Supporting Information

Data-Driven Machine Learning Prediction of Glass Transition Temperature and Glass-Forming Ability for Metallic Glasses

Jingzi Zhang[a,b†], Mengkun Zhao[a,b†], Chengquan Zhong[a,b], Jiakai Liu[a,b], Kailong Hu[a,b,c], and Xi Lin[a,b,c*]*

[a] School of Materials Science and Engineering, Harbin Institute of Technology, Shenzhen 518055, P. R. China

[b] Blockchain Development and Research Institute, Harbin Institute of Technology, Shenzhen 518055, P. R. China

[c] State Key Laboratory of Advanced Welding and Joining, Harbin Institute of Technology, Harbin 150001, P. R. China

† These authors contribute equally to this work.

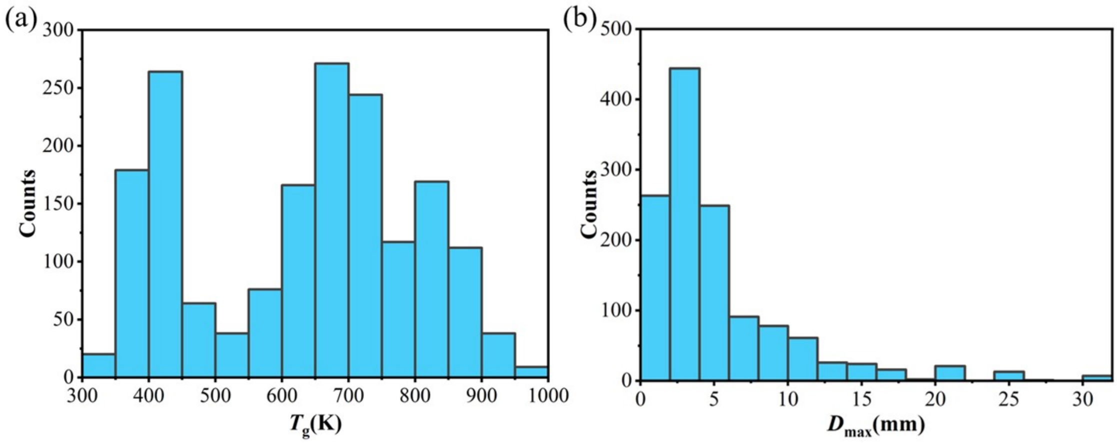

Fig. S1 The overall distribution of (a) T_g and (b) D_{max} values in dataset.

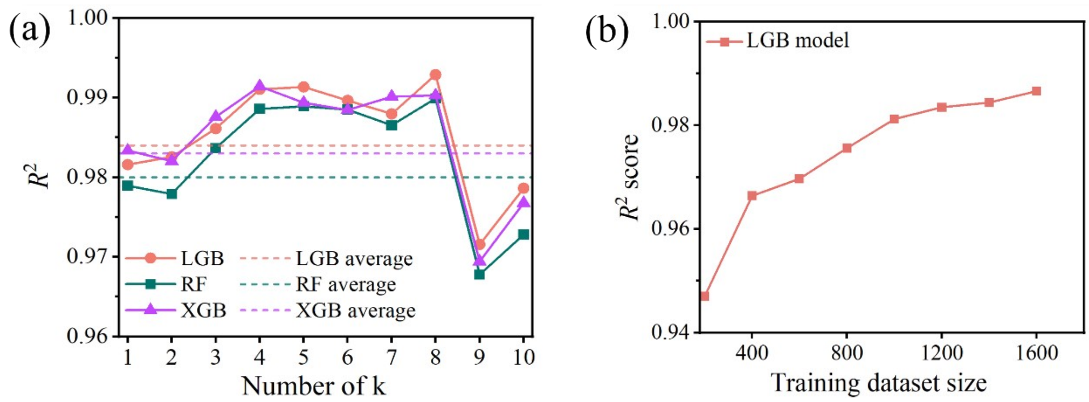

Fig. S2 (a) R^2 score of 10-fold cross-validation for three ML models on T_g dataset. (b) Training dataset size dependences of the R^2 score for ML model during T_g prediction.

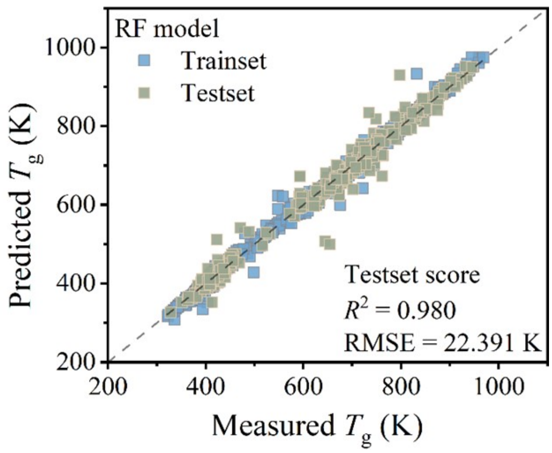

Fig. S3 Comparison of predicted and measured T_g values using the RF model.

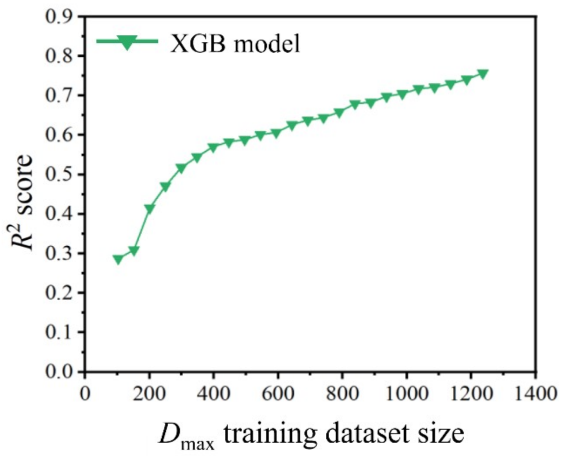

Fig. S4 Training dataset size dependences of the R^2 score for ML model during D_{max} prediction.

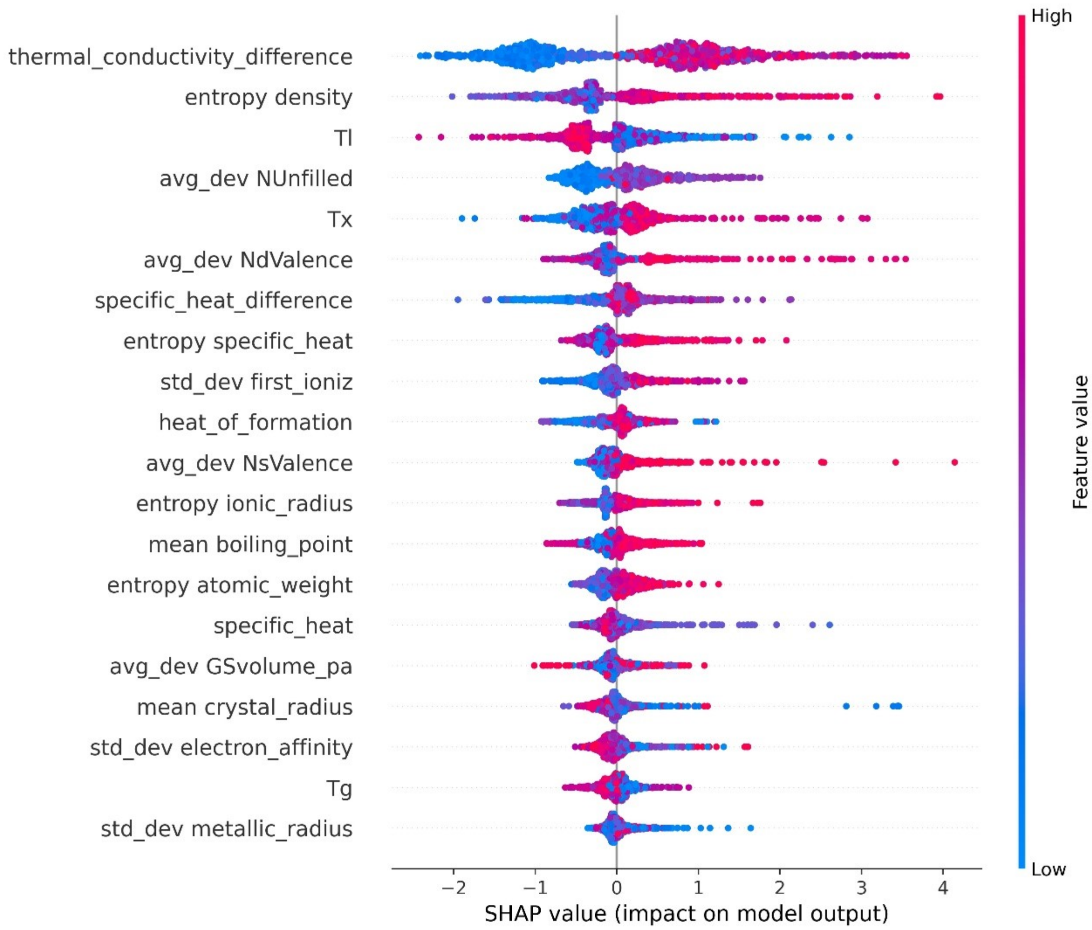

Fig. S5 Feature importance based on SHAP value of XGBoost model, in order of increasing importance (i.e., the sum of SHAP value magnitudes). The color corresponds to the value of each input feature and can demonstrate positive or negative correlation with D_{max} values. Red and blue color mean the values of listed feature on each data point, respectively.

Table S1. Three main types of duplicate data in the merged dataset.

Type	Example
The chemical formulas of MGs are the same.	$Fe_{36}Co_{36}B_{19.2}Si_{4.8}Nb_4$ $Fe_{36}Co_{36}B_{19.2}Si_{4.8}Nb_4$
The data format of the element composition in the chemical formula is inconsistent.	$Fe_{75}C_7Si_{3.3}B_5P_{8.7}Ga_1$ $Fe_{75.0}C_{7.0}Si_{3.3}B_{5.0}P_{8.7}Ga_{1.0}$; $(Fe_{75}Co_{25})_{83}P_{17}$ $(Fe_{0.75}Co_{0.25})_{83}P_{17}$; $Cu_{43}Zr_{43}Al_{17}Be_7$ $(Cu_{0.5}Zr_{0.5})_{86}Al_7Be_7$.
The order of the elements in the chemical formula is different.	$Zr_{63.5}Al_{10.7}Cu_{10.7}Ni_{15.1}$ $Al_{10.7}Cu_{10.7}Ni_{15.1}Zr_{63.5}$

Table S2. Calculation methods of element features in compounds. The “i” represents the element number of MGs. The “t” represents the property of element. The “p” represents the weighted score [1].

In the calculation of the weight entropy of mixing $\omega = t_i/(t_1 + ... + t_i)$.

Features description	Computational formula
The average	$= \mu = (t_1 + ... + t_i)/i$
Weighted mean	$= \nu = (p_1 \times t_1) + ... + (p_i \times t_i)$
Geometric mean	$= (t_1 \times ... \times t_i)^{1/i}$
Weighted geometric mean	$= (t_1)^{p1} \times ... \times (t_i)^{pi}$
The entropy of mixing	$=- \omega_1 In(\omega_1) - ... - \omega_i In(\omega_i)$
Weighted entropy of mixing	$=- \frac{p_1\omega_1}{p_1\omega_1 + ... + p_i\omega_i} In\left(\frac{p_1\omega_1}{p_1\omega_1 + ... + p_i\omega_i}\right) ... - \frac{p_i\omega_i}{p_1\omega_1 + ... +}$
Extreme value range	$= t_1 - t_2\ (t_1 > t_2)$
Weighted range	$= p_1 t_1 - p_2 t_2$
The standard deviation	$= [(1/2)((t_1 - \mu)^2 + (t_2 - \mu)^2)]^{1/2}$
Weighted standard deviation	$= [p_1(t_1 - \nu)^2 + p_2(t_2 - \nu)^2]^{1/2}$

Table S3 The explication of 16 features selected for T_g model by RFE method.

Features	Explanation
mean melting point	the mean value of melting point
std dev vdw radius	the standard deviation value of vdw radius
mean crystal radius	the mean value of crystal radius
mean electronegativity	the mean value of electronegativity
mean column	the mean value of columns in the periodic table
Gmix	mixing Gibbs free energy
gmean ionic radius	the geometric mean value of ionic radius
std dev atomic volume	the standard deviation value of atomic volume
avg dev MendeleevNumber	the average deviation value of MendeleevNumber
mean NdValence	the mean value of *d* atomical orbit electrons
std dev melting point	the standard deviation value of melting point
std dev fusion heat	the standard deviation value of fusion heat
avg dev electronegativity	the average deviation value of electronegativity
gmean second ionisation energy	the geometric mean value of second ionisation energy
std dev dipole polarizability	the standard deviation value of dipole polarizability
std dev heat of formation	the standard deviation value of heat of formation

Table S4 The explication of 22 features selected for D_{max} model by RFE method.

Features	Explanation
thermal conductivity difference	the mismatch of thermal conductivity
avg_dev NdValence	the average deviation value of *d* atomical orbit electrons
entropy density	the entropy value of density
avg_dev Nunfilled	the average deviation value of unfilled electrons
std_dev specific heat	the standard deviation value of specific heat
avg_dev NsValence	the average deviation value of *s* atomical orbit electrons
specific heat difference	the mismatch of specific heat
T_x	the crystallization temperature
entropy ionic radius	the entropy value of ionic radius
mean crystal radius	the mean value of crystal radius
entropy specific heat	the entropy value of specific heat
std_dev electron affinity	the standard deviation value of electron affinity
std_dev first ioniz	the standard deviation value of first ioniz
T_l	liquids temperature
std_dev heat of formation	the standard deviation value of heat of formation
mean boiling point	the mean value of boiling point
T_g	glass transition temperature
avg_dev GSvolume	the average deviation value of ground state volume
entropy atomic weight	the entropy value of atomic weight
entropy classical valence	the entropy value of classical valence
std_dev boiling point	the standard deviation value of boiling point
std_dev metallic radius	the standard deviation value of metallic radius

Table S5 The hyperparameter of the RF, XGBoost and LGB models.

Model	Parameter	Value
RF	n_estimators	500
	max_depth	17
	min_samples_split	2
	max_features	sqrt
XGBoost	n_estimators	180
	max_depth	10
	learning_rate	0.05
	min_child_weight	6
	colsample_bytree	0.68
	subsample	1.0
LGB	n_estimators	260
	max_depth	12
	num_leaves	20
	learning_rate	0.16

Table S6 The predicted and measured D_{max} values, thermal conductivity difference, and entropy density values of some samples.

MGs	Measured D_{max}(mm)	Predicted D_{max}(mm)	Thermal conductivity difference	Entropy density
$Zr_{42}Ni_6Cu_{36}Al_8Ag_8$	25	22.3	0.85	1.26
$Zr_{48}Pd_4Cu_{32}Al_8Ag_8$	25	24	0.93	1.25
$Zr_{44}Pd_4Cu_{36}Al_8Ag_8$	25	23.1	0.87	1.24
$Zr_{46}Ni_2Cu_{36}Al_8Ag_8$	25	21.7	0.87	1.17
$Zr_{48}Ni_2Cu_{34}Al_8Ag_8$	25	20.5	0.90	1.17
$Zr_{48}Ni_6Cu_{30}Al_8Ag_8$	25	20.5	0.94	1.27
$(Ti_{0.41}Zr_{0.25}Be_{0.26}Ni_{0.08})_{98}Cu_2$	25	21	1.11	1.37
$Zr_{46}Pd_2Cu_{36}Al_8Ag_8$	30	27.4	0.87	1.19
$Zr_{48}Pd_2Cu_{34}Ag_8Al_8$	30	25.3	0.91	1.19
$Zr_{48}Ni_4Cu_{32}Al_8Ag_8$	30	20.5	0.92	1.23
$Zr_{44}Ni_4Cu_{36}Al_8Ag_8$	30	26.3	0.86	1.22
$(Ti_{0.41}Zr_{0.25}Be_{0.26}Ni_{0.08})_{96}Cu_4$	30	23.9	1.14	1.43
$(Ti_{0.41}Zr_{0.25}Be_{0.26}Ni_{0.08})_{94}Cu_6$	30	26.7	1.15	1.46
$(Ti_{0.41}Zr_{0.25}Be_{0.26}Ni_{0.08})_{90}Cu_{10}$	30	20.3	1.15	1.50

Utilizing the SISSO method, we have selected eight critical features from the SHAP importance features to establish a formulaic relationship with D_{max}, resulting in a correlation coefficient (r) of 0.687—exceeding the 0.672 reported in Xiong's work [2]. The formulation of feature equations facilitates a more intuitive exploration of the connection between D_{max} and these pivotal features, thereby providing fresh insights into the design of metallic glasses with enhanced D_{max}.

$$D_{max} = 0.64D_1 + 0.482D_2 + 0.339D_3 + 3.777 \quad (1)$$

$$D_1 = \left(X_1^2 + X_6^2\right)\left(X_1 + X_3\right)exp(X_1) \quad (2)$$

$$D_2 = (X_6 - X_2)(X_0 + X_6) - X_4^2 exp(X_6) \tag{3}$$

$$D_3 = \frac{exp(X_1)(X_1 + X_6)}{\frac{X_1}{X_5} + \frac{X_7}{X_3}} \tag{4}$$

where the X_1 is the mean deviation of unfilled electrons; X_2 is the range value of thermal conductivity; X_3 is the thermal conductivity difference; X_4 is the entropy of first ionization energy; X_5 is the standard deviation of thermal conductivity; X_6 is the average deviation of *d* atomical orbit electrons; X_7 is the standard deviation of specific heat;

Reference

[1] Hamidieh K. A data-driven statistical model for predicting the critical temperature of a superconductor. Computational Materials Science, 2018, 154: 346-354.

[2] Xiong, J., Shi, S.-Q., Zhang, T.-Y. A Machine-Learning Approach to Predicting and Understanding the Properties of Amorphous Metallic Alloys. Mater. Des. 2020, 187, 108378.