Supplementary Materials

Transfer learning-based prediction of high-temperature fatigue life in Fe-based structural alloys with limited data

Qi Wang[1,#], Chunlei Shang[1,#], Hong-Hui Wu[1,2,3,*], Dexin Zhu[1], Shuize Wang[1,3], Junheng Gao[1,3], Haitao Zhao[1,3], Chaolei Zhang[1,3], Yuhe Huang[1,3], Jun Lu[1,3], Xinping Mao[1,3]

[1]Beijing Advanced Innovation Center for Materials Genome Engineering, Institute for Carbon Neutrality, University of Science and Technology Beijing, Beijing 100083, China.
[2]Institute of Materials Intelligent Technology, Liaoning Academy of Materials, Shenyang 110004, Liaoning, China.
[3]Institute of Steel Sustainable Technology, Liaoning Academy of Materials, Shenyang 110004, Liaoning, China.
[#]Authors contributed equally to this work.

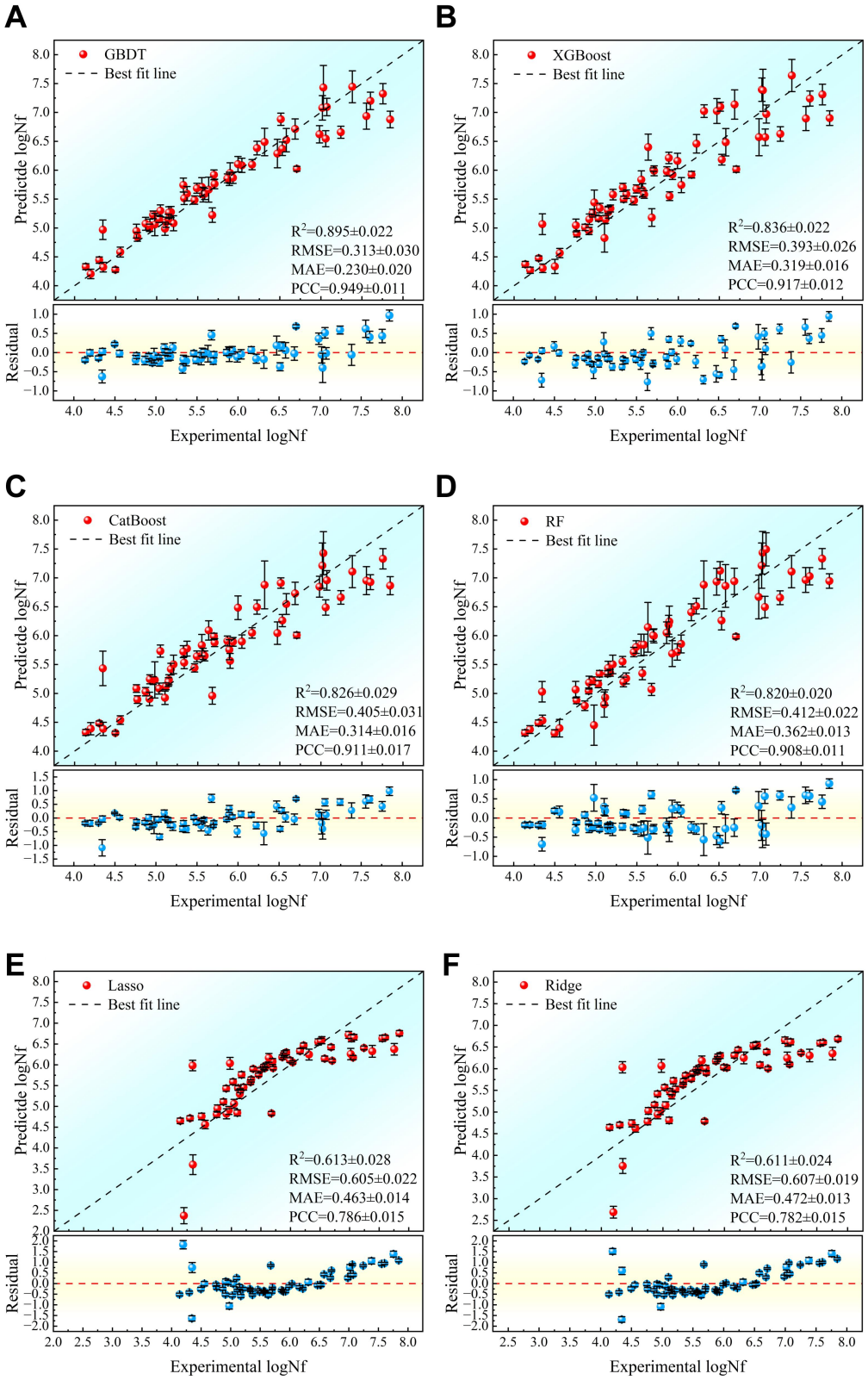
[*]**Correspondence to:** Prof. Hong-Hui Wu, Beijing Advanced Innovation Center for Materials Genome Engineering, Institute for Carbon Neutrality, University of Science and Technology Beijing, Beijing 100083, China. E-mail: wuhonghui@ustb.edu.cn

Supplementary Table 1. Definitions of abbreviations and variables used in the study

