

ARTICLE OPEN



Predicting temperature-dependent ultimate strengths of body-centered-cubic (BCC) high-entropy alloys

B. Steingrímsson¹, X. Fan², X. Yang³, M. C. Gao⁴, Y. Zhang^{5,6,7} and P. K. Liaw²

This paper presents a bilinear log model, for predicting temperature-dependent ultimate strength of high-entropy alloys (HEAs) based on 21 HEA compositions. We consider the break temperature, T_{break} , introduced in the model, an important parameter for design of materials with attractive high-temperature properties, one warranting inclusion in alloy specifications. For reliable operation, the operating temperature of alloys may need to stay below T_{break} . We introduce a technique of global optimization, one enabling concurrent optimization of model parameters over low-temperature and high-temperature regimes. Furthermore, we suggest a general framework for joint optimization of alloy properties, capable of accounting for physics-based dependencies, and show how a special case can be formulated to address the identification of HEAs offering attractive ultimate strength. We advocate for the selection of an optimization technique suitable for the problem at hand and the data available, and for properly accounting for the underlying sources of variations.

npj Computational Materials (2021)7:152; <https://doi.org/10.1038/s41524-021-00623-4>

INTRODUCTION

Metallic structural materials with excellent mechanical properties have been widely used in a variety of operating conditions and often applied under constant or static loads. Engineering components under either loading conditions are usually required to exhibit high strength. Thus, it is important to be able to design advanced materials with favorable strength properties. High-entropy alloys (HEAs) have drawn great attention in the recent decade due to their excellent mechanical properties and vast compositional space, which makes them suitable for joint optimization¹. A key objective is to suggest a framework for joint optimization of mechanical properties, to introduce—in context with such a framework—compositions of HEAs yielding high ultimate strengths (USs), and to conduct experimental verification of our findings.

Figure 1 outlines the multiple sources that impact the mechanical properties of HEAs, and highlights dependence between them. It is worth noting that improvements in the US may come at the expense of other properties (hence, framework for joint optimization). For example, there usually is a trade-off between the ductility and the strength of alloys. Sources of variations in US may involve difference in compositions, microstructures, parameters of post-fabrication processes, or defect levels.

In contrast to traditional alloys containing only one or two principal elements, multi-principal-element alloys, also referred to as HEAs, have been developed and studied in the recent decade^{1–7}. Carefully designed HEAs with either single or multiple phases have shown encouraging mechanical properties, compared to conventional alloys^{8–16}.

Data analytics and machine learning (ML) can help with rapid screening, i.e., expedite identification of HEAs exhibiting given properties of interest¹⁷. But as opposed to specifically applying ML, (narrowly) defined in terms of single-layer or multi-layer

neural networks¹⁷, Bayesian graphical models, support vector machines, or decision trees, to identification of HEA compositions of interest, we reformulate the task in the broader context of engineering optimization. We recommend picking an optimization technique suitable for the application at hand and data available. But we certainly include ML in the consideration. For background material on ML, refer to¹⁷.

Effective application of ML may require a large number of data points. If you have such data, then ML can help you organize the data in a meaningful fashion and extract complex, hidden relationships¹⁷. But in the case of experimental data on HEA compositions with attractive strength properties (the present state of affairs), we are working in a domain of relatively limited data, a domain where traditional ML may exhibit limitations. Producing high-quality experimental data is usually both time consuming and expensive. In case of such limited data, it is essential to make the most of the underlying physics, i.e., to account for underlying physical dependencies, in the prediction model. Occam's razor and Bayesian learning provide tools for quantifying the notion of limited data in this context¹⁷.

Our approach is in part based on observations of Agrawal et al.¹⁸. Table 2 and Figure 5 of¹⁸ illustrate that there is at most a difference of a few percentages between the techniques applied to predict the fatigue strength of stainless steels. Table 2 of¹⁸ shows that both simple linear regression and pace regression yield the coefficient of determination, R^2 , of 0.963, while an artificial neural network, a traditional ML technique, results in R^2 of 0.972.

In terms of an important contribution, this study presents a method capable of yielding consistency among the predictions of HEA compositions with attractive US, empirical rules of thermodynamics^{19,20}, and experimental results. We accomplish this goal, despite relatively limited data available, and the corresponding selection of a simple prediction algorithm (multi-variate regression).

¹Imagars LLC, 2062 Thorncroft Drive Suite 1214, Hillsboro, OR 97124, USA. ²Department of Materials Science and Engineering, The University of Tennessee, Knoxville, TN 37996, USA. ³University of Chinese Academy of Sciences, Center of Materials Science and Optoelectronics Engineering, Beijing 100049, China. ⁴National Energy Technology Laboratory, 1450 Queen Ave. SW, Albany, OR 97321, USA. ⁵Beijing Advanced Innovation Center of Materials Genome Engineering, State Key Laboratory for Advanced Metals and Materials, University of Science and Technology Beijing, Beijing 100083, China. ⁶Qinghai Provincial Engineering Research Center of High Performance Light Metal Alloys and Forming, Qinghai University, Xining 810016, China. ⁷Shunde Graduate School, University of Science and Technology Beijing, Foshan 528399, China. ✉email: baldur@imagars.com; drzhangy@ustb.edu.cn

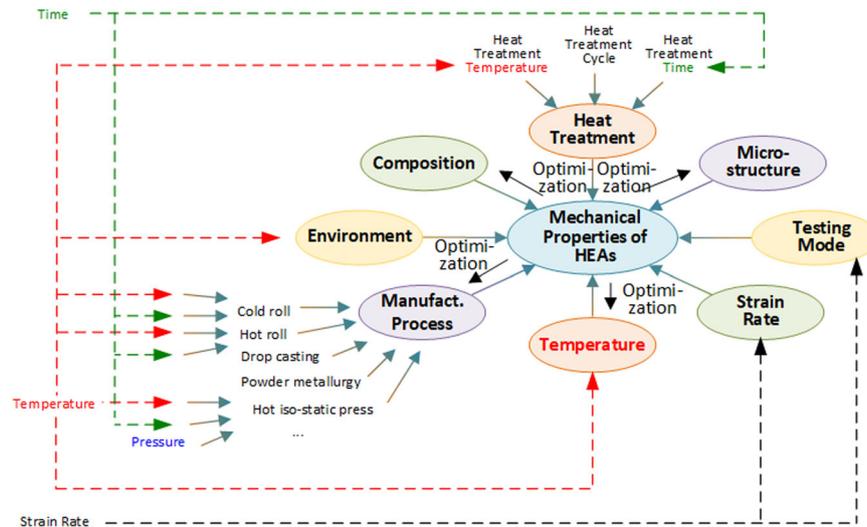


Fig. 1 High-level depiction of role of optimization techniques for inferring the features of HEAs, including the composition, microstructure, heat treatment, and process, from mechanical properties, including ultimate strength, desired. The color coding provides the insight into the extent to which the sources are separate, and yet interconnected to a certain extent^{34,35}.

Then, we present in Fig. 2 elements of a physics-based model for predicting the US, a model that accounts for physical dependencies as a priori information.

But more importantly, we introduce a bilinear log model for predicting USs across temperatures. The model consists of separate exponentials, for a low-temperature and a high-temperature regime, with a break temperature, T_{break} , in between. The model accounts for the underlying physics, in particular diffusion processes required to initiate phase transformations in the high-temperature regime²¹. Furthermore, we show how piecewise linear regression can be employed to extend the model beyond two exponentials and yield accurate fit, in case of a non-convex objective function caused by hump (s) in the data. Previous models for the temperature dependence of yield strengths (YSs) only accounted for a single exponential^{22,23}. Hence, there was no break temperature, T_{break} . We consider the break point critical for the optimization of the high-temperature properties of alloys. For reliable operation, the temperature of turbine blades made out of refractory alloys may need to stay below T_{break} . Once above T_{break} , materials can lose strength rapidly due to rapid diffusion, leading to easy dislocation motion and dissolution of strengthening phases²¹. We consider T_{break} an important parameter for the design of materials with attractive high-temperature properties, one warranting inclusion in alloy specifications. Hence, it is important to accurately estimate T_{break} , e.g., using the global optimization approach presented.

