the next challenge is to construct descriptors capable of capturing the intrinsic factors governing phase stability. Conventional descriptors are primarily derived from composition-averaged properties or elemental mismatches, which effectively characterize the overall chemical characteristics of alloys but do not explicitly describe the local chemical incompatibility responsible for secondary-phase formation. Since secondary phases originate from preferential interactions among subsets of constituent elements, we hypothesized that descriptors incorporating secondary-phase characteristics could provide additional predictive information beyond conventional composition-derived descriptors. Based on this hypothesis, we designed new descriptors (P1_b and EN_b) by integrating secondary phase information with modified Hume-Rothery rule. Steps of the descriptors are shown as follows:

Step 1: Calculate the weight averaged elemental property by using

$$\bar{x} = \sum_{i=1}^n c_i x_i \quad (1)$$

where x_i represent the elemental properties of the i_{th} constituent element, such as S, R_c , e/a, VEC, electronegativity (EN), and others. c_i refers to the atomic concentrations of the i_{th} element, and n is the total number of elemental species in an alloy. The elemental properties are also taken from ref[59].

Step 2: Divide the constituent elements into two sub-sets based on $\bar{x}$, and calculate the weighted average elemental properties of the two subsets, $\bar{x}_1$ and $\bar{x}_2$, using Eq. (1).

Step 3: Calculate the absolute difference between the mean values of the two subsets using

$$\Delta x = \text{abs}|\bar{x}_1 - \bar{x}_2| \quad (2)$$

MSO_dv was calculated using the standard deviation formula (Eq. 3), reflecting the chemical similarity among the constituent elements in alloys. The descriptors can be easily obtained from

$$X_{dv} = 100 \times \sqrt{\sum_{i=1}^n c_i \left(1 - \frac{x_i}{\bar{x}}\right)^2} \quad (3)$$

Hume-Rothery rules-based descriptors (VEC_dv, e/a_dv, EA_dv and R_c _dv) were also adopted and calculated using the standard deviation formula (Eq. 3). Additionally, we also made use of ΔH_{mix} and ΔS_{mix} as descriptors, both of which are known to be closely linked to SS formation in previous studies. ΔH_{mix} can be calculated by Miedema method[60]. The formula of ΔH_{mix} is shown as Eq (4):

$$\Delta H_{\text{mix}} = 4 \sum_{i=1, i < j}^n H_{ij} c_i c_j \quad (4)$$

where H_{ij} is the enthalpy of atomic pairs of the i_{th} and j_{th} elements.

The formula of S_{mix} is calculated as Eq (5):

$$\Delta S_{\text{mix}} = -R \sum_{i=1}^n c_i \ln(c_i) \quad (5)$$

where R is the gas constant.

The phase formation is already known to be influenced by the fabrication process and heat treatment conditions[61], both of which were considered during the dataset construction process.

Pearson correlation analysis was first employed to eliminate highly correlated descriptors and reduce multicollinearity prior to model training. Subsequently, the Mantel test was performed to evaluate the overall association between individual descriptors and phase constitution, thereby facilitating descriptor selection and physical interpretation.

4.2 Machine learning framework

Ten simple ML algorithms, including K-Nearest Neighbors (KNN), Extra Trees (ET), Support Vector Machine (SVM), Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), AdaBoost (ADB), Gaussian Process Classifier (GPC), Gradient Boosting (GB), and XGBoost (XGB), were employed for phase prediction. Besides the primary SPSS classifier, two auxiliary binary classifiers (SS+IM and multiphase SS) were additionally trained for mechanistic interpretation.

Hyperparameter optimization for each algorithm was initially performed using Optuna, an automated hyperparameter optimization framework. During the hyperparameter optimization process, SMOTE was applied to increase the volume and diversity of the training data, while StratifiedKFold was adopted to preserve the class distribution across the training and validation folds. Owing to the binary classification nature of the problem, the F1 score was used as the optimization objective. Using the optimal hyperparameters identified by Optuna, the models were subsequently trained and evaluated using repeated 5-fold cross-validation (CV). Specifically, the dataset was randomly partitioned into five folds, and the cross-validation procedure was repeated 40 times with different random seeds. After model selection and performance evaluation, the optimal hyperparameters were fixed, and the final model was retrained on the entire dataset for composition screening and model interpretation.

