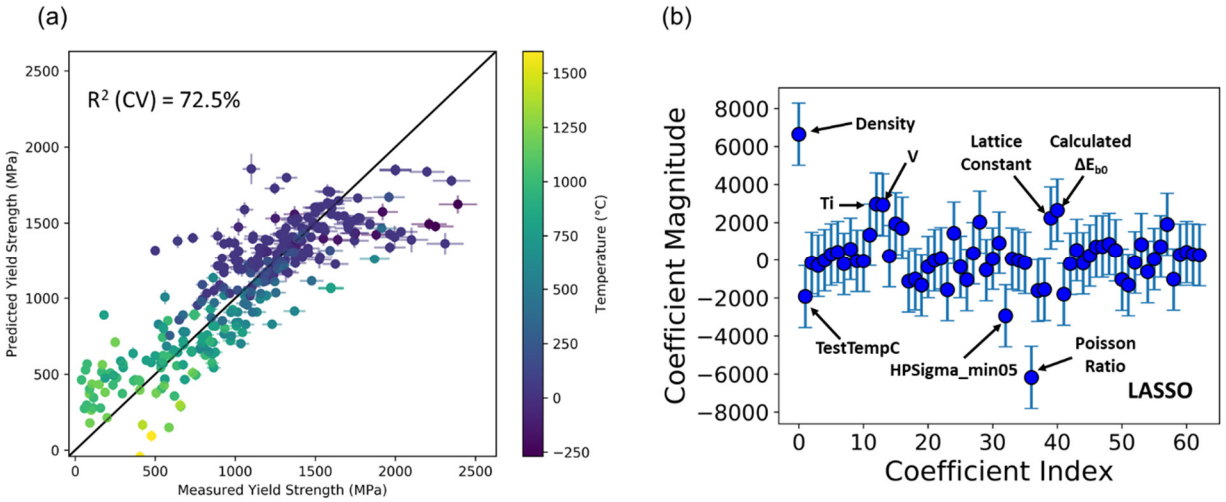
Machine-Learning-Based Intelligent Framework for Discovering High-Entropy Alloys with Improved High-Temperature Yield Strength

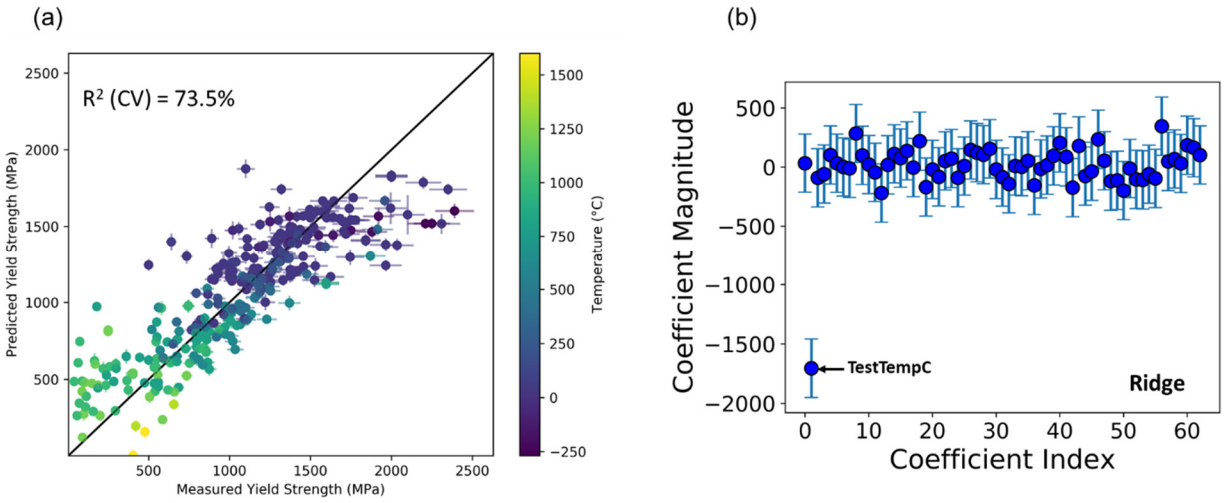
Stephen A. Giles[1], Debasis Sengupta[1], Scott R. Broderick[2], Krishna Rajan[2]

[1] CFD Research Corporation, 6820 Moquin Dr. NW, Huntsville 35806
[2] Department of Material Design and Innovation, University at Buffalo, 120 Bonner Hall, NY-14260

