Electronic Supplementary Material (ESI) for Journal of Analytical Atomic Spectrometry.

Table S1. The average 50-run results of PLS regression model on raw, SNV, 0–1 and L2 normalized LIBS coal data.

Carbon–1 (%)	RMSEC	R^2_C	RMSECV	R^2_{CV}	RMSEP	R^2_P	LVs
Raw	1.1036	0.9681	1.56	0.9357	1.4234	0.9462	9.82
SNV	0.9579	0.9759	1.295	**0.9561**	1.238	**0.9596**	9.26
0–1 scaling	0.837	0.9816	1.4276	0.9465	1.3537	0.9515	9.72
L2 normalization	1.0602	0.9706	1.4706	0.9432	1.3704	0.9506	9.8

Volatile matter–1 (%)	RMSEC	R^2_C	RMSECV	R^2_{CV}	RMSEP	R^2_P	LVs
Raw	0.9767	0.9785	1.3459	0.9589	1.2642	0.965	9.94
SNV	0.8004	0.9854	1.0588	**0.9746**	1.0537	**0.9756**	9.38
0–1 scaling	0.7225	0.9881	1.1893	0.968	1.1743	0.9696	9.58
L2 normalization	0.9158	0.981	1.2534	0.9644	1.1902	0.9688	9.82

Carbon–2 (%)	RMSEC	R^2_C	RMSECV	R^2_{CV}	RMSEP	R^2_P	LVs
Raw	0.9949	0.9592	1.5263	0.9042	1.518	0.8952	9.9
SNV	0.851	0.9701	1.5688	0.8988	1.5685	0.888	9.82
0–1 scaling	0.6667	0.9816	1.6087	0.8936	1.6068	0.8828	9.82
L2 normalization	0.9279	0.9645	1.519	**0.9051**	1.5055	**0.8966**	9.98

Volatile matter–2 (%)	RMSEC	R^2_C	RMSECV	R^2_{CV}	RMSEP	R^2_P	LVs
Raw	1.2545	0.9796	1.9325	**0.9517**	1.835	**0.9547**	9.98
SNV	1.1374	0.9833	1.9991	0.9483	1.914	0.9507	10
0–1 scaling	0.8986	0.9895	2.1076	0.9425	1.9665	0.948	9.88
L2 normalization	1.2137	0.981	1.9485	0.9509	1.847	0.954	10

Ash (%)	RMSEC	R^2_C	RMSECV	R^2_{CV}	RMSEP	R^2_P	LVs
Raw	0.5726	0.9709	0.8745	**0.9328**	0.8682	**0.9309**	9.68
SNV	0.5348	0.9748	0.9018	0.9287	0.9015	0.9257	9.7
0–1 scaling	0.4418	0.9823	0.9383	0.9228	0.9307	0.9211	9.62
L2 normalization	0.567	0.9717	0.8898	0.9306	0.8865	0.9282	9.76

Calorific value (MJ/kg)	RMSEC	R^2_C	RMSECV	R^2_{CV}	RMSEP	R^2_P	LVs
Raw	0.5617	0.957	0.8843	0.8937	0.8607	0.8905	9.74
SNV	0.508	0.9643	0.9006	0.8902	0.8885	0.8836	9.54
0–1 scaling	0.3921	0.9786	0.9207	0.8852	0.9069	0.8781	9.58
L2 normalization	0.5239	0.9625	0.8745	**0.8961**	0.8558	**0.8916**	9.88

Sulphur (%)	RMSEC	R^2_C	RMSECV	R^2_{CV}	RMSEP	R^2_P	LVs
Raw	0.0961	0.9818	0.1524	**0.9542**	0.1467	**0.9563**	9.98
SNV	0.0997	0.9804	0.1752	0.9394	0.1738	0.9388	10
0–1 scaling	0.0764	0.9884	0.1711	0.9423	0.165	0.9449	9.84
L2 normalization	0.0969	0.9815	0.1562	0.9519	0.1557	0.9507	10

Table S2. The 50-run average calibration and 5-fold cross-validation results of PLS regression on two LIBS data with full variables and subsets of variables selected by different methods.