RESULTS

Room temperature

Figure 26 of¹⁷ summarizes the rationale for initially restricting the analysis to room-temperature data. As illustrated in the figure, the US exhibits significant dependence on the temperature. While all compositions in Figure 26 of¹⁷ contained a body-centered-cubic (BCC) phase, and were subjected to some type of annealing, the US at 1000 °C can be ~1/8th (~12%) to ~1/3rd (~33%) of the US at room temperature. With this fact in mind, and to maintain consistency across compositions, we elected to first apply the optimization framework to US at room temperature only.

Our original data set, listed in Table 13 of¹⁷, contains some 24 compositions that yield relatively high US at room temperature.

To accommodate the elements involved, we derived two feature sets, hereafter referred to as A and B, from the original data set in Table 13 of¹⁷:

$$\text{Feature Vector A} = \mathbf{x}_A = [\%Al, \%Mo, \%Nb, \%Ti, \%V, \%Ta, \%Zr, \%Hf], \quad (1)$$

$$\text{Feature Vector B} = \mathbf{x}_B = [\%Al, \%Mo, \%Nb, \%Ti, \%V, \%Ta, \%Zr, \%Hf, \%Cr]. \quad (2)$$

We have available 19 instances of Feature Vector A, and 22 of Feature Vector B. While the set of input data may seem small, we will show that the data suffice for meaningful prediction, provided that a suitable optimization technique is selected^{17,24}.

In terms of data curation, we concluded that the US values, except for MoNbTiV_xZr (from²⁵), were recorded with high enough fidelity to warrant inclusion. For revised US measurements for MoNbTiV_{0.75}Zr, refer to Table 1. For catching suspicious recordings of the US, one can employ proportionality relationships with the YS as a guideline. At least (or about) half of the references associated an uncertainty interval with the US values reported, with ΔUS usually within the range of 1% of the US reported.

In order to develop insight into the causes of variations in the US for the pure elements comprising Feature Vectors, A and B, and for the identification of a model for predicting compositions yielding high US, we point to Figures 24 and 25 of¹⁷. Figure 24 of¹⁷ shows that processing conditions and purity can contribute to variations in US of Al of ~3x and of ~4x in the US of Co. Figure 25 of¹⁷ similarly illustrates that processing methods can have significant influence on the US of V and Cr. For V, the variations in the US are ~2x, and for Cr, we are looking at variations of up to ~5x. This trend suggests that the inputs listed in Eq. (22) are indeed able to account for the variations in the US observed. But for a relative comparison of US across compositions, for the same heat-treatment process and defect levels, and at a fixed (room) temperature, the model

$$US = h(\text{composition}) \quad (3)$$

may suffice. There may be additional sources of variations, such as the test mode applied. But according to our tentative analysis, presented in Supplementary Table 2, the variations in the YS observed, based on the test mode applied, tend to be relatively small. For the prediction presented in Supplementary Fig. 3, and

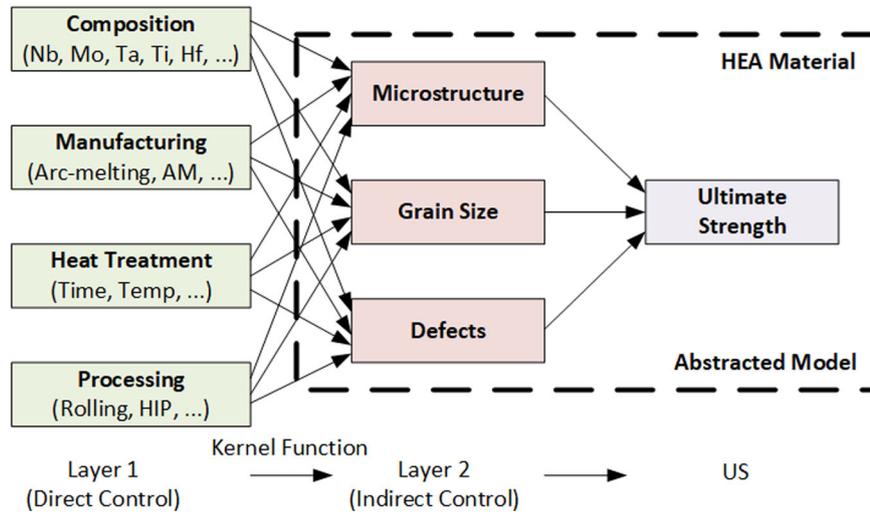


Fig. 2 Underlying physical dependencies structured in a fashion resembling a neural network. Our intent is to construct models capturing the underlying physics. This abstracted model shows that the microstructure formed depends on the heat-treatment process applied, manufacturing, processing as well as the composition³⁴. In case that an artificial neural network (ANN) is deemed suitable for the application at hand, we suggest employing custom kernel functions consistent with the underlying physics, for the purpose of attaining tighter coupling, better prediction, and extracting the most out of the—usually limited—input data available. Note that the composition can include trace-level elements (impurities), in addition to the principal components.

Table 1. Compression yield strength, σ_y , maximum strength, σ_{\max} , and fracture strain, ϵ_f , of the reference and predicted compositions at room temperature.

Composition	Sample diameter (mm)	Strain rate (s^{-1})	σ_y (MPa)	σ_{\max} (MPa)	ϵ_f (%)
$Al_{0.5}Mo_{0.5}NbTa_{0.5}TiZr$	3	1×10^{-3}	1786	1910	7.6
$Al_{0.5}Mo_{0.5}Nb_{1.5}Ta_{0.5}Zr_{1.5}$	3	1×10^{-3}	1791	2024	7.6
$MoNbTiV_{0.75}Zr$	3	1×10^{-3}	1675	2427	25.1
$MoNbTiV_{0.75}Zr$	5	2×10^{-4}	1599 ± 40	1780 ± 70	8.6 ± 2.8
$Mo_{1.25}Nb_{1.25}Ti_{0.5}V_{0.5}Zr_{1.25}$	3	1×10^{-3}	1705	2013	10.0

All the alloys show high strength and obvious plastic deformation. With the similar fracture strain (7.6%), the predicted $Al_{0.5}Mo_{0.5}Nb_{1.5}Ta_{0.5}Zr_{1.5}$ exhibit higher σ_y and σ_{\max} of 1791 and 2024 MPa, respectively, compared with the reference composition, $Al_{0.5}Mo_{0.5}NbTa_{0.5}TiZr$, which has σ_y and σ_{\max} as 1786 and 1910 MPa, respectively. $MoNbTiV_{0.75}Zr$ and $Mo_{1.25}Nb_{1.25}Ti_{0.5}V_{0.5}Zr_{1.25}$ have similar YSs (1675 and 1705 MPa, respectively), but the reference composition exhibits the higher maximum strength due to the high fracture strain and strain hardening. Note the corrected US values for $MoNbTiV_{0.75}Zr$ relative to²⁵. The uncertainty intervals in YS and US for the 5-mm diameter $MoNbTiV_{0.75}Zr$ were derived from measurements of three separate samples.

the experimental verification outlined in Fig. 5, we employ the prediction model of Eq. (3).

Given the relatively small size of the data set in Table 13 of¹⁷, it appears that we may not be ready for traditional ML models. Models, such as artificial neural networks, decision trees, support vector machines, Bayesian networks, or genetic algorithms, tend to be effective in organizing and extracting complex patterns from large sets of data¹⁷. But for the application and limited data set at hand, it makes sense to select a simple linear-prediction model, multi-variate linear regression, to begin with, and build from there. As suggested by Agrawal et al.¹⁸, changing the method may not vary the results that much. According to Figure 5 of¹⁸ and Table 2 of¹⁸, the linear regression yields R^2 of 0.963, when predicting the fatigue strength of a stainless steel, compared to R^2 of 0.972 for the artificial neural networks.

Our approach, outlined in¹⁷, assumes starting with a simple model, multi-variate linear regression, and accounting for the input sources that contribute to variations in the US observed. The approach then involves expanding the model, and adding non-linearities, based on the underlying physics, and as necessitated by the application at hand and data available.