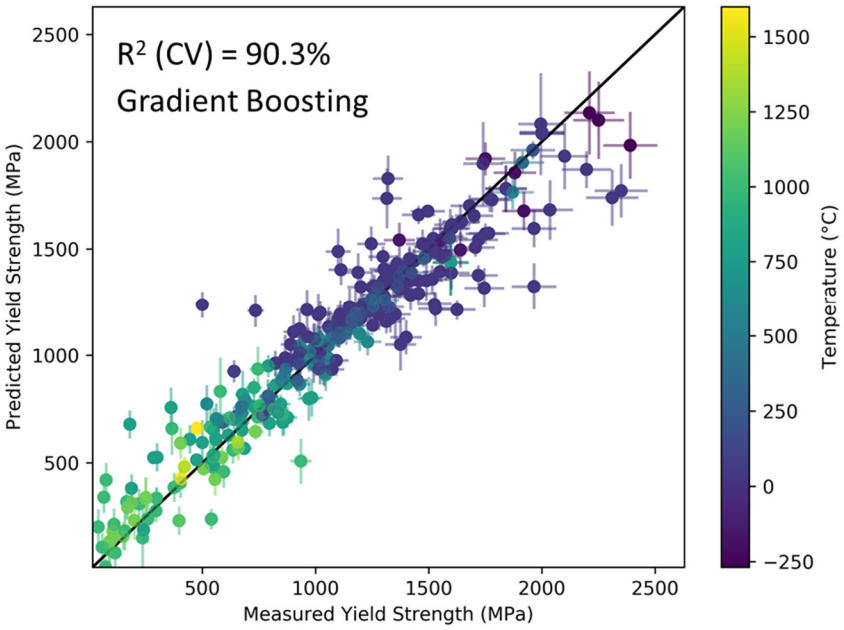
Supplementary Figures



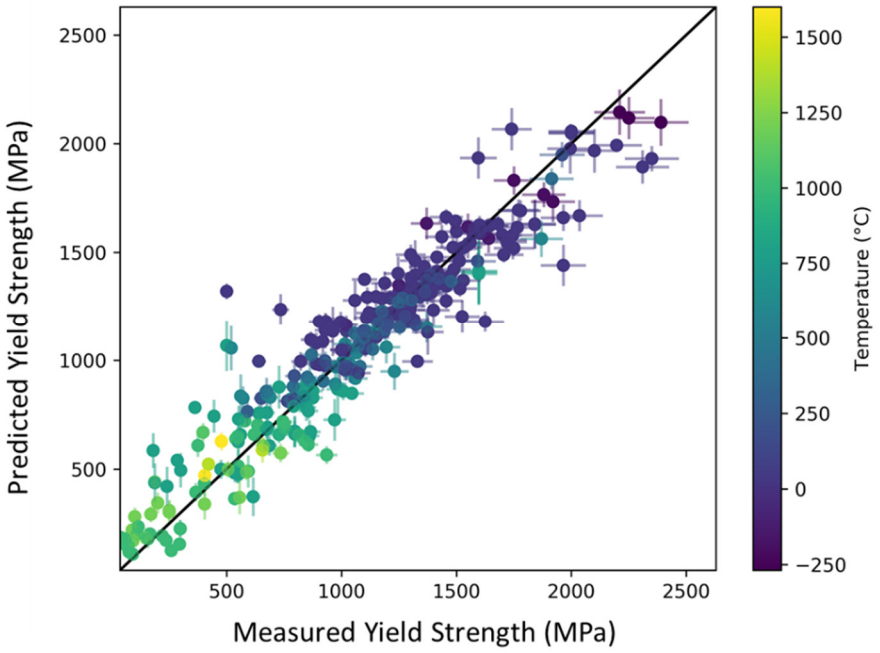
Supplementary Figure 1. LASSO regression: (a) parity plot comparison to experimental values; (b) descriptor coefficients following L1 regularization. Most important descriptors have been annotated.



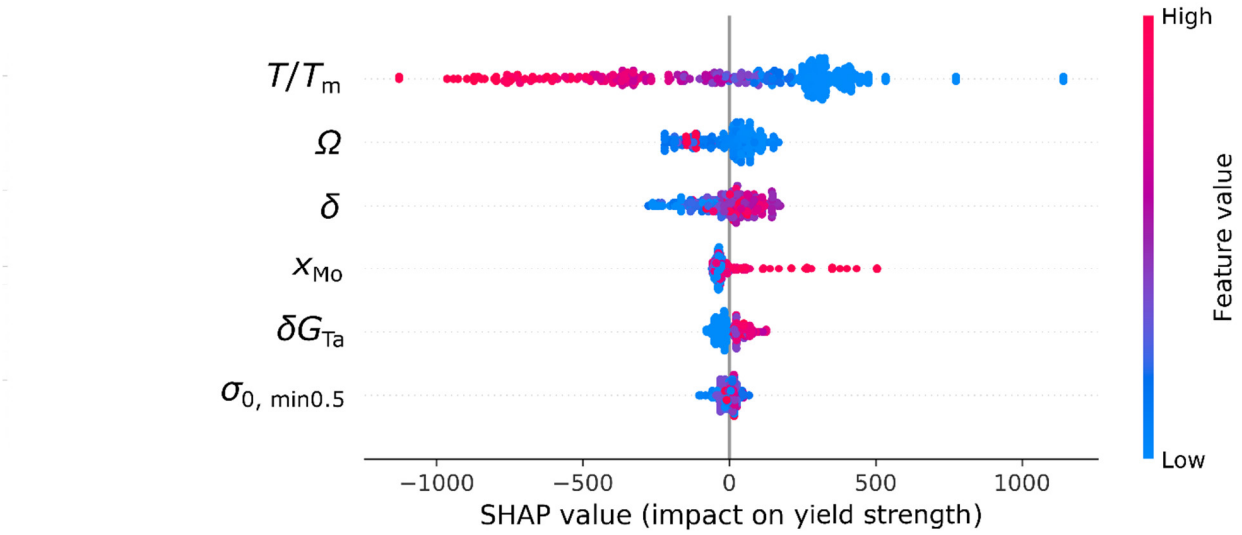
Supplementary Figure 2. Ridge regression: (a) parity plot comparison to experimental values; (b) descriptor coefficients following L2 regularization. The test temperature is the most important descriptor by far.