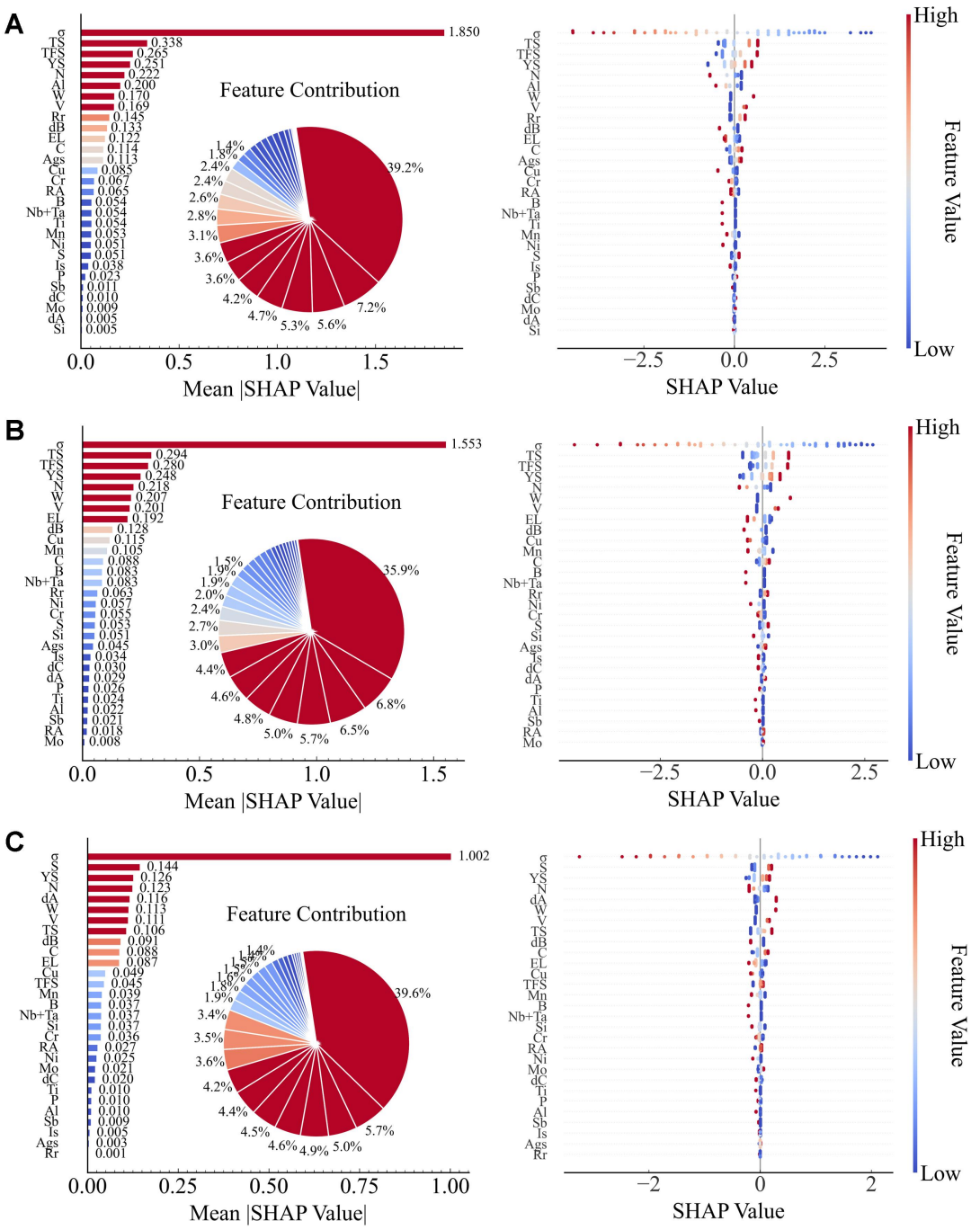
	Abbreviation	Description
Inputs	Is	Ingot size (t)
	Rr	Reduction ratio (%)
	dA	For inclusions deformed by plastic work (%)
	dB	For inclusions occuring in discontinuous (%)
	dC	For isolated inclusions (%)
	Ags	Austenite grain size number
	C	Mass fraction of C (%)
	Si	Mass fraction of Si (%)
	Mn	Mass fraction of Mn (%)
	P	Mass fraction of P (%)
	S	Mass fraction of S (%)
	Ni	Mass fraction of Ni (%)
	Cr	Mass fraction of Cr (%)
	Mo	Mass fraction of Mo (%)
	W	Mass fraction of W (%)
	V	Mass fraction of V (%)
	Cu	Mass fraction of Cu (%)
	Sb	Mass fraction of Sb (%)
	Ti	Mass fraction of Ti (%)
	Al	Mass fraction of Al (%)
	B	Mass fraction of B (%)
	N	Mass fraction of N (%)
	Nb+Ta	Mass fraction of Nb+Ta (%)
	YS	0.2% proof stress (MPa)
	TS	Tensile strength (MPa)
	TFS	True fracture stress (MPa)
	EL	Elongation (%)
	RA	Reduction of area (%)
	σ	Stress amplitude (MPa)
Output	logNf	Number of cycles to failure