When applying the multi-variate linear regression, we solve an unconstrained optimization problem of the form

$$\min_{\mathbf{b}} \|\mathbf{X}\mathbf{b} - \mathbf{y}\|_2^2. \quad (4)$$

Here, \mathbf{y} represents a vector of US values, \mathbf{b} denotes a vector of regression coefficients, and \mathbf{X} symbolizes a training set of compositions (a stacked version of \mathbf{x} vectors¹⁷, derived from Table 13 of¹⁷). This unconstrained optimization problem has a well-known, closed-form solution¹⁷:

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \quad (5)$$

When the training set is very small, the inverse $(\mathbf{X}^T\mathbf{X})^{-1}$ may not exist. In that case, we recommend replacing $(\mathbf{X}^T\mathbf{X})^{-1}$ with $(\mathbf{X}^T\mathbf{X})^+$, the pseudoinverse²⁶.

The observations reported in Supplementary Fig. 2—and more extensively in¹⁷—strengthen our belief in that the prediction accuracy, measured in terms of R^2 and the standard deviation normalized per data point, is primarily limited by the quality of (variance in) the input data. These limitations in the prediction accuracy are consistent with the variations observed in Table 13 and Figures 24 and 25 of¹⁷. These observations further suggest

that multi-variate regression is indeed a suitable technique for this application.

Elevated temperatures

In an effort to identify compositions exhibiting the ability to retain strength at high temperatures, we present Fig. 3. In case of high-temperature applications, we are looking to derive a model of the form

$$US = h(\text{composition}, T) \quad (6)$$

for the prediction of the US across temperatures. In addition to the temperature dependence of pure tungsten and the HEAs, the temperature dependence of the commercial alloys [the Mo-rich Titanium-Zirconium-Molybdenum alloy and the Nb-rich C-103 alloy] is also of interest.

Looking at Fig. 3, one first notices that the strength vs. temperature data definitely do not look linear. Hence, the multi-variate linear regression may no longer be the preferred approach. Second, the temperature dependence does come across as approximately exponential, but not exactly. It seems to entail a high-temperature and a low-temperature regime. Third, one may shy away from employing an automated ML suite, such as the Tree-Based Pipeline Optimization Tool²⁷, because of limited ability of such black-box models to provide much needed insights into the underlying physics. One is motivated to make the most of the limited data available, by incorporating important a priori information about the underlying physics into the model structure, for purpose of deriving such insights. Fourth, Fig. 3a, e.g., the high-temperature data points for MoNbTaW, highlights the need for data curation.

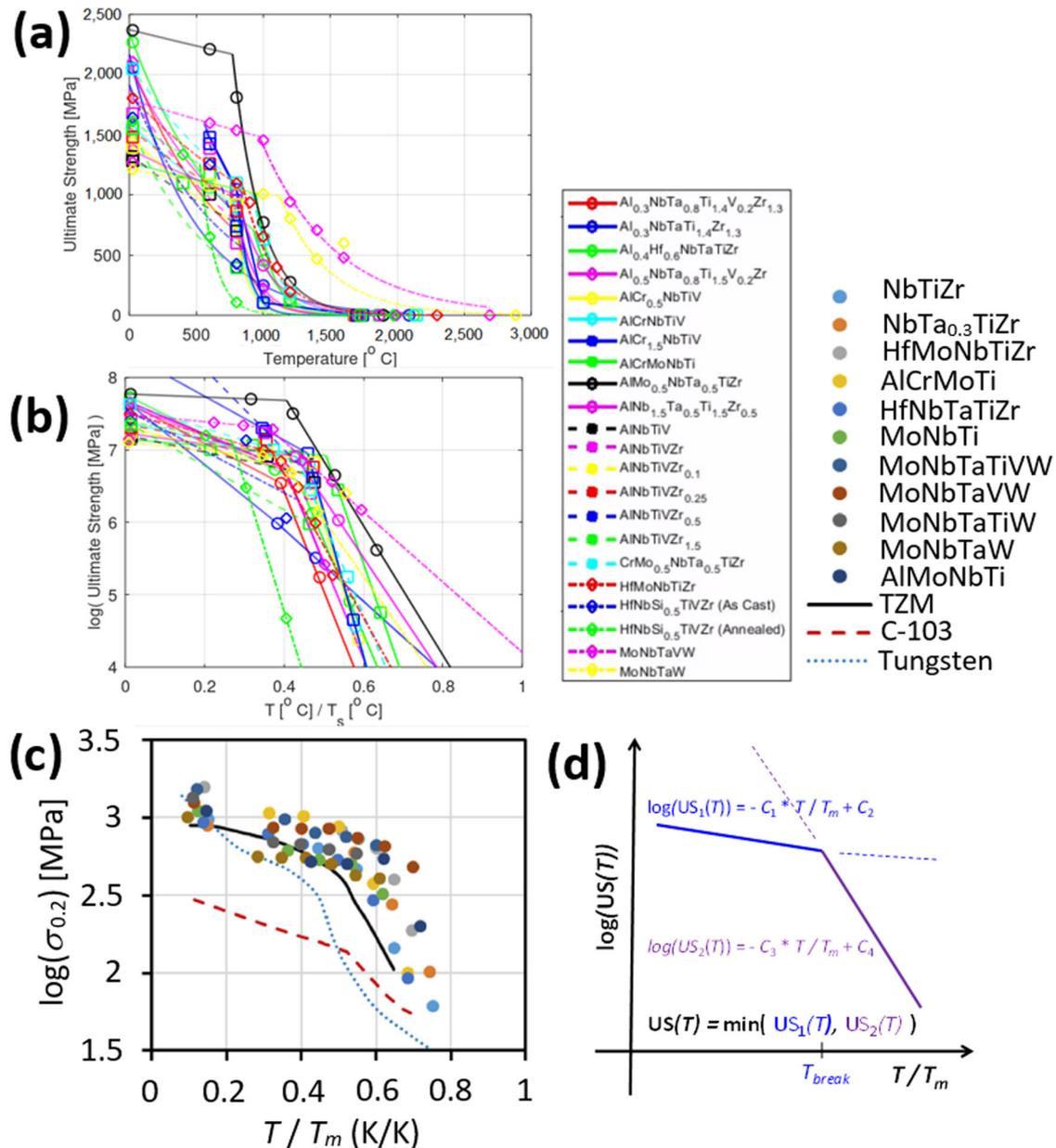


Fig. 3 Identification of compositions with the ability to retain strengths at high temperatures. Panel (a) shows the quality of fit for the bilinear log model in the linear domain. Panel (b) depicts the quality of fit for the bilinear log model in the logarithmic domain. Panels (c) and (d) illustrate how experimental measurements of temperature-dependent yield strength naturally conform to two linear regions, when viewed in the logarithmic domain. Panel (c) is reproduced from²¹ with permission.

Motivated by Fig. 3c, d, together with physics-based insights from²¹, we model the temperature dependence of the US(T), in terms of a bilinear log model, parametrized by the melting temperature, T_m , as follows:

$$US(T) = \min(\log(US_1(T)), \log(US_2(T))), \quad (7)$$

$$US_1(T) = \exp(-C_1 * T/T_m + C_2), \quad 0 < T < T_{\text{break}}, \quad (8)$$

$$US_2(T) = \exp(-C_3 * T/T_m + C_4), \quad T_{\text{break}} < T < T_m. \quad (9)$$

There is an additional physics (diffusion) induced constraint on T_{break} ²¹:

$$0.35 \lesssim T_{\text{break}}/T_m \lesssim 0.55, \quad (10)$$

and a continuity constraint between the low-temperature and the high-temperature regimes:

$$US_1(T_{\text{break}}) = US_2(T_{\text{break}}); \quad (11)$$

$$T_{\text{break}} = \frac{(C_4 - C_2)}{(C_3 - C_1)} T_m \quad (12)$$

A conceptually simple approach for fitting the model in Eqs. (7)–(12) to the US data available consists of first deriving the constant coefficients, C_1 and C_2 , by applying linear regression to data points available to the lowest temperature region ($0 < T < 0.35 T_m$) as well as to the intermediate region ($0.35 T_m \leq T \leq 0.55 T_m$). One can then derive the constants, C_3 and C_4 by applying linear regression to data points available to the intermediate ($0.35 T_m \leq T \leq 0.55 T_m$) and high-temperature ($T > 0.55 T_m$) regions. Note that T_{break} does not need to be known in advance. According to Eq. (12), this inherent property of a given alloy comes out of the model as the break point between the two linear regions. The model consists of only four independent parameters, C_1 , C_2 , C_3 , and C_4 , which simply can be estimated by applying linear regression separately to low-temperature and high-temperature regimes, even to a fairly small data set. Note, furthermore, that for a new alloy system, T_m does not need to be known experimentally in advance either; a rough estimate for T_m can be obtained, using “the rule of mixing”, and a more refined estimate obtained, employing Calculation of Phase Diagram (CALPHAD) simulations¹⁷.