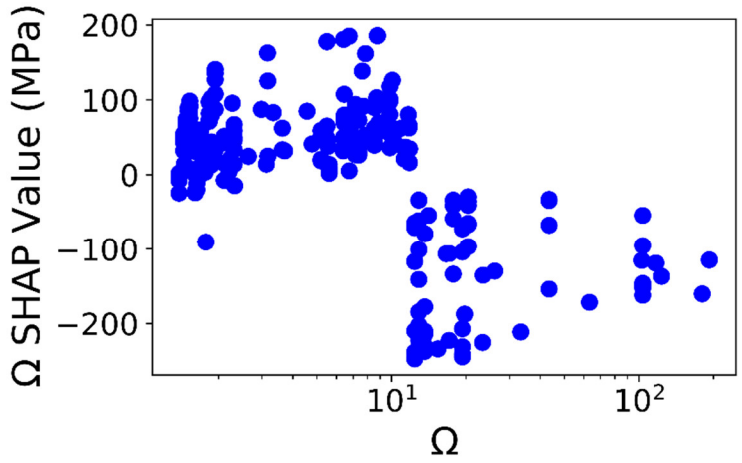
Supplementary Figure 3. Parity plot comparison for the gradient boosting model. The R^2 (CV) is approximately equivalent to what was obtained for the random forest model.



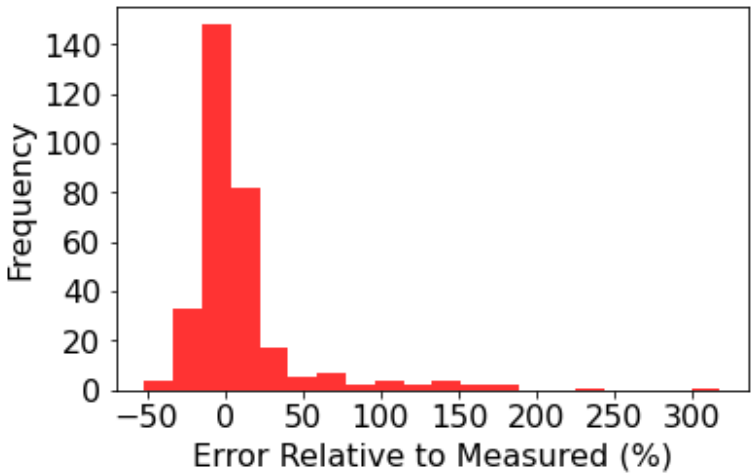
Supplementary Figure 4. Predicted vs. experimental yield strength for compression data when the test temperature T is replaced by the homologous temperature, T/T_m. R^2 (CV) = 88.8%. Note that we used a repeated k-fold method where 5-fold cross validation was performed 1000 times to compute the mean and standard deviation of each data point;



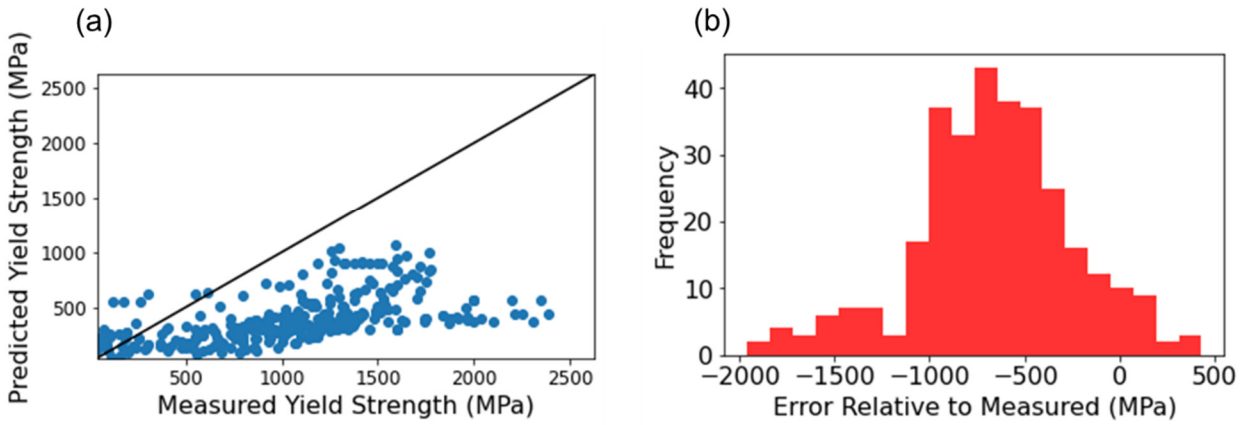
Supplementary Figure 5. SHAP analysis is performed to identify descriptor importance for model with T/T_m descriptor in place of T. Features on the left are ordered according to their importance, with the most important feature (temperature) shown on top. The data points correspond to the individual alloy data points, where each have been colored according to the magnitude (high or low) of the feature in question. Positive SHAP values indicate that the yield strength is increased as a result of the feature value, whereas negative SHAP values indicate the yield strength is decreased due to the feature value.