Supplementary Table 2. Hyperparameters of the evaluated models

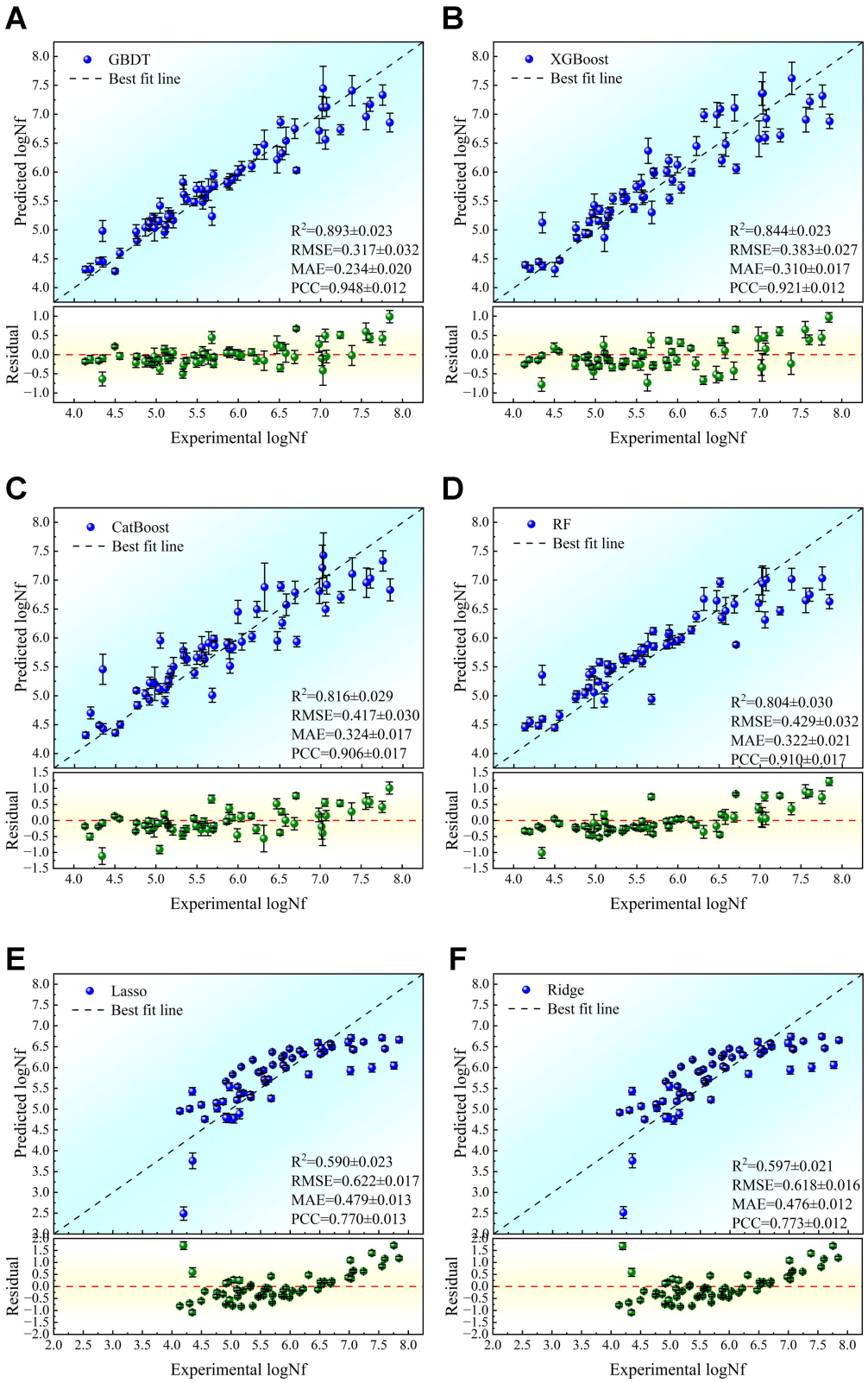
Models	Hyperparameters
GBDT	n estimators: 500 learning rate: 0.066 max depth: 10 min samples split: 4 min samples leaf: 1 subsample: 0.59
XGBoost	n estimators: 400 learning rate: 0.049 max depth: 10 subsample: 0.515 colsample bytree: 0.882 reg alpha: 0.0023 reg lambda: 0.161 min child weight: 1 gamma: 0.0053
CatBoost	n estimators: 400, learning rate: 0.0874 depth: 5 l2_leaf_reg: 0.0065 border count: 222 random strength: 0.155 bagging temperature: 0.0643
RF	n estimators: 350 max depth: 10 min samples split: 2 min samples leaf: 1
Lasso	alpha: 0.01 max iter: 1000
Ridge	alpha: 0.5 tol: 0.001



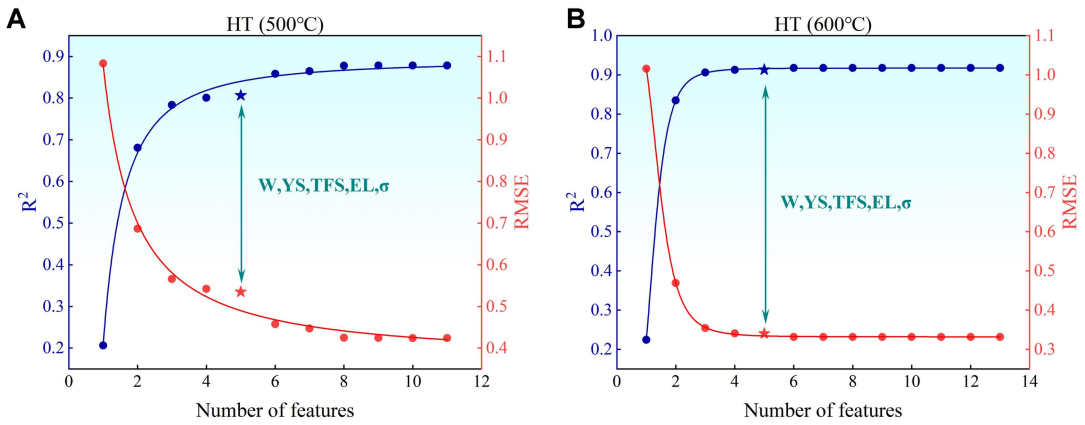
Supplementary Figure 1. The performance of the ML models on the full RT dataset, including scatter plots of actual versus predicted values and residual plots. (A) GBDT, (B) XGBoost, (C) CatBoost, (D) RF, (E) Lasso, and (F) Ridge. Test set sample size (*N*) = 60. Error bars represent the standard deviation of 50 independent replicates.



