Supplementary information

Rapid Quantitative Analysis of Raw Rocks by LIBS based on Feature-based Transfer Learning

Yu Rao[a,c], Wenxin Ren[a], Weiheng Kong[a], Lingwei Zeng[a], Mengfan Wu[a], Xu Wang[a], Jie Wang[c], Qingwen Fan[a], Yi Pan[b], Jiebin Yang[b,*] and Yixiang Duan[a,*]

[a] *Research Center of Analytical Instrumentation, School of Mechanical Engineering, Sichuan University, Chengdu 610065, China*

[b] *National Institute of Measurement and Testing Technology, Chengdu, Sichuan 610021, China*

[c] *School of Mechanical Engineering, Sichuan University, Chengdu 610065, China*

[**] Corresponding author.
E-mail address: 626056334@qq.com (Jiebin Yang), yduan@scu.edu.cn (Yixiang Duan).

Table of Contents

Table S1. The elemental content of all samples. The content of Si, Al, Fe, Ca and Mg are high as major elements and Ti, Mn, K, Na and Ba are less as trace elements in the rocks. The elemental content in the table is converted by the percentage of mass fraction.

Samples	Concentration (%)									
	Si	Al	Fe	Ca	Mg	Ti	Mn	K	Na	Ba
G1	0.289	0.053	0.028	21.443	13.080	0.009	0.008	0.032	0.022	0.004
G2	3.850	0.053	0.040	23.621	10.800	0.002	0.021	0.008	0.019	0.003
G3	3.929	0.635	0.333	20.521	11.856	0.022	0.015	0.032	0.024	0.288
G4	0.873	0.109	0.171	21.586	12.510	0.005	0.012	0.015	0.009	0.234
G5	1.008	0.132	0.174	21.536	12.546	0.007	0.012	0.022	0.008	0.223
G6	2.436	0.374	0.250	21.071	12.258	0.013	0.014	0.027	0.017	0.250
G7	33.987	7.094	2.291	1.107	0.252	0.172	0.046	4.157	2.322	0.034
G8	28.289	8.561	5.289	3.714	1.032	0.309	0.060	1.568	2.864	0.102
G9	20.832	7.322	15.291	6.293	4.662	1.420	0.131	1.925	2.508	0.053
G10	25.424	9.381	5.185	0.993	0.390	0.288	0.093	6.207	5.312	0.025
G11	29.428	8.524	3.305	1.764	0.504	0.480	0.069	4.290	2.270	0.105
G12	27.851	8.767	4.244	3.371	1.686	0.462	0.073	2.904	3.005	0.190
G13	16.655	7.486	17.321	7.043	3.150	4.614	0.150	0.124	1.565	0.009
G14	33.964	6.861	2.245	0.421	0.096	0.180	0.108	4.506	1.907	0.051
G15	30.926	8.645	3.428	1.900	0.978	0.180	0.043	2.157	3.932	0.114
G16	23.156	7.285	18.760	6.857	4.320	0.551	0.160	0.398	1.536	0.062
G17	23.324	13.908	7.385	0.093	0.276	0.420	0.040	0.656	0.045	--
G18	25.046	16.581	0.231	1.286	0.050	0.018	0.015	0.954	1.892	--
G19	31.099	7.031	3.870	2.307	1.104	0.396	0.068	2.074	1.343	--
G20	32.494	7.846	3.969	0.157	0.402	0.408	0.019	3.120	0.148	0.040
G21	27.953	15.125	0.602	0.500	0.180	0.726	--	1.278	1.291	--
G22	21.126	7.264	3.388	6.986	1.044	0.331	0.185	3.568	0.190	--
G23	27.641	9.964	5.320	0.429	1.206	0.395	0.017	3.452	0.260	0.045
G24	0.784	0.180	0.112	28.029	1.044	0.010	--	0.078	0.048	--
G25	0.294	0.074	0.077	23.071	1.482	0.006	--	0.022	0.010	0.001
G26	0.509	0.127	0.077	38.593	0.486	0.006	0.005	0.070	0.013	0.001
G27	0.975	0.175	0.119	36.864	1.350	0.009	0.007	0.141	0.013	0.003
G28	1.890	0.498	0.406	35.779	1.074	0.031	0.011	0.349	0.020	0.003
G29	1.050	0.318	0.266	33.621	3.486	0.018	0.009	0.166	0.012	0.002
G30	1.129	0.198	0.175	33.364	3.930	0.014	0.004	0.041	0.008	0.001
G31	1.097	0.287	0.292	29.864	6.222	0.019	0.005	0.061	0.008	0.001
G32	1.409	0.329	0.202	36.150	1.728	0.019	0.004	0.136	0.019	0.002
G33	2.147	0.150	0.131	32.921	3.588	0.009	0.004	0.032	0.007	0.005
G34	1.549	0.180	0.127	36.586	1.458	0.012	0.003	0.078	0.005	0.001
G35	0.616	0.119	0.106	38.421	0.702	0.008	0.003	0.055	0.004	0.001
G36	1.862	0.348	0.211	34.686	2.586	0.031	0.004	0.153	0.021	0.001
G37	1.101	0.429	0.207	37.107	0.738	0.022	0.006	0.137	0.007	0.002
G38	0.448	0.174	0.102	38.714	0.492	0.010	0.004	0.070	0.004	0.001