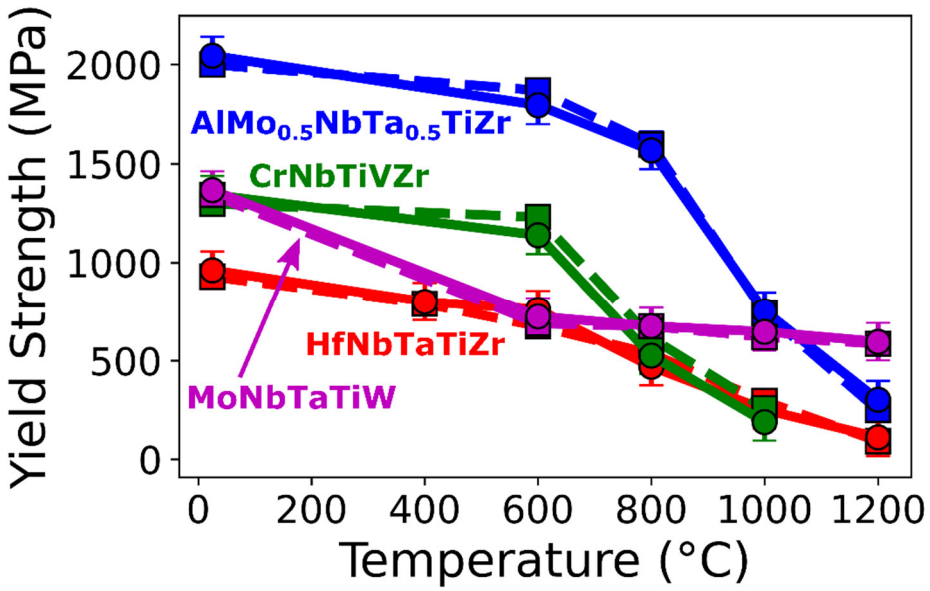
Supplementary Figure 6. Dependence of the SHAP value of Ω on the value of Ω.



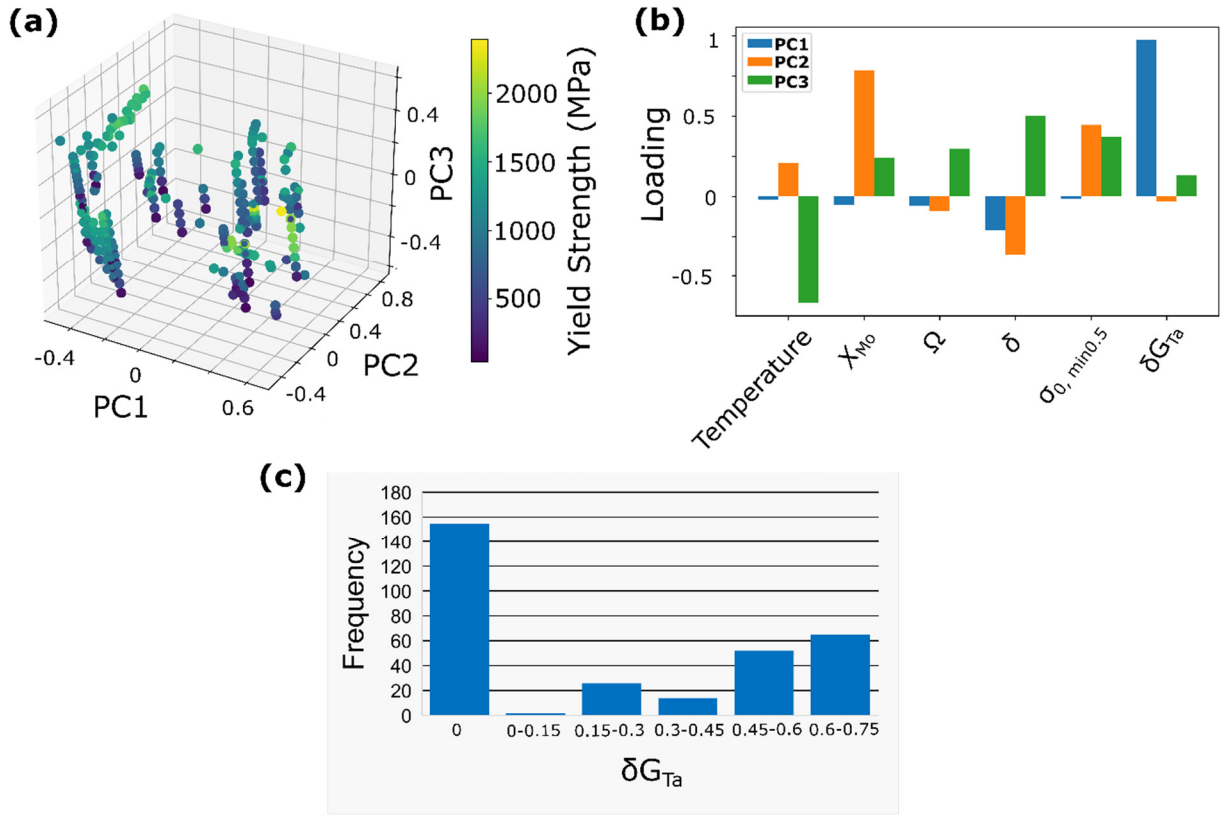
Supplementary Figure 7. Percentage error distribution of the model presented in this work relative to the measured values. MAE = 20.1%.



