Supplementary Materials

Prediction of temperature-dependent yield strength of refractory high entropy alloy based on stacking integrated framework

Liping Yu[1], Jiarui Zhai[2], Wenzhe Cao[2], Jingli Ren[2,*]

[1]School of Artificial Intelligence and Big Data, Henan University of Technology, Zhengzhou 450001, Henan, China.
[2]School of Mathematics and Statistics, Zhengzhou University, Zhengzhou 450001, Henan, China.

[*]**Correspondence to:** Prof. Jingli Ren, School of Mathematics and Statistics, Zhengzhou University, No.100, Science Avenue, Hi-Tech Zone, Zhengzhou 450001, Henan, China. E-mail: renjl@zzu.edu.cn

Supplementary Table 1. Raw features and their calculation formulas, where c_i in the formula represents the molar ratio of each element

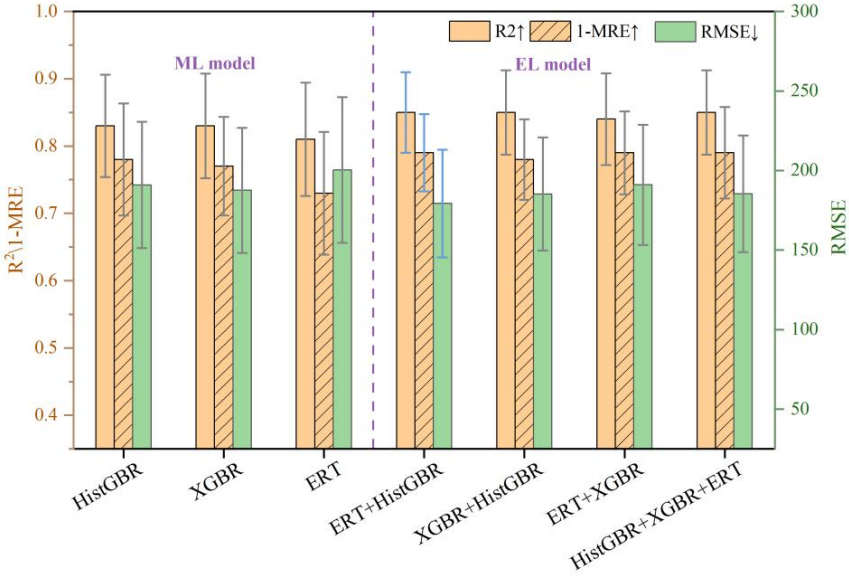
Symbol	Description	Equation for Feature Calculation	Symbol	Description	Equation for Feature Calculation
VEC	Valence electron concentration	$VEC=\sum_i c_i vec_i$	Ω	Solid solution phase formation parameter	$\Omega=\frac{T_m \Delta S_{min}}{\lvert \Delta H_{mix} \rvert}$
σVEC	Standard deviation of valence electron concentration	$\sigma VEC=\sqrt{\sum_i c_i (vec_i - VE($	Λ	Λ parameter	$\Lambda=\frac{\Delta S_{mix}}{\delta^2}$
T_m	Melting point	$T_m = c_i T_{m_i}$	K	Bulk modulus	$K = c_i K_i$
ΔH_{mix}	Enthalpy of mixing	$\Delta H_{mix} = \sum_{i=1,i<j} 4H_{ij}$	σK	Standard deviation of bulk modulus	$\sigma K=\sqrt{\sum_i c_i (K_i - K)^2}$
G	Shear modulus	$G=\sum_i c_i G_i$	Tb	Boiling point	$Tb=\sum_i c_i Tb_i$
σG	Difference in shear modulus	$\sigma G=\sum_i \frac{2(G_i - G)}{G_i + G}$	a	Lattice constant	$a=\sum_i c_i a_i$
T_{max}	Maximum melting point	Maximum melting point	AN	Atomic number	$AN=\sum_i c_i AN_i$
ΔS_{mix}	Configurational entropy	$\Delta S_{mix} = -R\sum_i c_i \ln c$	ρ	Density	$\rho=\sum_i c_i \rho_i$
ΔG_{mix}	Gibbs free energy	$\Delta G_{mix} = \Delta H_{mix} - \Delta S$	σT	Standard deviation of melting point	$\sigma T=\sqrt{\sum_i c_i (1-\frac{T_{m_i}}{T_m})^2}$
δ	Standard deviation of atomic radii	$\delta=\sqrt{\sum_i c_i (1-\frac{r_i}{\bar{r}})^2}$ $r=\sum_i c_i r_i$	Temperature	Experimental temperature	
$\Delta\chi$	Standard deviation of electronegativity	$\Delta\chi=\sqrt{\sum_i c_i (\chi_i - \bar{\chi})^2}$ $\bar{\chi}=\sum_i c_i \chi_i$			