G39	0.390	0.128	0.100	38.400	0.900	0.007	0.004	0.040	0.004	0.001
G40	7.280	2.663	3.040	25.479	3.114	0.196	0.043	0.647	0.059	0.012
G41	3.103	0.360	0.147	36.500	0.426	0.023	0.003	0.124	0.022	0.001
G42	0.257	0.090	0.252	34.207	4.056	0.007	0.007	0.036	0.016	0.001
G43	0.336	0.116	0.268	29.964	6.972	0.013	0.007	0.043	0.022	0.001
G44	0.140	0.079	0.054	39.636	0.144	0.004	0.023	0.010	0.010	0.001
G45	0.504	0.095	0.225	38.629	0.852	0.004	0.003	0.036	0.011	--
G46	45.971	0.445	0.065	0.055	0.040	0.012	0.001	0.051	0.016	0.032
G47	44.679	1.249	0.147	0.121	0.059	0.022	0.003	0.556	0.185	0.055
G48	41.809	2.901	0.336	0.243	0.096	0.061	0.008	1.718	0.809	0.067
G49	42.168	1.864	2.254	0.214	0.049	0.158	0.016	0.539	0.045	0.014
D1	2.357	0.471	0.329	36.379	0.408	--	--	0.232	0.074	0.994
D2	23.847	7.851	3.983	2.407	1.500	0.300	0.101	3.601	0.378	1.540
D3	24.729	8.333	4.431	1.286	1.404	0.306	0.046	3.676	0.378	1.083
D4	20.202	5.416	2.779	10.336	1.362	0.276	0.093	2.622	0.393	0.564
D5	19.129	5.966	3.017	10.086	1.434	0.330	0.093	3.103	0.349	0.376
D6	1.848	0.752	0.665	36.643	0.348	0.072	0.046	0.274	0.067	0.528
D7	2.081	0.826	0.546	36.600	0.360	--	0.070	0.290	0.074	0.421
D8	3.682	1.493	1.085	34.036	0.342	0.150	0.062	0.573	0.126	0.466
D9	4.219	1.599	1.918	32.707	0.678	0.108	0.209	0.581	0.141	0.510
D10	23.105	11.192	4.172	2.214	0.762	0.456	0.015	5.219	0.334	0.492
D11	2.030	0.228	0.175	36.143	1.182	--	--	0.124	0.074	0.949
D12	24.486	9.503	5.075	1.336	1.536	0.366	0.077	4.116	0.275	0.564
D13	22.115	6.983	3.304	9.029	1.506	0.444	0.070	3.659	0.282	0.439
D14	3.043	1.329	1.113	34.693	0.336	0.120	0.070	0.548	0.082	0.492
D15	4.405	1.901	1.512	32.607	0.378	0.162	0.046	0.846	0.126	0.412
D16	4.368	1.932	1.435	32.621	0.372	0.198	0.062	0.788	0.134	0.439
D17	2.464	0.926	2.261	35.307	0.438	0.072	0.077	0.249	0.126	0.851
D18	2.805	0.937	1.477	35.450	0.528	0.072	0.225	0.407	0.119	0.475

Table.S2 The optimal parameters in the RFR prediction model. The performance of the model is determined by the number and maximum depth of the decision trees. In this experiment, we employed the GridSearchCV function to conduct a grid search and cross-validation for the optimal parameters of the model, ultimately obtaining the optimal parameters for the prediction models of each element.