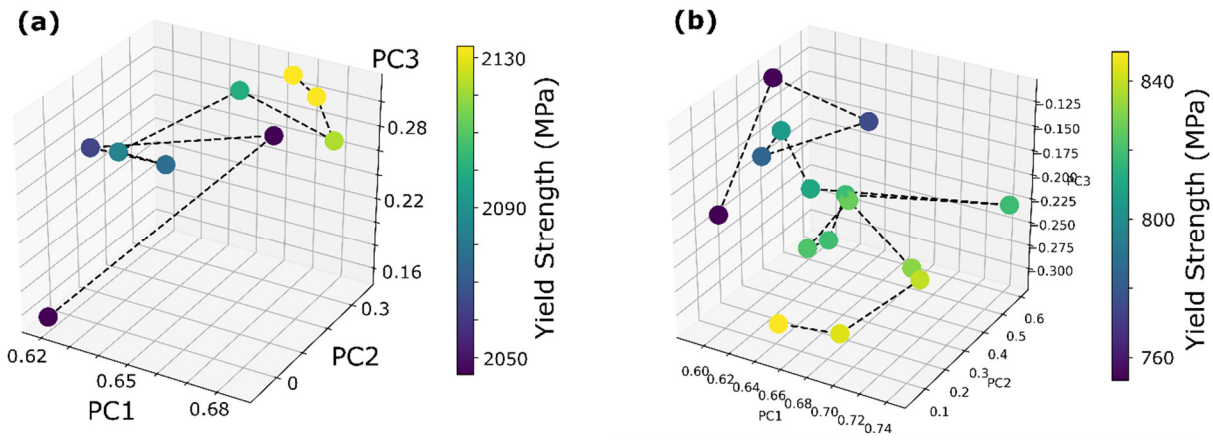
Supplementary Figure 8. Predictions obtained using the analytical temperature-dependent yield strength model proposed by Varvenne and Curtin. (a) parity plot of measured values and predicted values from the Varvenne and Curtin model. (b) distribution of error relative to measured values for the Varvenne and Curtin model. MAE = 683 MPa.



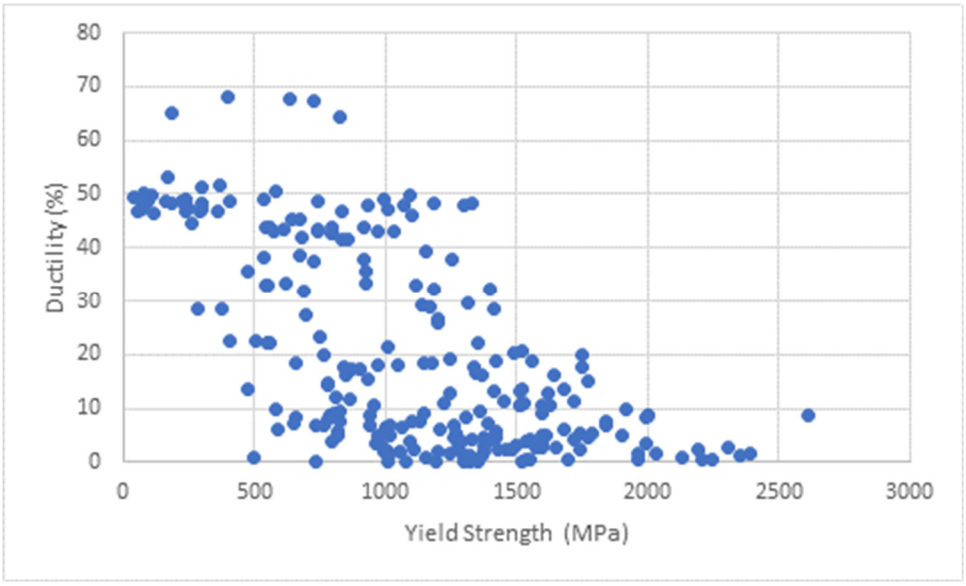
Supplementary Figure 9: Comparison of experimental (dashed lines, square symbols) and ML model predictions (solid lines, round symbols) of yield strength for four different RHEAs at various temperatures.



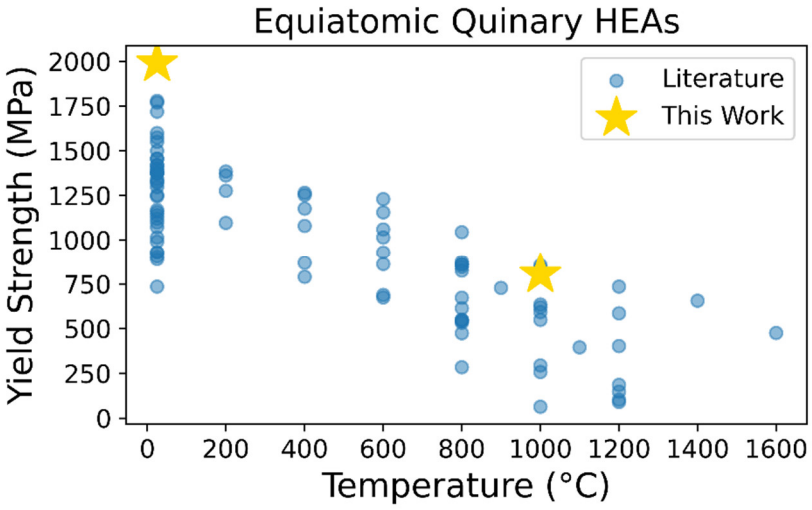
Supplementary Figure 10: (a) Three-dimensional score plot of compression yield strength data according to their first, second, and third principal component values; (b) Loadings of the six original descriptors on the first three principal components; (c) Histogram of tantalum modulus distortion values. The 155 alloys with δG_{Ta} equal to zero correspond to non-tantalum-containing alloys.