A superior approach for deriving the coefficients, C_1 , C_2 , C_3 , and C_4 , involves concurrent optimization over the low-temperature and high-temperature regimes using global optimization. Here, we seek to minimize

$$\min_{C_1, C_2, C_3, C_4} \text{norm}_{2,i} \left((US(T_i) - y_i)^2 \right), \quad (13)$$

where y_i represents the measured US values,

$$US(T_i) = \min(\log(US_1(T_i)), \log(US_2(T_i))), \quad (14)$$

and $US_1(T_i)$ and $US_2(T_i)$ are modeled through Eqs. (8) and (9), respectively. Matlab provides a function, `fminunc()`, for solving this type of unconstrained minimization over a generic function. The results in Fig. 3a, b were derived, using the global optimization approach, applied separately to individual alloys, for the purpose of obtaining a tighter fit and more accurate estimation of T_{break} , than for separate optimization over the low and high-temperature regimes.

It is worth noting that previous models for the temperature dependence of the YS only accounted for a single exponential^{22,23}. Hence, there was no break temperature, T_{break} . We consider the break point critical for the optimization of the high-temperature properties of alloys. For reliable operation, the temperature of turbine blades made out of refractory alloys may need to stay below T_{break} (not accounting for coatings¹⁷). Once the temperature of the turbine blades exceeds T_{break} , undesirable phase transformations (e.g., dissolution of strengthening precipitates) may start to take place²¹, and the alloy may begin to lose

structural integrity. Here, the second exponential, modeled through $US_2(T)$, may prove detrimental. Turbine blades made out of certain alloys, such as Ni-based superalloys, should only be operated above T_{break} , if supported by extensive experimental test results. We consider the break point, T_{break} , an important parameter for the design of materials with attractive high-temperature properties, one warranting inclusion in alloy specifications. Hence, it is important to be able to accurately estimate T_{break} , e.g., using the global optimization approach presented [Eq. (13)].

Senkov et al. provide physics-based foundation for the prediction model in Eqs. (7)–(12)²¹. According to Senkov et al., the diffusion processes required to initiate phase transformations generally become noticeable at temperatures $T > 0.4 T_m$, while at $T < 0.4 T_m$ the phase transformations are kinetically restricted²¹. The atoms cannot move out of the lattice, and no phase transformations can take place. This trend applies to low-entropy alloys, medium-entropy alloys, and HEAs, and serves to explain the two exponentials. A solid-solution strengthening model by Rao et al.^{28,29} does not take into account diffusion effects and agrees well with the experimental data only at relatively low temperatures, where diffusion-controlling deformation mechanisms can be ignored. Then, as the temperature increases, the chemical bonds between the elements become softer. The diffusion-controlled regime generally occurs above ~ 0.5 – $0.6 T_m$. It can be distinguished from the low-temperature regime by a more rapid drop in strength with increasing temperature, because dislocations are able to move more easily around obstacles³⁰.

A related model for the prediction of YS over temperature was presented by Wu et al.²². The authors separately analyzed the temperature dependencies of the YS and strain hardening of a family of equi-atomic binary, ternary, and quaternary alloys based on the elements, Fe, Ni, Co, Cr, and Mn, which had been shown to form single-phase FCC solid solutions. The authors presented a model with a single exponential for the overall YS, $\sigma_y(T)$, of the form²²

$$\sigma_y(T) = \sigma_a \exp\left(\frac{-T}{C}\right) + \sigma_b, \quad (15)$$

where σ_a , C , and σ_b were fitting coefficients. The authors showed that lattice friction appeared to be the predominant component of the temperature-dependent YS, possibly because of the Peierls barrier height decreasing with increasing temperature, due to a thermally induced increase in dislocation width. Note, while similar to the YS, we are here modeling the US.

According to Maresca et al., the YS of the solid-solution BCC matrix alloy constitutes the major part of the alloy and can be estimated by²³:

$$\tau_y(T, \dot{\epsilon}) = \tau_{y0} \exp\left[-\frac{1}{0.55} \left(\frac{k_B T}{\Delta E_b} \ln \frac{\dot{\epsilon}_0}{\dot{\epsilon}}\right)^{0.91}\right], \quad (16)$$

where τ_{y0} is the zero-temperature flow stress, ΔE_b is the energy barrier for dislocation movement, T is the absolute temperature, $\dot{\epsilon}$ is the strain rate, and k_B is the Boltzmann constant. For accurate modeling of the YS, it is important to consider dislocations, atomic and volume misfits.

Depending on the grain sizes and compositions involved, a trilinear log model may yield a better fit for certain alloys³¹:

$$US(T) = \min(US_1(T), US_2(T), US_3(T)), \quad (17)$$

$$US_1(T) = \exp(-C_1 * T/T_m + C_2), \quad 0 < T < T_{\text{break}1}, \quad (18)$$

$$US_2(T) = \exp(-C_3 * T/T_m + C_4), \quad T_{\text{break}1} < T < T_{\text{break}2}, \quad (19)$$

$$US_3(T) = \exp(-C_5 * T/T_m + C_6), \quad T_{\text{break}2} < T < T_m, \quad (20)$$

Table 2. Quantification of ability of compositions to retain ultimate strength at high temperatures.

No.	Alloy	Solvus temperature (°C)	T_{break} (°C)	C_3 : slope for high-temp. regime in Fig. 3b	MSE: two-exponential		MSE: single exponential		
					Log	Linear	Log	Linear	
1	Al _{0.3} NbTa _{0.8} Ti _{1.4} V _{0.2} Zr _{1.3}	2043	800	13.77	1.7e-05	31	0.130	31,363	
2	Al _{0.3} NbTaTi _{1.4} Zr _{1.3}	2088	800	4.95 (only 3 data points)	0.039	2588	0.034	2455	
3	Al _{0.4} Hf _{0.6} NbTaTiZr	2124	927.3	12.90	4.4e-05	102	0.146	72,312	
4	Al _{0.5} NbTa _{0.8} Ti _{1.5} V _{0.2} Zr	1992	800	13.25	9.2e-11	0	0.151	47,797	
5	AlCr _{0.5} NbTiV	1704	769.3	18.30	7.4e-09	0	0.504	186,447	
6	AlCrNbTiV	1725	798.9	19.35	6.0e-05	57	0.501	203,776	
7	AlCr _{1.5} NbTiV	1741	806.3	20.00	6.2e-08	0	0.211	143,719	
8	AlCrMoNbTi	1867	943.7	15.80	0.002	1706	0.230	106,796	
9	AlMo _{0.5} NbTa _{0.5} TiZr	1896	770.1	8.93	0.001	1195	0.228	502,343	
10	AlNb _{1.5} Ta _{0.5} Ti _{1.5} Zr _{0.5}	1863	779.5	8.19	4.2e-08	22	0.080	47,474	
11	AlNbTiV	1679	Need more data				0.007	38,994	
12	AlNbTiVZr	1714	Need more data				0.070	78,813	
13	AlNbTiVZr _{0.1}	1683	Need more data				0.001	737	
14	AlNbTiVZr _{0.25}	1689	Need more data				0.012	90,556	
15	AlNbTiVZr _{0.5}	1698	Need more data (only 2 data points)				0.000	0	
16	AlNbTiVZr _{1.5}	1727	Need more data (only 2 data points)				0.000	0	
17	CrMo _{0.5} NbTa _{0.5} TiZr	2145	923.5	12.86	5.5e-08	0	0.175	97,285	
18	HfMoNbTiZr	2297	948.0	13.96	0.002	969	0.162	94,093	
19	HfNbSi _{0.5} TiVZr	1973	527.5	17.81	1.2e-07	0	0.304	236,176	
20	MoNbTaVW	2690	970.2	4.85	0.001	1229	0.017	30,751	
21	MoNbTaW	2885	1124.8	2.75–7.82	1.3e-11	14	0.045	36,023	
Average (No. 1–10 and 17–21):						2.9e-3	528	0.195	122,587

T_{break} refers to the breaking point between bilinear log models, defined in Fig. 3.

