ARTICLE   OPEN

Composition design of high-entropy alloys with deep sets learning

Jie Zhang[1,3], Chen Cai[2,3], George Kim[1], Yusu Wang[2✉] and Wei Chen [1✉]

High entropy alloys (HEAs) are an important material class in the development of next-generation structural materials, but the astronomically large composition space cannot be efficiently explored by experiments or first-principles calculations. Machine learning (ML) methods might address this challenge, but ML of HEAs has been hindered by the scarcity of HEA property data. In this work, the EMTO-CPA method was used to generate a large HEA dataset (spanning a composition space of 14 elements) containing 7086 cubic HEA structures with structural properties, 1911 of which have the complete elastic tensor calculated. The elastic property dataset was used to train a ML model with the Deep Sets architecture. The Deep Sets model has better predictive performance and generalizability compared to other ML models. Association rule mining was applied to the model predictions to describe the compositional dependence of HEA elastic properties and to demonstrate the potential for data-driven alloy design.

INTRODUCTION

High-entropy alloys (HEA) are a new class of multi-principal element materials with diverse and fascinating structure-property relationships[1]. The term "high entropy" was coined based on the idea that a single solid-solution phase can be stabilized with a high configurational entropy associated with the random mixing of multiple elements at similar atomic fractions[2]. Numerous studies have revealed a wide variety of structural and functional properties[2–7] in this large space of materials, ranging from cryogenic ductility[8], high strength[9,10], corrosion resistance[11,12], to excellent wear behavior[13], and thermoelectric properties[14]. High-entropy materials represent a fast-growing field in materials research, covering both alloys and ceramics such as high-entropy nitrides, carbides, and oxides[15,16], and have many important applications such as protective coatings and energy storage[17].

Designing a HEA often involves painstaking experimental and computational studies using a trial-and-error approach to explore the astronomical composition space. The experimental approach involves expensive and time-consuming processes of synthesis, characterization, and analysis, which often studies only a few candidate compositions at a time. Even with the explosive growth of computing power, high-throughput density functional theory (DFT) calculations are still incapable of computing the complete multi-element alloy space. Empirical models that predict the general trends in the multi-principal element space are instrumental in accelerating the exploration of the HEA composition space. For example, Senkov et al. performed high-throughput CALPHAD calculations and concluded that more elements do not necessarily stabilize solid-solution phases in HEAs[18]. Lederer et al. proposed the "LVTC" model that can accurately predict the transition temperatures of solid solution HEAs[19]. By employing the state-of-the-art machine learning (ML) algorithms, it is possible to explore the high-dimensional composition space much more efficiently[20]. However, the application of ML on HEA studies is often hindered by the scarcity of HEA property data, especially quality experimental data. Two recently compiled experimental HEA datasets have the phase composition for 401 HEAs[4] and

mechanical properties for 630 HEAs[21]. These datasets have enabled the development of predictive models on the phase selection rules by ML methods such as artificial neural network[22] and Gaussian process classification[23]. Nonetheless, ML models trained with small datasets usually do not generalize well. In many cases, researchers have to limit the scope of their ML models to specific alloy systems[24,25].

The goal of this study is to integrate high-throughput first-principles calculations and the Deep Sets architecture to understand the effects of elemental combinations on the HEA properties over a broad composition space. We choose the elastic properties of HEAs as a case study. Elasticity describes the resistance for deformation between atoms before yielding, providing a critical starting point to study the mechanical properties of HEAs[26]. For instance, the ductility of an alloy can be estimated using the Pugh's ratio which can be tailored by doping HEAs[27]. A perfect elastic isotropy has also been achieved in HEAs by composition design with the aim of controlling the deformation behavior[28,29]. First-principles DFT calculations are predictive methods that can reliably compute the elastic constants of the HEAs[28,30–32]. DFT calculations with coherent potential approximation (CPA) and supercell methods usually give similar results on the elastic moduli of disordered HEA sytems[26]. Our recent study on the elasticity of $Al_{0.3}CoCrFeNi$ HEA also proved that first-principles and ML methods can give accurate predictions of HEA elasticity that are comparable to neutron diffraction measurements. Owning to the rapid development of high-throughput first-principles calculations[33], accurate elasticity data for ordered inorganic structures are readily available for data-driven materials research[34,35]. However, similar datasets are not available for HEAs yet. Many studies still rely on estimations with the Vegard's law to predict the HEA elastic modulus as a compositionally weighted average of elemental properties[36,37].

In this study, we quantitively map the elastic properties of a 14-element HEA space with the integration of high-throughput first-principles calculations, deep learning, and association-rule analysis (Fig. 1). We generate a dataset of 3579 quaternary HEA

[1]Department of Mechanical, Materials, and Aerospace Engineering, Illinois Institute of Technology, Chicago, IL 60616, USA. [2]Halıcıoğlu Data Science Institute, University of California San Diego, La Jolla, CA 92093, USA. [3]These authors contributed equally: Jie Zhang, Chen Cai. ✉email: yusuwang@ucsd.edu; wchen66@iit.edu
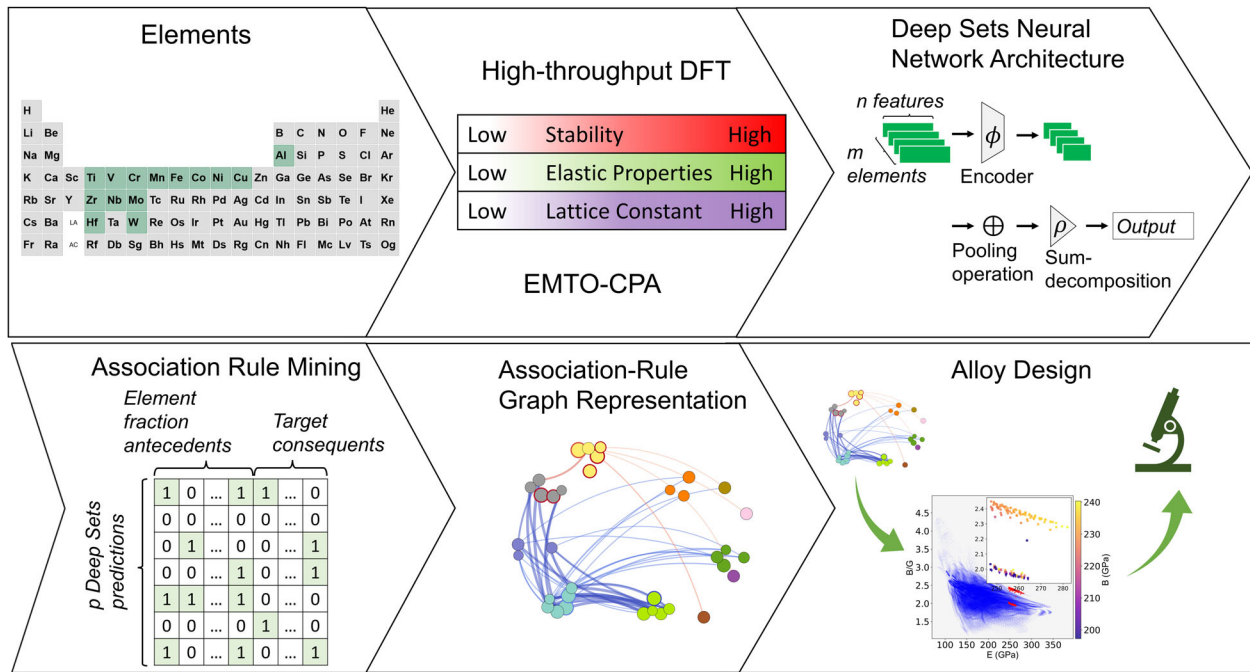
Fig. 1 Data-driven workflow to map the elastic properties of the high-entropy alloy space. A dataset of the elastic properties of quaternary HEA compositions, containing elements from the set of the 14 highlighted elements, is created using high-throughput EMTO-CPA calculations. After training Deep Sets models on this dataset, association rule mining is used on Deep Sets model predictions to discover trends between elemental combinations and elastic properties. These trends are in the form of association rules which are then visualized using network graphs. Deep Sets neural network architecture explanation: Each element feature vector in the input set is transformed by the same mapping function, ϕ. The resulting vectors are summed in a pooling operation which ensures permutation invariance. Finally, the result of the pooling operation is passed to a MLP (multi-layer perceptron), ρ, which maps the input to the prediction values.

compositions from high-throughput DFT calculations with the exact muffin-tin orbitals and coherent potential approximation (EMTO-CPA) method[38,39]. Recent progress in graph representation learning[40–44] for link prediction[45], graph classification[46], physics simulation[47], and combinatorial optimization[48,49], has produced a growing body of literature applying graph neural networks on property prediction for materials[50,51] and molecules[52,53]. One notable challenge when applying graph representation learning to HEAs is that they usually contain simple geometrical lattice but randomness in the elemental site occupancy. Representing HEAs as neighborhood graphs[50] is inefficient because the final representation will be the specific instantiation of the underlying random configurations. Conventional ML architectures use either engineered elemental properties (e.g., different types of means) or compound properties as features. Directly using elemental properties as features introduces permutation variance in the model making predictions dependent on the specific order of the elements in the feature vector. The feature engineering of elemental properties can also be tedious and inefficient. If there were data for a very large number of HEA configurations, a neural network would theoretically be able to capture the permutation invariance of elements. However, the amount of available material property data is often insufficient in reality.