The model of element	Si	Al	Fe	Ca	Mg
Number of estimators	100	50	50	50	50
Max depth	10	20	20	10	30

Table.S3 The optimal parameters in the XGboost prediction model. The performance of the model is determined by the learning factors, the number and maximum depth of the decision tree.In this experiment, we use the GridSearchCV function to perform a grid search and cross-validation of the optimal parameters of the model, and finally obtain the optimal parameters of the prediction model of each element.

The model of element	Si	Al	Fe	Ca	Mg
Learning Rate	0.01	0.01	0.01	0.01	0.01
Number of estimators	100	100	100	100	100
Max depth	10	20	10	30	10

Table.S4 The comparison of predictive performance of models based on transfer learning.

Test Set		GBW Samples					
		Si	Al	Fe	Ca	Mg	Mean
PLSR	RSD	0.0860	0.1639	0.3918	0.0485	0.2290	0.1839
	MAE	5.1243	1.7406	1.5335	6.0842	0.6110	3.0187
	RMSE	38.2007	5.7529	3.7643	60.2634	0.6316	21.7226
SVR	RSD	0.0503	0.0508	0.0776	0.0189	0.0894	0.0574
	MAE	3.8565	1.5914	0.4408	2.5776	0.8213	1.8575
	RMSE	28.8695	12.7702	1.0774	14.2983	2.6972	11.9425
RFR	RSD	0.0249	0.0447	0.1107	0.0260	0.0590	0.0530
	MAE	0.2862	0.0711	0.0794	0.1426	0.0658	0.1290
	RMSE	0.2148	0.0083	0.0111	0.0730	0.0109	0.0636
XGBoost	RSD	0.0173	0.0292	0.0744	0.0212	0.0667	0.0418
	MAE	0.2115	0.0656	0.0656	0.1826	0.0841	0.1219
	RMSE	0.0656	0.0087	0.0078	0.1122	0.0127	0.0414

Table.S5 The XGBoost model was trained with the raw rock data and validated with the validation set data

Data Set		Elements					
		Si	Al	Fe	Ca	Mg	Mean
Test Set	RSD	0.0011	0.0003	0.0003	0.0037	0.012	0.0035
	MAE	0.7428	0.3995	0.2801	0.5114	0.0591	0.3986
	RMSE	0.6792	0.2112	0.1001	0.4379	0.0047	0.2866
Validation Set	RSD	0.2059	0.2605	0.2527	0.1525	0.1957	0.2135
	MAE	11.1526	3.3086	0.9704	10.2568	0.3068	5.199
	RMSE	143.8686	13.3744	1.1716	174.6432	0.1082	66.6332