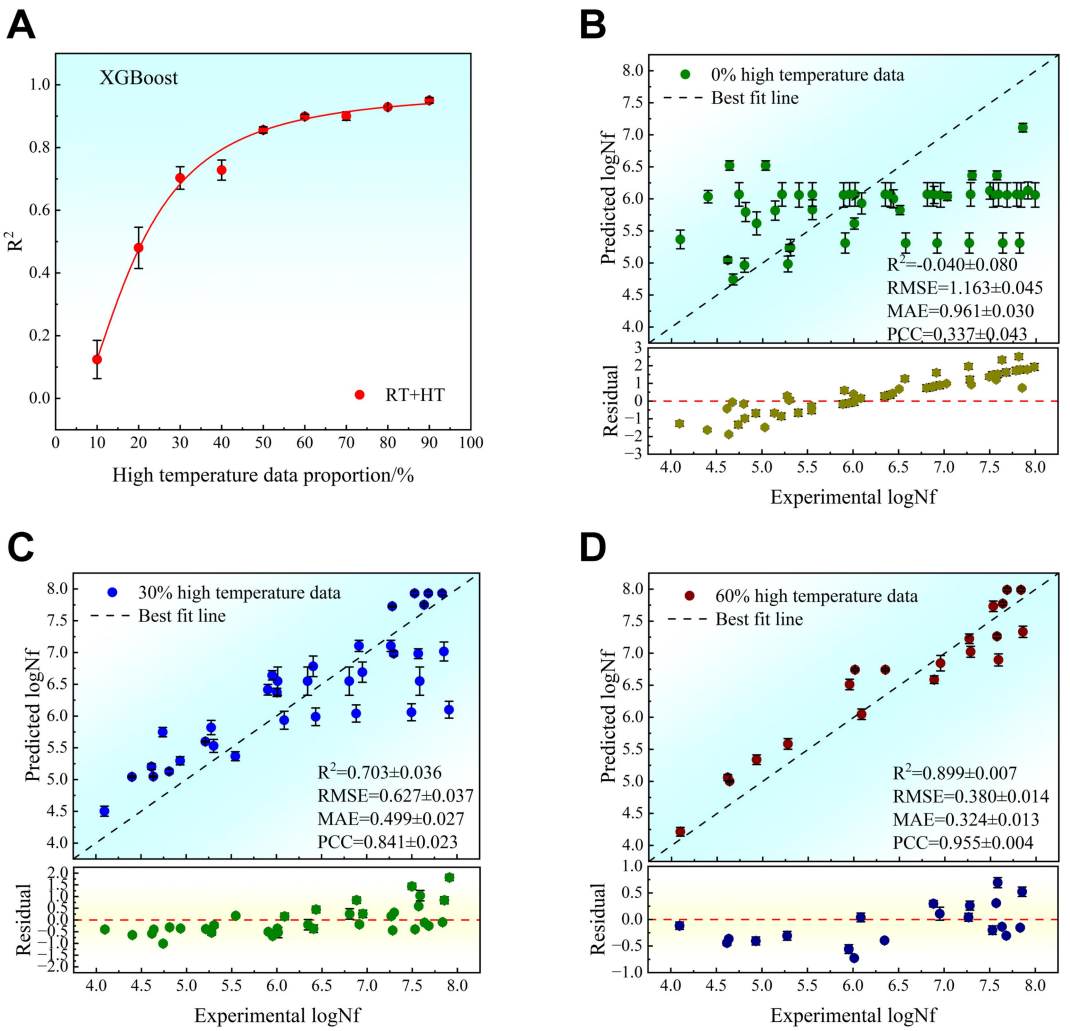
Supplementary Figure 2. Feature importance under high-temperature dataset. (A-C) correspond to the rankings of feature importance based on SHAP values at 400 °C, 500 °C, and 600 °C, respectively.



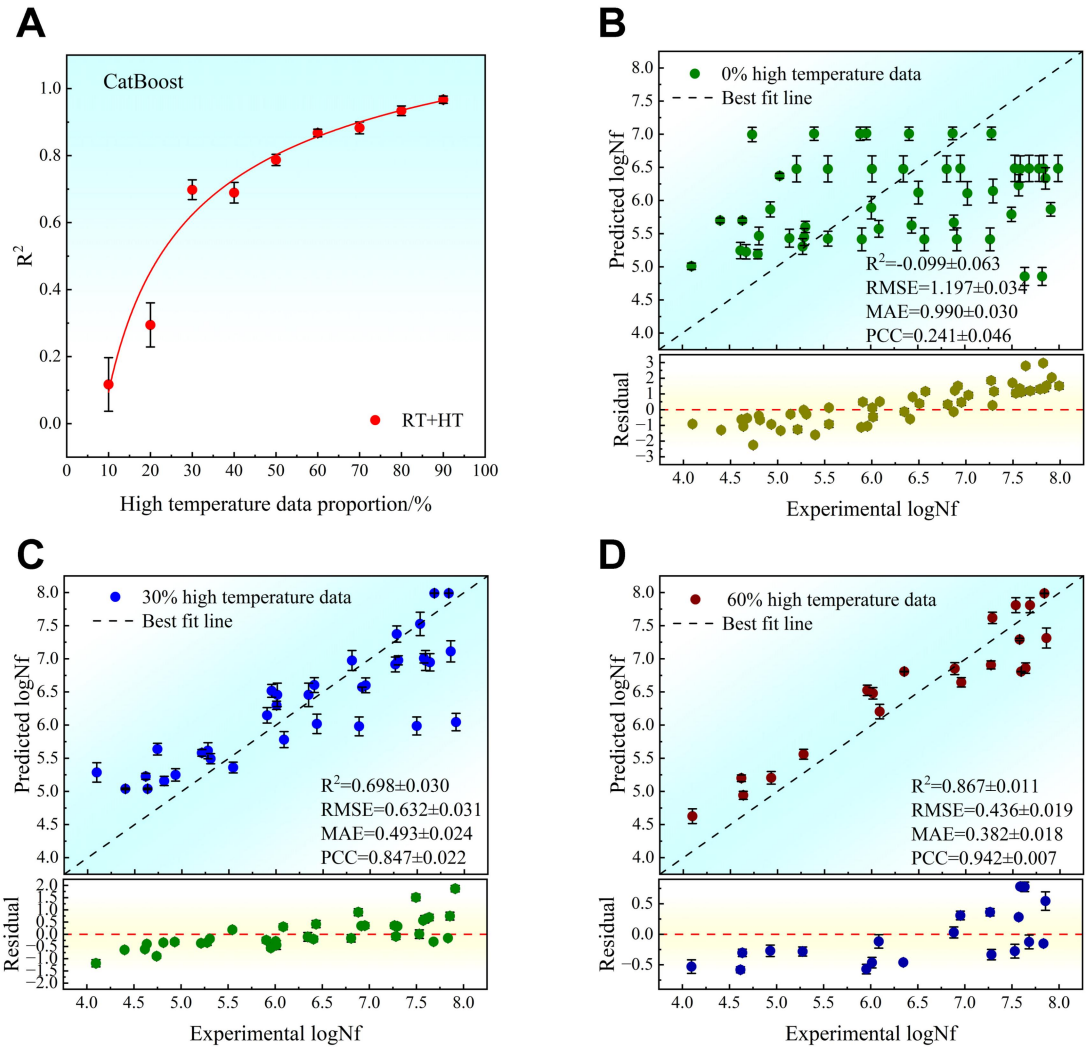
Supplementary Figure 3. The performance of the ML models on the RT dataset after feature selection, including scatter plots of actual versus predicted values and residual plots. (A) GBDT, (B) XGBoost, (C) CatBoost, (D) RF, (E) Lasso, and (F) Ridge. Test set sample size (*N*) = 60. Error bars represent the standard deviation of 50 independent replicates.