Visualization and analysis tools, including Partial dependence plots (PDP) and the corresponding individual conditional expectation (ICE) plots and SHapley Additive exPlanations (SHAP) were combined to evaluate how the descriptors influence the “black box” model. All ML analyses were implemented using the scikit-learn library in Python.

4.3 Density-functional theory calculation

The Vienna Ab-initio Simulation Package (VASP) based density functional theory (DFT), has been employed to calculate ternary and quaternary alloys systems. The electron-ion interactions are described by projector augmented wave (PAW) method and the Perdew-Burke-Ernzerhof (PBE) functional in generalized gradient approximation (GGA) is used for the exchange-correlation contributions. Random solid solution structures are obtained by the special quasi-random structures (SQS) approach using the mcsqs code as implemented in the ATAT software. 128 (or 160) atoms special quasi-random BCC structures have been used for ternary and quaternary alloys. The electron wavefunction was expanded using a plane wave basis set with a kinetic energy cutoff of 500 eV. A 4×4×4 *k*-point mesh was used to ensure accuracy, and the energy convergence criterion was set to 10⁻⁶ eV. All the alloy structures are thoroughly optimized to obtain ground state structural parameters and cohesive energies. The formation enthalpy $\Delta H_{\text{mix}}^{\text{DFT}}$ of refractory multicomponent alloys can be calculated as below:

$$\Delta H_{\text{mix}}^{\text{DFT}} = E_{\text{alloy}} - \frac{1}{n} E_{\text{coh}}^{\text{bulk}}(\text{A}) - \frac{1}{n} E_{\text{coh}}^{\text{bulk}}(\text{B}) - \frac{1}{n} E_{\text{coh}}^{\text{bulk}}(\text{C}) - \frac{1}{n} E_{\text{coh}}^{\text{bulk}}(\text{D}) - \dots - \frac{1}{n} E_{\text{coh}}^{\text{bulk}}(\text{i}) \quad (6)$$

where E_{alloy} represents the total energy of the alloys, $E_{\text{coh}}^{\text{bulk}}(\text{i})$ represents the cohesion energy of element in ground state, and *n* is the number of elements in crystal lattice, respectively.

4.4 Experimental validation

The raw materials of the synthesized RHEAs were tungsten powder (99.95% purity), molybdenum powder (99.95% purity), vanadium powder (99.95% purity), niobium powder (99.95% purity), titanium powder (99.9% purity), tantalum powder (99.95% purity), hafnium powder (99.9% purity), zirconium powder (99.5% purity) and chromium powder (99.9% purity) with average particle sizes of 1 ~10 μm. The powders were mixed by planetary ball mill (mass ratio of the ball material: 10:1; mixed for 2 h). The mixture was then shaped using cold isostatic pressing. Subsequently, each pellet was melted using vacuum arc melting. The melting process was carried

out under an environment of high-purity argon, and mechanical and molecular pumps were used to achieve a vacuum degree of 3×10^{-3} Pa to prevent oxidation of the ingots. To guarantee chemical uniformity, each ingot was re-melted more than six times, flipping them over between each melt. Finally, the ingots were annealed at 1400 °C for 4 h in vacuum, to eliminate the dendrite structures of the alloys. X-ray diffraction with Cu-K α radiation (XRD, Model: DX-2700 from Dandong Haoyuan) was applied to examine the phase of RHEAs. A scanning electron microscope (SEM, Thermo Fisher Scientific Apreo 2S) equipped with energy dispersive spectrometer (EDS, Oxford ULTIM Max 100) was employed to examine the microstructures and phase compositions of the specimens.

Data availability

Data will be made available on request.