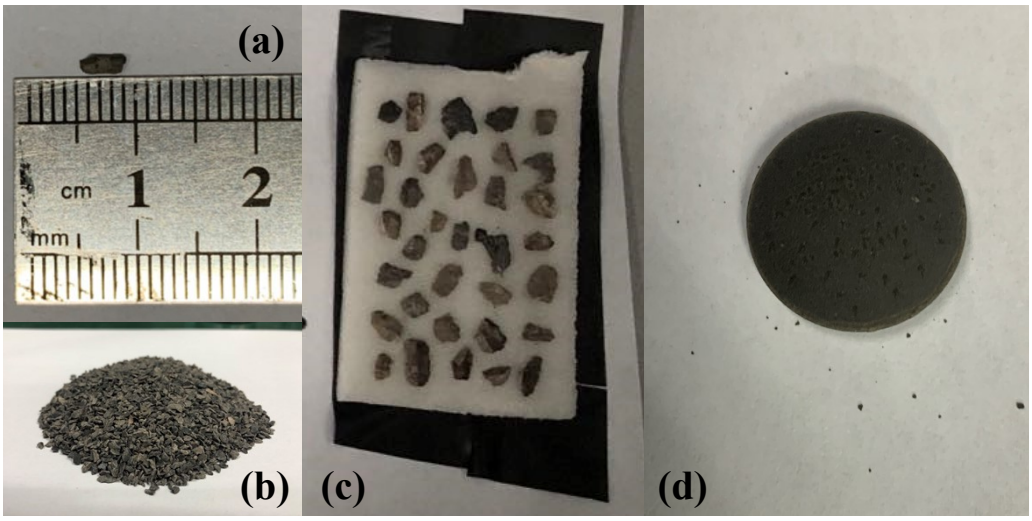

Fig.S1. In this experiment, two forms of samples (rock debris and pressed pellet) were used. (a) is the schematic of rock debris size, which ranges from 1 mm to 4 mm; (b) is all the screened rock debris of sample number D1; (c) is the 30 randomly selected rock debris, which was held in place on the sample table with double-sided adhesive tape; and (d) is the pressed pellet sample G1 of the Chinese national standard sample, 2g of each sample was placed in a pellet press and held under a pressure of 20 MPa for 5 minutes, producing pellets with a diameter of 20 mm and a thickness of 2 mm.

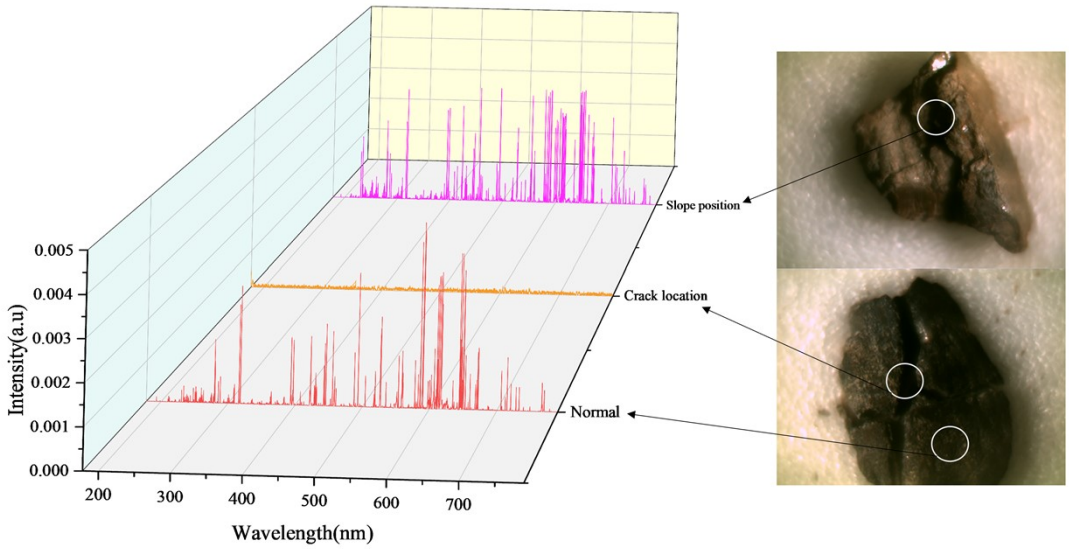

Fig.S2. The LIBS spectral data of the normal position, cracked position and slope position were compared. The cosine similarity scores were 0.9782 for the normal position, 0.1985 for the cracked position and 0.83 for the slope position.