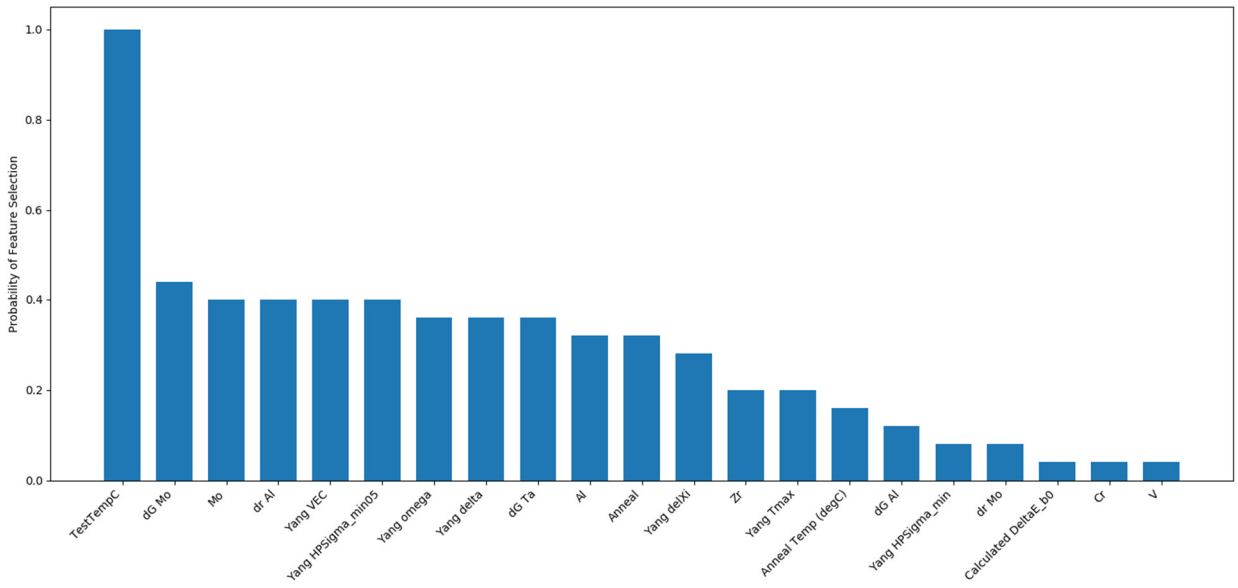
Supplementary Figure 11: Optimization trajectories in principal component space for the (a) 25 °C optimization and the (b) 1,000 °C optimization of the $AlMo_{0.5}NbTa_{0.5}TiZr$ base alloy.



Supplementary Figure 12. Experimental values of yield strength and ductility at different temperatures from the Couzinie et al. dataset[8].



Supplementary Figure 13. Comparison of temperature-dependent yield strength of equiatomic quinary HEAs reported in the literature to the two equiatomic HEAs (AlMoTaTiZr at 25 °C, and AlMoTaTiHf at 1000 °C) predicted by our model.



Supplementary Figure 14. Histogram of feature selection probability for a six-feature random forest model.

Supplementary Notes

Dependence of Yield Strength on Ω from SHAP Analysis

As was noted in the discussion alongside Figure 2 in regards to the SHAP explainability analysis that was performed for the forward model, Ω has a complex, but important effect on the predicted yield strength. As shown in Supplementary Figure 6, small values of Ω result in positive SHAP values (i.e., an increase in the predicted yield strength), whereas large values of Ω result in a negative impact on the yield strength. Notably, this shift in the effect of Ω undergoes a transition across a small regime. It is theorized that this could result from the interplay of the entropy and enthalpy of mixing. Large values of Ω result when the enthalpy of mixing approaches zero, which may be detrimental to improving yield strength.

Examination of Temperature-Dependence of Yield Strength for Training Data

Supplementary Figure 9 shows the temperature dependent prediction of four different RHEAs along with their experimental yield strength. Readily apparent from Supplementary Figure 9 is an excellent agreement between predicted and experimental yield strengths over the entire temperature range. Two of these RHEAs have garnered some attention recently. The $AlMo_{0.5}NbTa_{0.5}TiZr$ HEA is of interest due to its exceptional high-temperature yield strength (745 MPa at 1,000 °C)[1–4]. It has been reported in multiple studies, and is among the state-of-the-art RHEAs. The HfNbTaTiZr alloy is also multiply attested, and is of interest due to its high ductility

under tension[5–7]. However, HfNbTaTiZr has a relatively poor yield strength (929 MPa at 25 °C). In the following section, $AlMo_{0.5}NbTa_{0.5}TiZr$ and HfNbTaTiZr were used as the base alloys for subsequent improvement of the yield strength through compositional optimization. Finally, CrNbTiVZr and MoNbTaTiW were chosen to complete the remaining elements (i.e., Cr, V, and W) that are covered in the Couzinie dataset[8].

Interpretation Using Principal Component Analysis

As mentioned above, the yield strength model provided here employs six descriptors: test temperature, Ω, atomic size mismatch (δ), modulus distortion of tantalum (δG_{Ta}), molybdenum atomic fraction, x_{Mo}, and a base strength quantity, $\sigma_{0,\ min0.5}$. Given that the input to the model has been limited to only these six descriptors, an unsupervised approach such as principal component analysis (PCA) can be used to further reduce the dimensionality and allow for clustering of the data in the descriptor space to be visualized.