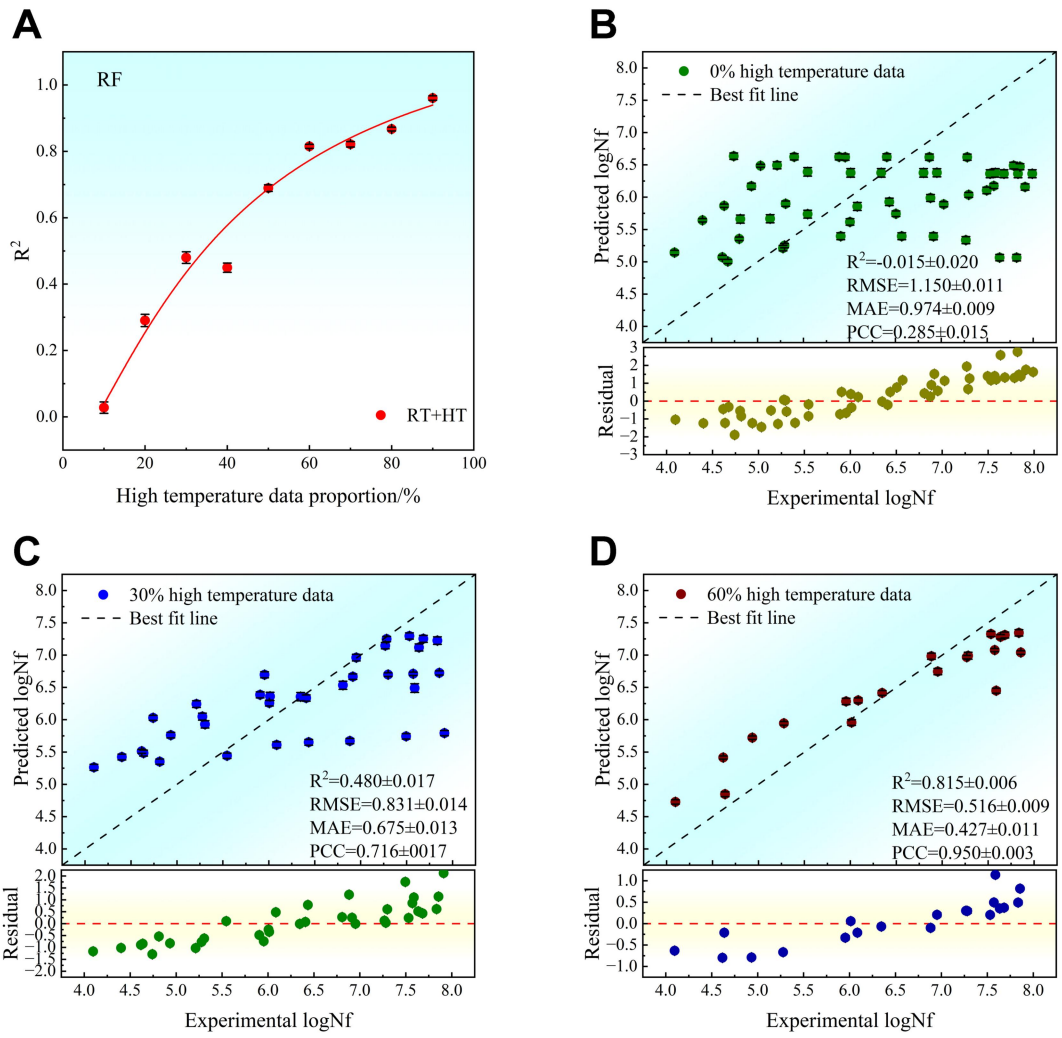
Supplementary Figure 4. Feature selection under high-temperature dataset. (A and B) represent the optimal subset selection at 500 ℃ and 600 ℃, respectively.