To overcome this problem, we represent HEAs as *sets* of elements and employ the *Deep Sets*[54] architecture for predicting elastic properties. Deep Sets is a recently developed deep learning architecture that can represent any invariant function over a set. Compared with other ML models, our Deep Sets models show superior predictive performance in the broad HEA space. We further perform association-rule analysis to understand the trends of elemental effects on the elastic properties of HEAs and leverage these insights on the composition design of HEAs with targeted properties. Our study showcases an efficient, accurate, and generalizable approach to study multi-element materials systems.

RESULTS

Validations of HEA property predictions from EMTO-CPA calculations

In total, the equation of state (EOS) and bulk modulus (B) of 7086 cubic quaternary HEA phases were successfully calculated, corresponding to 3579 compositions, 75 of which contained Ta. Due to the lack of enough Ta containing compositions being successfully calculated, Ta containing compositions were not used in the training of the Deep Sets model. Of the 996 equimolar compositions completed, 962 were more stable in the body-centered cubic (BCC) phase. The elastic properties of all 996 compositions were calculated using the stable phase for each composition. It should be noted that there are 1001 possible quaternary equimolar compositions given 14 elements, but the following compositions AlCuMnNi, CoCuNiZr, AlCrMoV, CoCrNbNi, and CoCuFeHf are missing from the dataset due to convergence issues. Of the 2508 non-equimolar compositions that were completed, 2331 compositions were more stable in the BCC phase. The cubic elastic constants were calculated for 840 non-equimolar compositions whose formation energy is lower than 0.15 eV/atom. To our knowledge, this is the largest dataset of HEA structures with calculated stability and elastic property information with 7086 structures and over 3579 compositions. Additionally, a set of 264 structures (132 compositions) were separately calculated for validation. The elastic property, crystal structure information, and total energy of each structure is organized into JSON files using key-value pairs; Supplementary Table 1 shows the labels used as keys, and a description of the values.

The high-throughput EMTO-CPA results were validated by comparing with reported experimental and computational results in literature. As shown in Supplementary Table 2, the EMTO-CPA predicted preferred cubic-phase type and lattice parameters agree well with reported experimental results[55–64]. EMTO-CPA gives the correct phase for all HEA systems and a small mean absolute error
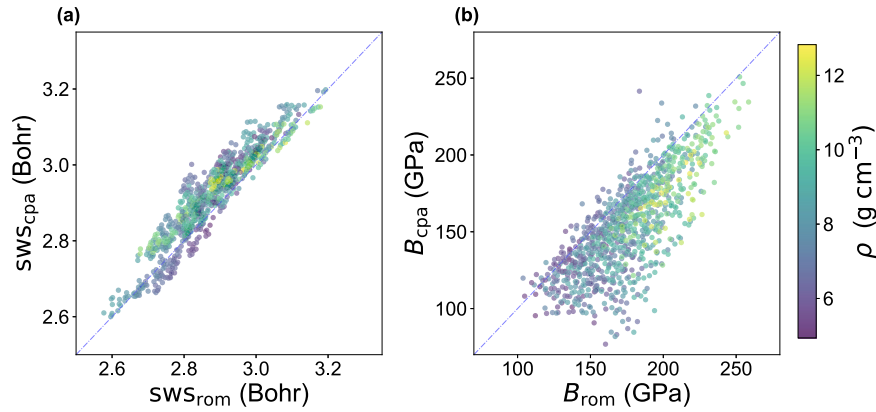
Fig. 2 Comparison of the EMTO-CPA predictions of HEA properties with the rule-of-mixture (ROM) estimates. a It compares the Wigner–Seitz radius, and **b** compares the bulk modulus. Density is described using the color bar.

Table 1.	Property prediction MAE for equimolar quaternary HEAs.				
Method	B (GPa)	sws (Bohr)	C_{11} (GPa)	C_{12} (GPa)	C_{44} (GPa)
Deep Sets	**4.596 ± 0.639**	**0.004 ± 0.001**	**5.199 ± 0.760**	**4.230 ± 0.412**	3.162 ± 0.188
GBT	6.033 ± 0.349	0.005 ± 0.001	8.959 ± 0.682	5.370 ± 0.391	3.162 ± 0.126
KNN	14.953 ± 0.620	0.044 ± 0.002	22.325 ± 1.594	10.999 ± 0.689	7.098 ± 0.497
LR	7.891 ± 0.461	0.007 ± 0.001	15.844 ± 0.955	7.464 ± 0.401	5.634 ± 0.285
RF	7.488 ± 0.574	0.008 ± 0.001	11.369 ± 0.925	6.447 ± 0.347	3.391 ± 0.154
SVM	5.829 ± 0.536	0.009 ± 0.001	8.675 ± 0.807	5.191 ± 0.467	**2.923 ± 0.243**
Bold values represent the best prediction of each property.					

(MAE) of 1.1% for the lattice parameters. In general, the elastic moduli predicted from DFT calculations are comparable to experiments, with an error smaller than 15% for most inorganic phases[34]. For random alloys, EMTO-CPA calculations do not consider local lattice distortions, which can slightly overestimate the elastic moduli than results from the supercell method, but the quantitative trends in elastic moduli are similar[31]. Supplementary Table 3, Supplementary Table 4, and Supplementary Fig. 1 compare EMTO-CPA predicted HEA elastic properties from our high-throughput calculations with literature[27,29,65–71]. Good agreements are found between these EMTO-CPA results. For elastic constants C_{11} and C_{12}, the MAEs are about 5%. C_{44} shows a larger MAE of about 10%. The discrepancies can be attributed to different exchange-correlation functionals and numerical uncertainties. The MAEs for all polycrystalline elastic moduli are about 5%. For Poisson's ratio and Pugh's ratio, the MAEs are 1.8% and 4.0%, respectively. Owning to difficulties in measuring the elastic properties of HEAs, published experimental HEA elasticity data are scarce. Our results indicate first-principles tools can be an efficient tool to generate reliable fundamental HEA data over a large composition space.

In literature, Vegard's law, or the rule of mixture (ROM), is a popular method to estimate the elastic moduli of HEAs. Our EMTO-CPA dataset allows a quantitative assessment of the accuracy of ROM for HEAs. A ROM estimation of elastic moduli can be made using Eq. (1)[37] or Eq. (2)[72], corresponding to the upper and lower limit of the estimation.

$$M = \frac{\sum c_i V_i M_i}{\sum V_i M_i} \tag{1}$$

$$M^{-1} = \frac{\sum c_i V_i M_i^{-1}}{\sum V_i M_i} \tag{2}$$

where c_i is the molar fraction, V_i is the molar volume, and M_i is the elastic moduli of ith element.

We used EMTO-CPA to calculate the equilibrium volume and bulk modulus of pure elements, which were then used to make ROM estimations of the Wigner–Seitz radius (sws) and bulk moduli of HEAs. Compared with other ROM estimations, we found the average value of Eqs. (1) and (2) gives slightly better agreement with the EMTO-CPA values, especially when Al or Ti are in the HEA. Figure 2 compares sws and B from the ROM estimations for quaternary equimolar HEAs with the EMTO-CPA predictions. The ROM estimation does not consider chemical interactions between different species, so a certain degree of discrepancy in B is expected, as is in the ROM estimated sws. However, the discrepancies in the ROM estimation of B are still quite significant. For many systems, the ROM estimates are significantly higher than the EMTO-CPA predicted B. Due to the disregard of local atomic relaxations, the EMTO-CPA elastic moduli are already overestimated. When taking this factor into account, the ROM estimation can overestimate true HEA elastic modulus by a large fraction. While it is convenient to use ROM to estimate the elastic properties[36,37], our results show reliable predictions should be made with DFT calculations or data-driven predictive modeling.

Performance of deep sets prediction of HEA elastic properties

Tables 1 and 2 compare the performance of HEA property predictions between Deep Sets and other ML models. Regardless of the model chosen, the prediction MAE for non-equimolar quaternary HEAs is slightly larger than equimolar systems, indicating that predicting properties for non-equimolar systems is more challenging. For all properties, simple models such as k-nearest neighbor (KNN) and linear regression (LR) generally perform much worse than random forest (RF), support vector machine (SVM), gradient boosting tree (GBT), and Deep Sets models. Among the ML models, the Deep Sets model

Table 2. Property prediction MAE for non-equimolar quaternary HEAs.