Carbon–1 (%)	RMSEC	R^2_C	RMSECV	R^2_{CV}	LVs
All variables	0.9395	0.9772	1.2843	0.9574	9.38
Peak area	1.2198	0.9615	1.8209	0.9142	9.64
RC	0.8833	0.9798	1.1158	0.9678	9.18
LASSO	0.4941	0.9937	0.7641	0.9849	9.96
CARS	0.4862	0.9939	0.6863	0.9878	9.84
rPLS	0.8249	0.9824	1.0187	0.9732	8.62
sMC	0.8553	0.9811	1.0989	0.9688	9.34
mRMR	1.0596	0.971	1.3471	0.953	9.86
Ensemble	0.6529	0.9888	0.925	0.9778	9.74

Volatile matter–1 (%)	RMSEC	R^2_C	RMSECV	R^2_{CV}	LVs
All variables	0.807	0.9853	1.0612	0.9746	9.42
Peak area	1.1214	0.9717	1.6312	0.94	9.56
RC	0.7665	0.9867	0.9623	0.9791	9.18
LASSO	0.4115	0.9962	0.6193	0.9913	10
CARS	0.422	0.996	0.5765	0.9925	9.72
rPLS	0.7074	0.9887	0.8417	0.984	8.46
sMC	0.7471	0.9873	0.9323	0.9804	8.94
mRMR	0.8896	0.9822	1.0926	0.973	9.94
Ensemble	0.5035	0.9941	0.7188	0.9882	9.88

Carbon–2 (%)	RMSEC	R^2_C	RMSECV	R^2_{CV}	LVs
All variables	1.0168	0.9561	1.553	0.898	9.82
Peak area	1.0023	0.9573	1.6057	0.891	8.82
RC	1.0417	0.9538	1.3802	0.9194	8.94
LASSO	0.7128	0.9785	1.099	0.9485	10
CARS	0.6361	0.9828	0.908	0.965	9.88
rPLS	0.9608	0.961	1.1876	0.9404	9.04
sMC	1.2661	0.9312	1.5706	0.8944	8.12
mRMR	1.1554	0.9434	1.6478	0.8853	9.52
Ensemble	0.7478	0.976	1.1156	0.9471	9.9

Volatile matter–2 (%)	RMSEC	R^2_C	RMSECV	R^2_{CV}	LVs
All variables	0.5656	0.9723	0.8864	0.9321	9.76
Peak area	0.563	0.9722	0.9348	0.9244	9.26
RC	0.5866	0.9701	0.7798	0.9476	8.5
LASSO	0.3806	0.9875	0.6152	0.9673	9.9
CARS	0.4056	0.9857	0.5511	0.9737	9.44
rPLS	0.5537	0.9735	0.6814	0.9597	8.34
sMC	0.6266	0.9661	0.8051	0.9439	8.16
mRMR	0.6497	0.9637	0.9044	0.9292	9.56
Ensemble	0.4118	0.9851	0.6445	0.9639	9.82

Ash (%)	RMSEC	R^2_C	RMSECV	R^2_{CV}	LVs
All variables	1.2595	0.9788	1.9212	0.9507	9.96
Peak area	1.2649	0.9786	2.0812	0.9422	9.44
RC	1.2721	0.9783	1.7185	0.9606	8.82
LASSO	0.8701	0.9899	1.3279	0.9764	9.98
CARS	0.8348	0.9907	1.1651	0.9819	9.6
rPLS	1.166	0.9819	1.4365	0.9722	9.08
sMC	1.3439	0.9759	1.8019	0.9565	9.22
mRMR	1.4325	0.9727	2.0261	0.9452	9.82
Ensemble	0.8729	0.9897	1.3194	0.9766	9.94