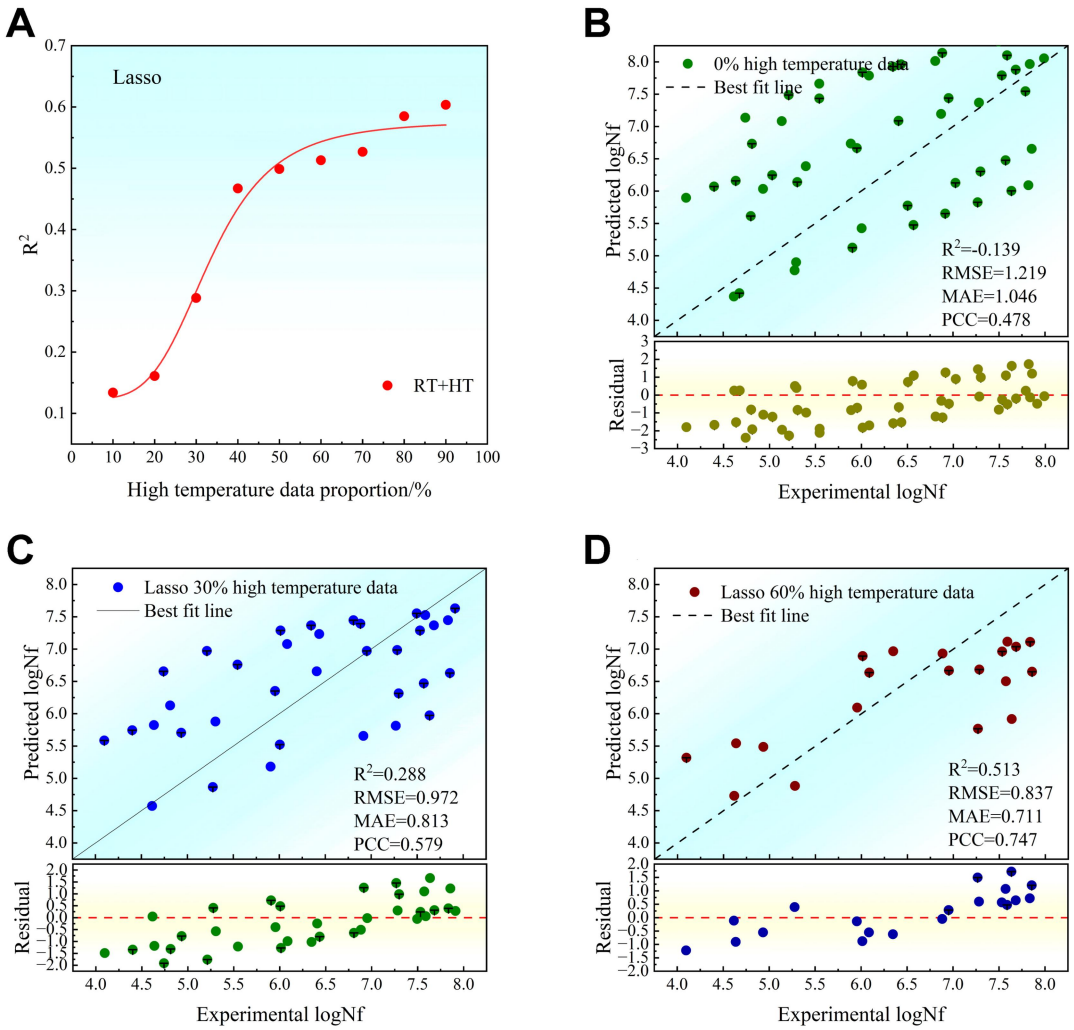
Supplementary Figure 5. Transfer learning prediction results of the XGBoost model from room temperature to high temperature. (A) R^2 values of the model under different proportions of high-temperature data. (B–D) scatter plots and residual plots for the remaining high-temperature test sets at proportions of 0 %, 30 %, and 60 %, respectively. Test set sample size (N): 0% HT = 50, 30% HT = 35, 60% HT = 20. Error bars represent the standard deviation of 50 independent replicates.



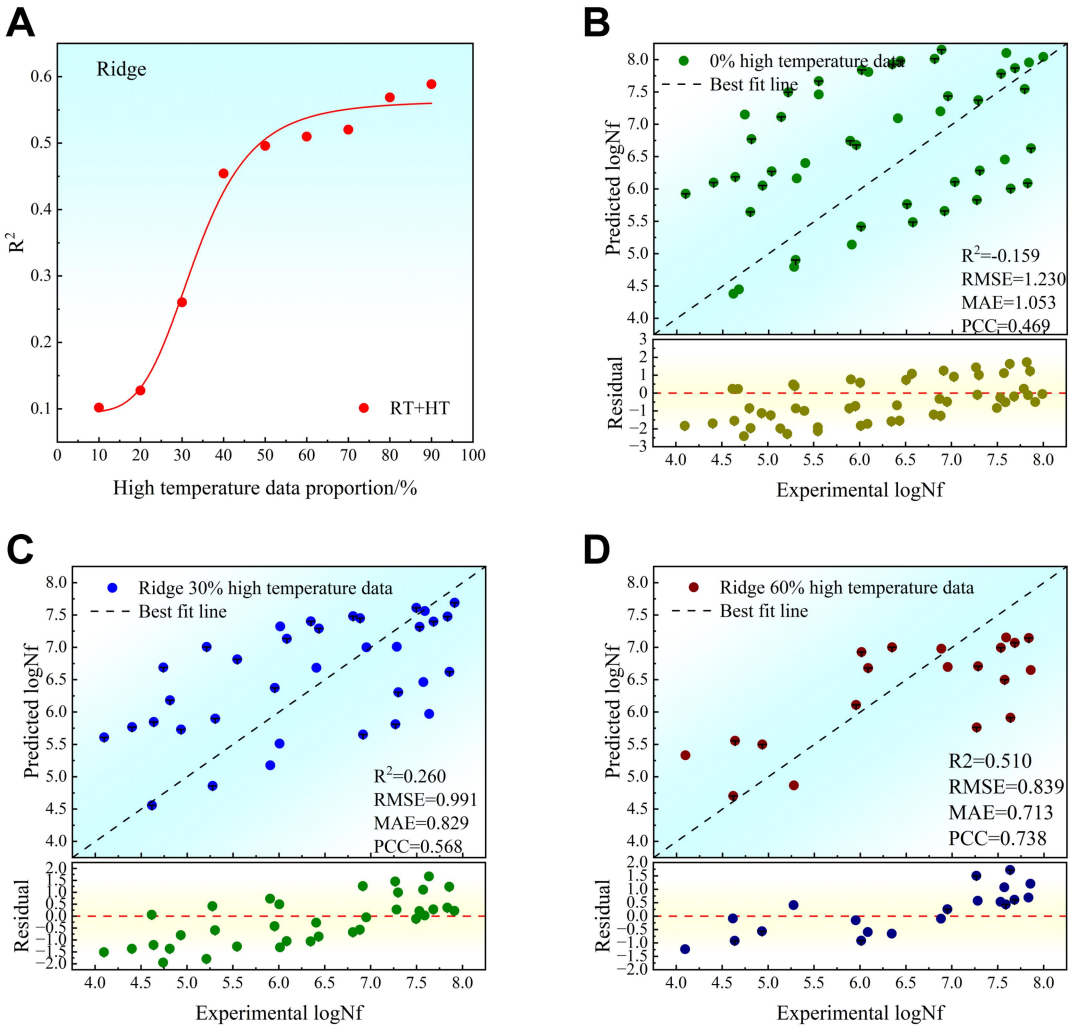
Supplementary Figure 6. Transfer learning prediction results of the CatBoost model from room temperature to high temperature. (A) R^2 values of the model under different proportions of high-temperature data. (B–D) scatter plots and residual plots for the remaining high-temperature test sets at proportions of 0%, 30%, and 60%, respectively. Test set sample size (N): 0% HT = 50, 30% HT = 35, 60% HT = 20. Error bars represent the standard deviation of 50 independent replicates.