Method	B (GPa)	sws (Bohr)	C_{11} (GPa)	C_{12} (GPa)	C_{44} (GPa)
Deep Sets	**6.025 ± 0.415**	**0.004 ± 0.001**	**11.613 ± 0.845**	8.398 ± 0.899	**2.447 ± 0.157**
GBT	8.622 ± 0.485	0.009 ± 0.001	15.402 ± 1.124	8.662 ± 0.857	4.183 ± 0.277
KNN	12.508 ± 0.424	0.037 ± 0.001	21.492 ± 1.445	10.466 ± 0.521	6.579 ± 0.577
LR	11.179 ± 0.414	0.027 ± 0.001	28.069 ± 7.561	14.151 ± 10.0	10.232 ± 3.072
RF	9.607 ± 0.531	0.015 ± 0.001	16.876 ± 1.536	8.93 ± 0.810	5.218 ± 0.476
SVM	7.636 ± 0.506	0.011 ± 0.001	12.998 ± 1.126	**7.571 ± 0.537**	3.538 ± 0.213

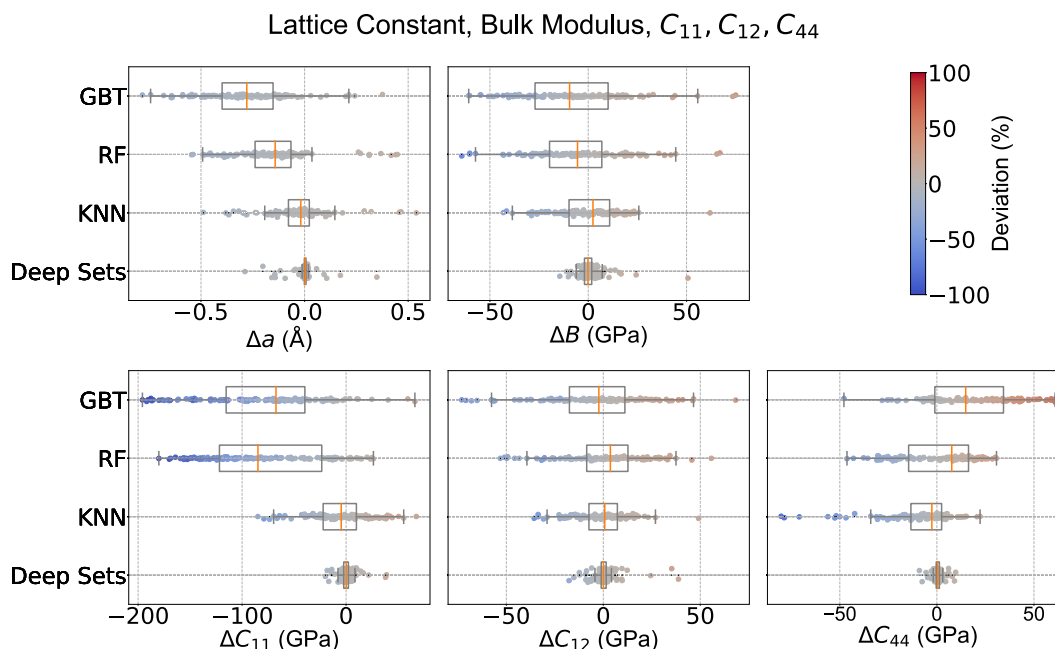Bold values represent the best prediction of each property.

Fig. 3   Boxplots comparing the accuracy of four models (GBT, RF, KNN, and Deep Sets) predicting (from top to bottom, left to right) lattice constant, bulk modulus, and elastic constants C_{11}, C_{12}, and C_{44}. The orange line in the boxplot represents the median value; the lower and upper limits of the box represent the 25th and 75th percentiles respectively, and the whiskers extend to 1.5 times the interquartile range. The color of the scatterplot points corresponds to the colorbar which represents the percentage deviation of the predicted value from the EMTO-CPA calculated value.

overwhelmingly outperforms the other methods, achieving better results in almost all predicted properties. For the elastic constant predictions of equimolar compositions, the Deep Sets method achieves the best result for C_{11} and C_{12}, whereas for C_{44} it is slightly worse than the SVM model, although the difference is small considering the standard error. The difference between the directly predicted B and the derived B from predicted C_{11} and C_{12} was also compared. The highest deviation between these values is 8%. This deviation is acceptable considering the uncertainty of DFT elastic constants, which can deviate from experimental values by ±15%[34]. For the same reason, the predicted polycrystalline shear modulus (G), Young's modulus (E), Poisson's ratio (v), Pugh's ratio (B/G), and the Zener ratio (A_z) were all derived from predicted cubic elastic constants via mathematical operation for further analysis. It is difficult to directly compare the performance of published ML models on elastic properties as they are often trained on different datasets. For reference, the AFLOW-ML and the JARVIS-ML model has a MAE of 8.68 GPa and 10.5 GPa for predicting the bulk modulus of inorganic structures, respectively[73,74]. To further validate the performance of our Deep Sets model, we selected 132 additional quaternary non-equimolar HEA compositions and performed EMTO-CPA calculations. The

compositions in this validation group were randomly selected and were different from the compositions in the training set. The Deep Sets model without the stable lattice-type feature was used for this validation set. Figure 3 compares the accuracy of different models, showing the Deep Sets model generalizes much better than the other methods. The reason that the MAE for the unseen HEAs is even better than the test MAE can be attributable to the procedure of averaging the prediction of 10 models that are trained to 10 subsets (60%) of the original data. This procedure is effectively equivalent to the ensemble method. The LR and SVM perform poorly on the unseen HEAs so their MAE is not included. The Deep Sets model predicts all compositions in the training set and validation group to be mechanical stable, i.e., $C_{11} > 0$, $C_{44} > 0$, $C_{11} - C_{12} > 0$, and $C_{11} + 2C_{12} > 0$, in agreement with the EMTO-CPA predictions.

Lastly, we tested the generalization performance of our Deep Sets model on a widely studied HEA system $Al_x CoCrFeNi$. Table 3 shows a comparison between the Deep Sets predictions and the EMTO-CPA calculations reported in the literature[65]. The result shows a very encouraging agreement especially considering that the Deep Sets model was not trained by any quinary HEA composition.

Table 3. Predicted Wigner–Seitz radius (sws) and bulk modulus (B) of Al$_x$CoCrFeNi from Deep Sets and comparisons with results from EMTO-CPA calculations[76] for the preferred phase.

x	sws$_{Deep Sets}$ (Bohr)	sws$_{CPA}$ (Bohr)	$B_{Deep Sets}$ (GPa)	B_{CPA} (GPa)	Phase
0.1	2.620	2.611	199	200	FCC
0.25	2.628	2.619	195	197	FCC
0.3	2.631	2.622	194	196	FCC
0.375	2.635	2.626	192	194	FCC
0.5	2.641	2.632	189	190	FCC
1.25	2.680	2.667	166	171	BCC
1.3	2.678	2.670	166	170	BCC
1.5	2.689	2.675	161	167	BCC
2.0	2.706	2.690	151	159	BCC
2.5	2.721	2.701	145	153	BCC

Elemental effects on HEA elasticity uncovered from association rule mining

The rules generated by association rule mining (ARM) are visualized in graph representations in Fig. 4. The nodes represent element types. The size of the node represents element fractions: The larger the node, the larger the fraction is represented. When a pair of nodes are connected by lines, referred to as edges, the edge colors and widths represent the consequents and lift values of the association rules in which the pair of nodes makes up the antecedent. When the antecedent of an association rule is a single element, the elastic property consequent and lift value of the rule is represented by the color and width of the node outline. The redder (bluer) the color of the edge or node outline, the higher (lower) the value of the elastic property consequent. The node outlines and connections for Fig. 4(f) are mapped to a different colorbar to emphasize rules that predict Zener ratios close to 1.0 for the isotropic case. Zener ratios lower than 1.0 are blue, those close to 1.0 are green, and those that are greater are mapped to yellow, orange, and red.

From Fig. 4(a) it is observed that most element pairs that decrease (increase) B involve elements with low (high) elemental B. For reference, the elements ranked from the lowest to the highest B (obtained for BCC using EMTO-CPA) are Al, Zr, Ti, Fe, Hf, Cu, Nb, Co, V, Ni, Mn, Mo, Cr, and W. However, it is observed that while elemental Mn has a relatively high B, Mn can lower B in an HEA on its own as well as when combined with Zr or Hf. Cr also has high B, but it decreases B when paired with Zr or Hf. These trends show that expectations based on ROM are sometimes contradicted and depends on the specific combination. B is related to the bonding strength among atoms. W may be expected to have an attractive force on the other elements due to its high electronegativity, thus increasing B when paired with the $3d$ transition metals. The decrease in B that accompanies the (Mn, Zr), (Mn, Hf), (Cr, Zr), and (Cr, Hf) combinations may be due to the low B of Zr and Hf, and the half-filled $3d$ orbitals of Mn and Cr. A composition containing such a combination would have an element with low B (Hf or Zr) as well as an element (Cr or Mn) that is less likely to be attracted to other constituent elements. In the case of Cr, high B can still be achieved when it is paired with an element with very high electronegativity such as W.