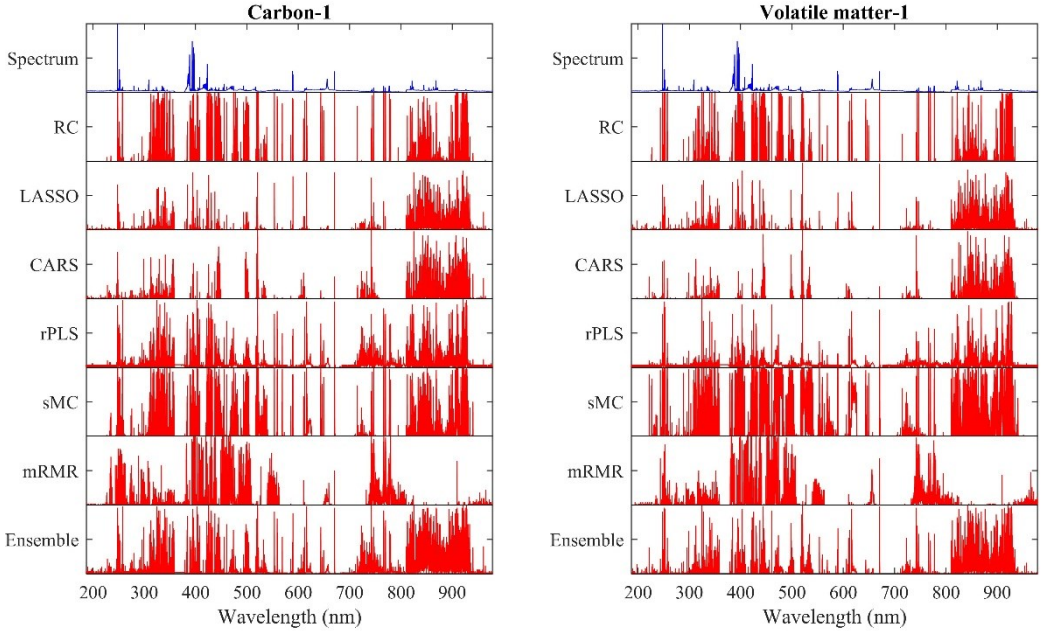
Calorific value (MJ/kg)	RMSEC	R^2_C	RMSECV	R^2_{CV}	LVs
All variables	0.5601	0.9558	0.8911	0.888	9.84
Peak area	0.524	0.9611	0.8841	0.8897	9.06
RC	0.5721	0.9539	0.7772	0.9148	8.8
LASSO	0.3819	0.9794	0.6212	0.9451	10
CARS	0.3396	0.9837	0.5057	0.9639	9.86
rPLS	0.5226	0.9617	0.6525	0.9398	9.16
sMC	0.6946	0.9308	0.8721	0.8905	7.92
mRMR	0.639	0.9428	0.9079	0.8838	9.48
Ensemble	0.393	0.9776	0.6171	0.9457	9.92

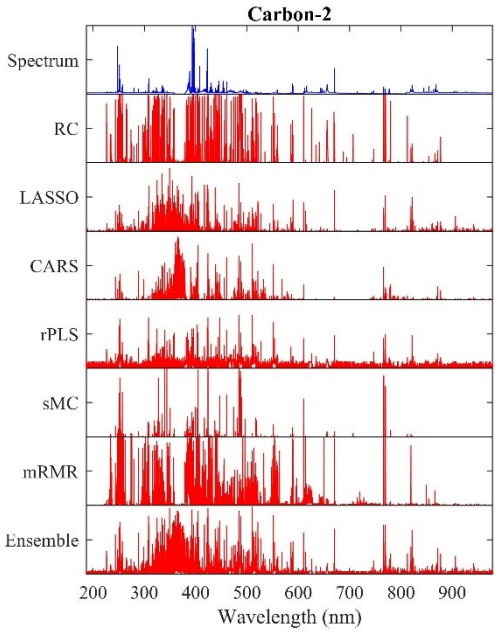
Sulphur (%)	RMSEC	R^2_C	RMSECV	R^2_{CV}	LVs
All variables	0.0958	0.9821	0.1531	0.9542	10
Peak area	0.0821	0.9868	0.1464	0.9581	9.66
RC	0.0914	0.9837	0.1261	0.969	9.68

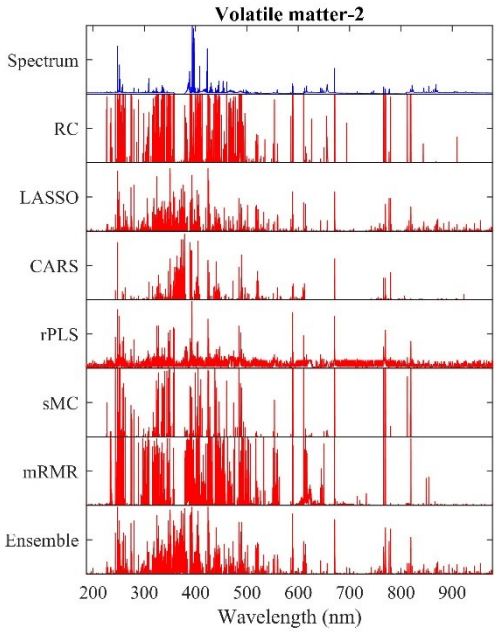
LASSO	0.0608	0.9928	0.0911	0.9837	10
CARS	0.063	0.9922	0.0853	0.9858	9.58
rPLS	0.0911	0.9838	0.1169	0.9733	8.96
sMC	0.1036	0.979	0.1381	0.9628	9.02
mRMR	0.0987	0.981	0.1376	0.9631	9.7
Ensemble	0.0717	0.9898	0.1038	0.9788	9.94

