Support Information

Enhancing LIBS Analysis of Heavy Metals in Fly Ash: A Hybrid Strategy Integrating Data Augmentation with Selective Ensemble

Chunhua Yan[a,*], Haiting Zhang[a], Xuedong Chen[a], Maogang Li[a], Tianlong Zhang[b], Hua Li[a,b,*]

[a] *College of Chemistry and Chemical Engineering, Xi'an Shiyou University, Xi'an, 710065, China*

[b] *College of Chemistry & Materials Science, Northwest University, Xi'an, 710069, China*

* Corresponding author: Chunhua Yan, E-mail: huahua18254@163.com; Hua Li, E-mail: huali@nwu.edu.cn

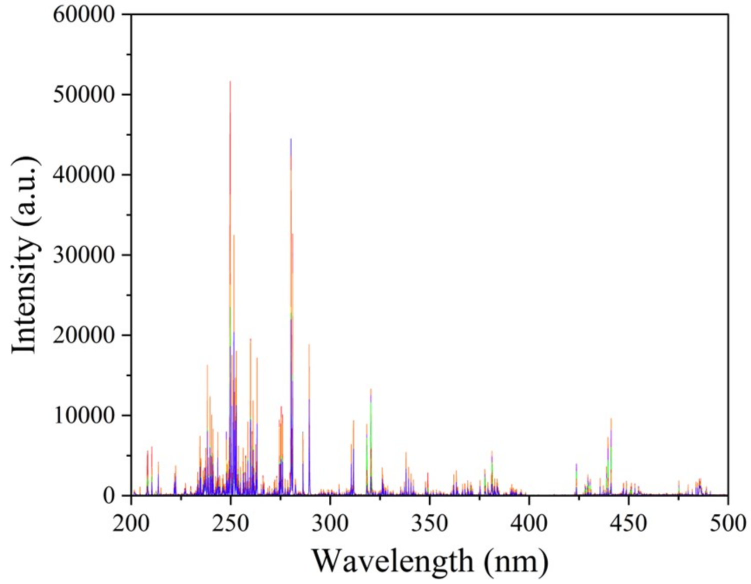

Fig. S1 LIBS spectrum of all samples fly ash

Fig. S2 Optimization results for SMOTE hyperparameters: (a) Number of nearest neighbors (k), (b) Oversampling ratio

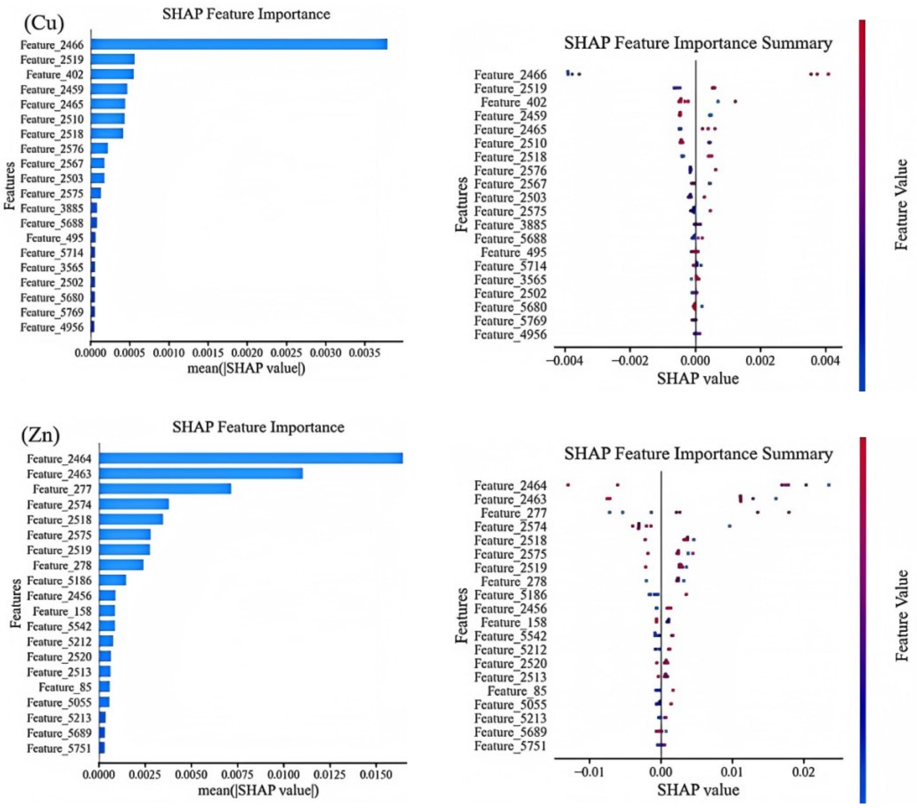

Fig. S3 SHAP analysis of the WOA-SE model for (a) Cu and (b) Zn

Note:

Cu		Zn	
Feature_ID	Wavelength (nm)	Feature ID	Wavelength (nm)
2466	327.87	2464	327.77
2519	330.62	2463	327.71
402	220.8	277	214.31
2459	327.51	2574	333.47
2465	327.82	2518	330.57
2510	330.15	2575	333.52
2518	330.57	2519	330.62
2576	333.58	278	214.36
2567	333.11	5186	468.97
2503	329.79	2456	327.35
2575	333.52	158	208.14
3835	401.48	5542	487.44
5638	495.01	5212	470.36
495	225.62	2520	330.67
5714	496.36	2513	330.31
3565	384.88	85	204.35
2502	329.74	5055	462.18
5680	494.6	5213	470.37
5769	499.22	5689	495.07

Fig. S4 Residual distribution of the training and test sets for (a) Cu and (b) Zn

Table S1 Hyperparameter settings and predictive performance for each individual model

Model	Element	Hyperparameter	5-fold CV	
			R_{CV}^2	$RMSE_{CV}$ (wt%)
RF	Cu	n_estimators=568, max_depth=34, min_samples_split=6, min_samples_leaf=1	0.9320	0.0017

	Zn	n_estimators=987，max_depth=18，min_samples_split=2，min_samples_leaf=1	0.9272	0.0119
GBDT	Cu	n_estimators=215，learning_rate=0.164，subsample=0.915	0.9426	0.0017
	Zn	n_estimators=458，learning_rate=0.134，subsample=0.9	0.9216	0.0132
SVR	Cu	C=9.8976，epsilon =0.001, gamma ='scale'	0.9101	0.0018
	Zn	C=0.1，epsilon=0.001, svrgamma ='scale'	0.9065	0.0151
KNN	Cu	n_neighbors=7, leaf_size=10, p=1	0.8729	0.0026
	Zn	n_neighbors=5, leaf_size=10, p=1	0.8201	0.0246
DT	Cu	'min_samples_leaf'=1, 'min_samples_split'=10, 'min_weight_fraction_leaf'=0.1,	0.9282	0.0019
	Zn	'min_samples_leaf'=2, 'min_samples_split'=5, 'min_weight_fraction_leaf'=0.1,	0.9149	0.0136
LR	Cu	—	0.9149	0.0020
	Zn	—	0.9310	0.0120

Table S2 The predicted values of Cu for different models

Reference values (wt%)	Predicted values (wt%)						
	RF	GBDT	SVR	LR	DT	KNN	WOA-SE
0.00824	0.00784	0.00884	0.0079	0.0091	0.00928	0.00958	0.00943
0.0104	0.0088	0.0098	0.01319	0.0138	0.009	0.008999	0.00984
0.0244	0.0258	0.02676	0.0252	0.0265	0.02237	0.02137	0.02580
0.0116	0.0085	0.00986	0.00898	0.0101	0.00919	0.00919	0.01013

0.02192	0.02164	0.021676	0.02558	0.02118	0.02434	0.02434	0.02144
0.0192	0.02167	0.020698	0.0159	0.0196	0.01647	0.01647	0.01857
0.00584	0.00859	0.00896	0.0075	0.0088	0.00917	0.00917	0.00805

Table S3 The predicted values of Zn for different models

Reference (wt%)	Predicted content(wt%)						
	RF	GBDT	SVR	LR	DT	KNN	WOA-SE
0.1986	0.1806	0.16656	0.1606	0.1606	0.185225	0.1606	0.17168
0.1249	0.1223	0.1204	0.1213	0.1213	0.15595	0.1513	0.11276
0.1021	0.10543	0.10134	0.10643	0.10643	0.106024	0.11643	0.10556
0.1132	0.12524	0.11155	0.12724	0.12724	0.117195	0.12724	0.12096
0.0991	0.11331	0.1271	0.10331	0.11331	0.117195	0.10331	0.10872
0.0484	0.04216	0.0566	0.04116	0.04216	0.049652	0.05116	0.05150
0.033	0.06734	0.05839	0.06334	0.06734	0.077431	0.08734	0.04945

Table S4 Performance comparison of different data augmentation methods

Element	Method	R_{CV}^2	$RMSE_{CV}$ (wt%)	R_P^2	$RMSE_P$ (wt%)	MRE_P(%)
Cu	None	0.1918	0.0065	0.1732	0.0072	28.6
	Duplication	0.4523	0.0051	0.4315	0.0056	24.3
	Noise addition	0.6237	0.0042	0.6018	0.0046	20.5
	GAN	0.7625	0.0034	0.7456	0.0037	16.8
	VAE	0.7246	0.0037	0.7023	0.0041	18.2
	SMOTE	0.8481	0.0029	0.8413	0.0032	14.2
Zn	None	0.2833	0.0355	0.2612	0.0386	22.7
	Duplication	0.4386	0.0328	0.4125	0.0357	19.6
	Noise addition	0.5521	0.0312	0.5823	0.0338	17.4
	GAN	0.6102	0.0298	0.5864	0.0321	14.8
	VAE	0.5834	0.0307	0.5592	0.0332	15.9
	SMOTE	0.6834	0.0273	0.6751	0.0293	12.6