Figure 4(b) shows that Young's modulus, E, is decreased when Zr is combined with almost any other element. Even a small fraction of W results in lowered Young's modulus with Zr. Only when the composition contains Cr or a larger fraction of W is an increase in E predicted with Zr. E is related to bulk and shear moduli assuming a cubic symmetry and elastic isotropy (Eq. 14). W and Cr are expected to increase both B and G, so an increase in E due to W and Cr alloying is expected. An increase in E may be

attributed to increased B and/or G which is in turn due to mechanisms such as an increased interatomic attraction, or an increase in the covalent characteristic of the bonding that results in stiffer bonds. Some differences between the figures for B and E are that a combination of Zr with Ni, V, or W can decrease E, but not B, while a combination of Zr and Cr can decrease B, but not E. These trends demonstrate the potential of tuning elastic properties by predictive models such as the Deep Sets model and descriptive models such as ARM.

Figure 4(c) shows that Zr combined with Cu, Al, Hf, or Ni, and Hf combined with Cu, Al, or Ni lower G. On the other hand, Cr or W alone increases G. Notably, combinations involving Mn which are present in the plots for E and B are missing in the plot for G. This may be attributed to the half-filled $3d$ orbitals of Mn diminishing the effect Mn has on the characteristic nature of interatomic bonding, which in turn means that Mn does not affect G as much in the systems we considered.

Figure 4(d) represents the association rules for Pugh's ratio B/G. Most notably, combinations of Cu and Ni increase B/G. The Pugh's condition predicts that when the ratio B/G is greater than 1.75, the material will be ductile. A related criterion known as Pettifor's criterion states that when Cauchy pressure $C_{12}-C_{44}$ (Eq. 3) is positive, the bonding characteristic is predicted to be metallic, and the material is predicted to be intrinsically ductile. Otherwise, a negative Cauchy pressure suggests the bonding characteristic is predicted to be covalent and the material is predicted to be intrinsically brittle[75].

$$C_{12} - C_{44} = B\left(1 - \frac{5G_v}{3B}\right) \qquad (3)$$

where G_v is the Voight averaged shear modulus, and B is bulk modulus[75]. The elements involved in combinations that increase B/G such as Cu, Ni, Co are late transition metals with greater valence electron concentration (VEC). Barring strong interatomic interactions that cause directional bonding with angular characteristics, the increased electron density due to the increased VEC might be expected to increase the metallic character of interatomic bonding.

Poisson's ratio which is represented in Fig. 4(e) has mostly similar trends as Pugh's ratio, with a few differences. This observation is expected since Poisson's ratio can be expressed in terms of Pugh's ratio, $v = \frac{\frac{3B}{G}-2}{\frac{6B}{G}+2}$. Therefore, the same physics that underlies the B/G trends might be applicable to Poisson's ratio. Figure 4(f) shows that combinations of V, W, and Cr are predicted to produce near isotropic Zener ratios. Senkov and Miracle derived a modified expression for the Pugh's condition to take elastic
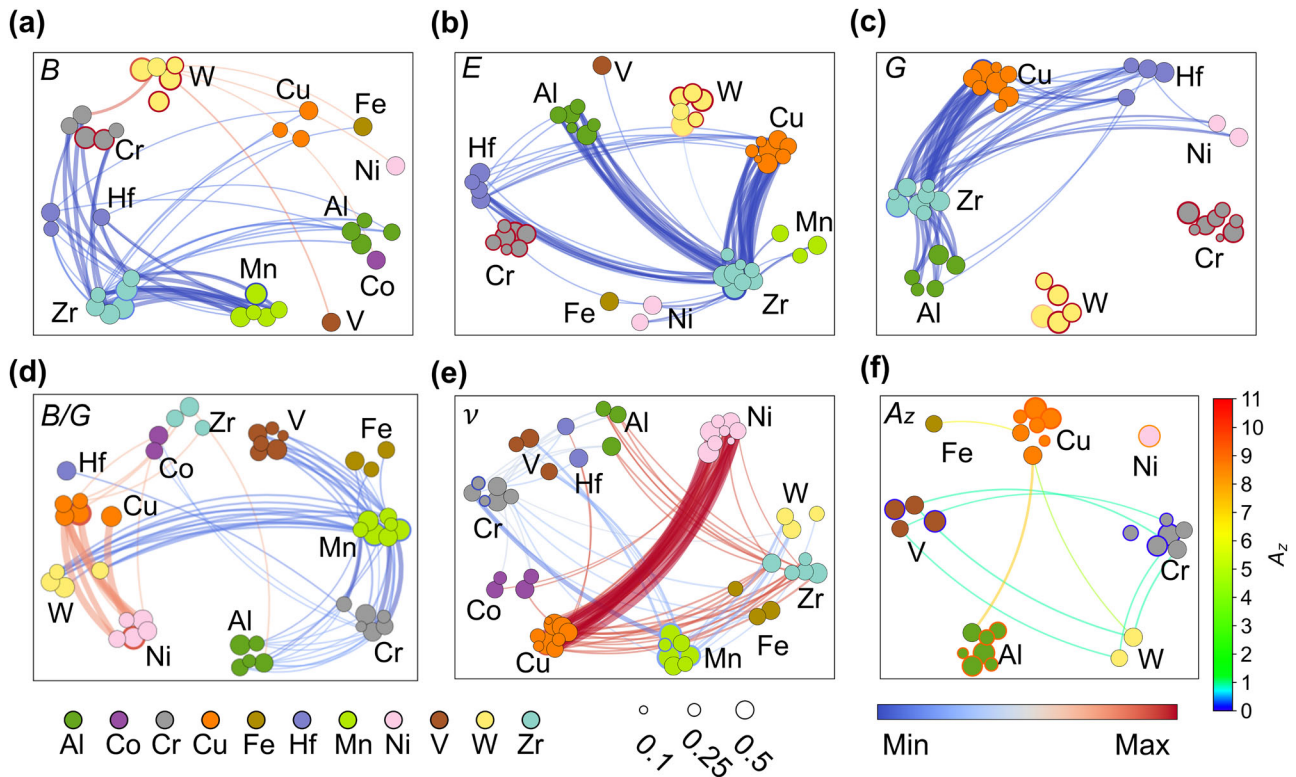
Fig. 4 Graph representations of association rules between elements and elastic properties of HEAs. Results for (**a**) bulk modulus, (**b**) Young's modulus, (**c**) shear modulus, (**d**) Pugh's ratio, (**e**) Poisson' ratio, and (**f**) Zener ratio respectively. Node colors and sizes represent different elements (as shown in the legend) and fractions. The redder (bluer) the color of the node outlines and connections, the higher (lower) the value of the elastic property is predicted to be. The thicker the node outline or connection, the higher the lift value of the rule is. The node outlines and connections for the Zener ratio are mapped to a separate color bar to emphasize rules that predict Zener ratios that are close to 1.0 for the isotropic case.

anisotropy into account (Eq. 4)[75]:

$$\frac{B}{G} < \frac{5(2A_z + 3)(2 + 3A_z)}{3(3A_z^2 + 19A_z + 3)} \qquad (4)$$

Interestingly, from Eq. (4), increasing the elastic anisotropy also increases the Pugh's ratio threshold for ductile materials which means that the more elastically anisotropic a material is, the greater the Pugh's ratio is needed for the material to be ductile. This observation was confirmed with an analysis of 308 intermetallic compounds and 24 metals[75]. Relating this result to the rules in Fig. 4(f), compositions containing Al, and Cu will need larger Pugh's ratios to satisfy the modified Pugh's condition, but as seen in Fig. 4(d) Cu does increase the Pugh's ratio, as does Al combined with Zr.

DISCUSSION

Discovery of HEA compositions with tailored elasticity

Predictions from the Deep Sets model can be used to screen the HEA space in search of compositions with desired elastic properties. In combination with the association rules, one can further explain the predicted properties in the context of the general property trends in the whole composition space. However, we note that the predictive power of any ML model is limited by the training data. For example, we only consider single cubic phase solid solutions in this study. Additional thermodynamic modeling or experimental studies are necessary to determine the phase selection of the compositions.