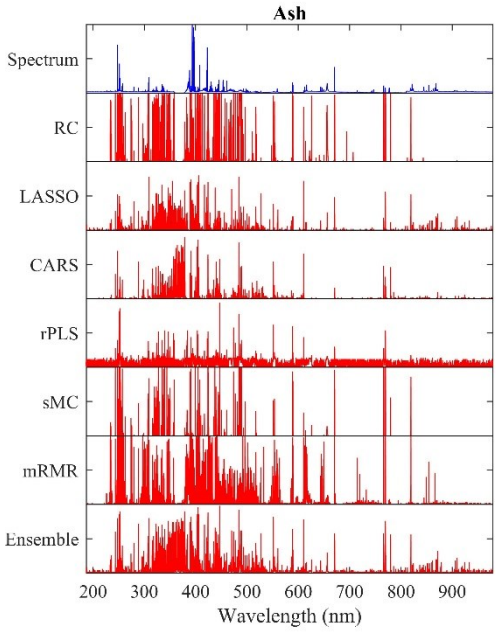
Table S3. The 50-run average standard deviation of all variable selection methods for different tasks in prediction.

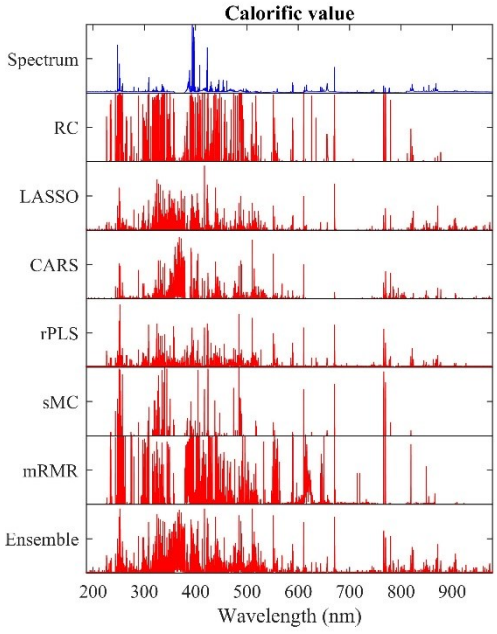
	Carbon-1	Volatile matter-1	Carbon-2	Volatile matter-2	Ash	Calorific value	Sulphur
All variables	0.0916	0.0795	0.1575	**0.0908**	0.2027	0.0941	0.0161
Peak area	0.1761	0.1901	**0.1348**	0.1045	0.1807	**0.0683**	0.0173
RC	0.1074	0.0809	0.1997	0.095	0.2192	0.107	0.0131
LASSO	0.0918	0.0577	0.2081	0.11	0.2102	0.0934	0.0127
CARS	0.1089	0.097	0.1466	0.11	0.2168	0.1229	0.0151
rPLS	0.1022	0.0664	0.1871	0.1039	0.2091	0.1136	0.0155
sMC	**0.0893**	0.0878	0.2429	0.098	0.2482	0.1381	0.0165
mRMR	0.1094	0.0854	0.2161	0.1166	0.2191	0.1079	0.0126
Ensemble	0.0981	**0.0526**	0.1664	0.0998	**0.1745**	0.1052	**0.0121**




Carbon-2
Spectrum
RC
LASSO
CARS
rPLS
sMC
mRMR
Ensemble
200 300 400 500 600 700 800 900
Wavelength (nm)


Volatile matter-2
Spectrum
RC
LASSO
CARS
rPLS
sMC
mRMR
Ensemble
200 300 400 500 600 700 800 900
Wavelength (nm)


Ash
Spectrum
RC
LASSO
CARS
rPLS
sMC
mRMR
Ensemble
200 300 400 500 600 700 800 900
Wavelength (nm)


Calorific value
Spectrum
RC
LASSO
CARS
rPLS
sMC
mRMR
Ensemble
200 300 400 500 600 700 800 900
Wavelength (nm)