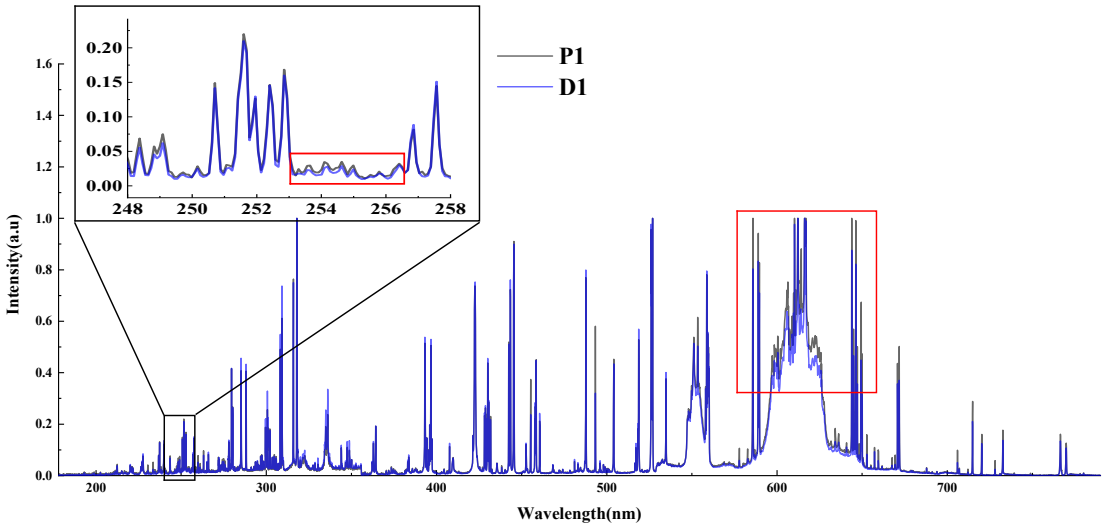

Fig.S3. The comparison between the normalized average spectrum of rock sample D1 and the corresponding spectrum of pressed sample P1. It is shows that normalization can eliminate some mechanical errors and noise interference, but it still cannot completely eliminate the differences in spectra collected under different forms (indicated by the red box).

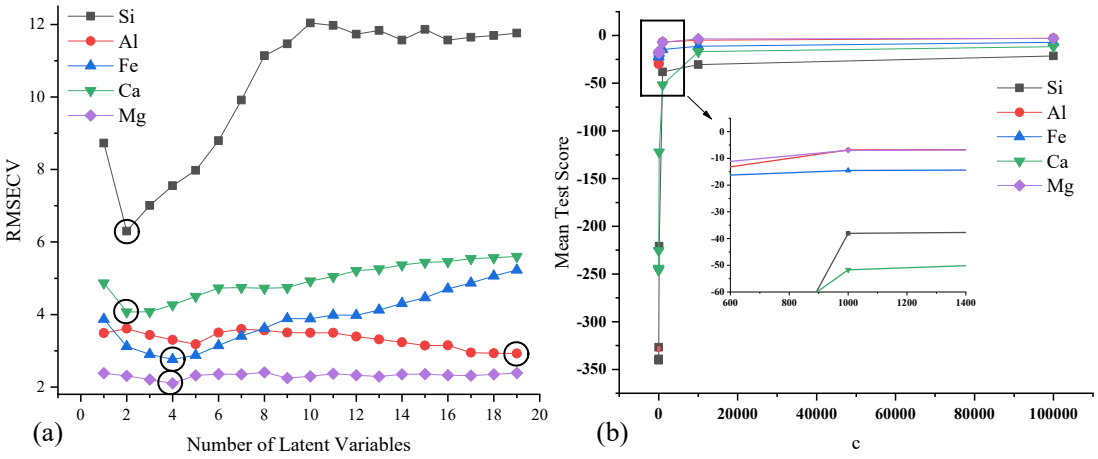

Fig.S4. The parameter optimization in PLSR and SVR algorithms. In this experiment, a ten-fold cross-validation method was employed to validate the models. (a) For PLSR, training and predicting models for Si, Al, Fe, Ca, and Mg, the impact of latent variable quantity on the root mean square error of cross-validation (RMSECV) was investigated. A smaller RMSECV value indicates better predictive capability. Hence, the chosen latent variable quantities were 2, 19, 4, 2, and 4 for Si, Al, Fe, Ca, and Mg, respectively. (b) The optimization process of SVR model parameters was conducted, considering the similarity between feature dimensions and sample quantities. The Linear kernel function was deemed more suitable for this regression task due to its advantages of fewer parameters and faster computation speed. Within the Linear kernel function, the penalty coefficient "C" was optimized. A larger "C" enhances model accuracy but weakens generalization, while a smaller "C" diminishes accuracy but improves generalization. In this experiment, the penalty factor "C" for each element in the training model was set to 1000.