Table S5 Performance comparison of different optimization algorithms

Element	Models	R_{CV}^2	$RMSE_{CV}$ (wt%)	R_P^2	$RMSE_P$ (wt%)	MRE_P(%)
Cu	PSO-SE	0.9432	0.0017	0.8697	0.0024	21.51
	GA-SE	0.9462	0.0017	0.9107	0.0020	17.92
	WOA-SE	0.9529	0.0016	0.9639	0.0013	11.65

Zn	PSO-SE	0.9409	0.0116	0.8792	0.0148	14.38
	GA-SE	0.9376	0.0109	0.9178	0.0122	11.85
	WOA-SE	0.9460	0.0109	0.9614	0.0099	9.62

Table S6 Performance comparison of different ensemble mechanisms

Element	Models	R_{CV}^2	$RMSE_{CV}$ (wt%)	R_P^2	$RMSE_P$ (wt%)	MRE_P (%)
Cu	RF + GBDT	0.9158	0.0020	0.9402	0.0016	13.58
	RF + GBDT + LR	0.9286	0.0018	0.9487	0.0015	12.84
	RF + GBDT + LR + SVR	0.9389	0.0017	0.9553	0.0014	12.17
	RF + GBDT + LR +SVR + DT	0.9448	0.0016	0.9581	0.0014	11.92
	Equal weight ensemble	0.9221	0.0019	0.9305	0.0017	14.76
	WOA-SE	0.9529	0.0016	0.9639	0.0013	11.65
Zn	RF + GBDT	0.9132	0.0146	0.9217	0.0142	17.35
	RF + GBDT + LR	0.9245	0.0138	0.9304	0.0135	15.82
	RF + GBDT + LR + SVR	0.9368	0.0126	0.9412	0.0123	13.47
	RF + GBDT + LR +SVR + DT	0.9439	0.0118	0.9486	0.0116	11.83
	Equal weight ensemble	0.9184	0.0143	0.9128	0.0148	18.29
	WOA-SE	0.9460	0.0109	0.9614	0.0099	9.62

Table S7 Performance comparison of different ensemble frameworks

Element	Method	R_{CV}^2	$RMSE_{CV}$ (wt%)	R_P^2	$RMSE_P$ (wt%)	MRE_P(%)
Cu	Bagging	0.9104	0.0021	0.9466	0.0015	13.44
	Stacking	0.9395	0.0017	0.8926	0.0022	19.70
	Boosting	0.9257	0.0018	0.9137	0.0020	17.92
	WOA-SE	0.9529	0.0016	0.9639	0.0013	11.65
Zn	Bagging	0.9047	0.0141	0.8822	0.0146	14.19
	Stacking	0.9420	0.0112	0.8690	0.0154	14.97
	Boosting	0.9324	0.0110	0.8284	0.0176	17.11
	WOA-SE	0.9460	0.0109	0.9614	0.0099	9.62

Table S8 Ablation study results evaluating the necessity of data augmentation, preprocessing, and feature selection

Element	Method	R_{CV}^2	$RMSE_{CV}$ (wt%)	R_P^2	$RMSE_P$ (wt%)	MRE_P (%)
Cu	SMOTE	0.8481	0.0029	0.8260	0.0028	24.63
	D2nd	0.6042	0.0057	0.5930	0.0054	35.17
	CARS	0.6948	0.0046	0.6940	0.0046	29.69
	SMOTE + CARS	0.9156	0.0021	0.8909	0.0022	20.50

	SMOTE + D2nd	0.9037	0.0022	0.9060	0.0023	19.62
	D2nd + CARS	0.7989	0.0032	0.8710	0.0025	22.85
	SMOTE + D2nd + CARS	0.9529	0.0016	0.9639	0.0013	11.65
Zn	SMOTE	0.6834	0.0273	0.5721	0.0328	26.92
	D1st	0.5762	0.0304	0.5978	0.0294	24.68
	CARS	0.5431	0.0358	0.5229	0.0347	29.68
	SMOTE + CARS	0.8616	0.0177	0.7722	0.0239	18.74
	SMOTE + D1st	0.9033	0.0148	0.8895	0.0162	15.82
	D1st + CARS	0.9368	0.0126	0.9412	0.0123	13.47
	SMOTE + D1st + CARS	0.9460	0.0109	0.9614	0.0099	9.62