Ductile HEAs with high Young's moduli

One of the most desirable property combinations of structural HEAs is high Young's modulus (or bulk modulus), high-temperature softening resistance, good ductility, and low weight density. An example of a refractory HEA with low-density, high-ductility, and temperature resistance reported in literature is $Nb_{40}Ti_{25}Al_{15}V_{10}Ta_5Hf_3W_2$[76]. However, this HEA is a two-phase alloy with a BCC matrix and B2 nanoprecipitates. With our materials screening it may be possible to discover a single-phase solid solution with comparable properties. When a system's Pugh's ratio B/G is greater than 1.75, the system can be considered ductile. Figure 5 visualizes the trends of B/G vs. E for the 14-element quaternary HEA space. Detailed screening results are in Supplementary Table 5. It is obvious the B/G generally decreases with E, indicating high E HEAs usually do not have good ductility. The inset of Fig. 5 highlights a region where a balance of high B/G and E can be achieved. In particular, the CrMnTiV system can have high Young's moduli and good ductility. The pale blue edges for (Cr, Mn) and (Mn, V) pairs in Fig. 4(d) indicate these element pairs tend to produce intermediate to low Pugh's ratios, which is consistent with the screening. The Pugh's ratio of CrMnTiV also has a similar value of 2.530 from DFT calculations using a 64-atom special quasirandom structure. There is no publication reporting the experimental synthesis and elastic constants of the CrMnTiV HEA. Senkov et al reported that CALPHAD modeling predicts the equimolar system CrMnTiV to be a single BCC phase[18], but Yoav et al indicated the phase composition as inconclusive using their "LTVC" approach[19].
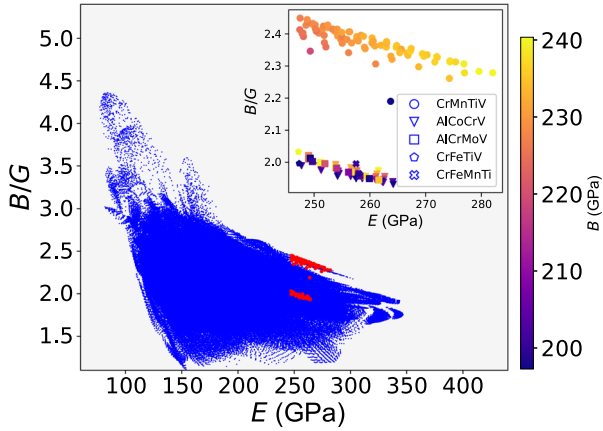
Fig. 5 Distribution of Young's modulus E vs. Pugh's ratio B/G of the 14-element HEA composition space. The inset shows the highlighted HEA systems with targeted elastic properties, where the B values are represented using the colorbar.

Fig. 6 Distribution of Young's modulus E vs. Zener ratio A_z of the 14-element HEA composition space. The inset shows the highlighted HEA systems with targeted elastic properties, where density ρ are represented using the colorbar.

Elastically isotropic HEAs with low Young's moduli

Another group of interesting HEAs are elastically isotropic systems with low Young's moduli. These HEAs can be potentially used in biomedical application[77]. The inset of Fig. 6 shows a composition window that satisfies these conditions. Detailed screening results are in Supplementary Table 6. The (Cr, V) pair identified as related to elastically isotropic systems from ARM in Fig. 4(f) also appears in HEAs within the composition window. From the screening, we notice that the densities of most HEAs in the composition window are relatively low, which can be another advantage for biomedical applications. The EMTO-CPA method usually overestimates the elastic moduli, suggesting that those selected compositions may even have lower E in real world applications.

In summary, we present an efficient, generalizable, and accurate Deep Sets model that can predict the energetic, structural, and elastic properties of HEA compositions. To our knowledge, the Deep Sets model was trained on the largest dataset of first-principles HEA elastic properties. The present work also analyzed the elastic properties predicted by our Deep Sets model using ARM to demonstrate correlations between compositional trends and properties that may assist the ultimate goal of alloy design. Elasticity is the underlying drive for mechanical responses, and the huge composition space that HEAs span gives rise to the potential for tuning elastic properties for targeted applications. Effective alloy design must also be efficient, and our methodology represents a step forward in improving our ability to design HEAs.

METHODS

High-throughput DFT calculations

High-throughput DFT calculations were performed using the EMTO-CPA method to predict the stability and elastic constants of selected HEA compositions. Fourteen elements (Al, Co, Cr, Cu, Fe, Hf, Mn, Mo, Nb, Ni, Ti, V, and W) were selected to create a large composition space of elements commonly found in 3d transition metal and refractory HEAs. All possible quaternary equimolar compositions of the 14 elements, and more than 2000 sampled quaternary non-equimolar compositions of these elements were calculated. Non-equimolar compositions that are close to the center of the composition space, i.e., $A_1B_xC_yD_z$ ($0.6 \le x, y, z \le 1$), were sampled more frequently. For each composition, we considered the face-centered cubic (FCC) and BCC random solid-solution phases for calculations. While realistic HEAs can crystalize as multi-phase alloys and with different lattices, the focus of the study is to understand the trends of elemental combinations on the properties of single-phase random alloys. The high-throughput workflow can be extended with additional DFT calculations and CALPHAD modeling to treat structurally complex HEAs[78].
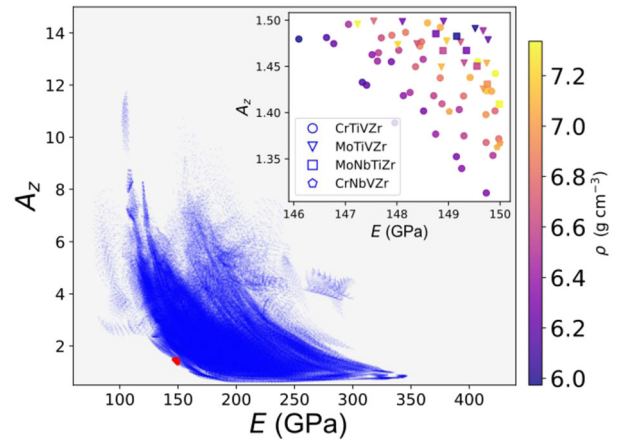
Two sets of EMTO-CPA calculations were performed for each HEA composition. First, the EOS was calculated to determine the relative phase stability and the lattice parameters of the HEA. At least fifteen volumes were dynamically generated for both lattices, and the total energy was calculated at each volume[35]. The energy, E, vs. volume, V, pairs were used to fit the Birch-Murnaghan EOS[79] in Eq. (5), where the subscript 0 represents the equilibrium condition and B_0' is the derivative of the bulk modulus with respect to pressure. The bulk modulus B_0 was obtained from the EOS.

$$E(V) = E_0 + \frac{9V_0B_0}{16}\left\{\left[\left(\frac{V_0}{V}\right)^{\frac{2}{3}}-1\right]^3 B_0' + \left[\left(\frac{V_0}{V}\right)^{\frac{2}{3}}-1\right]^2 \left[6-4\left(\frac{V_0}{V}\right)^{\frac{2}{3}}\right]\right\}$$

(5)

Next, EMTO-CPA calculations were performed to predict the elastic constants for the cubic phase with lower energy at the equilibrium volume. For non-equimolar compositions, the complete cubic elastic constant calculations were only calculated for relatively stable compositions with formation energy < 0.15 eV/atom. The elastic constants were found by fitting the energy changes with respect to externally applied deformations to the lattice. The energy change with the orthorhombic deformation (Eq. 6) gives the tetragonal shear modulus c' (Eq. 7). Elastic constants C_{11} and C_{12} can be derived using Eqs. (8) and (9). The energy change with the monoclinic deformation (Eq. 10) gives the elastic constant C_{44} (Eq. 11). The deformation to the equilibrium cell was applied at three steps of δ_o (or δ_m) = 0.00, 0.03, 0.05.

$$\mathcal{D}_t + \mathcal{J} = \begin{pmatrix} 1+\delta_o & 0 & 0 \\ 0 & 1-\delta_o & 0 \\ 0 & 0 & \frac{1}{1-\delta_o^2} \end{pmatrix}$$

(6)

$$\Delta E(\delta_o) = 2Vc'\delta_o^2 + \mathcal{O}(\delta_o^4)$$

(7)

$$B = \frac{(C_{11}+2C_{12})}{3}$$

(8)

$$c' = \frac{(C_{11}-C_{12})}{2}$$

(9)

$$\mathcal{D}_m + \mathcal{J} = \begin{pmatrix} 0 & \delta_m & 0 \\ \delta_m & 1 & 0 \\ 0 & 0 & \frac{1}{1-\delta_m^2} \end{pmatrix}$$

(10)

$$\Delta E(\delta_m) = 2VC_{44}\delta_m^2 + \mathcal{O}(\delta_m^4)$$

(11)