Supplementary Table 2. Representative models including XGBR, RFR and AdaBoost are chosen for modeling without feature filtering, where R^2 and MRE are used as assessment indicators

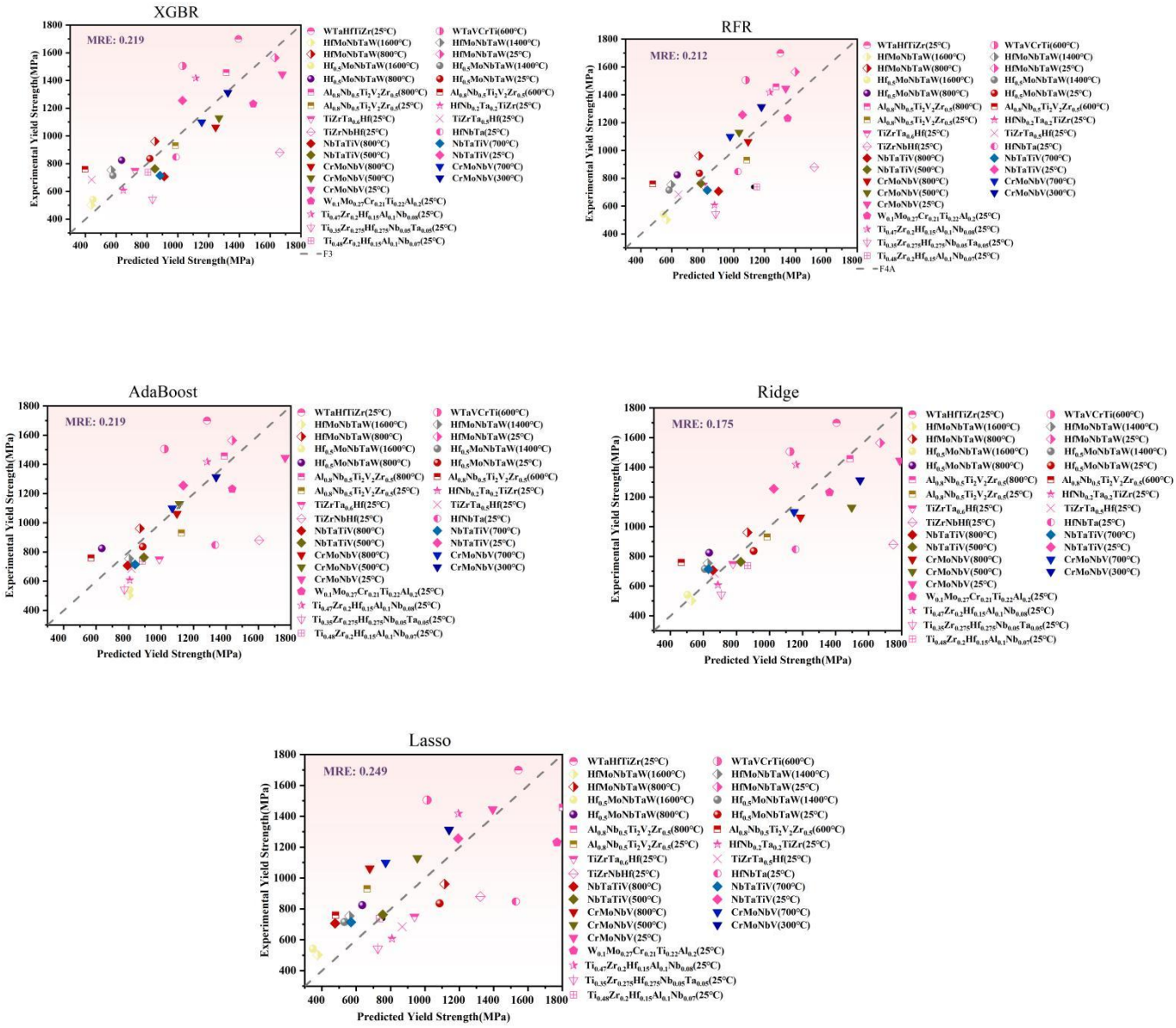
	XGBR	RFR	AdaBoost	Ridge	Lasso
R^2	0.817	0.794	0.720	0.693	0.696
MRE	0.239	0.357	0.516	0.506	0.511