Given an anomalous yield stress phenomenon in a CMSX-4, single-crystal, Nickel-based superalloy, three exponentials are needed for accurate modeling in case of Heat Treatment A, but four exponentials in case of Heat Treatment B, according to Supplementary Fig. 4. This phenomenon manifests itself as a hump between the low-temperature and high-temperature regimes, found in superalloys strengthened by L1₂-ordered intermetallics. Here, the increased strength of γ' phase with temperature is explained by thermally activated cross-slip of dislocations from {111} planes to {100} planes. Supplementary Tables 4 and 5 present a practical approach to model selection suitable for this case. We stop increasing the model order, once the mean squared error (MSE) starts to taper off. Supplementary Figs. 5 and 6 capture an application of piecewise linear regression needed to address challenges imposed by non-convexity of the objective function possible in this case. Here we expand the parameter set such as to explicitly include the break temperatures. Supplementary Figs. 7 and 8 contain Matlab pseudo-code for the bilinear log model (a convex case) and a trilinear log model (a possibly non-convex case).

In Supplementary Note 7, we provide physics-based reasons explaining why a bilinear model will likely suffice for refractory HEAs³². We also address the number of data points needed for modeling. Since the hump between the low-temperature and high-temperature regimes originates from the γ' phase (which is a L1₂ phase, i.e., ordered FCC structure), and since most refractory HEAs contain BCC or hexagonal-closed-packed phases, with totally different dislocation systems, it is unlikely that cross-slip from {111} to {100} planes will happen in refractory HEAs³². Hence, we expect the bilinear log model (two exponentials) to suffice for most refractory HEAs.

Table 2 further characterizes the ability of the 21 compositions under consideration to retain strength at high temperatures, both in terms of a high break temperature, T_{break} , and a small slope, C_3 . Table 2 also compares the modeling accuracy of the bilinear log model to that of a single exponential. It is not surprising that the composition, MoNbTaW, which consists of strongly refractory elements (Mo, Nb, Ta, and W), i.e., elements with the melting point above 2200 °C³³, yields the highest T_{break} of 1124 °C. This evidence serves to validate the model. MoNbTaVW, which includes one weakly refractory element (V), i.e., an element with a melting point above 1850 °C, results in the smallest slope, C_3 , of 4.85, compared to 7.82 for MoNbTaW. But this observation assumes omitting the data point at $T = 1600$ °C, as a result of data curation. MoNbTaW will result in the lowest slope ($C_3 = 2.75$), if this data point is included. This trend underscores the importance of considering dislocations, interactions between elements, volume or lattice misfit, and atomic mismatch²³, when designing materials with attractive high-temperature properties, in addition to the melting points of the constituent elements. Similarly, it is not surprising that the composition, AlMo_{0.5}NbTa_{0.5}TiZr, which additionally contains the weakly refractory elements, Ti and Zr, also results in a small slope ($C_3 = 4.95$). While AlMo_{0.5}NbTa_{0.5}TiZr does seem to offer relatively favorable high-temperature properties, it is worth noting that the estimation of its slope is only based on three data points.

In terms of the modeling accuracy, the bilinear log model yields the average MSE of 0.003 in the log domain, for composition No. 1–10 and 17–21 from Table 2, compared to 0.195 for the model with a single exponential. In the linear domain, this trend translates to MSE of 528 for the bilinear log model, compared to that of 122,588 for the single-exponential model. For the case of composition No. 18 from Table 2 (HfMoNbTiZr), Fig. 4 provides

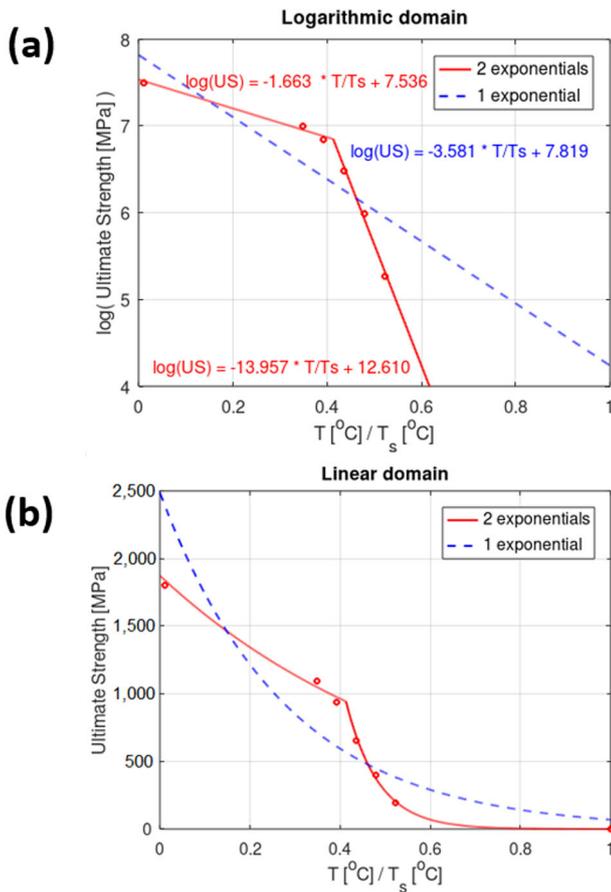


Fig. 4 Quantification of modeling accuracy of the bilinear log model, for the composition, No. 18, from Table 2 (HfMoNbTiZr), and comparison of the modeling accuracy to that of a model with a single exponential. Panel (a) compares the modeling accuracy of the bilinear log model to that of a single exponential in the logarithmic domain, whereas panel (b) presents corresponding comparison for the linear domain.

graphical insight as to why the bilinear log model yields a better match to the data available than a model consisting of a single exponential. Supplementary Figs. 9–23 provide similar diagrams for the other 20 alloy compositions from Table 2.

DISCUSSION AND CONCLUSION

For interpretation of the prediction results, we refer to Section 4.7 of¹⁷. To analyze consistency with experimental verification, we point to Supplementary Table 2, which captures the outcomes from applying the empirical rules of^{19,20} to the formation of the compositions predicted, $\text{Al}_{0.5}\text{Mo}_{0.5}\text{Nb}_{1.5}\text{Ta}_{0.5}\text{Zr}_{1.5}$, $\text{Mo}_{1.25}\text{Nb}_{1.25}\text{Ti}_{0.5}\text{V}_{0.75}\text{Zr}$, $\text{Mo}_{1.25}\text{Nb}_{1.25}\text{Ti}_{0.5}\text{V}_{0.5}\text{Zr}_{1.25}$, and MoNbZr . We expect $\text{Al}_{0.5}\text{Mo}_{0.5}\text{Nb}_{1.5}\text{Ta}_{0.5}\text{Zr}_{1.5}$ to be a stable composition with two types of phases (BCC1 + BCC2). Supplementary Table 2 suggests that the compositions have high chance of forming a solid-solution main phase with ordered solid-solution precipitates.