The polycrystalline elastic moduli were obtained from the calculated cubic elastic constants. The Voigt bound and Reuss bound of the polycrystalline bulk modulus of cubic structure are the same, $B_R = B_V = B_0$[80]. We used the arithmetic Hill average for the polycrystalline shear

modulus: $G = \frac{G_V + G_R}{2}$, where G_R is the Reuss bound (Eq. 12) and G_V is the Voigt bound (Eq. 13). The Young's modulus E and Poisson's ratio v were obtained from Eqs. (14) and (15). The Zener ratio A_z (Eq. 16) was employed to assess the elastic anisotropy. When $A_z = 1$, the crystal lattice is elastically isotropic.

$$G_V = \frac{C_{11} - C_{12} + 3C_{44}}{5} \tag{12}$$

$$G_R = \frac{5(C_{11} - C_{12})C_{44}}{4C_{44} + 3(C_{11} - C_{12})} \tag{13}$$

$$E = \frac{9BG}{3B + G} \tag{14}$$

$$v = \frac{3B - 2G}{2(3B + G)} \tag{15}$$

$$A_z = \frac{2C_{44}}{C_{11} - C_{12}} \tag{16}$$

For all DFT calculations, the exchange-correlation energy was defined by the generalized gradient approximation (GGA) in the Perdew-Burke-Ernzerhof (PBE) parameterization[81] with the full charge density (FCD) techniques[39,82,83]. The Monkhorst-Pack k-point grid was set to $17 \times 17 \times 17$ and the energy was converged to 10^{-6} eV/atom. The screened impurity model parameter of CPA was 0.6 with the soft-core approximation. The magnetic moment was initialized as ferromagnetic. We developed a workflow based on pyEMTO[84], performed and analyzed more than 160,000 high-throughput EMTO-CPA calculations[85]. The complete dataset is included in the supplementary information. Due to numerical issues, a small fraction of the calculations (especially for compositions with Ta) did not converge to the required accuracy. Results from all successful calculations are included in the reported dataset, but our ML modeling and analysis does not include data with Ta to avoid introducing bias.

Deep sets model

The main idea of Deep Sets is that any permutation invariant function f over a set X (e.g., the set of elements x in an equimolar HEA) can be represented as $\rho\left(\sum_{x \in X} \phi(x)\right)$, where ρ and ϕ are two functions that are usually parametrized by deep neural networks. In this study, we used MLP (multilayer perceptron) with ELU (exponential linear unit) activation function for ρ and MLP with ReLU (rectified linear unit) for ϕ. In particular, $\mathrm{ELU}(x) = \begin{cases} x & \text{if } x > 0 \\ \alpha(e^x - 1) & \text{if } x < 0 \end{cases}$, where α was set as 1, and $\mathrm{ReLU}(x) = \begin{cases} x & \text{if } x > 0 \\ 0 & \text{if } x < 0 \end{cases}$. We used the Adam optimizer[86] with batch size 32, learning rate 10^{-3} and weight decay rate 10^{-4}. In the case of quaternary HEAs with non-equimolar ratios, we incorporated the weights by replacing $\rho\left(\sum_{x \in X} \phi(x)\right)$ with $\rho\left(\sum_{x \in X} \phi(w_x x)\right)$ where w_x stands for the weight of element x in the HEA. Since we consider HEAs as a collection of atoms randomly decorated on a lattice, it is natural to represent HEAs as a set of weighted atoms, with weights representing the elemental concentration.

We predicted the following properties using the Deep Sets model: sws, lattice parameter (a), elastic constants (C_{11}, C_{12}, C_{44}), B, G, E, v, and B/G. Each quaternary HEA system was represented as a set of size four, $\{u_1, u_2, u_3, u_4\}$, where u_i is the feature vector for an element of the HEA. Elemental properties were encoded in u_i using one-hot encoding[50]. Table 4 lists the elemental features we selected to train the model. For discrete properties, the property values were encoded according to the category the value belongs to; for continuous properties, the range of property values was evenly divided into 10 categories and the vectors were encoded accordingly. For instance, if we use the group number and period number as the elemental features, the atom feature vector for H will be a 27-dimensional vector with the 1st and 19th element being 1 and the other elements being 0. If the atomic radius is 0.7 pm, then the atomic size feature vector will be a 10-dimensional vector with the 1st element being 1 and the other elements being 0. The training data consisted of EMTO-CPA computed property data for 1911 equimolar and non-equimolar quaternary HEAs with complete cubic elastic constants. We randomly split the dataset into 60%, 20%, 20% as training, validation, and test set. We repeated the experiments 10 times and report the average MAE and the standard deviation. We compared our Deep Sets model with KNN, LR, RF,

Table 4. Features for Deep Sets training.

Feature	Unit	Range	# of categories
Group number	–	1, 2, …, 18	18
Period number	–	1, 2, …, 9	9
Electronegativity	–	0.5~4.0	10
Covalent radius	pm	25~250	10
Valence electrons	–	1, 2, …, 12	12
First ionization energy	eV	1.3~3.3	10
Electron affinity	eV	–3~3.7	10
Block	–	s, p, d, f	4
Atomic volume	$cm^3 mol^{-1}$	1.5~4.3	10

Table 5. Tested space for hyper-parameters.

Method	Hyper-parameters
Deep Sets	Hidden dimension for ϕ: 94; hidden dimension for ρ: {100, 50}
GBT	Number of estimators: {10,100,300,500,1000}
KNN	Number of neighbors: {3,5,7,10,20}
LR	No hyperparameters.
RF	Number of estimators: {10,20,50,100,200}
SVM	Kernel type: {linear,rbf}; bandwidth:{10^{-2},10^{-1},1,10^1,10^2}, C: {10^{-2},10^{-1},1,10^1,10^2,10^3,10^4}.

GBT, and SVM models. Table 5 lists the hyper-parameter space that was explored for each algorithm. We first used the training set that contains the features listed in Table 4 along with the stable lattice type (FCC or BCC) as an optional feature.

After training, we applied the Deep Sets model to predict the elastic properties of 369,369 HEA compositions for further analysis. The larger composition pool contains all quaternary compositions $A_1B_xC_yD_z$ ($0.6 \le x, y, z \le 1$ with an 0.1 increment) of the 14 elements. Because the relative cubic phase stability is unknown for most compositions in the space, we used a Deep Sets model that does not include the stable lattice type for property prediction.

Association rule mining

ARM was utilized to identify meaningful correlations between the elemental combinations and the Deep Sets predicted HEA elastic properties. ARM is a descriptive data mining method to discover underlying relationships between different items in a dataset[87]. In this study, the items are element fractions and target properties. The relationships or associations are represented by *If-Then* rules consisting of an antecedent and a consequent. For example, the rule shown below says: If a HEA composition contains low fractions of elements A and B, then the target property will have a relatively high elastic constant value.

$$X \Rightarrow Y$$

X : low fractions of elements A and B

Y : high value predicted for target elastic property

The condition of low fractions of the A and B elements is the antecedent, X, which is correlated with the consequent, Y, the high value of the target property. While these rules represent correlations and do not imply causation, they have directionality meaning that while antecedent X leads to consequent Y it does not necessarily mean that Y leads to X.

The first step in ARM is rule generation. Due to the sheer number of feature combinations that can make up both the antecedents and consequents, it is impossible to evaluate the entire combinatorial space of rules. Therefore, Apriori algorithm was used to generate rules from frequently occurring item sets[88]. The Apriori algorithm searches for rules based on the assumption that all subsets of a frequently occurring item set

will also be frequent. In other words, a rule with fewer conditions, which is more general, will be satisfied more frequently. From this assumption it follows that if an item set is infrequent, its supersets will also be infrequent. The Apriori algorithm performs a breadth-first search of rules starting with item sets with few items and adding items to the item set, while only considering item sets that exceed a threshold of support in the dataset. The support of an item set is defined as the frequency of the item set within the dataset. If an item set does not have enough support, none of its supersets are considered. After the frequent item sets are found, relevant rules are evaluated based on metrics such as support, confidence, and lift.

$$\text{support}(X \Rightarrow Y) = P(X \text{ and } Y)$$

$$\text{confidence}(X \Rightarrow Y) = P(Y|X) = \frac{P(X \text{ and } Y)}{P(X)}$$

$$\text{lift}(X \Rightarrow Y) = \frac{P(X \text{ and } Y)}{P(X) \cdot P(Y)}$$

Confidence can be used to find rules with high probability of the consequent Y conditional on the antecedent X. Confidence has a drawback that if the consequent is frequent, the confidence for the rule may be high even though a true relation does not exist between the antecedent and the consequent. Lift is the ratio of the probability of events X and Y co-occurring to the product of the probability of event X and event Y. In other words, lift is a metric that shows whether the co-occurrence of events X and Y is more frequent than would be expected if the two events are statistically independent. Lift does not suffer from the same drawback as confidence.