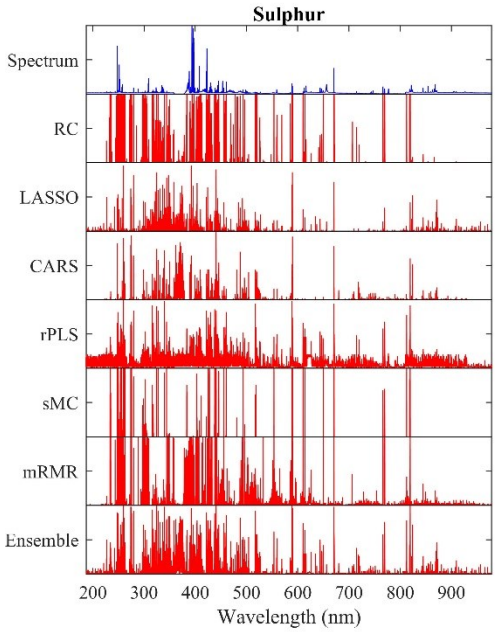
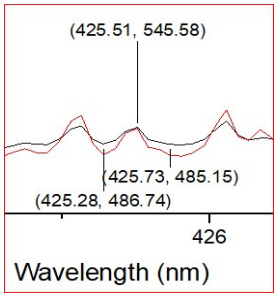
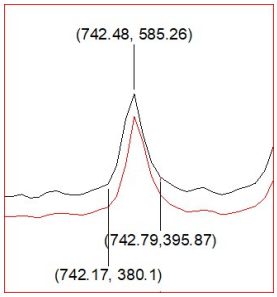
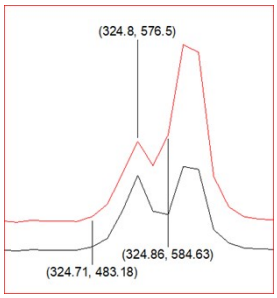
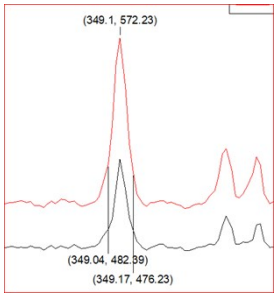
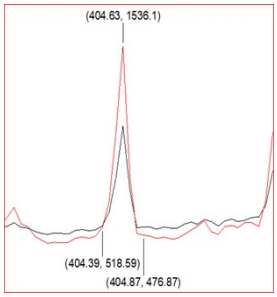
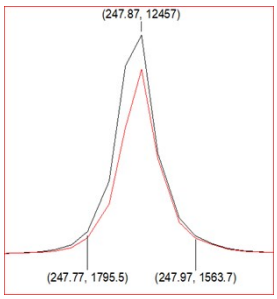
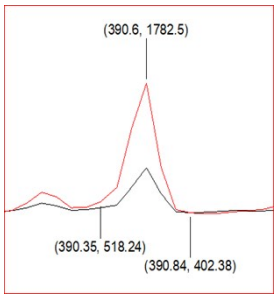
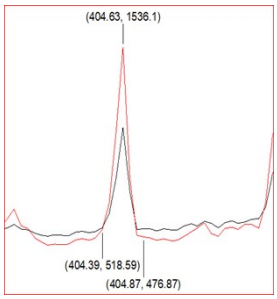
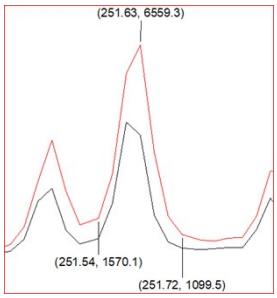
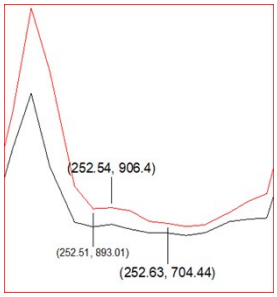
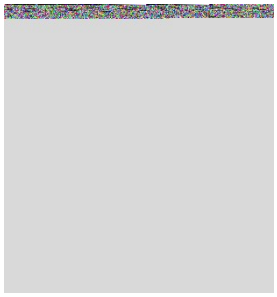
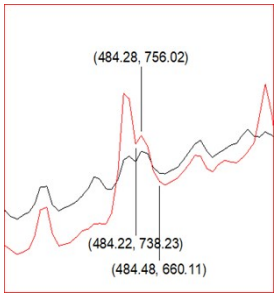
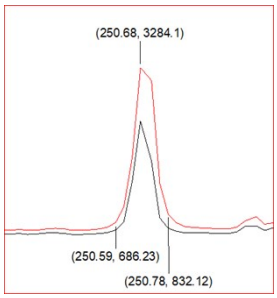
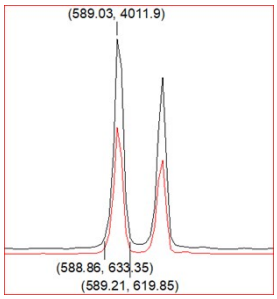