Supplementary Table 3. Hyperparameters selection and performance of different models. Grid search and cross-validation are used to search for and optimize the hyperparameters of various models in order to improve the model prediction effect

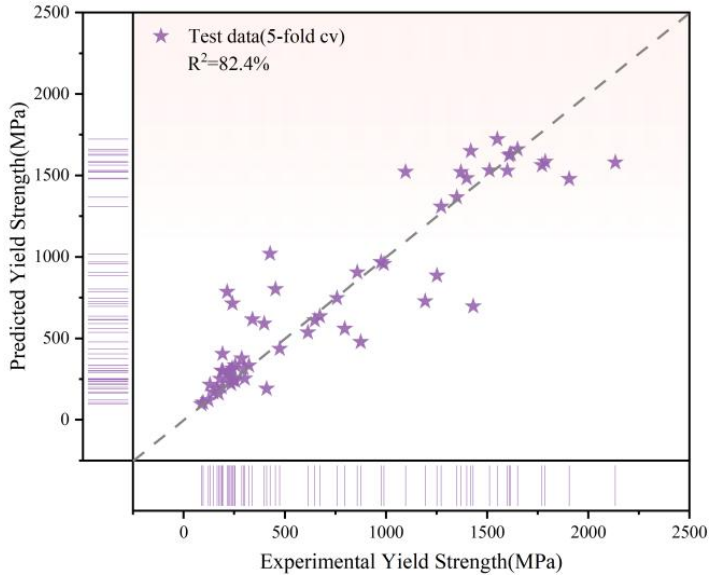
Model	R^2	1-MRE	-RMSE	Hyperparameters
HistGBR	0.83	0.78	-187.56	max_iter=200, learning_rate=0.17, min_samples_leaf=10
XGBR	0.83	0.77	-190.93	max_depth=4, learning_rate=0.2, n_estimators=160
ERT	0.81	0.72	-200.31	max_depth=16, n_estimators=200
LGBMR	0.78	0.69	-218.6	num_leaves=10, learning_rate=0.5, n_estimators=100
GBR	0.77	0.7	-220.45	max_depth=10, max_leaf_nodes=10
RFR	0.76	0.68	-217.5	n_estimators=200, max_depth=9
AdaBoost	0.72	0.46	-252.55	n_estimators=200, learning_rate=0.9
GPR	0.69	0.69	-259.18	alpha=0.05
KNN	0.66	0.46	-275.26	n_neighbors=2
LR	0.61	0.43	-294.63	n_jobs=1



Supplementary Figure 1. The outcomes of 10 times 10-fold cross-validation comparing conventional machine learning and multiple stacking ensemble models on the training set.

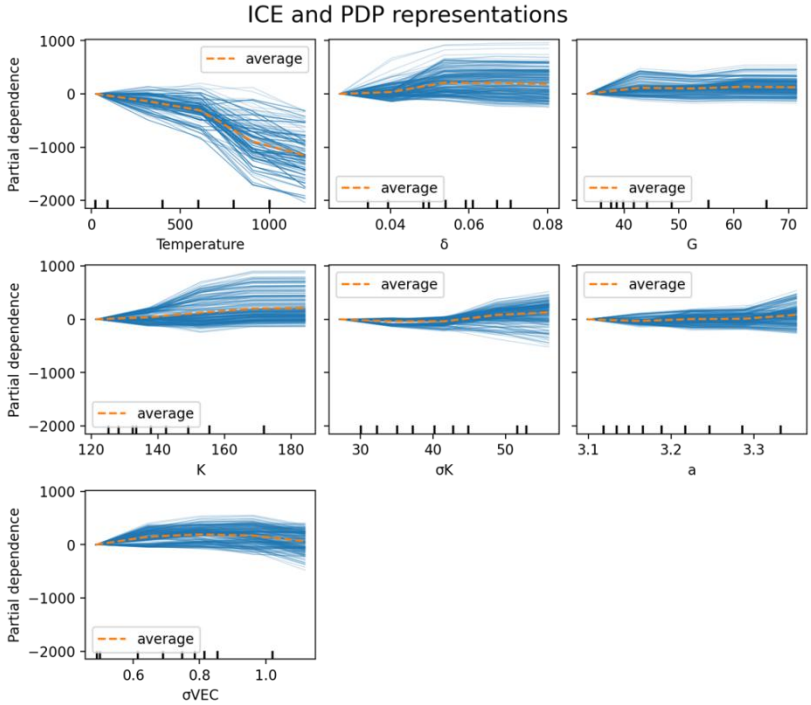


Supplementary Figure 2. Fitted scatterplot for experimental data by the model based on XGBR, RFR, AdaBoost, Ridge, Lasso.



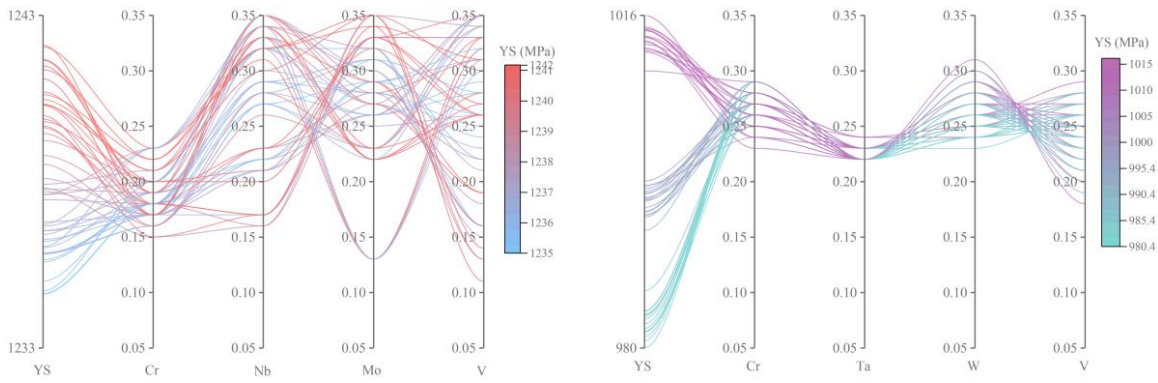
Supplementary Figure 3. Area map of 59 out-of-packet data predictions based on the same features.

The orange area is where the predicted value exceeds the experimental value, and the white area is when the anticipated value is less than the experimental value. The model has a stronger prediction effect on data with yield strengths less than 1000 MPa, according to the area comparison.



Supplementary Figure 4. Individual conditional expectation (ICE) and Partial Dependence (PDP) plots for 7 key features.

PDP plot depicts the average effect of a set of features, whereas ICE plot eliminates non-uniform effects and visualizes the prediction's dependence on features for each sample separately, with each sample represented by a separate line.



Supplementary Figure 5. Parallel coordinate plots illustrate the ratio of alloy elements with outstanding mechanical characteristics discovered by DBO in the CrNbMoV system at 1000 degrees Celsius and the CrTaWV system at 1200 degrees Celsius. Each line in the picture depicts a RHEA with element ratio data, the redder the hue, the higher the alloy's yield strength(YS) value.

Supplementary Table 4. The coefficients and intercept of second-order polynomial regression.

y \ *x*	1	Temperature	Zr	Ti	Hf	Nb	Mo	V	Ta	W	Al	Cr
1	-0.008	-0.038	6.264	-1.905	-3.273	-1.391	-1.256	-2.475	-0.191	0.938	0.555	2.734
Temperature		-0.005	-0.178	-0.039	0.032	0.078	0.131	0.055	0.008	0.008	-0.049	-0.084
Zr			-10.168	-16.300	-1.015	0.284	8.682	-5.962	12.521	0.000	7.003	11.219
Ti				2.233	8.816	-4.289	-3.801	1.538	1.327	5.974	6.984	-4.386
Hf					-10.636	6.827	0.332	-1.421	-0.0882	0.000	-10.716	5.423
Nb						3.566	3.187	10.499	-11.824	-2.042	7.482	-15.082
Mo							-3.023	4.300	2.819	-11.666	-0.386	-1.699
V								-2.459	6.787	-11.312	-3.493	-0.951
Ta									-4.798	5.981	-1.654	-10.466
W										5.981	0.000	8.022
Al											-18.985	14.321
Cr												-3.667
Intercept	2.294											

The first row in the table represents the coefficient for the primary term, and the first from left to right is the error term.The remaining positions represent the coefficients before the quadratic term, and the product of the corresponding horizontal and vertical table heads of the coefficients is its quadratic term.