References

- [1] E.P. George, D. Raabe, R.O. Ritchie, High-entropy alloys, *Nat. Rev. Mater.* 4(8) (2019) 515-534.
- [2] O.N. Senkov, G.B. Wilks, D.B. Miracle, C.P. Chuang, P.K. Liaw, Refractory high-entropy alloys, *Intermetallics* 18(9) (2010) 1758-1765.
- [3] Q.F. He, J.G. Wang, H.A. Chen, Z.Y. Ding, Z.Q. Zhou, L.H. Xiong, J.H. Luan, J.M. Pelletier, J.C. Qiao, Q. Wang, L.L. Fan, Y. Ren, Q.S. Zeng, C.T. Liu, C.W. Pao, D.J. Srolovitz, Y. Yang, A highly distorted ultraelastic chemically complex Elinvar alloy, *Nature* 602(7896) (2022) 251-257.
- [4] O. El Atwani, H.T. Vo, M.A. Tunes, C. Lee, A. Alvarado, N. Krienke, J.D. Poplawsky, A.A. Kohnert, J. Gigax, W.Y. Chen, M. Li, Y.Q. Wang, J.S. Wróbel, D. Nguyen-Manh, J.K.S. Baldwin, O.U. Tukac, E. Aydogan, S. Fensin, E. Martinez, A quinary WTaCrVHf nanocrystalline refractory high-entropy alloy withholding extreme irradiation environments, *Nat. Commun.* 14(1) (2023) 2516.
- [5] J. Qi, X. Fan, D.I. Hoyos, M. Widom, P.K. Liaw, J. Poon, Integrated design of aluminum-enriched high-entropy refractory B2 alloys with synergy of high strength and ductility, *Sci. Adv.* 10(49) eadq0083.
- [6] S. Wei, S.J. Kim, J. Kang, Y. Zhang, Y. Zhang, T. Furuhashi, E.S. Park, C.C. Tasan, Natural-mixing guided design of refractory high-entropy alloys with as-cast tensile ductility, *Nat. Mater.* 19(11) (2020) 1175-1181.
- [7] J. Liu, B.-S. Li, H. Gardner, Y. Gong, F. Liu, G. He, M. Moorehead, C. Parkin, A. Couet, A.J. Wilkinson, D.E.J. Armstrong, Origin of age softening in the refractory high-entropy alloys, *Sci. Adv.* 9(49) eadj1511.
- [8] T. Tsuru, S. Han, S. Matsuura, Z. Chen, K. Kishida, I. Iobzenko, S.I. Rao, C. Woodward, E.P. George, H. Inui, Intrinsic factors responsible for brittle versus ductile nature of refractory high-entropy alloys, *Nat. Commun.* 15(1) (2024) 1706.

- [9] D.H. Cook, P. Kumar, M.I. Payne, C.H. Belcher, P. Borges, W. Wang, F. Walsh, Z. Li, A. Devaraj, M. Zhang, M. Asta, A.M. Minor, E.J. Lavernia, D. Apelian, R.O. Ritchie, Kink bands promote exceptional fracture resistance in a NbTaTiHf refractory medium-entropy alloy, *Science* 384(6692) (2024) 178-184.
- [10] J. Gou, G. Liu, T. Yang, X. Liu, Y. Pan, C. Liu, Y. Qian, Y. Liu, Y. Chen, X. Zhang, T. Ma, X. Ren, A high-entropy alloy showing gigapascal superelastic stress and nearly temperature-independent modulus, *Nat. Commun.* 16(1) (2025) 1227.
- [11] Y. Zhang, D. Wang, S. Wang, High-Entropy Alloys for Electrocatalysis: Design, Characterization, and Applications, *Small* 18(7) (2022) 2104339.
- [12] J. Guo, H. Wang, F. von Rohr, Z. Wang, S. Cai, Y. Zhou, K. Yang, A. Li, S. Jiang, Q. Wu, R.J. Cava, L. Sun, Robust zero resistance in a superconducting high-entropy alloy at pressures up to 190 GPa, *Proc. Natl. Acad. Sci. U. S. A* 114(50) (2017) 13144-13147.
- [13] L. Sun, R.J. Cava, High-entropy alloy superconductors: Status, opportunities, and challenges, *Phys. Rev. Mater.* 3(9) (2019) 090301.
- [14] C. Cheng, R. Feng, T. Lyu, Y. Zou, Accelerated Discovery of (TiZrHf)(NbTa)1– High-Entropy Alloys With Superior Thermal Stability and a New Crystallization Mechanism, *Adv. Mater.* 36(31) (2024) 2403632.
- [15] J. Wang, L. Zhu, M. Zhou, L. Kong, J. Li, Ti₃₅Zr₃₅Nb₁₀Hf_{20-x}Ta_x high-entropy alloys with exceptional performance as biomedical load-bearing materials, *J. Mater. Sci. Technol* 241 (2026) 298-310.
- [16] T. Zhang, H.-F. Zhao, Z.-J. Chen, Q. Yang, N. Gao, L. Li, N. Luo, J. Zheng, S.-D. Bao, J. Peng, X. Peng, X.-W. Liu, H.-B. Yu, High-entropy alloy enables multi-path electron synergism and lattice oxygen activation for enhanced oxygen evolution activity, *Nat. Commun.* 16(1) (2025) 3327.
- [17] K. Glazyrin, K. Spektor, M. Bykov, W. Dong, J.-H.Y. Yu, S. Yang, J.-S.L. Lee, S.V. Divinski, M. Hanfland, K.V. Yusenkov, High-Entropy Alloys and Their Affinity with Hydrogen: From Cantor to Platinum Group Elements Alloys, *Adv. Sci.* 11(31) (2024) 2401741.
- [18] L. Han, S. Zhu, Z. Rao, C. Scheu, D. Ponge, A. Ludwig, H. Zhang, O. Gutfleisch, H. Hahn, Z. Li, D. Raabe, Multifunctional high-entropy materials, *Nat. Rev. Mater.* 9(12) (2024) 846-865.
- [19] E. Halpren, X. Yao, Z.W. Chen, C.V. Singh, Machine learning assisted design of BCC high entropy alloys for room temperature hydrogen storage, *Acta Mater.* 270 (2024) 119841.
- [20] Z. An, A. Li, S. Mao, T. Yang, L. Zhu, R. Wang, Z. Wu, B. Zhang, R. Shao, C. Jiang, B. Cao, C. Shi, Y. Ren, C. Liu, H. Long, J. Zhang, W. Li, F. He, L. Sun, J. Zhao, L. Yang, X. Zhou, X. Wei, Y. Chen, Z. Lu, F. Ren, C.-T. Liu, Z. Zhang, X. Han, Negative mixing enthalpy solid solutions deliver high strength and ductility, *Nature* 625(7996) (2024) 697-702.
- [21] D. Xu, X. Wang, Y. Lu, Heterogeneous-Structured Refractory High-Entropy Alloys: A Review of State-of-the-Art Developments and Trends, *Adv. Funct. Mater.* 34(49) (2024) 2408941.

- [22] P. Sathiyamoorthi, H.S. Kim, High-entropy alloys with heterogeneous microstructure: Processing and mechanical properties, *Prog. Mater. Sci.* 123 (2022) 100709.
- [23] R.K. Nutor, Q. Cao, R. Wei, Q. Su, G. Du, X. Wang, F. Li, D. Zhang, J.-Z. Jiang, A dual-phase alloy with ultrahigh strength-ductility synergy over a wide temperature range, *Sci. Adv.* 7(34) (2021) eabi4404.
- [24] H. Luan, X. Zhang, H. Ding, F. Zhang, J.H. Luan, Z.B. Jiao, Y.-C. Yang, H. Bu, R. Wang, J. Gu, C. Shao, Q. Yu, Y. Shao, Q. Zeng, N. Chen, C.T. Liu, K.-F. Yao, High-entropy induced a glass-to-glass transition in a metallic glass, *Nat. Commun.* 13(1) (2022) 2183.
- [25] Y. Cao, M. Yang, Q. Du, F.-K. Chiang, Y. Zhang, S.-W. Chen, Y. Ke, H. Lou, F. Zhang, Y. Wu, H. Wang, S. Jiang, X. Zhang, Q. Zeng, X. Liu, Z. Lu, Continuous polyamorphic transition in high-entropy metallic glass, *Nat. Commun.* 15(1) (2024) 6702.
- [26] Z. An, S. Mao, T. Yang, C.T. Liu, B. Zhang, E. Ma, H. Zhou, Z. Zhang, L. Wang, X. Han, Spinodal-modulated solid solution delivers a strong and ductile refractory high-entropy alloy, *Mater. Horiz.* 8(3) (2021) 948-955.
- [27] Y. Yan, Z. Pei, M.C. Gao, S. Misture, K. Wang, Data-driven discovery of a formation prediction rule on high-entropy ceramics, *Acta Mater.* 253 (2023) 118955.
- [28] L. Tan, K. Ali, P.S. Ghosh, A. Arya, Y. Zhou, R. Smith, P. Goddard, D. Patel, H. Shahmir, A. Gandy, Design principles of low-activation high entropy alloys, *J. Alloys Compd.* 907 (2022) 164526.
- [29] X. Liu, J. Zhang, Z. Pei, Machine learning for high-entropy alloys: Progress, challenges and opportunities, *Prog. Mater. Sci.* 131 (2023) 101018.
- [30] D. Sobieraj, J.S. Wróbel, T. Rygier, K.J. Kurzydłowski, O. El Atwani, A. Devaraj, E. Martinez Saez, D. Nguyen-Manh, Chemical short-range order in derivative Cr–Ta–Ti–V–W high entropy alloys from the first-principles thermodynamic study, *PCCP* 22(41) (2020) 23929-23951.
- [31] W. Chen, A. Hilhorst, G. Bokas, S. Gorsse, P.J. Jacques, G. Hautier, A map of single-phase high-entropy alloys, *Nat. Commun.* 14(1) (2023) 2856.
- [32] Y. Zhang, Y.J. Zhou, J.P. Lin, G.L. Chen, P.K. Liaw, Solid-Solution Phase Formation Rules for Multi-component Alloys, *Adv. Eng. Mater.* 10(6) (2008) 534-538.
- [33] M.C. Tropsch, J.R. Morris, P.R.C. Kent, A.R. Lupini, G.M. Stocks, Criteria for Predicting the Formation of Single-Phase High-Entropy Alloys, *Phys. Rev. X* 5(1) (2015) 011041.
- [34] L. Zhichao, M. Dong, L. Xiongjun, Z. Lu, High-throughput and data-driven machine learning techniques for discovering high-entropy alloys, *Commun. Mater.* 5(1) (2024) 76.
- [35] A.K. Shargh, C.D. Stiles, J.A. El-Awady, Deep learning accelerated phase prediction of refractory multi-principal element alloys, *Acta Mater.* 283 (2025) 120558.
- [36] L. Wang, B. Ouyang, Phase Selection Rules of Multi-Principal Element Alloys, *Adv. Mater.* 36(16) (2024) 2307860.

- [37] M.G. Poletti, L. Battezzati, Electronic and thermodynamic criteria for the occurrence of high entropy alloys in metallic systems, *Acta Mater.* 75 (2014) 297-306.
- [38] W. Steurer, Single-phase high-entropy alloys – A critical update, *Mater. Charact.* 162 (2020) 110179.
- [39] M. Li, F. Lin, S. Zhang, R. Zhao, L. Tao, L. Li, J. Li, L. Zeng, M. Luo, S. Guo, High-entropy alloy electrocatalysts go to (sub-)nanoscale, *Sci. Adv.* 10(23) (2024) eadn2877.
- [40] R. Chen, G. Qin, H. Zheng, L. Wang, Y. Su, Y. Chiu, H. Ding, J. Guo, H. Fu, Composition design of high entropy alloys using the valence electron concentration to balance strength and ductility, *Acta Mater.* 144 (2018) 129-137.
- [41] A. Walsh, A.A. Sokol, J. Buckeridge, D.O. Scanlon, C.R.A. Catlow, Oxidation states and ionicity, *Nat. Mater.* 17(11) (2018) 958-964.
- [42] C.F. Melius, On the role of d electrons in chemisorption and catalysis on transition metal surfaces, *Chem. Phys. Lett.* 39(2) (1976) 287-290.
- [43] L. Gu, Y. Zhao, Y. Li, R. Hou, F. Liang, R. Zhang, Y. Wu, Y. Fan, N. Liang, B. Zhou, Y. Chen, G. Sha, G. Chen, Y. Wang, X. Chen, Ultrastrong and ductile medium-entropy alloys via hierarchical ordering, *Sci. Adv.* 10(22) (2024) eadn7553.
- [44] G. Leitner, G.E.R. Schulze, Crystal chemical stability conditions for intermetallic compounds (I) frequency distributions of laves-phases, *Krist. Tech.* 6(4) (1971) 449-463.
- [45] C.T. Liu, J.O. Stiegler, Ductile Ordered Intermetallic Alloys, *Science* 226(4675) (1984) 636-642.
- [46] U. Mizutani, H. Sato, T.B. Massalski, The original concepts of the Hume-Rothery rule extended to alloys and compounds whose bonding is metallic, ionic, or covalent, or a changing mixture of these, *Prog. Mater. Sci.* 120 (2021) 100719.
- [47] O. Johnson, Itinerant and Localized d-Electrons in Transition Metals from Positron Annihilation Data, *physica status solidi (b)* 99(2) (1980) 745-754.
- [48] V. Heine, s– d interaction in transition metals, *Physical Review* 153(3) (1967) 673.
- [49] G. Frenking, N. Fröhlich, The Nature of the Bonding in Transition-Metal Compounds, *Chem. Rev.* 100(2) (2000) 717-774.
- [50] U. Mizutani, Hume-Rothery rules for structurally complex alloy phases, *MRS Bulletin-Materials Research Society* 37(2) (2012) 169.
- [51] Y. Zhang, S. Guo, C.T. Liu, X. Yang, Phase Formation Rules, in: M.C. Gao, J.-W. Yeh, P.K. Liaw, Y. Zhang (Eds.), *High-Entropy Alloys: Fundamentals and Applications*, Springer International Publishing, Cham, 2016, pp. 21-49.
- [52] C. Acemi, B. Vela, E. Norris, W. Trehern, K.C. Atli, C. Cleek, R. Arróyave, I. Karaman, Multi-objective, multi-constraint high-throughput design, synthesis, and characterization of tungsten-containing refractory multi-principal element alloys, *Acta Mater.* 281 (2024) 120379.