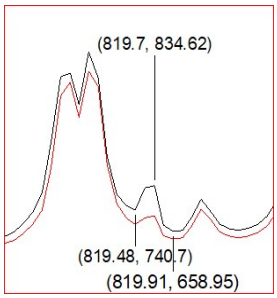
Fig. S1. The frequency of each variable selected by different methods based on the results of the 50-run. For each task plot, the Y-axis of the top subplot represents intensity, and that of the remaining subplots represents frequencies from 0 to 50.

Table S4. The emission lines with high frequency for different coal properties. The black spectral line represents an averaged spectrum for set–1 and red spectra shows averaged spectrum for set–2.

Line Shape	Property	Wavelengths (nm)			Main reference NIST (Spectrometer resolution: 0.1 nm)
		Selected	Peak	Range	
	Carbon–1	425.51	425.51	[425.28, 425.73]	Strong Fe (I) line with a couple of C (III) lines around this wavelength. However, C (III) lines may already been decayed at 1 μs delay. Multiple C (I) lines in the vicinity are also present which are not resolved due to the limited spectrometer resolution. So, safely this can be attributed to Fe (I) lines.
	Carbon–1	742.38	742.48	[742.17, 742.79]	No C line in the wavelength range. In fact, it is a strong N (I) line, the correlation of N (I) line to C concentration may be result of matrix effect, and is not reasonable on physical basis.

	Vad–1 Carbon–1	324.8	324.8	[324.71, 324.86]	Cu (I) 324.75 nm, or combination of this with other metals/heavy metals including V (I). Matrix effect.
	Carbon–2	349.13	349.1	[349.04, 349.17]	Very strong Fe (I) at 349.057 nm. Matrix Effect.
	Carbon–2	404.79	404.63	[404.39, 404.87]	Very strong Fe (I) at 404.581 nm. Matrix effect.
	Vad–2 Carbon–2	247.97	247.87	[247.77, 247.97]	**C (I) 247.856 nm. Rational Correlation.**
	Vad–2 Carbon–2	390.76	390.6	[390.35, 390.84]	Strong Fe (I) 390.65 nm, V (I) 390.675 nm. Matrix effect.
	Vad–2 Carbon–2	404.79	404.63	[404.39, 404.87]	Very strong Fe (I) at 404.581 nm. Matrix effect.

	Ash*	251.66	251.63	[251.54, 251.72]	Zn (I) 251.580, Fe (II) 251.713, V (I) 251.715 nm and other Fe lines. Reasonable as heavy metals are a considerable part of ash.
	Ash	252.6	252.54	[252.51, 252.63]	A mix of multiple Co (II), Fe (II) and V (I) lines.
	Ash	404.79	404.63	[404.39, 404.87]	Very strong Fe (I) 404.581 nm.
	Ash	484.41	484.28	[484.22, 484.48]	Fe (I) 484.314, V (I) 484.300 nm are strong lines in this range.
	Calorific value**	250.72	250.68	[250.59, 250.78]	Very strong Cu (II) 250.627, Fe (II) 250.609, Fe (II) 250.643 nm. Matrix effect.
	Sulphur	589.00	589.03	[588.86, 589.21]	The strong famous peak of Na (I) with a very weak S (I) line. The Na (I) emission can relate to S and combination of Na (I) and S (I) lines are responsible for this correlation. At least, this correlation has a logical basis and different machine learning algorithms determined a right correlation.

	Sulphur	819.7	819.7	[819.48, 819.91]	As the peak is clear, it may be the V (I) line if we see the previous correlations. A strong V (I) line is + 0.17 nm apart. Possibility that line since resolution is lower at higher wavelengths. Matrix effect.

*** Ash content may be determined by heavy meatal lines**
**** C, H, O and S lines important for Calorific heat value**