Compression tests were conducted on both the reference and predicted compositions in the as-cast condition. Figure 5 summarizes the engineering stress–strain curves of the predicted compositions, including $\text{Al}_{0.5}\text{Mo}_{0.5}\text{Nb}_{1.5}\text{Ta}_{0.5}\text{Zr}_{1.5}$ and $\text{Mo}_{1.25}\text{Nb}_{1.25}\text{Ti}_{0.5}\text{V}_{0.5}\text{Zr}_{1.25}$, in comparison to the respective references, $\text{Al}_{0.5}\text{Mo}_{0.5}\text{NbTa}_{0.5}\text{TiZr}$ and $\text{MoNbTiV}_{0.75}\text{Zr}$. The compression properties of these alloys, such as the YS, σ_y , maximum strength, σ_{max} , and fracture strain, ϵ_f , are listed in Table 1. We conclude from the experimental results in Fig. 5 that the candidate compositions,

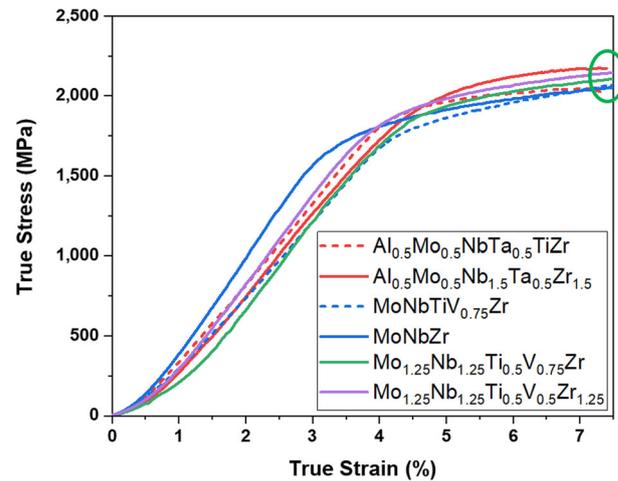


Fig. 5 Engineering stress–strain curves of the reference (dashed lines) and predicted (solid lines) compositions. For estimation of the error margins, refer to Table 1.

$\text{Al}_{0.5}\text{Mo}_{0.5}\text{Nb}_{1.5}\text{Ta}_{0.5}\text{Zr}_{1.5}$, $\text{Mo}_{1.25}\text{Nb}_{1.25}\text{Ti}_{0.5}\text{V}_{0.75}\text{Zr}$, and $\text{Mo}_{1.25}\text{Nb}_{1.25}\text{Ti}_{0.5}\text{V}_{0.5}\text{Zr}_{1.25}$, indeed exhibit higher strengths than the respective reference, hence confirming the outcome of our two sets of prediction.

Figure 6 shows the energy dispersive X-Ray spectroscopy (EDX) mapping for the predicted alloys. It can be noted that both of the predicted compositions feature typical dendrite-inter-dendrite microstructure, which indicates elemental segregation during solidification with high cooling rates. $\text{Al}_{0.5}\text{Mo}_{0.5}\text{Nb}_{1.5}\text{Ta}_{0.5}\text{Zr}_{1.5}$ exhibits segregation in all the five elements. The X-ray diffraction (XRD) results in Fig. 7 indicate that both of the predicted alloys contain two BCC phases. These results are consistent with known properties of BCC and FCC phases, in terms of the BCC phases usually helping improve material strength, but the FCC phases helping improve ductility.

In this study, we proposed a bilinear log model for predicting the US of HEAs across temperatures and evaluated its effectiveness for 21 compositions. We considered the break temperature, T_{break} , an important parameter for design of materials with attractive high-temperature properties, one warranting inclusion in alloy specifications. For reliable operation, the operating temperature for the corresponding alloys may need to stay below T_{break} . Previous models for temperature dependence of the YS only accounted for a single exponential. Hence, there was no break temperature.

We, further, suggested a general methodology for joint optimization, a methodology capable of accounting for physics-based dependencies, and presented the maximization of the US as an initial step toward the joint optimization of mechanical properties. We applied an optimization technique suitable for the problem under study, linear regression analysis, to a data set of modest size from the literature, to predict HEA compositions yielding the exceptional US at room temperature. For accurate prediction, we recommended picking an optimization technique appropriate for the application at hand and the data available and carefully accounting for the underlying sources of variations. Despite relatively limited data and a simple prediction algorithm¹⁷, we were able to attain the goal of successfully predicting HEA compositions, exhibiting superior strength, compared to previous work, and to demonstrate consistency of our prediction, both with empirical rules (in Table 16 of¹⁷) and with an experimental finding (per Fig. 5). In this way, we successfully addressed the research objective of predicting compositions of HEAs yielding the best strength, and conducting experimental verification of our findings. Next, one needs to account for the ductility. In case of the

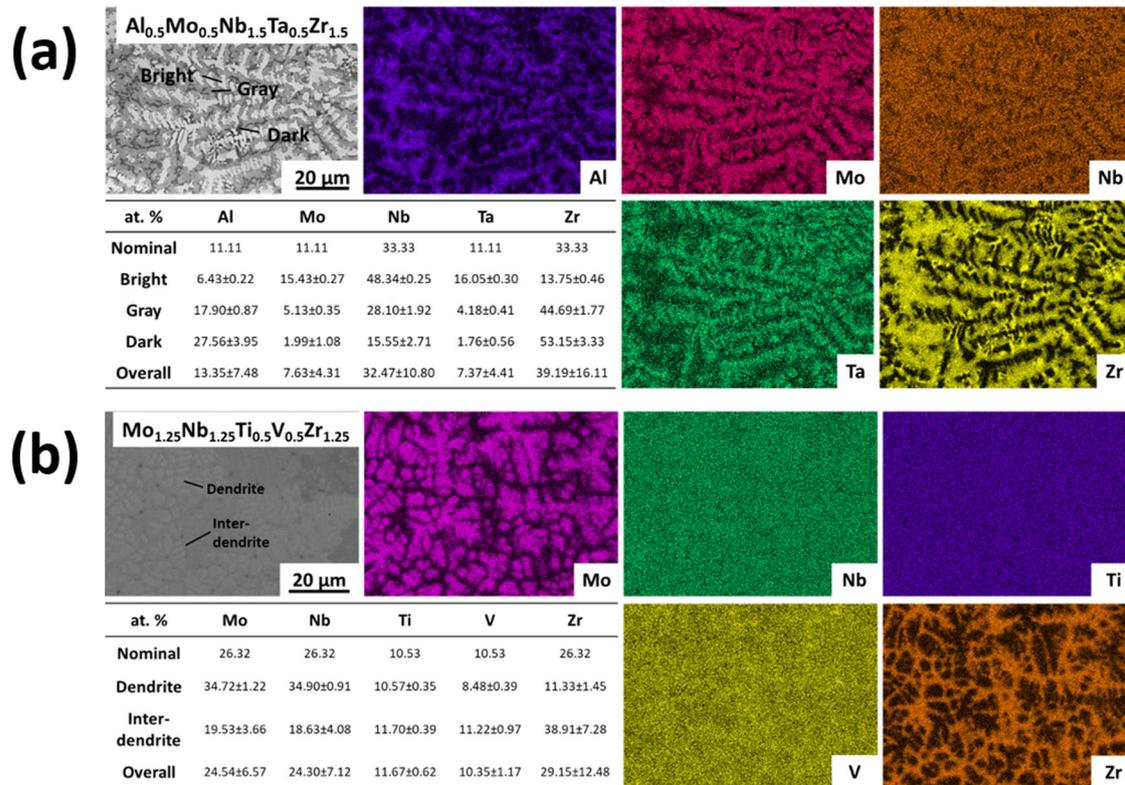


Fig. 6 EDX mapping and chemical compositions of the predicted compositions, $\text{Al}_{0.5}\text{Mo}_{0.5}\text{Nb}_{1.5}\text{Ta}_{0.5}\text{Zr}_{1.5}$ and $\text{Mo}_{1.25}\text{Nb}_{1.25}\text{Ti}_{0.5}\text{V}_{0.5}\text{Zr}_{1.25}$. For panel (a), the brighter regions are rich in Mo, Nb, and Ta, while the darker regions are Al and Zr enriched. The other predicted composition, $\text{Mo}_{1.25}\text{Nb}_{1.25}\text{Ti}_{0.5}\text{V}_{0.5}\text{Zr}_{1.25}$, in panel (b) exhibits a nearly homogenized distribution of Ti and V. The dendrite regions are rich in Mo and Nb, and the inter-dendrite regions are Zr enriched. The detailed elemental concentrations are also listed. A Zeiss EVO MA15 scanning electron microscope with a back-scattered electron detector and Bruker xFlash 6130 energy dispersive X-ray spectroscopy was used for microstructural and chemical composition analysis.