Analysis of the corresponding loading coefficients of the original six descriptors on the three principal components reveals the origin of the clustering observed in Supplementary Figure 10a. The loadings are provided in Supplementary Figure 10b. The loadings reveal that the first principal component is primarily composed of, and positively correlated to, the tantalum modulus distortion (δG_{Ta}). As a result, the clustering observed in Supplementary Figure 10 (a) is in fact distinguishing between Ta and non-Ta HEAs. The Couzinie data[8] is approximately balanced with regards to Ta alloys, with 159 alloys (50.6%) containing Ta, and 154 alloys (49.4%) not containing Ta. The data clustering in Supplementary Figure 10a based on Ta content is further evident upon analyzing the effect of Ta on modulus distortion. The Ta modulus distortion values for the RHEA dataset is provided in Supplementary Figure 10(c). As can be seen, in the 155 alloys which do not contain

Ta, δG_{Ta} is equal to zero. When Ta is present in the alloy, however, δG_{Ta} always has a positive value. This is due to Ta having a high elemental shear modulus (69 GPa) relative to most of the other elements of interest. From a physical point of view, a positive modulus distortion implies that the presence of Ta usually leads to an increase in the shear modulus of the alloy. Another conclusion which can be drawn from the loading plot is that an increase in test temperature is primarily responsible for decreasing PC3 which has negative effects on the yield strength (Supplementary Figure 10(a)).

To analyze the optimization progress, we have performed principal component analysis (PCA) to reduce dimensionality such that the descriptor values for a given alloy composition can be visualized in the three dimensions. We first performed PCA on the original Couzinie *et al.* dataset[8] that was used to train and validate the model (Supplementary Figure 7). Using the same PCA transformation, Supplementary Figure 11 shows the values of the principal components for both the 25 °C and the 1,000 °C optimal alloy along the optimization path. Considering first the 25 °C optimal alloy in Supplementary Figure 11a, the optimization proceeds in the direction of increasing yield strength, ultimately optimizing primarily through an increase in the value of the second and third principal component. Evident from the PCA loading plot in Supplementary Figure 10 is that the second principal component (PC2) is primarily composed of positive loading of x_{Mo}, whereas the third principal component (PC3) mainly of positive loadings of x_{Mo}, δ, and δG_{Ta}. Optimization occurring through an increase in PC2 and PC3 is, therefore, consistent with the observation made in the Manuscript for Supplementary Figure 10 that an increase in x_{Mo} is frequently correlated to an improvement of yield strength. For the optimization trajectory performed at 1,000 °C in Supplementary Figure 11b, the initial and final point are noticeably closer together in principal component space, indicating that the base alloy and optimized alloy are

relatively similar materials. Nevertheless, the optimal alloy has a noticeably smaller value of PC3, which corresponds in an observed decrease of δ. We also want to emphasize that the protocol presented here does not sample the composition space randomly to find the composition with a maximized yield strength, but instead maximizes the yield strength through an intelligent navigation of the composition space informed by previous iterations.

Example Calculation of Yield Strength Using Varvenne and Curtin Model

We have utilized the temperature-dependent yield strength model formulated by Varvenne and Curtin[9] to augment the experimentally available HEA dataset at temperatures between 25 °C and 600 °C. The accuracy of ML approaches is always fundamentally limited by the availability of adequate data, and HEA mechanical properties at temperatures between 25 °C and 600 °C are typically not reported. Using the experimental yield strengths at 25 °C and 600 °C and the temperature-dependent yield strength model proposed by Varvenne and Curtin[9], the yield strength for 200 °C and 400 °C were computed. An example calculation is provided below for the $AlMo_{0.5}NbTa_{0.5}TiZr$ HEA, which has a yield strength of 2000 MPa at 25 °C and 1870 MPa at 600 °C.

The thermally-activated finite-temperature yield strength is given by Varvenne and Curtin as,

$$\tau_y(T,\dot{\varepsilon}) = \tau_{y0}\exp\left(-\frac{1}{0.51}\frac{kT}{\Delta E_{b0}}\ln\frac{\dot{\varepsilon}_0}{\dot{\varepsilon}}\right)$$

, where τ_{y0} is the zero-temperature flow stress, k is the Boltzmann constant, T is the temperature, ΔE_{b0} is the dislocation energy barrier, $\dot{\varepsilon}$ is the strain rate (a value that is always reported experimentally when generating the stress-strain curve), and $\dot{\varepsilon}_0$ is the reference strain rate (taken as 10^4 s^{-1}). Let T_1 = 25 °C = 298.15 K and T_2 = 600 °C = 873.15 K, where $\tau_y(T_1,\dot{\varepsilon})$ and $\tau_y(T_2,\dot{\varepsilon})$

are known experimentally. Then, τ_{y0} cancels when calculating the ratio of the two known yield strengths:

$$\frac{\tau_y(T_1,\dot{\varepsilon})}{\tau_y(T_2,\dot{\varepsilon})} = \frac{\exp\left(-\frac{1}{0.51}\frac{kT_1}{\Delta E_{b0}}\ln\frac{\dot{\varepsilon}_0}{\dot{\varepsilon}}\right)}{\exp\left(-\frac{1}{0.51}\frac{kT_2}{\Delta E_{b0}}\ln\frac{\dot{\varepsilon}_0}{\dot{\varepsilon}}\right)}$$

Thus, the only unknown is the zero-temperature energy barrier, ΔE_{b0}. Taking the natural logarithm of both sides and solving for ΔE_{b0} yields:

$$\Delta E_{b0} = \frac{1}{\ln\left(\frac{\tau_y(T_1,\dot{\varepsilon})}{\tau_y(T_2,\dot{\varepsilon})}\right)}\left(-\frac{1}{0.51}k(T_1 - T_2)\ln\frac{\dot{\varepsilon}_0}{\dot{\varepsilon}}\right)$$

The yield strength measurements for the $AlMo_{0.5}NbTa_{0.5}TiZr$ HEA were performed at a strain rate of 10^{-3} s^{-1}. Therefore, ΔE_{b0} is calculated from the experimental data as:

$$\Delta E_{b0} = \frac{1}{\ln\left(\frac{2000\text{ MPa}}{1870\text{ MPa}}\right)}\left(-\frac{1}{0.51}(8.6713\times10^{-5}\text{ eV K}^{-1})(298.15 - 873.15\text{ K})\ln\frac{10^4\text{ s}^{-1}}{10^{-3}\text{ s}^{-1}}\right)$$

$$\Delta E_{b0} = 21.74\text{ eV} = 2085\text{ kJ mol}^{-1}$$

The value of ΔE_{b0} calculated from the experimental yield strength values can then be used to calculate the yield strength at a new temperature, T. Calculating for T = 400 °C:

$$\tau_y(T,\dot{\varepsilon}) = \tau_y(T_1,\dot{\varepsilon})\left(-\frac{1}{0.51}\frac{k(T - T_1)}{\Delta E_{b0}}\ln\frac{\dot{\varepsilon}_0}{\dot{\varepsilon}}\right)$$

$$\tau_y(T,\dot{\varepsilon}) = 2000\text{ MPa}\left(-\frac{1}{0.51}\frac{(8.6713\times10^{-5}\text{ eV K}^{-1})(673.15\text{ K} - 873.15\text{ K})}{21.74\text{ eV}}\ln\frac{10^4\text{ s}^{-1}}{10^{-3}\text{ s}^{-1}}\right)$$

$$\tau_y(T,\dot{\varepsilon}) = 1959.5\text{ MPa}$$

Repetition of Feature Selection Process

Sequential feature selection (SFS)[10,11] was performed 25 times to assess the variability of which features are included in the model. The result of the SFS replication is provided in Supplementary Figure 14. The test temperature is selected as one of the six features in every model. Yield strength has a well-known inverse relationship with the temperature as materials generally weaken with increasing temperature. The relationship of yield strength with the other composition-based descriptors are generally less understood. From the SFS replication, the most frequently selected descriptors, in addition to test temperature, were: dG Mo (molybdenum modulus distortion), Mo (element fraction of molybdenum), dr Al (lattice distortion of aluminum), Yang VEC (valence electron concentration), Yang HPsigma_min05 (half the average base strength, plus half the minimum elemental base strength), Yang omega (defined in Equation 4), Yang delta (defined in Equation 1), dG Ta (modulus distortion of tantalum), Al (aluminum element fraction), Anneal (0/1 hot-encoded variable indicated whether the synthesized alloy was annealed following casting), and Yang delXi (the standard deviation of the electronegativity).

Supplementary References

1. Senkov, O. N., Isheim, D., Seidman, D. N. & Pilchak, A. L. Development of a refractory high entropy superalloy. *Entropy* **18**, (2016).

2. Senkov, O. N., Jensen, J. K., Pilchak, A. L., Miracle, D. B. & Fraser, H. L. Compositional variation effects on the microstructure and properties of a refractory high-entropy superalloy AlMo0.5NbTa0.5TiZr. *Mater. Des.* **139**, 498–511 (2018).

3. Senkov, O. N., Woodward, C. & Miracle, D. B. Microstructure and Properties of

Aluminum-Containing Refractory High-Entropy Alloys. *JOM* **66**, 2030–2042 (2014).

4. Senkov, O. N., Senkova, S. V. & Woodward, C. Effect of aluminum on the microstructure and properties of two refractory high-entropy alloys. *Acta Mater.* **68**, 214–228 (2014).

5. Senkov, O. N., Scott, J. M., Senkova, S. V., Miracle, D. B. & Woodward, C. F. Microstructure and room temperature properties of a high-entropy TaNbHfZrTi alloy. *J. Alloys Compd.* **509**, 6043–6048 (2011).

6. Senkov, O. N. *et al.* Microstructure and elevated temperature properties of a refractory TaNbHfZrTi alloy. *J. Mater. Sci.* **47**, 4062–4074 (2012).

7. Juan, C. C. *et al.* Simultaneously increasing the strength and ductility of a refractory high-entropy alloy via grain refining. *Mater. Lett.* **184**, 200–203 (2016).

8. Couzinié, J. P., Senkov, O. N., Miracle, D. B. & Dirras, G. Comprehensive data compilation on the mechanical properties of refractory high-entropy alloys. *Data Br.* **21**, 1622–1641 (2018).

9. Varvenne, C., Luque, A. & Curtin, W. A. Theory of strengthening in fcc high entropy alloys. *Acta Mater.* **118**, 164–176 (2016).

10. Pudil, P., Novovičová, J. & Kittler, J. Floating search methods in feature selection. *Pattern Recognit. Lett.* **15**, 1119–1125 (1994).

11. Ferri, F. J., Pudil, P., Hatef, M. & Kittler, J. Comparative study of techniques for large-scale feature selection. *Pattern Recognit. Pract. IV* **16**, 403–413 (1994).