The ARM methodology works on datasets with binary data rather than the continuous data that the Deep Sets model predicts. Our aim of the ARM analysis is to produce descriptive trends rather than to train a separate quantitative predictive model. Therefore, the continuous data was first discretized into 10 bins, which transformed the continuous dataset into an ordinal dataset[89]. Then, the ordinal data was further transformed by one-hot encoding. The labels of the one-hot encoded data correspond to the rank of the ordinal data. For example, the one-hot encoded feature labels 'Cr0' and 'Cr9' corresponds to low and high Cr fractions respectively. A low threshold support value of 0.006 was used to filter the item sets, to ensure that meaningful rules were not missed due to the relative rarity of an item set. An item set with the support of 0.006 corresponds to 2,216 occurrences in our dataset. The rules were then filtered using a lift threshold of 2.4. The choice of this lift threshold was to keep the number of rules low. Higher lift thresholds can result in too few rules.

DATA AVAILABILITY
All data used to train and validate the Deep Sets model are available in the Supplementary Materials.

CODE AVAILABILITY
The Deep Sets code is available at GitHub (github.com/Chen-Cai-OSU/hea-ml).

REFERENCES
1. George, E. P., Raabe, D. & Ritchie, R. O. High-entropy alloys. *Nat. Rev. Mater.* **4**, 515–534 (2019).
2. Yeh, J. W. et al. Nanostructured high-entropy alloys with multiple principal elements: pagovel alloy design concepts and outcomes. *Adv. Eng. Mater.* **6**, 299–303 (2004).
3. Miracle, D. et al. Exploration and development of high entropy alloys for structural applications. *Entropy* **16**, 494–525 (2014).
4. Miracle, D. B. & Senkov, O. N. A critical review of high entropy alloys and related concepts. *Acta Mater.* **122**, 448–511 (2017).
5. Zhang, Y., Zuo, T., Cheng, Y. & Liaw, P. K. High-entropy alloys with high saturation magnetization, electrical resistivity, and malleability. *Sci. Rep.* **3**, 1455 (2013).
6. Senkov, O. N., Miracle, D. B., Chaput, K. J. & Couzinie, J.-P. Development and exploration of refractory high entropy alloys—a review. *J. Mater. Res.* **33**, 3092–3128 (2018).
7. Cantor, B., Chang, I. T. H., Knight, P. & Vincent, A. J. B. Microstructural development in equiatomic multicomponent alloys. *Mater. Sci. Eng. A* **375–377**, 213–218 (2004).
8. Zhang, Y., Yang, X. & Liaw, P. K. Alloy design and properties optimization of high-entropy alloys. *JOM* **64**, 830–838 (2012).
9. Yang, X., Zhang, Y. & Liaw, P. K. Microstructure and compressive properties of NbTiVTaAlx high entropy alloys. *Procedia Eng.* **36**, 292–298 (2012).
10. Wong, S. K., Shun, T. T., Chang, C. H. & Lee, C. F. Microstructures and properties of Al$_{0.3}$CoCrFeNiMn$_x$ high-entropy alloys. *Mater. Chem. Phys.* **210**, 146–151 (2018).
11. Shi, Y., Yang, B. & Liaw, P. Corrosion-resistant high-entropy alloys: a review. *Metals* **7**, 43 (2017).
12. Chen, Y. Y., Duval, T., Hung, U. D., Yeh, J. W. & Shih, H. C. Microstructure and electrochemical properties of high entropy alloys-a comparison with type-304 stainless steel. *Corros. Sci.* **47**, 2257–2279 (2005).
13. Chuang, M. H., Tsai, M. H., Wang, W. R., Lin, S. J. & Yeh, J. W. Microstructure and wear behavior of Al$_x$Co$_{1.5}$CrFeNi$_{1.5}$Ti$_y$ high-entropy alloys. *Acta Mater.* **59**, 6308–6317 (2011).
14. Shafeie, S. et al. High-entropy alloys as high-temperature thermoelectric materials. *J. Appl. Phys.* **118**, (2015).
15. Kaufmann, K. et al. Discovery of high-entropy ceramics via machine learning. *npj Comput. Mater.* **6**, 42 (2020).
16. Sarker, P. et al. High-entropy high-hardness metal carbides discovered by entropy descriptors. *Nat. Commun.* **9**, 4980 (2018).
17. Sarkar, A. et al. High entropy oxides for reversible energy storage. *Nat. Commun.* **9**, 3400 (2018).
18. Senkov, O. N., Miller, J. D., Miracle, D. B. & Woodward, C. Accelerated exploration of multi-principal element alloys with solid solution phases. *Nat. Commun.* **6**, 6529 (2015).
19. Lederer, Y., Toher, C., Vecchio, K. S. & Curtarolo, S. The search for high entropy alloys: a high-throughput ab-initio approach. *Acta Mater.* **159**, 364–383 (2018).
20. Hart, G. L. W., Mueller, T., Toher, C. & Curtarolo, S. Machine learning for alloys. *Nat. Rev. Mater.* **6**, 730–755 (2021).
21. Borg, C. K. H. et al. Expanded dataset of mechanical properties and observed phases of multi-principal element alloys. *Sci. Data* **7**, 430 (2020).
22. Islam, N., Huang, W. & Zhuang, H. L. Machine learning for phase selection in multi-principal element alloys. *Comput. Mater. Sci.* **150**, 230–235 (2018).
23. Pei, Z., Yin, J., Hawk, J. A., Alman, D. E. & Gao, M. C. Machine-learning informed prediction of high-entropy solid solution formation: beyond the Hume-Rothery rules. *npj Comput. Mater.* **6**, 50 (2020).
24. Li, J. et al. High-throughput simulation combined machine learning search for optimum elemental composition in medium entropy alloy. *J. Mater. Sci. Technol.* **68**, 70–75 (2021).
25. Jafary-Zadeh, M., Khoo, K. H., Laskowski, R., Branicio, P. S. & Shapeev, A. V. Applying a machine learning interatomic potential to unravel the effects of local lattice distortion on the elastic properties of multi-principal element alloys. *J. Alloy. Compd.* **803**, 1054–1062 (2019).
26. Huang, S., Tian, F. & Vitos, L. Elasticity of high-entropy alloys from ab initio theory. *J. Mater. Res.* **33**, 2938–2953 (2018).
27. Huang, S., Vida, Á., Heczel, A., Holmström, E. & Vitos, L. Thermal expansion, elastic and magnetic properties of FeCoNiCu-based high-entropy alloys using first-principle theory. *JOM* **69**, 2107–2112 (2017).
28. Lee, C. et al. Temperature dependence of elastic and plastic deformation behavior of a refractory high-entropy alloy. *Sci. Adv.* **6**, eaaz4748 (2020).
29. Tian, F., Varga, L. K., Chen, N., Shen, J. & Vitos, L. Ab initio design of elastically isotropic TiZrNbMoVx high-entropy alloys. *J. Alloy. Compd.* **599**, 19–25 (2014).
30. Chen, S. Y. et al. Phase transformations of HfNbTaTiZr high-entropy alloy at intermediate temperatures. *Scr. Mater.* **158**, 50–56 (2019).
31. Kim, G. et al. First-principles and machine learning predictions of elasticity in severely lattice-distorted high-entropy alloys with experimental validation. *Acta Mater.* **181**, 124–138 (2019).
32. Lee, C. et al. Lattice-distortion-enhanced yield strength in a refractory high-entropy alloy. *Adv. Mater.* **32**, 2004029 (2020).
33. Curtarolo, S. et al. The high-throughput highway to computational materials design. *Nat. Mater.* **12**, 191–201 (2013).
34. De Jong, M. et al. Charting the complete elastic properties of inorganic crystalline compounds. *Sci. Data* **2**, 150009 (2015).
35. Jain, A. et al. Commentary: the materials project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
36. Couzinié, J.-P., Senkov, O. N., Miracle, D. B. & Dirras, G. Comprehensive data compilation on the mechanical properties of refractory high-entropy alloys. *Data Brief* **21**, 1622–1641 (2018).
37. Gorsse, S., Nguyen, M. H., Senkov, O. N. & Miracle, D. B. Database on the mechanical properties of high entropy alloys and complex concentrated alloys. *Data Brief* **21**, 2664–2678 (2018).
38. Vitos, L., Abrikosov, I. A. & Johansson, B. Anisotropic lattice distortions in random alloys from first-principles theory. *Phys. Rev. Lett.* **87**, 156401 (2001).
39. Vitos, L. The EMTO-CPA method. In *Computational Quantum Mechanics for Materials Engineers* (Springer London, 2007).

40. Wu, Z. et al. A Comprehensive Survey on Graph Neural Networks. *IEEE Trans. Neural Netw. Learn. Syst.* **32**, 4–24 (2021).

41. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. In *5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings* (2017).

42. Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A. & Vandergheynst, P. Geometric deep learning: going beyond euclidean data. *IEEE Signal Process. Mag.* **34**, 18–42 (2017).