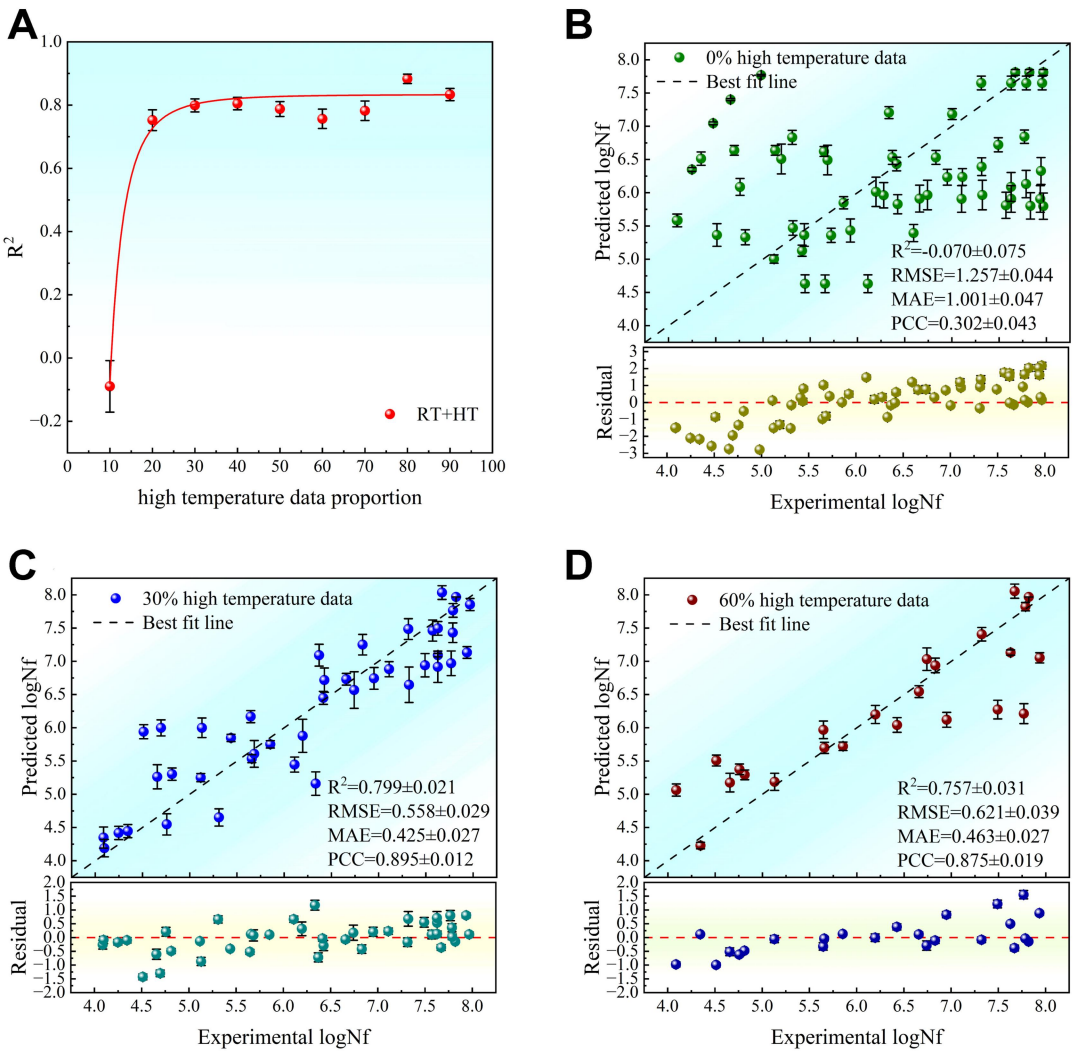
Supplementary Figure 7. Transfer learning prediction results of the RF model from room temperature to high temperature. (A) R^2 values of the model under different proportions of high-temperature data. (B–D) scatter plots and residual plots for the remaining high-temperature test sets at proportions of 0%, 30%, and 60%, respectively. Test set sample size (N): 0% HT = 50, 30% HT = 35, 60% HT = 20. Error bars represent the standard deviation of 50 independent replicates.



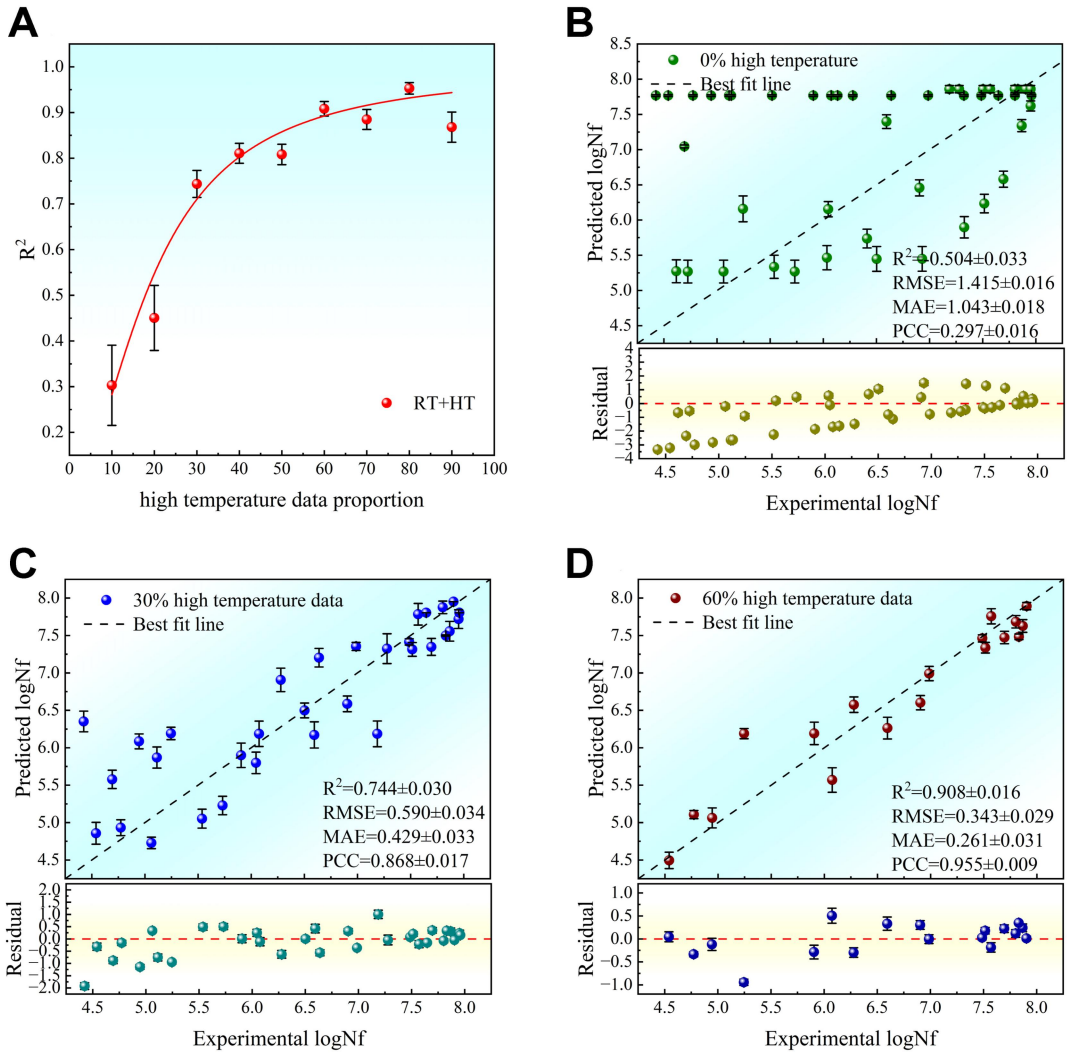
Supplementary Figure 8. Transfer learning prediction results of the Lasso model from room temperature to high temperature. (A) R^2 values of the model under different proportions of high-temperature data. (B–D) scatter plots and residual plots for the remaining high-temperature test sets at proportions of 0 %, 30 %, and 60 %, respectively. Test set sample size (N): 0% HT = 50, 30% HT = 35, 60% HT = 20. Error bars represent the standard deviation of 50 independent replicates.



Supplementary Figure 9. Transfer learning prediction results of the Ridge model from room temperature to high temperature. (A) R^2 values of the model under different proportions of high-temperature data. (B–D) scatter plots and residual plots for the remaining high-temperature test sets at proportions of 0%, 30%, and 60%, respectively. Test set sample size (N): 0% HT = 50, 30% HT = 35, 60% HT = 20. Error bars represent the standard deviation of 50 independent replicates.



Supplementary Figure 10. Transfer learning prediction results of the GBDT model from room temperature to high temperature of 500 ℃. (A) R^2 values of the model under different proportions of high-temperature data. (B-D) scatter plots and residual plots for the remaining high-temperature test sets at proportions of 0%, 30%, and 60%, respectively. Test set sample size (N): 0% HT = 59, 30% HT = 41, 60% HT = 24. Error bars represent the standard deviation of 50 independent replicates.



Supplementary Figure 11. Transfer learning prediction results of the GBDT model from room temperature to high temperature of 600 ℃. (A) R^2 values of the model under different proportions of high-temperature data. (B-D) scatter plots and residual plots for the remaining high-temperature test sets at proportions of 0%, 30%, and 60%, respectively. Test set sample size (N): 0% HT = 45, 30% HT = 32, 60% HT = 18. Error bars represent the standard deviation of 50 independent replicates.