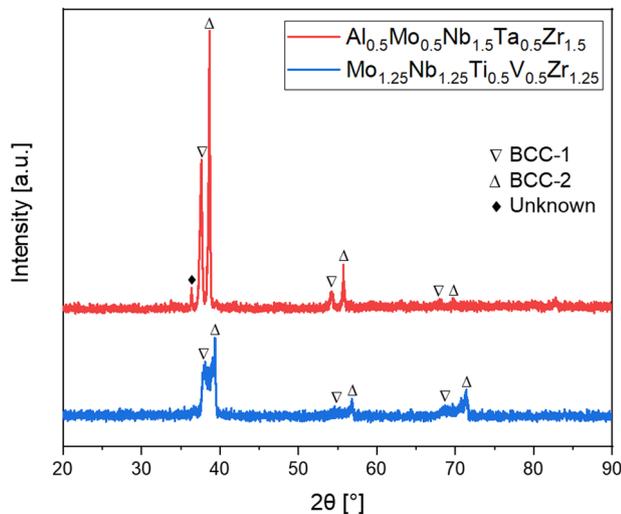


Fig. 7 XRD patterns for the predicted compositions. The two BCC phases could be related to the segregated microstructures observed in the EDX maps, which may contribute to the high strengths of the alloys due to the solid-solution strengthening effect and the second phase strengthening effects. For comparison and consistency, note that (according to Table 13 of¹⁷) the training set consisted of a combination of single-phase and multi-phase—mostly BCC compositions. A Panalytical Empyrean X-ray diffractometer, at Cu $K\alpha$ radiation, was used to identify the crystal structure of the alloys.

maximization of US and presence of a relatively small data set, we recommended multi-variate linear regression as the method of choice. In this case, the prediction rule is fairly general: One can extrapolate in a direction of the gradient from the data point in the training set exhibiting the highest US. As long as the step size is selected as sufficiently small (only aiming for 5–10% increase in US for a single step), the resulting prediction is considered much superior to a trial-and-error approach. Sequential learning¹⁷ is expected to greatly expedite the identification of alloys exhibiting given mechanical properties of interest.

METHODS

While the primary emphasis here is on the US, the optimization of the mechanical properties is assumed to take place within a framework for joint optimization. We essentially take the US to represent the ultimate compression strength, since Supplementary Table 1 contains 34 compression measurements, but only 1 tensile measurement. For further explanations, refer to Supplementary Table 1.

The spider chart in Fig. 8 provides a high-level depiction of the content of the database used for the present research. The database currently captures mechanical properties of some 218 US recordings for HEA configurations, many of which contain refractory elements, with 147 configurations measured at room temperature and 71 at elevated temperatures. The US was measured using tensile testing for 29 of the recordings. But the remaining 189 recordings were obtained through compression testing.

Figure 2 captures an abstracted model of physical dependencies for the prediction of the US. This model is an extension of the input sources modeled in Eq. (21). Capturing of the physical dependencies helps greatly in terms of the incorporation of a priori knowledge, derived from the

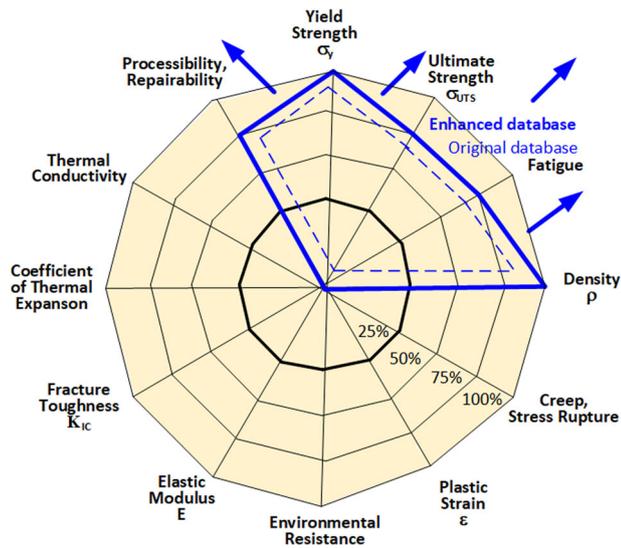


Fig. 8 Characterization of the HEA data contained in the database used for the present research.

underlying physics, and in terms of making the most of the—usually limited—input data available.

Our intent is to accurately capture the input sources that contribute to variations in the US observed (to variations in the output). In the present research, we model the input combination as

$$\text{Input} = (\text{composition}, T, \text{process}, \text{defects}, \text{grain size}, \text{microstructure}). \quad (21)$$

Here, “defects” are defined broadly such as to include inhomogeneities, impurities, dislocations, or unwanted features. “ T ” represents temperature. Similarly, the term “microstructure” broadly represents microstructures, at nano- or micro-scale, as well as phase properties. The term “process” broadly refers both to manufacturing processes and post-processing. Correspondingly, the term “grain size” refers to the distribution in grain sizes. Section 4.4 of¹⁷ allows for dependence between input sources, and Section 4.5 outlines the expected dependence of the US on the individual input sources listed. Dependence amongst the inputs is further addressed in Supplementary Note 2.

Methodology for maximization of the US

The overall methodology for predicting the US is presented in Supplementary Fig. 2. We summarize the prediction model as follows:

$$\text{US} = h[\text{composition}, T, \text{process}, \text{defects}(\text{process}, T), \text{grains}(\text{process}, T), \text{microstructure}(\text{process}, T)]. \quad (22)$$

If the US corresponding to a given input combination is known, one can simply look up the known value. If the US corresponding to a given input combination is not known, then a prediction step can be applied (e.g., interpolation or extrapolation).

The purpose of the data curation step in Supplementary Fig. 2 is to ensure that input data to the prediction step are of the highest quality possible¹⁷. Here, the intent is to look for outliers, suspected cases of discrepancy, or incorrect data (data that one may not fully trust). Generally, it is recommended to filter out data that have no relevance to the application domain or the task at hand¹⁷.

The step in our methodology for maximization of the predicted US, \bar{y} , assumes a generic model of the type

$$\bar{y} = h(\bar{\mathbf{x}}). \quad (23)$$

Here, the input vector, $\bar{\mathbf{x}}$, can be considered as the definition of a feature set comprising of parameters related to the compositions, temperature, heat-treatment process, defect property, grain size, microstructure (phase properties), manufacturing process, or post-processing, essentially all the source parameters that impact the output quantity of interest. The transformation, $h(\cdot)$, can be a non-linear function of the input, $\bar{\mathbf{x}}$. Artificial intelligence and supervised ML are presented as one of the alternatives for

deriving the system model¹⁷. For a parametrized description of the terms in Eq. (22), including composition, manufacturing processes, microstructures, and defects, and for support for the model in context with theory on ML, refer to Supplementary Notes 3–5.

Experimental validation

The alloys predicted were prepared by arc-melting a mixture of pure elements [purity >99.9 weight percent (wt.%)] in a Ti-gettered argon atmosphere. The ingots were flipped and remelted for at least five times to achieve homogenized elemental distributions. The ingots were cast into a water-cooled copper hearth and then cut into desired shapes for further experiments. Compression tests were performed on a computer-controlled uniaxial mechanical testing system with a servo hydraulic load frame at default strain rate of $1 \times 10^{-3} \text{ s}^{-1}$.

DATA AVAILABILITY

The data in this paper, including those in the Supplementary Figures, can be requested by contacting the corresponding authors (baldur@imagars.com or drzhang@ustb.edu.cn).

CODE AVAILABILITY

Matlab comprises the software package primarily used for this study. Supplementary Figs. 7 and 8 contain Matlab source code for the objective functions optimized in case of the bilinear or the trilinear log model, respectively.