43. Gurukar, S. et al. Network representation learning: consolidation and renewed bearing. Preprint at https://arxiv.org/abs/1905.00987 (2019).

44. Hamilton, W. L., Ying, R. & Leskovec, J. Inductive representation learning on large graphs. In *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems* (Curran Associates, Red Hook, NY, 2017).

45. Dettmers, T., Minervini, P., Stenetorp, P. & Riedel, S. Convolutional 2d knowledge graph embeddings. In *Thirty-Second AAAI Conference on Artificial Intelligence* (AAAI Press, Palo Alto, CA, 2018).

46. Ying, Z. et al. Hierarchical graph representation learning with differentiable pooling. In *Advances in Neural Information Processing Systems 31* (Curran Associates, Red Hook, NY, 2018).

47. Sanchez-Gonzalez, A. et al. Learning to simulate complex physics with graph networks. *Proc. Mac. Learn*, **119**, 8459–8468 (2020).

48. Dai, H., Khalil, E., Zhang, Y., Dilkina, B. & Song, L. Learning combinatorial optimization algorithms over graphs. In *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems* (Curran Associates, Red Hook, NY, 2017).

49. Li, Z., Chen, Q. & Koltun, V. Combinatorial optimization with graph convolutional networks and guided tree search. In *NIPS'18: Proceedings of the 32nd International Conference on Neural Information Processing Systems* (Curran Associates, Red Hook, NY, 2018).

50. Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2017).

51. Schütt, K. T., Sauceda, H. E., Kindermans, P.-J., Tkatchenko, A. & Müller, K.-R. SchNet – a deep learning architecture for molecules and materials. *J. Chem. Phys.* **148**, 241722 (2018).

52. Duvenaud, D. et al. Convolutional networks on graphs for learning molecular fingerprints. In *NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems* (Curran Associates, Red Hook, NY, 2015).

53. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. In *ICML'17: Proceedings of the 34th International Conference on Machine Learning* (JMLR.org, 2017).

54. Zaheer, M. et al. Deep sets. In *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems* (Curran Associates, Red Hook, NY, 2017).

55. Salishchev, G. A. et al. Effect of Mn and V on structure and mechanical properties of high-entropy alloys based on CoCrFeNi system. *J. Alloy. Compd.* **591**, 11–21 (2014).

56. Adomako, N. K., Kim, J. H. & Hyun, Y. T. High-temperature oxidation behaviour of low-entropy alloy to medium-and high-entropy alloys. *J. Therm. Anal. Calorim.* **133**, 13–26 (2018).

57. Senkov, O. N., Senkova, S. V., Woodward, C. & Miracle, D. B. Low-density, refractory multi-principal element alloys of the Cr–Nb–Ti–V–Zr system: microstructure and phase analysis. *Acta Mater.* **61**, 1545–1557 (2013).

58. Zuo, T. T., Li, R. B., Ren, X. J. & Zhang, Y. Effects of Al and Si addition on the structure and properties of CoFeNi equal atomic ratio alloy. *J. Magn. Magn. Mater.* **371**, 60–68 (2014).

59. Liu, L., Zhu, J. B., Zhang, C., Li, J. C. & Jiang, Q. Microstructure and the properties of FeCoCuNiSn x high entropy alloys. *Mater. Sci. Eng. A* **548**, 64–68 (2012).

60. Chen, H. et al. Contribution of lattice distortion to solid solution strengthening in a series of refractory high entropy alloys. *Metall. Mater. Trans. A* **49**, 772–781 (2018).

61. Stepanov, N. D., Shaysultanov, D. G., Salishchev, G. A. & Tikhonovsky, M. A. Structure and mechanical properties of a light-weight AlNbTiV high entropy alloy. *Mater. Lett.* **142**, 153–155 (2015).

62. Wu, Y. D. et al. A refractory $Hf_{25}Nb_{25}Ti_{25}Zr_{25}$ high-entropy alloy with excellent structural stability and tensile properties. *Mater. Lett.* **130**, 277–280 (2014).

63. Chen, S. Y., Yang, X., Dahmen, K. A., Liaw, P. K. & Zhang, Y. Microstructures and crackling noise of AlxNbTiMoV high entropy alloys. *Entropy* **16**, 870–884 (2014).

64. Zhang, M., Zhou, X. & Li, J. Microstructure and mechanical properties of a refractory CoCrMoNbTi high-entropy alloy. *J. Mater. Eng. Perform.* **26**, 3657–3665 (2017).

65. Tian, F. et al. Structural stability of $NiCoFeCrAl_x$ high-entropy alloy from ab initio theory. *Phys. Rev. B - Condens. Matter Mater. Phys.* **88**, 85128 (2013).

66. Tian, F., Varga, L. K., Shen, J. & Vitos, L. Calculating elastic constants in high-entropy alloys using the coherent potential approximation: current issues and errors. *Comput. Mater. Sci.* **111**, 350–358 (2016).

67. Ge, H., Song, H., Shen, J. & Tian, F. Effect of alloying on the thermal-elastic properties of 3d high-entropy alloys. *Mater. Chem. Phys.* **210**, 320–326 (2018).

68. Tian, F., Varga, L. K. & Vitos, L. Predicting single-phase CrMoWX high entropy alloys from empirical relations in combination with first-principles calculations. *Intermetallics* **83**, 9–16 (2017).

69. Tian, L. Y. et al. Alloying effect on the elastic properties of refractory high-entropy alloys. *Mater. Des.* **114**, 243–252 (2017).

70. Ge, H., Tian, F. & Wang, Y. Elastic and thermal properties of refractory high-entropy alloys from first-principles calculations. *Comput. Mater. Sci.* **128**, 185–190 (2017).

71. Li, X., Tian, F., Schönecker, S., Zhao, J. & Vitos, L. Ab initio-predicted micro-mechanical performance of refractory high-entropy alloys. *Sci. Rep.* **5**, 12334 (2015).

72. Wang, W. H. The elastic properties, elastic models and elastic perspectives of metallic glasses. *Prog. Mater. Sci.* **57**, 487–656 (2012).

73. Isayev, O. et al. Universal fragment descriptors for predicting properties of inorganic crystals. *Nat. Commun.* **8**, 1–12 (2017).

74. Choudhary, K., DeCost, B. & Tavazza, F. Machine learning with force-field inspired descriptors for materials: fast screening and mapping energy landscape. *Phys. Rev. Mater.* **2**, 083801 (2018).

75. Senkov, O. N. & Miracle, D. B. Generalization of intrinsic ductile-to-brittle criteria by Pugh and Pettifor for materials with a cubic crystal structure. *Sci. Rep.* **11**, 10–13 (2021).

76. Pang, J. et al. A ductile $Nb_{40}Ti_{25}Al_{15}V_{10}Ta_5Hf_3W_2$ refractory high entropy alloy with high specific strength for high-temperature applications. *Mater. Sci. Eng. A* **831**, 142290 (2022).

77. Popescu, G. et al. New TiZrNbTaFe high entropy alloy used for medical applications. *IOP Confer. Ser. Mater. Sci. Eng.* **400**, 022049 (2018).

78. Zhang, C. & Gao, M. C. CALPHAD modeling of high-entropy alloys. in *High-Entropy Alloys* 399–444 (Springer International Publishing, 2016).

79. Vitos, L. *Computational Quantum Mechanics for Materials Engineers: The EMTO Method and Applications* (Springer Science & Business Media, 2007).

80. Hill, R. The elastic behaviour of a crystalline aggregate. *Proc. Phys. Soc. Sect. A* **65**, 349–354 (2002).

81. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

82. Vitos, L. Total-energy method based on the exact muffin-tin orbitals theory. *Phys. Rev. B* **64**, 014107 (2001).

83. Vitos, L. Full charge density technique. In *Computational Quantum Mechanics for Materials Engineers* 59–81 (Springer London, 2007).

84. Levämäki, H. & Hpleva. pyemto. https://pyemto.readthedocs.io/en/latest/features.html (2017).

85. Jain, A. et al. FireWorks: a dynamic workflow system designed for high-throughput applications. *Concurr. Comput. Pract. Exp.* **27**, 5037–5059 (2015).

86. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015* (2015).

87. Hipp, J., Güntzer, U. & Nakhaeizadeh, G. Algorithms for association rule mining — a general survey and comparison. *SIGKDD Explor. Newsl.* **2**, 58–64 (2000).

88. Agrawal, R., Mannila, H., Srikant, R., Toivonen, H. & Verkamo, A. I. Fast discovery of association rules. *Adv. Knowl. Discov. Data Min.* **12**, 307–328 (1996).

89. Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

W.C. and Y.W. conceived the study and managed the project. J.Z. performed the first-principles calculations. C.C. developed the Deep Sets machine learning model. G.K. performed the association rule mining. All authors participated in the results analysis and wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION