energies

MDPI

Article

Rapid Quantitation of Coal Proximate Analysis by Using Laser-Induced Breakdown Spectroscopy

Yulin Liu, Dongming Wang and Xiaohan Ren *

Institute of Thermal Science and Technology, Shandong University, Jinan 250061, China; 202014243@mail.sdu.edu.cn (Y.L.); 202134238@mail.sdu.edu.cn (D.W.)
* Correspondence: xiaohan09126@gmail.com

Abstract: Proximate analysis of coal is of great significance to ensure the safe and economic operation of coal-fired and biomass-fired power generation units. Laser-induced breakdown spectroscopy (LIBS) assisted by chemometric methods could realize the prediction of coal proximate analysis rapidly, which makes up for the shortcomings of the traditional method. In this paper, three quantitative models were proposed to predict the proximate analysis of coal, including principal component regression (PCR), artificial neural networks (ANNs), and principal component analysis coupled with ANN (PCA-ANN). Three model evaluation indicators, such as the coefficient of determination (R^2), root-mean-square error of cross-validation (RMSECV), and mean square error (MSE), were applied to measure the accuracy and stability of the models. The most accurate and stable prediction of coal proximate analysis was achieved by PCR, of which the average R^2, RMSECV, and MSE values were 0.9944, 0.39%, and 0.21, respectively. Although the R^2 values of ANN and PCA-ANN were greater than 0.9, the higher RMSECV and MSE values indicated that ANN and PCA-ANN were inferior to PCR. Compared with the other two models, PCR could not only achieve accurate prediction, but also shorten the modeling time.

Keywords: LIBS; coal; proximate analysis; PCR; ANN

Citation: Liu, Y.; Wang, D.; Ren, X. Rapid Quantitation of Coal Proximate Analysis by Using Laser-Induced Breakdown Spectroscopy. *Energies* **2022**, *15*, 2728. https://doi.org/10.3390/en15082728

Academic Editor: Dimitrios A. Georgakellos

Received: 22 February 2022
Accepted: 31 March 2022
Published: 8 April 2022

1. Introduction

In the next decades, coal is expected to remain an important energy source and comprise a large proportion of energy consumption worldwide; for instance, more than half of the electric power is supplied by coal-fired power plants in China [1]. Coal-fired burners face an urgent problem in that the flame combustion stability needs to be settled, which is not conducive to ignition, combustion efficiency, extinction, and pollutant control. If the physical and chemical properties of coal could be determined in time, it could lead to extensive advances in combustion control and optimization. Proximate analysis is a key parameter that could preliminarily distinguish the types of coal and assess the coal quality, contributing well to the knowledge of the changes in fuel characteristics in boilers [2]. Traditional chemical measurement is time-consuming and labor-intensive, as characterizing coal requires involved sample preparation and cannot obtain multi-component information simultaneously, which is unfavorable for real-time boiler combustion optimization.

Many rapid sample analysis techniques exist, such as prompt gamma neutron activation analysis (PGNAA) [3], X-ray fluorescence (XRF) [4], and inductively coupled plasma-optical emission spectroscopy (ICP-OES) [5]. However, the neutron source of PGNAA has potential radiation hazards; XRF cannot analyze C, H, and other low-atomic-number elements; and ICP–OES has several weaknesses; for example, the costly analysis is derived from a large amount of argon consumption and the occurrence of several matrix effects is difficult to avoid or mitigate. As many limitations exist in practical applications, searching for an efficient and convenient method to assess coal quality is necessary. Laser-induced breakdown spectroscopy (LIBS) is an atomic emission analysis approach that is

based on plasma techniques [6]. Under the action of an intense pulsed laser, the ablation point of the sample is ionized instantaneously to form high-temperature plasma. The plasma radiates different characteristic spectral emission lines during the cooling process, providing conditions for qualitative and quantitative analysis. Less than 10 min is needed to complete the whole process, including sample preparation and the prediction of chemometrics. Owing to its simple pretreatment process [7], rapid response, non-destructive testing, and simultaneous multi-element measurement [8], LIBS has unique value in some rapid detection fields, including coal analysis [9], soil testing [10], and food safety [11].

After LIBS spectra are obtained, a chemometric process is needed to further predict the properties of coal and ash samples [12]. The existing chemometric methods, including principal component analysis (PCA) [13], partial least squares (PLS) [14], support vector machine (SVM) [15], and ANNs [16], could improve the accuracy of consequence. Zhang et al. [16] investigated the coal fusion characteristics directly from coal instead of coal ash by using a LIBS experimental setup that was assisted by PLS, SVM, and an ANN. The ash content, volatiles, fixed carbon, and moisture were integrated into the model to improve the prediction accuracy of coal ash fusion temperatures (AFTs). The RMSECV values were 4.88 °C and 9.11 °C, and the corresponding root-mean-square error of prediction (RMSEP) values were 8.15 °C and 11.3 °C, respectively. The relationships between the AFTs and the intensities of the spectral lines of the LIBS spectra were further explored, which demonstrated the variation trend of AFTs with elemental content. In addition to the single chemometric method, some applications of hybrid methods, such as independent component analysis–wavelet neural network (ICA-WNN) and genetic algorithm-ANN (GA-ANN), have been widely used in industrial fields. Zhang et al. [17] employed PCA to extract the LIBS spectra of coal ash, and then independent component analysis (ICA) was used to optimize the input variables for the wavelet neural network (WNN). LIBS coupled with ICA-WNN, with 100% classification accuracy, was confirmed to classify coal ash by principal components. Lu et al. [18] developed LIBS assisted by GA-ANN to estimate the gross calorific value of continuous coal particle flows. As a result, the mean standard deviation (MSD) and the mean absolute error (MAE) of the gross calorific value for the prediction set were 0.38 and 0.39 MJ/kg, which could meet the Chinese national standard (GB/T 29161-2012) in the neutron activation method of coal analysis.

This work aimed at offering a more suitable quantitative method that was based on LIBS to predict the proximate analysis of coal expediently and precisely. Pulverized coal with added KBr binder was mixed and then pressed into a tablet, and sixteen samples were obtained in total. The LIBS experimental setup was established to investigate the atomic and molecular emissions related to the proximate analysis. The PCR, ANN, and PCA-ANN models, namely linear, nonlinear, and a combination of these two, were applied to predict coal proximate analysis, in which their quantitative accuracy and stability were analyzed and compared.

2. Materials and Methods

2.1. Coal Samples

Coal is a significant basic energy source around the world, with abound reserves, low cost, and high heating values. Two types of coal were investigated in this work, including bituminous coal and anthracite from Hebei Province, China. The raw coal was ground into finely ground powders, with a particle size of less than 75 μm, because coal is a heterogeneous mixture with a complex chemical composition. The powders mixed with KBr binder were placed in an agate mortar to make the sample uniform and eliminate the matrix effect of coal. Then, the mixture was pressed into tablets using a tablet press with a diameter of 20 mm and a thickness of 10 mm under 15 MPa pressure. The function of KBr was to enhance the intensities of the spectra and bond coal powders closely [19]. Bromide could not only inhibit the gas-phase radicals to control the chemical reaction, but could also consume H/OH radicals in the reaction zone to contribute to a stable condition [20].

According to the literature [21], the appropriate mass fraction of KBr was determined as 90 wt. %.

On the basis of the Chinese national standard (GB/T 212-2008), coal proximate analysis was carried out in an electric oven (5E-DHG) and muffle furnace (5E-MF6000). The heating values of the coal were measured with a calorimeter (5E-AC/PL) in accordance with standard GB/T 213-2008. All the instruments of coal proximate analysis mentioned above were produced by Changsha Kaiyuan Instrument, Ltd., Changsha, China. The proximate analyses were detailed in Ref. [22], and the results are shown in Table 1.

Table 1. Proximate analysis of 16 coal samples (air-dried basis).

No.	Moisture	Volatiles	Ash	Fixed Carbon [1]	Heating Value	Coal Type [2]
	(wt. %)	(wt. %)	(wt. %)	(wt. %)	(MJ/kg)	
1	9.15	29.15	9.31	52.39	29.42	Bituminous coal
2	5.61	31.13	6.35	56.91	29.65	Bituminous coal
3	3.71	30.49	6.81	58.99	30.18	Bituminous coal
4	2.12	31.33	9.00	57.55	29.84	Bituminous coal
5	3.74	30.94	3.84	61.48	30.55	Bituminous coal
6	3.95	29.86	5.55	60.64	30.24	Bituminous coal
7	8.18	32.25	3.23	56.34	29.75	Bituminous coal
8	4.50	31.28	4.97	59.25	30.00	Bituminous coal
9	3.71	30.35	6.87	59.07	30.09	Bituminous coal
10	4.04	29.68	5.59	60.69	30.42	Bituminous coal
11	0.14	17.63	10.54	71.69	32.16	Bituminous coal
12	2.11	7.07	17.61	73.21	27.19	Anthracite
13	2.25	7.38	11.83	78.54	30.23	Anthracite
14	0.59	25.65	8.72	65.04	31.93	Bituminous coal
15	0.80	12.52	27.01	59.67	23.77	Bituminous coal
16	1.64	6.51	21.87	69.98	25.48	Anthracite

[1] Determined by difference; [2] Classified by Chinese National Standard (GB5751-86).

2.2. Experimental Setup

The coal samples were assessed using a mobile LIBS system with a coaxial structure, as shown in Figure 1. A Q-switched Nd: YAG laser was used for a laser source of 1064 nm in wavelength, 1–20 Hz in repetition rate, and 7 ns in pulse width. The parameters of the linearly polarized pulses were optimized to 1 Hz in repetition rate and 10 mJ/pulse in energy (less than 1%). The LIBS signal, collected by two plano-convex lenses with 50 mm focal length, was accessed by a 50 µm single-channel spectrometer (AvaSpec-ULS4096-EVO, Avantes). The single-channel spectrometer could record the plasma emission in the wavelength range of 200–1100 nm with a resolution of 0.05–0.24 nm. Under the circumstance of a high signal-to-noise ratio (SNR), the delay time and integration time were set to 1.2 and 30 µs, respectively. The coal sample was placed on a 2D stage, which could be shifted continuously in a serpentine (S-shaped) path relative to the laser, to increase the reproducibility of the measurement. High-purity air (21% O_2 and 79% N_2) at a flow rate of 1 L/min was applied to purge the plasma region, which increased the intensity of the emission lines and decreased the limit of detection [23].

Some differences exist in the intensities of the spectral emission lines between ablation points. Every coal tablet was divided into 16 parts (4 × 4), and each part was placed at 16 ablation points (4 × 4) with a spacing of 500 µm. The spectral intensities of the 256 ablation points were averaged to represent the spectrum of each coal sample.

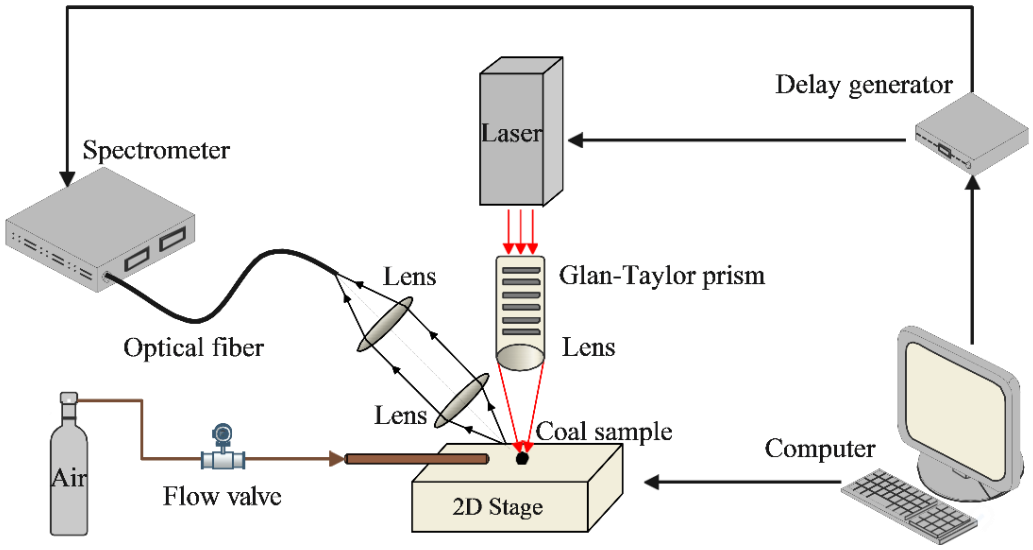

Figure 1. Schematic of the LIBS experimental setup.

2.3. Spectral Pretreatment

Taking the average spectrum of 256 spectra could enhance the repeatability of the LIBS signal and improve the accuracy of quantitative analysis. For example, the calculated relative standard deviation (RSD) of Fe 461.88 nm emission (one of 256 spectra selected in Section 3.1) was 0.09. However, the baseline of the average spectrum was prone to drift. Removing the baseline has a significant effect on increasing the SNR. The baseline was removed by outlier elimination and the first-order derivative [24], which could enhance the implicit peaks and separate the overlapped peaks [25].

$$I = I_{raw} - I_{baseline} \quad (1)$$

where I is the intensity of the corrected spectrum, I_{raw} is the intensity of the raw spectrum, and $I_{baseline}$ is the intensity of the estimated baseline.

2.4. Chemometric Methods

2.4.1. PCR

PCA combines the independent variables into a few principal components that could fully reflect the overall information through linear transformation to avoid the collinearity problem between variables. PCA could not only distinguish and remove the redundant spectral emission lines, but also extract the most related information [26]. Then, the principal components are used for the multiple linear regression (MLR) of dependent variables, namely, PCR.

Herein, the independent variables are the intensities of the selected spectra, referring to the characteristic emission lines in Section 3.1, and the dependent variable is the proximate analysis. The cumulative explained variance of the selected principal components was set to 90% or 95% [27]. In accordance with the component matrix generated by PCA, the principal components were expressed by the original variables for linear regression. Finally, MLR was used to map the selected principal components to the coal proximate analysis. Compared with MLR, PCR reduces the number of input variables and improves the stability and fitness of the regression. The key to the accuracy and stability of the PCR model is the number of principal components.

2.4.2. ANNs

Originating from biology, ANNs combine mathematics with physics to abstract the human brain's neural network from the perspective of information processing. ANNs are usually composed of an input layer, a hidden layer, and an output layer, with basic units

called neurons. The input data are inputted into the activation function in the hidden layer after weighted summation to calculate the output value. Meanwhile, the value of the loss function is calculated to update the weight. ANNs have strong input-output nonlinear mapping, self-adaptability, and learning ability. However, ANNs have a low study rate and local minimum solution, leading to the inaccuracy of predictive performance. ANNs cannot easily study the contribution of input variables to the output [28]. Though traditional ANNs could predict quantitative analysis [29], a large number of input variables could increase the difficulty of the training model and reduce the predictive precision [30].

Trained with a backpropagation algorithm, the ANN model consists of sigmoid hidden layer neurons and linear output neurons. The more neurons in the hidden layer, the higher the prediction accuracy. If the network is too complex, it may not converge. Therefore, the trade-offs between network complexity and training efficiency are worthy of attention. The number of hidden layer neurons is roughly determined by the following equation [31]:

$$N^{(h)} = \sqrt{N^{(i)} \cdot N^{(o)}} \quad (2)$$

where $N^{(h)}$ is the number of hidden layer neurons, $N^{(i)}$ is the number of input neurons, and $N^{(o)}$ is the number of output neurons.

2.4.3. PCA–ANN

The selected principal components with large variances were first used to replace the original variables to eliminate information redundancy in the raw data. An ANN assisted by PCA was used to reduce the input variables and improve the efficiency of training models. Compared with the ANN model, the PCA-ANN model simplifies the network structure and computational process.

The principal component values calculated by the component matrix and the coal proximate analysis were treated as independent and dependent variables for training the ANN model, respectively. When the principal component scores of a coal sample are inputted into the trained ANN model, a predicted proximate analysis result is outputted. The feature extraction of the LIBS spectrum performed by PCA-ANN can be found in previous research [32].

2.5. Model Evaluation Indicators

Leave-one-out cross-validation (LOO-CV) was adopted to verify the generalization effect of these three models. Three indicators were used to evaluate the accuracy and stability of the models. R^2 represents the fitting degree of a model (a value equal to one is a perfect fit). RMSECV describes the deviation between the predicted and real values (a value equal to zero is a perfect fit). MSE enlarges the value with a large deviation and compares the stability of different models (a value equal to zero is a perfect fit).

$$R^2 = 1 - \frac{\sum_i^n (y_i - \hat{y}_i)^2}{\sum_i^n (y_i - \overline{y}_i)^2} \quad (3)$$

$$RMSECV = \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n-1}} \quad (4)$$

$$MSE = \frac{1}{n}\sum_i^n (y_i - \hat{y}_i)^2 \quad (5)$$

where y_i, $\overline{y}_i$, and $\hat{y}_i$ are the real, mean, and predicted values from the training set, respectively; and n is the number of samples in the training set.

3. Results and Discussion

3.1. Selection of Characteristic Emission Lines

Five indices of proximate analysis, which consisted of moisture, ash, volatiles, fixed carbon, and heating value, were utilized to assess the coal quality in this context. The organic emission lines (e.g., C, H, and O) associated with these five indices were suitable to integrate into noise; thus, setting up a database alone was difficult [15]. The molecular spectra of CN and C_2 were chosen to complement the reduction in atomic C emissions, because the collisions of O and N with C in the air enhanced the formation of molecular species, such as CN, C_2, and CO [33]. Several mineral elements correlated with volatiles and ash were also taken into consideration to cover all elements in coal as much as possible. Therefore, the database was composed of 63 emissions, including C, H, O, N, Li, Na, Mg, Al, Si, Ca, Ti, Fe, C_2, and CN [34], as shown in Table 2. Owing to the addition of 90 wt. % KBr, the intensity of the K emission line (766.49 nm) was the highest in all wavelengths, as shown in Figure 2. Thus, the K emission lines were eliminated to prevent the high concentration from affecting the prediction accuracy. The removal of the baseline not only retained the original appearance of spectral peaks, but also improved the SNR, which contributed to the selection of characteristic emission lines. The basic considerations for selecting the emission lines were as follows [35]:

- High intensity to obtain a high SNR;
- High probability of excitation to ensure the repeatability of the experiment;
- Interference-free spectrum to exclude the influence of spectral line overlap.

Table 2. Emission lines related to coal proximate analysis.

Type	Name	Emission Lines [1] (nm)
Organic atom	C	247.86 (I [2])
	H	656.29 (I)
	O	777.19 (I), 844.68 (I)
	N	744.23 (I), 746.83 (I)
Inorganic atom	Si	220.80 (I), 221.67 (I), 250.69 (I), 251.43 (I), 251.61 (I), 251.92 (I), 252.85 (I), 288.16 (I)
	Li	670.78 (I)
	Na	589.00 (I), 589.59 (I)
	Mg	279.55 (II [3]), 280.27 (II), 285.21 (I)
	Al	226.91 (I), 237.31 (I), 308.22 (I), 309.27 (I), 394.40 (I), 396.15 (I)
	Ca	315.89 (II) 317.93 (II), 393.37 (II), 396.85 (II), 422.67 (I), 443.50 (I), 445.48 (I), 610.27 (I), 612.22 (I), 616.13 (I), 643.91 (I), 646.26 (I), 854.21 (II), 866.21 (II)
	Ti	334.94 (II), 336.12 (II), 398.18 (I), 398.96 (I), 399.90 (I)
	Fe	234.35 (II), 238.18 (I), 239.33 (II), 240.48 (II), 258.55 (I), 260.65 (II), 261.11 (II), 273.92 (I), 274.89 (II), 275.55 (I), 404.58 (I), 461.88 (I)
Molecule	C_2	469.76, 471.52, 473.71
	CN	386.19, 387.14, 388.34

[1] Data from NIST; [2] Atomic spectral line; [3] Ion spectral line.

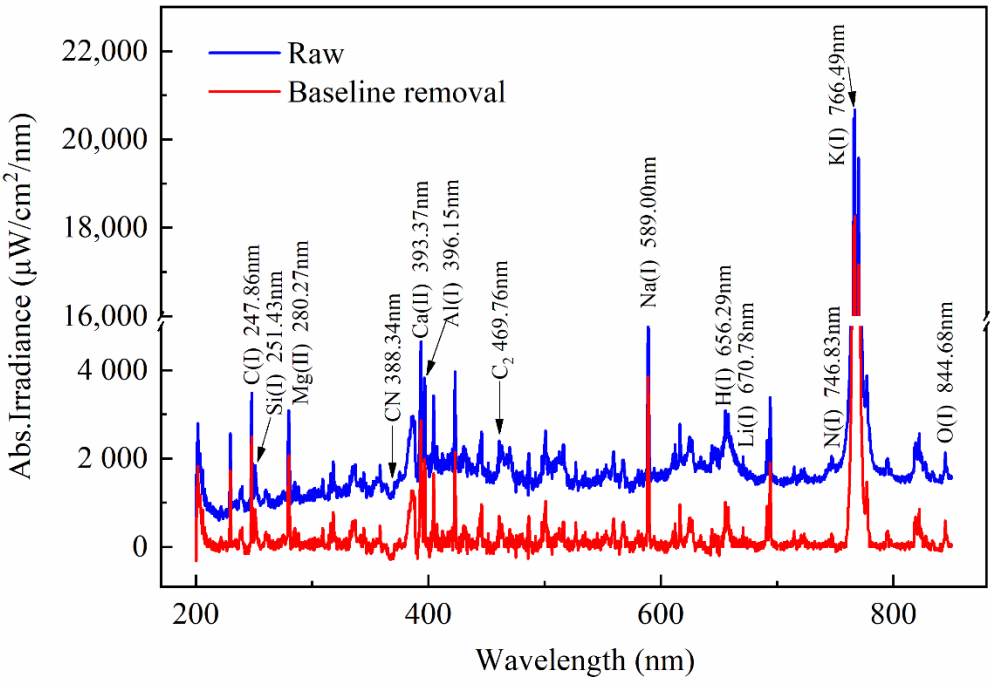

Figure 2. Average emission spectrum of coal sample 1.

3.2. PCR

Several uncorrelated principal components were obtained by dimensional reduction of PCA. According to Figure 3a, four principal components could account for 92.68% of the cumulative explained variance. Even if the cumulative explained variance was over 90%, the model evaluation indicators, such as R^2 and RMSECV, as shown in Figure 3b, were still essential. When the number of principal components was selected to be 14, the RMSECV value was the smallest, at only 0.53%, and the cumulative explained variance met the convention of 90% cumulative explained variance [27]. Then, 14 principal components of each coal sample were determined and mapped to the coal proximate analysis by MLR.

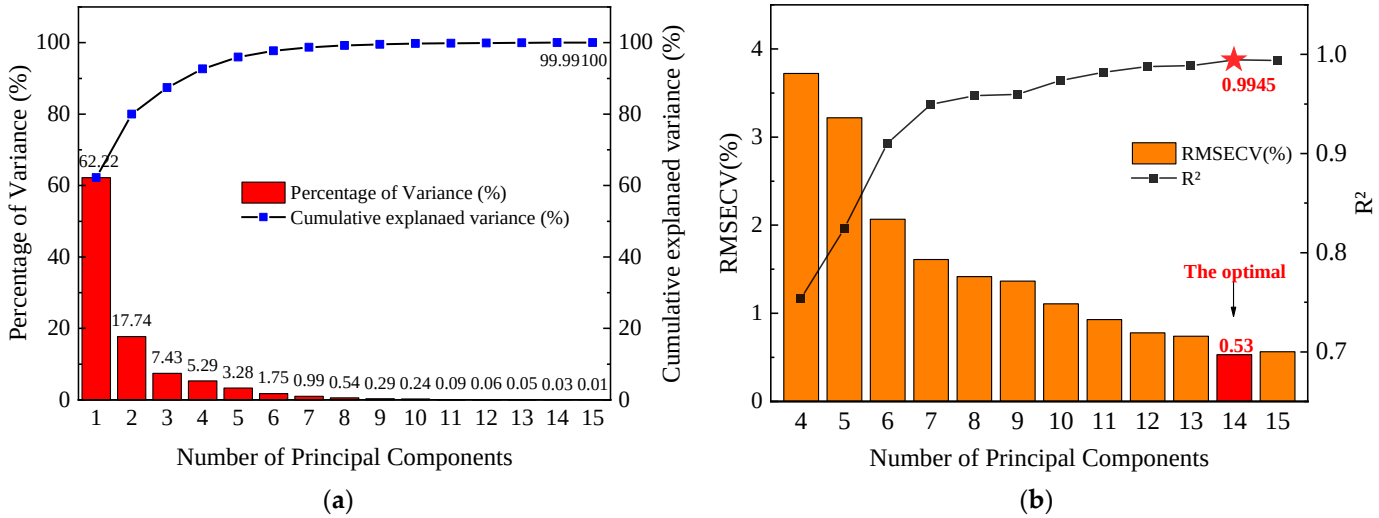

Figure 3. Scree plot (**a**) and RMSECV and R^2 values (**b**) of PCR with different numbers of principal components.

The correlation between the input emission lines and the first two principal components was shown in Figure 4. Principal component 1 was positively correlated with the intensities of the C, O, Li, Na, Mg, Ca, Fe, C_2, and CN emission lines. As principal

components 1 and 2 were orthogonal, some emission lines were negatively correlated with principal component 2, such as O, Na, Mg, Ca, and CN. The scoring trends of the C, Si, Li, Fe, and C_2 emission lines were generally similar, whereas those of the H, O, Na, Mg, Ca, and CN emission lines were the opposite. Although each principal component had no clear physical meaning, the loadings of the first two principal components were devoted to determining which potential emission lines were vital to select the emission lines [36].

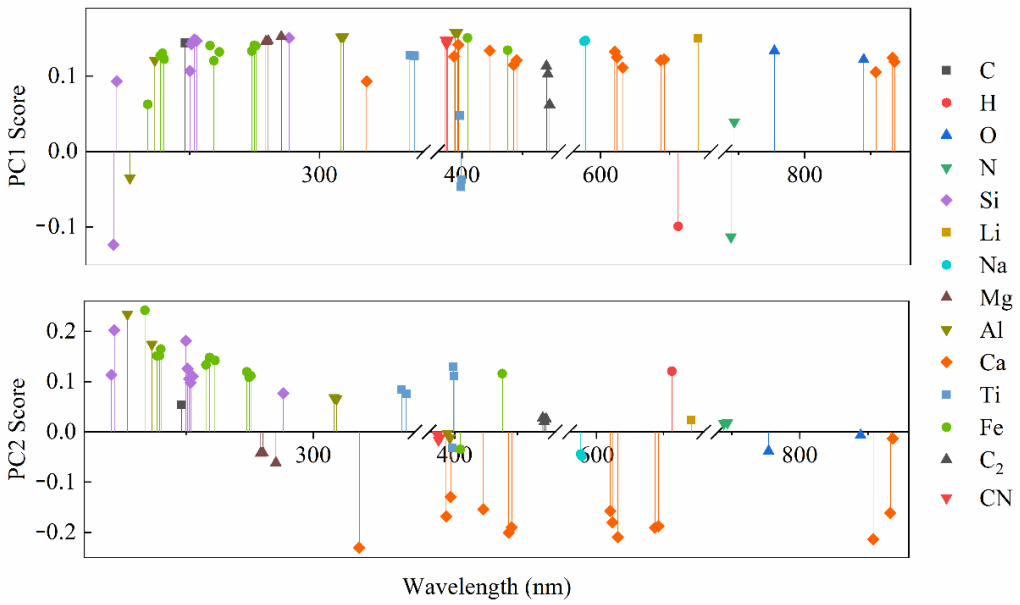

Figure 4. Loading plot of spectral emission lines on the first two principal components.

A clustering phenomenon could be observed in Figure 5, and the first two principal component scores of each coal sample corresponded to the types of coal, especially in bituminous coal. The scores of principal components calculated by PCA were widely used in sample classification, implying that similar samples could obtain close scores of principal components [37]. Thus, PCA could be applied to roughly classify a large number of coal samples into different types in accordance with the scores. Furthermore, the predictive model could be trained by different groups to predict the coal proximate analysis accurately.

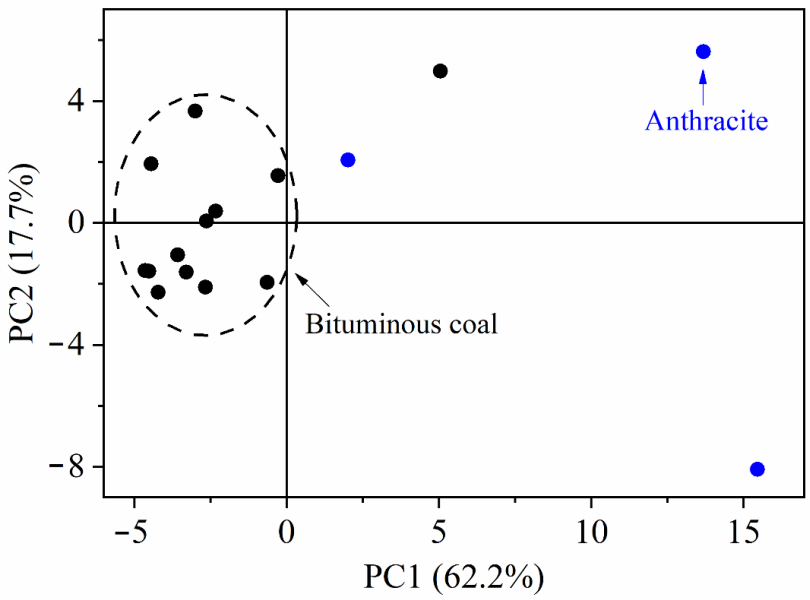

Figure 5. Score plot of 16 coal samples on the first two principal components.

3.3. ANNs

ANNs are devoted to the mapping of multi-inputs and multi-outputs; thus, they are suitable for solving nonlinear problems with complex internal mechanisms. However, an overfitted situation, where the ANN model performs well in training but poorly in prediction, may occur. The most direct method to avoid overfitting is to add the number of spectra for each coal sample [38]. Thus, the 256 spectra of each coal sample were used to cover more changes and enlarge the training set.

The intensities of the 63 emission lines selected in Table 2 were inputted as independent variables, and the proximate analysis was inputted as the dependent variable to train the model. In this work, 12 coal samples were selected as the training set, two samples were regarded as the validation set, and the remaining two were used as the prediction set. The proportions of the training set, validation set, and prediction set were approximately 70%, 15%, and 15%, respectively. The most suitable number of hidden layer neurons was 11, fixed by the minimum RMSECV and the maximum R^2, as shown in Figure 6. When the number of hidden layer neurons was 13, the ANN model was overfitted, which indicated that ANNs may need to conduct data regularization for enhanced stability [38].

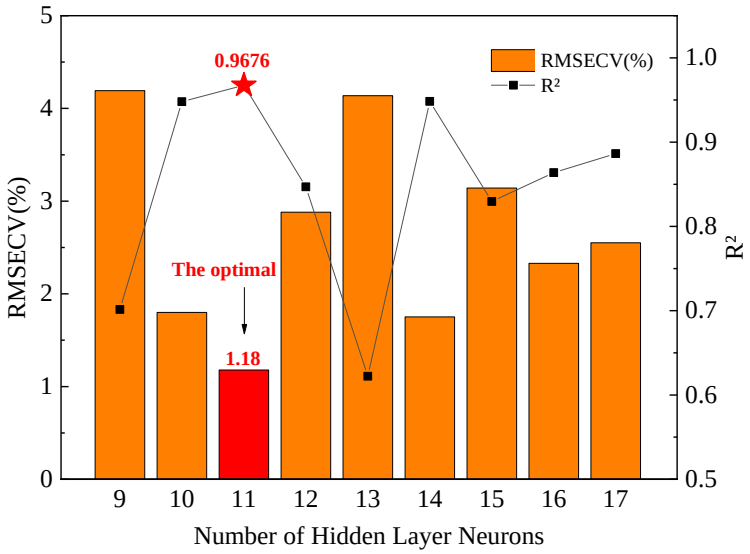

Figure 6. The RMSECV and R^2 values of ANN with different numbers of hidden layer neurons.

3.4. PCA-ANN

The PCA-ANN model contained the nonlinear regression of ANN. Thus, the problem of overfitting in Section 3.3 also existed when the number of hidden layer neurons was nine, as shown in Figure 7. If a sufficient number of principal components was used, the accuracy of classification could be improved [39], indicating the importance of selecting principal component input for ANN. According to Figure 3a, almost 100% of the variance could be explained by the first 15 principal components, which were inputted from large to small by eigenvalue (EV-PCA-ANN) to build an ANN model. The maximum R^2 and minimum RMSECV optimized the number of hidden layer neurons to five, at which point the performance of the PCA-ANN model was the best.

3.5. Comparison and Analysis

Due to the complex and diverse composition of coal, three chemometric methods were proposed to predict the coal proximate analysis after spectral pretreatment (See Table 3). PCR is regarded as a linear algorithm, and ANN is a nonlinear algorithm. PCA-ANN first performed linear dimensionality reduction (PCA) and then mapped the principal component scores to coal proximate analysis through nonlinear transformation (ANN). The optimal parameters of each algorithm were determined by cross-validation. The

comparison of predicted and real values of coal proximate analysis by three different chemometric methods was shown in Figure 8.

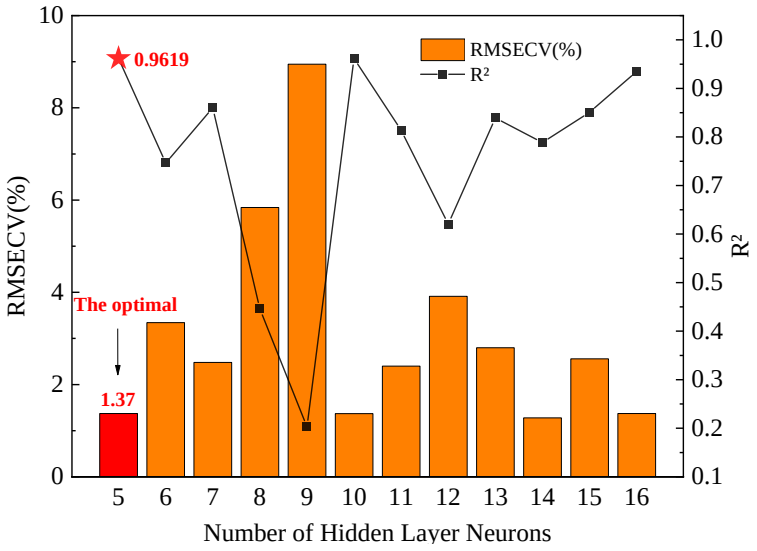

Figure 7. The RMSECV and R^2 values of PCA-ANN with different numbers of hidden layer neurons.

Table 3. Quantitative analysis of the three chemometric methods.

Proximate Analysis	Methods	R^2	RMSECV (%)	MSE	Modeling Time (s)
Moisture	PCR	0.9904	0.2458	0.0565	10
	ANN	0.9816	0.3386	0.1082	89
	PCA-ANN	0.9513	0.5555	0.2887	28
Ash	PCR	0.9986	0.2455	0.0564	10
	ANN	0.9770	1.0060	0.9496	85
	PCA-ANN	0.9179	1.8955	11.6299	25
Fixed carbon	PCR	0.9891	0.6620	0.4097	10
	ANN	0.9330	1.6334	2.5081	90
	PCA-ANN	0.9770	0.5765	1.7799	30
Volatiles	PCR	0.9958	0.6857	0.4399	10
	ANN	0.9928	0.8973	0.7579	79
	PCA-ANN	0.9917	1.6452	24.5579	22
Heating value	PCR	0.9979	0.0980	0.0900	10
	ANN	0.9342	0.5482	0.2837	83
	PCA-ANN	0.9438	0.5120	0.2452	24

A slight difference was found among the three predictions of moisture. Among them, the PCR model performed best, and its R^2, RMSECV, and MSE values were 0.9904, 0.2458%, and 0.0565, respectively. For ash, the prediction of PCR was the most accurate and stable. The R^2 values of ash between the predicted and real values, as calculated by PCR, ANN, and PCA-ANN, were 0.9986, 0.9770, and 0.9179, respectively; the corresponding RMSECV values obtained by LOO-CV were 0.2455%, 1.0060%, and 1.8955%, respectively. The MSE values of PCR and ANN were 0.0564 and 0.9496, while the MSE value of PCA-ANN was 11.6299, indicating that the prediction model was not as stable as the other two. In terms of fixed carbon, the best prediction model was PCR. The R^2 and RMSECV values of the three models were close. The MSE value of PCR was far lower than that of the other two, indicating that PCR was more robust. The PCR model had the best performance for the quantification of volatiles. Though the R^2 values of all models were greater than 0.99, the robustness of the model still needed to be considered. The RMSECV values of PCR, ANN,

and PCA-ANN were 0.6857%, 0.8973%, and 1.6452%, respectively. Compared with PCR and ANN, the MSE value of PCA-ANN reached 24.5579, which showed that PCA–ANN was not suitable for the prediction of volatiles. When predicting the heating value, the highest R^2, and the lowest RMSECV and MSE demonstrated that the PCR was still the prime model for prediction.

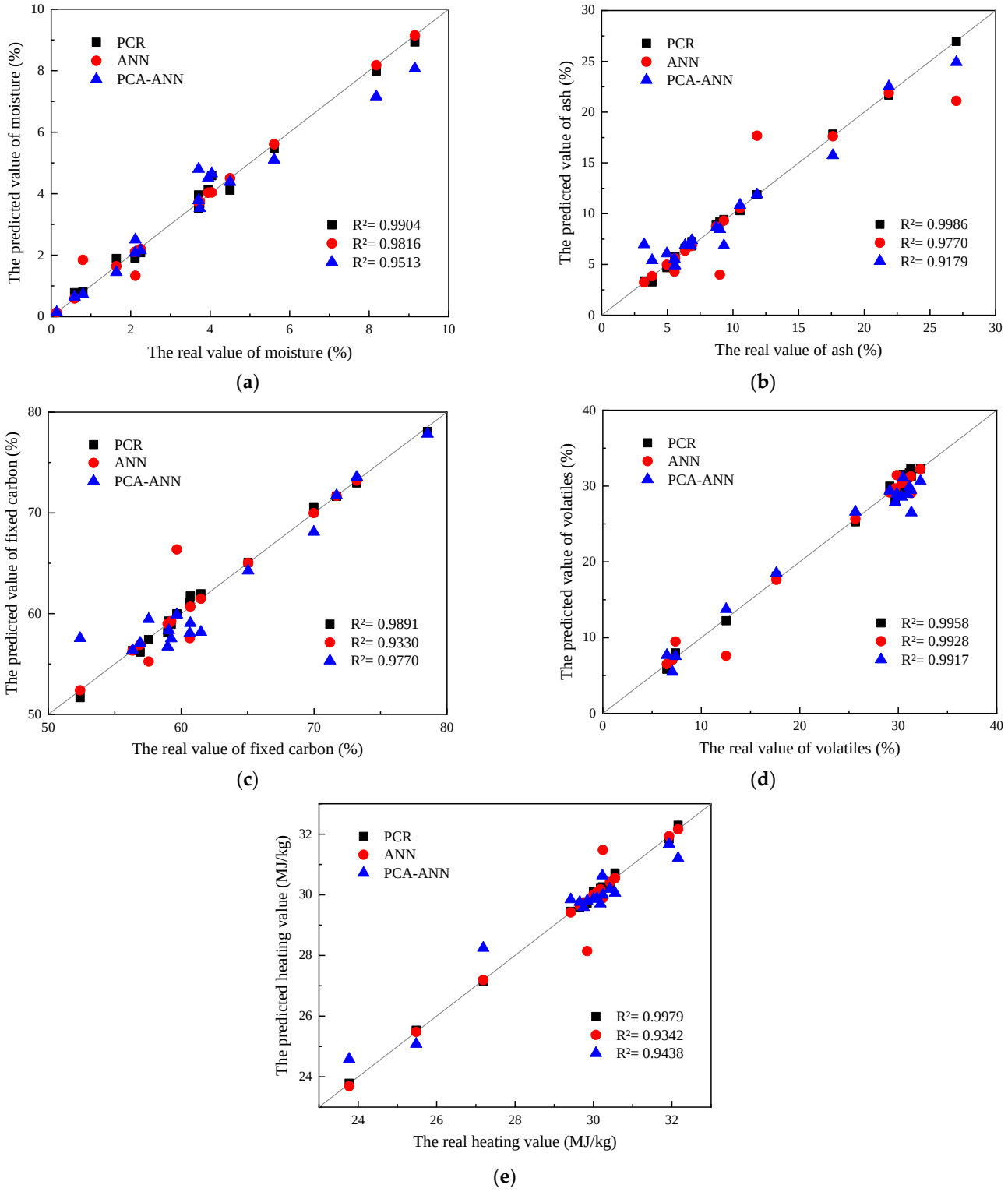

Figure 8. Comparison of predicted and real values of moisture (**a**), ash (**b**), fixed carbon (**c**), volatiles (**d**), and heating value (**e**) on an air-dries basis by three different chemometric methods.

The modeling time was considered as an important index of the modeling efficiency. When the same algorithm used different parameters, its modeling time was similar, which was consistent with Zhang's work [15]. The modeling time of different algorithm models varied greatly; for instance, the modeling times of PCR, ANN, and PCA-ANN to predict moisture were 10, 89, and 28 s, respectively. The modeling time of PCR was generally the shortest, PCA-ANN ranked second, and the modeling efficiency of ANN was the lowest. The prediction results of the cross-validation and modeling time were shown in Table 3.

As a result, the PCR should be considered first when coal proximate analysis was predicted under similar experimental conditions. After dimensional reduction and linear regression, PCR had good accuracy and stability, and its training time was the shortest. The intensities of the selected emission lines had a great linear correlation with the coal proximate analysis. First, the principal components were extracted directly from the 63 emission lines, non-reabsorbed and non-overlapping, which provided the possibility of linear correlation. With the addition of KBr binder, the generation and disappearance of shock waves were at the same level during the ablation process of coal [40], which suppressed the matrix effect of coal. Meanwhile, the laser-induced ablation pyrolysis of volatiles may have been strongly inhibited to provide a good environment for forming plasmas with stable emissions. Thus, a strong corresponding relationship existed between the component concentration and spectral intensity in coal.

Due to the large number of variables to be inputted, the training of the ANN was time consuming, while its prediction accuracy and stability were slightly lower than those of PCR. Some of the selected emission lines may not have mattered, which wasted computing resources and resulted in inaccurate results. Although the correlation between the component concentration and elemental concentration in coal was nonlinear, the selected emission lines weakened some nonlinear factors, to an extent. Another reasonable explanation was that the temporal and spatial uniformity of the plasma, provided by the KBr, improved the linear relation, causing the ANN to not be as applicable as PCR.

In contrast, PCA-ANN was not suitable for the prediction of coal proximate analysis. The reason why PCA-ANN lacked accuracy and stability was that the principal components generated by PCA and the input data for ANN were unrelated, consistent with the results of Drezga et al. [41]. The principal components obtained by PCA were usually sorted by EV–PCA-ANN for training ANN, similar to the present work. The results revealed that the principal components, obtained by PCA and then selected by GA, for training ANN (PCA-GA-ANN) had a better performance than the EV-PCA-ANN in literature [32].

4. Conclusions

In this work, the spectra of 16 coal samples with added KBr binder were investigated by a LIBS experimental setup. Three types of chemometric methods were applied to predict the coal proximate analysis: PCR, ANN, and PCA-ANN, representing linear, nonlinear, and a combination of linear and nonlinear methods, respectively. Three indicators that expressed the accuracy and stability of the model were employed to fix the optimal parameters and evaluate the performance of the model. With the highest R^2 and the lowest RMSECV and MSE values, the PCR method was deemed the most effective approach to quantitative proximate analysis. Although the R^2 values of ANN and PCA-ANN were both above 0.9, their RMSECV and MSE values were much higher than those of PCR. The training time of ANN was the longest, and a gap was found between the prediction and training ability. PCA reduced the number of input variables of ANN, whereas the nonlinear fitting of the unrelated variables reduced the robustness of PCA-ANN.

On the basis of the advantages of multi-element rapid measurement, simple sample preparation, and remote measurement, LIBS could be able to provide in-situ and online measurement and analysis for coal-fired boilers in the future. The three chemometric methods could be used to predict coal proximate analysis to provide an expedient method for obtaining the proximate analysis of coal burning in power plant boilers.

Author Contributions: Conceptualization, X.R. and Y.L.; methodology, Y.L.; software, Y.L.; validation, D.W.; formal analysis, Y.L.; investigation, D.W.; resources, X.R.; data curation, D.W.; writing—original draft preparation, Y.L.; writing—review and editing, X.R.; supervision, X.R.; project administration, X.R. All authors have read and agreed to the published version of the manuscript.

Funding: Scientific research program of Hebei administration for market regulation (No. 2022ZD09).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors are grateful to the anonymous reviewers whose comments have helped to improve the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Statistical Review of World Energy 2021. Available online: https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2021-full-report.pdf (accessed on 14 March 2022).
2. Ghetti, P. DTG combustion behaviour of coal: Correlations with proximate and ultimate analysis data. *Fuel* **1986**, *65*, 636–639. [CrossRef]
3. Charbucinski, J.; Nichols, W. Application of spectrometric nuclear borehole logging for reserves estimation and mine planning at Callide coalfields open-cut mine. *Appl. Energy* **2003**, *74*, 313–322. [CrossRef]
4. Parus, J.; Kierzek, J.; Malozewska-Bucko, B. Determination of the carbon content in coal and ash by XRF. *X-ray Spectrom. Int. J.* **2000**, *29*, 192–195. [CrossRef]
5. Ctvrtnickova, T.; Mateo, M.P.; Yañez, A.; Nicolas, G. Application of LIBS and TMA for the determination of combustion predictive indices of coals and coal blends. *Appl. Surf. Sci.* **2011**, *257*, 5447–5451. [CrossRef]
6. Carranza, J.E.; Hahn, D.W. Sampling statistics and considerations for single-shot analysis using laser-induced breakdown spectroscopy. *Spectrochim. Acta Part B At. Spectrosc.* **2002**, *57*, 779–790. [CrossRef]
7. Singh, J.P.; Almirall, J.R.; Sabsabi, M.; Miziolek, A.W. Laser-induced breakdown spectroscopy (LIBS). *Anal. Bioanal. Chem.* **2011**, *400*, 3191–3192. [CrossRef]
8. Rosenwasser, S.; Asimellis, G.; Bromley, B.; Hazlett, R.; Martin, J.; Pearce, T.; Zigler, A. Development of a method for automated quantitative analysis of ores using LIBS. *Spectrochim. Acta Part B At. Spectrosc.* **2001**, *56*, 707–714. [CrossRef]
9. Zhang, Y.; Dong, M.; Cheng, L.; Wei, L.; Cai, J.; Lu, J. Improved measurement in quantitative analysis of coal properties using laser induced breakdown spectroscopy. *J. Anal. At. Spectrom.* **2020**, *35*, 810–818. [CrossRef]
10. Senesi, G.S.; Senesi, N. Laser-induced breakdown spectroscopy (LIBS) to measure quantitatively soil carbon with emphasis on soil organic carbon. A review. *Anal. Chim. Acta* **2016**, *938*, 7–17. [CrossRef]
11. Sezer, B.; Bilge, G.; Boyaci, I.H. Capabilities and limitations of LIBS in food analysis. *Trends Analyt. Chem.* **2017**, *97*, 345–353. [CrossRef]
12. Hahn, D.W.; Omenetto, N. Laser-Induced Breakdown Spectroscopy (LIBS), Part I: Review of Basic Diagnostics and Plasma-Particle Interactions: Still-Challenging Issues within the Analytical Plasma Community. *Appl. Spectrosc.* **2010**, *64*, 335A–366A. [CrossRef]
13. Yang, Y.; Li, C.; Liu, S.; Min, H.; Yan, C.; Yang, M.; Yu, J. Classification and identification of brands of iron ores using laser-induced breakdown spectroscopy combined with principal component analysis and artificial neural networks. *Anal. Methods* **2020**, *12*, 1316–1323. [CrossRef]
14. Zhang, W.; Zhuo, Z.; Lu, P.; Tang, J.; Tang, H.; Lu, J.; Xing, T.; Wang, Y. LIBS analysis of the ash content, volatile matter, and calorific value in coal by partial least squares regression based on ash classification. *J. Anal. At. Spectrom.* **2020**, *35*, 1621–1631. [CrossRef]
15. Zhang, Y.; Xiong, Z.; Ma, Y.; Zhu, C.; Zhou, R.; Li, X.; Li, Q.; Zeng, Q. Quantitative analysis of coal quality by laser-induced breakdown spectroscopy assisted with different chemometric methods. *Anal. Methods* **2020**, *12*, 3530–3536. [CrossRef] [PubMed]
16. Zhang, W.; Zhuo, Z.; Lu, P.; Lu, J.; Sun, T.; Tang, J.; Tang, H.; Zhou, T.; Li, L. Laser-induced breakdown spectroscopy for quantitative and qualitative analysis of the ash fusion temperatures of coal in power plants. *J. Anal. At. Spectrom.* **2021**, *36*, 576–589. [CrossRef]
17. Zhang, T.; Yan, C.; Qi, J.; Tang, H.; Li, H. Classification and discrimination of coal ash by laser-induced breakdown spectroscopy (LIBS) coupled with advanced chemometric methods. *J. Anal. At. Spectrom.* **2017**, *32*, 1960–1965. [CrossRef]
18. Lu, Z.; Mo, J.; Yao, S.; Zhao, J.; Lu, J. Rapid Determination of the Gross Calorific Value of Coal Using Laser-Induced Breakdown Spectroscopy Coupled with Artificial Neural Networks and Genetic Algorithm. *Energy Fuels* **2017**, *31*, 3849–3855. [CrossRef]
19. Yao, S.; Zhao, J.; Xu, J.; Lu, Z.; Lu, J. Optimizing the binder percentage to reduce matrix effects for the LIBS analysis of carbon in coal. *J. Anal. At. Spectrom.* **2017**, *32*, 766–772. [CrossRef]
20. Dixon-Lewis, G.; Marshall, P.; Ruscic, B.; Burcat, A.; Goos, E.; Cuoci, A.; Frassoldati, A.; Faravelli, T.; Glarborg, P. Inhibition of hydrogen oxidation by HBr and Br2. *Combust. Flame* **2012**, *159*, 528–540. [CrossRef]

21. Zhu, W.; Li, X.; Sun, R.; Zhang, Y.; Yan, Y.; Yu, X.; Ren, X. Investigation of mineral-element migration upon pyrolysis and quantitative prediction of volatiles in coal using laser-induced breakdown spectroscopy. *J. Anal. At. Spectrom.* **2021**, *36*, 1399–1409. [CrossRef]
22. Ren, X.; Rokni, E.; Sun, R.; Meng, X.; Levendis, Y.A. Evolution of Chlorine-Bearing Gases during Corn Straw Torrefaction at Different Temperatures. *Energy Fuels* **2017**, *31*, 13713–13723. [CrossRef]
23. Dong, M.; Mao, X.; Gonzalez, J.J.; Lu, J.; Russo, R.E. Time-resolved LIBS of atomic and molecular carbon from coal in air, argon and helium. *J. Anal. At. Spectrom.* **2012**, *27*, 2066–2075. [CrossRef]
24. Liu, H.; Shah, S.; Jiang, W. On-line outlier detection and data cleaning. *Comput. Chem. Eng.* **2004**, *28*, 1635–1647. [CrossRef]
25. Chen, D.; Huang, Z.; Wang, T.; Ma, Y.; Zhang, Y.; Wang, G.; Zhang, P. High-throughput analysis of single particles by micro laser induced breakdown spectroscopy. *Anal. Chim. Acta* **2020**, *1095*, 14–19. [CrossRef]
26. Yu, H.; Xu, Y.; Lv, N.; Chen, H.; Chen, S. Arc spectral processing technique with its application to wire feed monitoring in Al–Mg alloy pulsed gas tungsten arc welding. *J. Mater. Process Technol.* **2013**, *213*, 707–716. [CrossRef]
27. Jolliffe, I. Principal Component Analysis. In *Encyclopedia of Statistics in Behavioral Science*; Springer Press: New York, NY, USA, 2002; ISBN 9783642048975.
28. Sjöberg, J.; Zhang, Q.; Ljung, L.; Benveniste, A.; Delyon, B.; Glorennec, P.Y.; Hjalmarsson, H.; Juditsky, A. Nonlinear black-box modeling in system identification: A unified overview. *Automatica* **1995**, *31*, 1691–1724. [CrossRef]
29. El Haddad, J.; Bruyère, D.; Ismaël, A.; Gallou, G.; Laperche, V.; Michel, K.; Canioni, L.; Bousquet, B. Application of a series of artificial neural networks to on-site quantitative analysis of lead into real soil samples by laser induced breakdown spectroscopy. *Spectrochim. Acta Part B At. Spectrosc.* **2014**, *97*, 57–64. [CrossRef]
30. Nourani, V.; Komasi, M.; Alami, M.T. Hybrid Wavelet–Genetic Programming Approach to Optimize ANN Modeling of Rainfall–Runoff Process. *J. Hydrol. Eng.* **2012**, *17*, 724–741. [CrossRef]
31. Shibata, K.; Yusuke, I. Effect of number of hidden neurons on learning in large-scale layered neural networks. In Proceedings of the ICROS-SICE International Joint Conference 2009, Fukuoka, Japan, 18–21 August 2009; IEEE: New York, NY, USA, 2009; pp. 5008–5013.
32. Hemmateenejad, B. Correlation ranking procedure for factor selection in PC-ANN modeling and application to ADMETox evaluation. *Chemometr. Intell. Lab. Syst.* **2005**, *75*, 231–245. [CrossRef]
33. Sheta, S.; Afgan, M.S.; Hou, Z.; Yao, S.C.; Zhang, L.; Li, Z.; Wang, Z. Coal analysis by laser-induced breakdown spectroscopy: A tutorial review. *J. Anal. At. Spectrom.* **2019**, *34*, 1047–1082. [CrossRef]
34. Liu, K.; He, C.; Zhu, C.; Chen, J.; Zhan, K.; Li, X. A review of laser-induced breakdown spectroscopy for coal analysis. *Trends Analyt. Chem.* **2021**, *143*, 327–334. [CrossRef]
35. Ferreira, E.C.; Anzano, J.M.; Milori, D.M.; Ferreira, E.J.; Lasheras, R.J.; Bonilla, B.; Montull-Ibor, B.; Casas, J.; Martin Neto, L. Multiple response optimization of laser-induced breakdown spectroscopy parameters for multi-element analysis of soil samples. *Appl. Spectrosc.* **2009**, *63*, 1081–1088. [CrossRef] [PubMed]
36. Pořízka, P.; Klus, J.; Képeš, E.; Prochazka, D.; Hahn, D.W.; Kaiser, J. On the utilization of principal component analysis in laser-induced breakdown spectroscopy data analysis, a review. *Spectrochim. Acta Part B At. Spectrosc.* **2018**, *148*, 65–82. [CrossRef]
37. Chatterjee, S.; Singh, M.; Biswal, B.P.; Sinha, U.K.; Patbhaje, S.; Sarkar, A. Application of laser-induced breakdown spectroscopy (LIBS) coupled with PCA for rapid classification of soil samples in geothermal areas. *Anal. Bioanal. Chem.* **2019**, *411*, 2855–2866. [CrossRef]
38. Li, L.N.; Liu, X.F.; Yang, F.; Xu, W.M.; Wang, J.Y.; Shu, R. A review of artificial neural network based chemometrics applied in laser-induced breakdown spectroscopy analysis. *Spectrochim. Acta Part B At. Spectrosc.* **2021**, *180*, 106183. [CrossRef]
39. Pokrajac, D.; Lazarevic, A.; Kecman, V.; Marcano, A.; Markushin, Y.; Vance, T.; Reljin, N.; McDaniel, S.; Melikechi, N. Automatic classification of laser-induced breakdown spectroscopy (LIBS) data of protein biomarker solutions. *Appl. Spectrosc.* **2014**, *68*, 1067–1075. [CrossRef]
40. Fu, Y.T.; Gu, W.L.; Hou, Z.Y.; Muhammed, S.A.; Li, T.Q.; Wang, Y.; Wang, Z. Mechanism of signal uncertainty generation for laser-induced breakdown spectroscopy. *Front. Phys.* **2020**, *16*, 22502. [CrossRef]
41. Drezga, I.; Rahman, S. Input variable selection for ANN-based short-term load forecasting. *IEEE Trans. Power Syst.* **1998**, *13*, 1238–1244. [CrossRef]