- [53] R. Goodall, Data of the maximum solid solubility limits of binary systems of elements, Data Brief 26 (2019) 104515.
- [54] P.S. Matsumoto, Trends in Ionization Energy of Transition-Metal Elements, J. Chem. Educ. 82(11) (2005) 1660.
- [55] C. Hatler, I. Robin, H. Kim, N. Curtis, B. Sun, E. Aydogan, S. Fensin, A. Couet, E. Martinez, D.J. Thoma, O. El Atwani, The path towards plasma facing components: A review of state-of-the-art in W-based refractory high-entropy alloys, Curr. Opin. Solid State Mater. Sci. 34 (2025) 101201.
- [56] T. Hori, T. Nagase, M. Todai, A. Matsugaki, T. Nakano, Development of non-equiatomic Ti-Nb-Ta-Zr-Mo high-entropy alloys for metallic biomaterials, Scr. Mater. 172 (2019) 83-87.
- [57] L. Wu, J. Si, S. Guan, H. Zhang, J. Dou, J. Luo, J. Yang, H. Yu, J. Zhang, X. Ma, P. Yang, R. Zhou, M. Liu, F. Hong, X. Yu, Record-High T_c and Dome-Shaped Superconductivity in a Medium-Entropy Alloy TaNbHfZr under Pressure up to 160 GPa, Phys. Rev. Lett. 132(16) (2024) 166002.
- [58] A.B. Peters, D. Zhang, S. Chen, C. Ott, C. Oses, S. Curtarolo, I. McCue, T.M. Pollock, S. Eswarappa Prameela, Materials design for hypersonics, Nat. Commun. 15(1) (2024) 3328.
- [59] Periodic Table <http://periodictable.com>.
- [60] A. Takeuchi, A. Inoue, Classification of Bulk Metallic Glasses by Atomic Size Difference, Heat of Mixing and Period of Constituent Elements and Its Application to Characterization of the Main Alloying Element, Mater. Trans. 46 (2005) 2817-2829.
- [61] X.D. Xu, S. Guo, T.G. Nieh, C.T. Liu, A. Hirata, M.W. Chen, Effects of mixing enthalpy and cooling rate on phase formation of $Al_xCoCrCuFeNi$ high-entropy alloys, Materialia 6 (2019) 100292.

Acknowledgements

This work was supported by the Natural Science Foundation of Sichuan Province, China (Grant No. 2026NSFSC0006, 2025ZNSFSC0063) and the Fundamental Research Funds for the Central Universities (No.YJ202423). The authors also thank Dr. Xing Gao for providing valuable advice for DFT calculations.

Author contributions

Yonggang Yan, Dahuan Zhu and Xunxiang Hu designed and supervised the project. Yonggang Yan performed all the experiments. Yonggang Yan constructed the dataset. Yonggang Yan and Yalin Liao performed the machine learning studies. Nan Zhou performed the DFT calculations.

641 Yonggang Yan, Nan Zhou, Yalin Liao, Lijuan Cui, Dahuan Zhu and Xunxiang Hu authored the
642 manuscript.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterial.docx](#)