Received: 27 February 2021; Accepted: 5 August 2021;

Published online: 24 September 2021

REFERENCES

- Miracle, D. B. & Senkov, O. N. A critical review of high entropy alloys and related concepts. *Acta Mater.* **122**, 448–511 (2017).
- Yeh, J.-W. et al. Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. *Adv. Eng. Mater.* **6**, 299–303 (2004).
- Cantor, B., Chang, I., Knight, P. & Vincent, A. Microstructural development in equiatomic multicomponent alloys. *Mater. Sci. Eng.: A* **375**, 213–218 (2004).
- Gao, M. C., Yeh, J.-W., Liaw, P. K. & Zhang, Y. *High Entropy Alloys – Fundamentals and Applications* (Springer, 2016).
- Zhang, Y. et al. Microstructures and properties of high-entropy alloys. *Prog. Mater. Sci.* **61**, 1–93 (2014).
- Senkov, O. N., Wilks, G. B., Miracle, D. B., Chuang, C. P. & Liaw, P. K. Refractory high-entropy alloys. *Intermetallics* **18**, 1758–1765 (2010).
- George, E. P., Raabe, D. & Ritchie, R. O. High-entropy alloys. *Nat. Rev. Mater.* **4**, 515–534 (2019).
- Hemphill, M. A. et al. Fatigue behavior of Al 0.5 CoCrCuFeNi high entropy alloys. *Acta Mater.* **60**, 5723–5734 (2012).
- Tang, Z. et al. Fatigue behavior of a wrought Al 0.5 CoCrCuFeNi two-phase high-entropy alloy. *Acta Mater.* **99**, 247–258 (2015).
- Shukla, S., Wang, T., Cotton, S. & Mishra, R. S. Hierarchical microstructure for improved fatigue properties in a eutectic high entropy alloy. *Scr. Mater.* **156**, 105–109 (2018).
- Liu, K., Nene, S. S., Frank, M., Sinha, S. & Mishra, R. S. Metastability-assisted fatigue behavior in a friction stir processed dual-phase high entropy alloy. *Mater. Res. Lett.* **6**, 613–619 (2018).
- Liu, K., Nene, S. S., Frank, M., Sinha, S. & Mishra, R. S. Extremely high fatigue resistance in an ultrafine grained high entropy alloy. *Appl. Mater. Today* **15**, 525–530 (2019).
- Suzuki, K., Koyama, M. & Noguchi, H. Small fatigue crack growth in a high entropy alloy. *Procedia Struct. Integr.* **13**, 1065–1070 (2018).
- Kim, Y.-K., Ham, G.-S., Kim, H. S. & Lee, K.-A. High-cycle fatigue and tensile deformation behaviors of coarse-grained equiatomic CoCrFeMnNi high entropy alloy and unexpected hardening behavior during cyclic loading. *Intermetallics* **111**, <https://doi.org/10.1016/j.intermet.2019.106486> (2019).
- Kashaev, N. et al. Fatigue behaviour of a laser beam welded CoCrFeNiMn-type high entropy alloy. *Mater. Sci. Eng.: A* **766**, <https://doi.org/10.1016/j.msea.2019.138358> (2019).
- Guenec, B. et al. Four-point bending fatigue behavior of an equimolar BCC HfNbTaTiZr high-entropy alloy: macroscopic and microscopic viewpoints. *Materialia* **4**, 348–360 (2018).

17. Steingrímsson, B., Fan, X., Kulkarni, A., Gao, M. C. & Liaw, P. K. Machine learning and data analytics for design and manufacturing of high-entropy materials exhibiting mechanical or fatigue properties of interest, (eds Liaw, P. K. & Brechtel, J.) Chapter 4 In *Fundamental Studies in High-Entropy Materials* (Springer, 2021).
18. Agrawal, A. et al. Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters. *Integr. Mater. Manuf. Innov.* **3**, 8 (2014).
19. Zhang, Y., Zhou, Y. J., Lin, J. P., Chen, G. L. & Liaw, P. K. Solid-solution phase formation rules for multi-component alloys. *Adv. Eng. Mater.* **10**, 534–538 (2008).
20. Feng, R. et al. Design of light-weight high-entropy alloys. *Entropy* **18**, 333 (2016).
21. Senkov, O. N., Gorsse, S. & Miracle, D. B. High temperature strength of refractory complex concentrated alloys. *Acta Mater.* **175**, 394–405 (2019).
22. Wu, Z., Bei, H., Pharr, G. M. & George, E. P. Temperature dependence of the mechanical properties of equiatomic solid solution alloys with face-centered cubic crystal structures. *Acta Mater.* **81**, 428–441 (2014).
23. Maresca, F. & Curtin, W. A. Mechanistic origin of high strength in refractory BCC high entropy alloys up to 1900K. *Acta Mater.* **182**, 235–249 (2020).
24. MacKay, D. J. Bayesian methods for neural networks: theory and applications, Neural Networks Summer School, University of Cambridge Programme for Industry <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.47.6409> (1995).
25. Zhang, Y., Yang, X. & Liaw, P. K. Alloy design and properties optimization of high-entropy alloys. *Jom* **64**, 830–838 (2012).
26. Ben-Israel, A. & Greville, T. N. *Generalized Inverses: Theory and Applications*, Vol. 15 (Springer Science & Business Media, 2003).
27. Le, T. T., Fu, W. & Moore, J. H. Scaling tree-based automated machine learning to biomedical big data with a feature set selector. *Bioinformatics* **36**, 250–256 (2020).
28. Rao, S. I. et al. Solution hardening in body-centered cubic quaternary alloys interpreted using Suzuki's kink-solute interaction model. *Scr. Mater.* **165**, 103–106 (2019).
29. Rao, S. I. et al. Modeling solution hardening in BCC refractory complex concentrated alloys: NbTiZr, Nb1.5TiZr0.5 and Nb0.5TiZr1.5. *Acta Mater.* **168**, 222–236 (2019).
30. Caillard, D. & Martin, J.-L. *Thermally Activated Mechanisms in Crystal Plasticity*. (Elsevier, 2003).
31. Otto, F. et al. The influences of temperature and microstructure on the tensile properties of a CoCrFeMnNi high-entropy alloy. *Acta Mater.* **61**, 5743–5755 (2013).
32. Diao, H. Y., Feng, R., Dahmen, K. A. & Liaw, P. K. Fundamental deformation behavior in high-entropy alloys: an overview. *Curr. Opin. Solid St. Mater. Sci.* **21**, 252–266 (2017).
33. Wilson, J. General behaviour of refractory metals. *Behavior and Properties of Refractory*. (Stanford University Press, 1965).
34. Steingrímsson, B. A., Liaw, P. K., Fan, X. & Kulkarni, A. A. (United States Patent Application Publication, 2020).
35. Miracle, D. B. et al. *ASM Handbook*, Vol. 21 (ASM international Materials Park, OH, 2001).

ACKNOWLEDGEMENTS

X.F. and P.K.L. very much appreciate the support of the U.S. Army Research Office Project (W911NF-13-1-0438 and W911NF-19-2-0049) with the program managers, Drs M.P. Bakas, S.N. Mathaudhu, and D.M. Stepp, as well as the support from the Bunch Fellowship. XF and PKL also would like to acknowledge funding from the State of Tennessee and Tennessee Higher Education Commission (THEC) through their support of the Center for Materials Processing (CMP). P.K.L., furthermore, thanks the

support from the National Science Foundation (DMR-1611180 and 1809640) with the program directors, Drs J. Yang, G. Shiflet, and D. Farkas. B.S. very much appreciates the support from the National Science Foundation (IIP-1447395 and IIP-1632408), with the program directors, Dr G. Larsen and R. Mehta, from the U.S. Air Force (FA864921P0754), with J. Evans as the program manager, and from the U.S. Navy (N6833521C0420), with Drs D. Shifler and J. Wolk as the program managers. M.C.G. acknowledges the support of the US Department of Energy's Fossil Energy Crosscutting Technology Research Program. The authors also want to thank Dr. G. Tewksbury for bringing to their attention suspicious recordings of the US from the literature, which have prompted the data curation effort.

AUTHOR CONTRIBUTIONS

B.S., X.F., and P.K.L. conceived the project. B.S. performed the ML predictions, but consulted with M.C.G. in the process. X.F. prepared the database and conducted experimental verification of the ML predictions. All authors edited and proofread the final manuscript and participated in discussions.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41524-021-00623-4>.

Correspondence and requests for materials should be addressed to B. Steingrímsson or Y. Zhang.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021