Infrared Physics and Technology 93 (2018) 34–40

Contents lists available at ScienceDirect

Infrared Physics & Technology

journal homepage: www.elsevier.com/locate/infrared

Regular article

Coal analysis based on visible-infrared spectroscopy and a deep neural network

Ba Tuan Le[a,b], Dong Xiao[a,c,*], Yachun Mao[c], Dakuo He[a]

[a] Information Science and Engineering School, Northeastern University, 110004 Shenyang, China
[b] Control Technology College, Le Quy Don Technical University, 100000 Hanoi, Viet Nam
[c] Intelligent Mine Research Center, Northeastern University, 110819 Shenyang, China

ARTICLE INFO

Keywords:
Coal
Proximate analysis
Visible-infrared spectroscopy
Convolutional neural network (CNN)
Extreme learning machine (ELM)
Artificial bee colony (ABC)

ABSTRACT

The proximate analysis of coal is the umbrella term for the six indexes that include the moisture, ash, volatile matter, fixed carbon, and sulphur contents and the heating value. Burning of coal creates carbon dioxide, sulphur dioxide and nitrogen dioxide which are main reasons causing air pollution. Therefore, before utilizing coal, it is indispensable to analyse coal. The traditional proximate analysis of coal mainly relies on chemical analysis, which is time-consuming and costly. Hence, a method to construct a coal analysis is introduced in this paper. By using the method to analyse moisture (%), ash (%), volatile matter (%), fixed carbon (%), and sulphur (%) contents and the low heating value (J/g). We first obtained different coal sample from different coal areas in China. Then, measured the spectral data through the spectral analysis instrument and extracted spectral features through a convolutional neural network. Finally, we applied the extreme learning machine algorithm to construct the prediction and analysis model of the spectral feature data. The experimental result shows that the model in the study can predict the components of coal. Compared with the chemical analysis method, this method has unparalleled advantages in terms of financial efficiency, speed and accuracy.

1. Introduction

Coal is the main source of energy for both China and the world. In 2015, the estimated total amount of extracted coal in the world is estimated to be 850 billion tons. Countries with ample coal production include the USA (245 billion tons), Russia (150 billion tons) and China (120 billion tons) [1]. As industries develop, global requirements regarding the quality of coal have increased. High quality coal has a great impact on production rates and the environmental pollution problem, which means that the proximate analysis before utilizing coal is indispensable. The traditional proximate analysis of coal mainly relies on chemical analysis. Despite its high accuracy, it is time-consuming and costly. Thus, in regard to modern proximate analysis, a significant task is to quickly and accurately identify the components of coal. It is vital to decrease analysis costs and increase classification efficiency.

Spectral analysis has the advantages of high speed, low costs and high efficiency. Therefore, in recent years, spectral analysis has been widely applied in many fields, such as ore analysis, grade appraisal and food inspection [2–7]. For coal, numerous studies haves proven that spectral characteristics can be used to measure the moisture, ash, volatile matter, fixed carbon, sulphur contents and the heating value. Andres et al. [8,9] used the near-infrared spectrum to measure moisture, volatile matter and ash in coal. Dong et al. [10] used spectral characteristics to quickly measure sulphur, carbon and nitrogen in coal. Li et al. [11] measured the sulphur content in coal using near-infrared spectroscopy. Likewise, Zhai et al. [12] a method of quickly measuring ash in coal based on near-infrared spectroscopy.

The decisive factors that influence the coal spectrum are the oxides of carbon, sulphur, silicon and the like. Therefore, visible-infrared spectral data usually include a lot of chemical information that is irrelevant to coal analysis. This makes the spectral data of coal bear some negative characteristics such as high dimensions and great redundancy. Thus, it is necessary to perform an effective extraction of the spectral data of coal. In recent years, deep learning using the convolutional neural network (CNN) [13–15] has been widely applied in prediction models. The CNN is a new method of extracting features, and it contains many layers in the hidden neural network. There are two kinds of hidden layers, including convolutional and sampling layers. The shared weight network structure of the CNN makes it similar to biological neural networks. Therefore, the CNN reduces the model's complexity, decreases the number of weights and more efficiently extracts data features.

* Corresponding author.
E-mail address: xiaodong@ise.neu.edu.cn (D. Xiao).

https://doi.org/10.1016/j.infrared.2018.07.013
Received 25 May 2018; Received in revised form 10 July 2018; Accepted 10 July 2018
Available online 11 July 2018

In 2006, Huang [16] proposed the extreme learning machine (ELM), which is a feed-forward neural network with a single hidden layer. The ELM is widely used due to its advantages such as a fast training speed, high generalization and high prediction accuracy [17–21]. Since the CNN is a good features extraction machine rather than a good prediction machine, the ELM can serve as a more ideal prediction machine. Therefore, in this study, we combined the two to construct a prediction model with even higher accuracy. We integrate this model with spectral techniques and apply it to the proximate analysis of coal. It provides a new, fast and accurate proximate analysis method for coal.

2. Experimental

2.1. Data collection and processing

2.1.1. Spectral data collection

The coal sampling collected in this study comes from the Fushun coal mine in Liaoning Province, the Zhijin coal mine in Guizhou Province, the Datong coal mine in Shanxi Province, the Jiajinkou coal mine in Henan Province, the Shendong coal mine in Shanxi Province and the Yimin coal mine in Neimenggu Province. These coal mines are widely distributed in all parts of China, from south to north. There are 100 coal samples in total, including anthracite, bituminous coal and lignite. The experimental apparatus was the, SVC HR-1024 from the American company Spectra Vista, which is a portable spectrometer for ground targets. The spectral range of this apparatus was: 350–2500 nm. The built-in storage could hold 500 scans. The device's weight was 3 kg. The number of channels was 1024. The spectral resolution (FWHM $\leqq$ 8.5 nm) ranged from 1000 to 1850 nm. Finally, the minimum integral time was 1 ms.

First, we washed the samples, and then, we performed the grinding process. In this experiment, the scanning time of the apparatus was always one second. The probe of the spectrometer was 480 mm away from the surface of the coal sample, and it was perpendicular to the surface of the sample. The halogen light was 320 mm away from the coal sample, and it formed a 45° angle with the sample surface. In the process of the experiment, in order to reduce the influence that the surroundings could have on the spectral test, the experiment had to be conducted in an enclosed interior space to avoid exposure to sunshine or other light sources. The experimenters could not walk around, and they were all dressed in dark clothing. We used the spectrometer (SVC HR-1024) to perform five spectral tests on the samples and then took the average as the spectral datum of every sample. Fig. 1 shows the experimental procedure of collecting samples and data.

Table 1
The proximate analysis results of some of the coal samples.

Samples	Ash (%)	Volatile matter (%)	Fixed carbon (%)	Low heating value (J/g)	Moisture (%)	Sulphur (%)
1	35.72	36.61	25.91	16.98	1.76	0.294
2	28.36	31.76	38.5	19.84	1.38	0.327
3	30.83	31.92	35.6	18.93	1.65	0.325
4	6.85	38.67	52.23	25.93	2.25	0.586
5	13.39	37.63	47.02	24.01	1.96	0.501
6	4.69	27.47	60.05	25.26	7.79	0.18
7	7.53	34.1	51.68	24.31	6.69	0.29
8	2.71	28.03	61.21	25.76	8.05	0.254
9	12.59	28.25	53.4	23.45	5.77	0.194
10	4.08	34.39	53.22	24.77	8.31	0.197
11	12.84	7.69	78.35	26.63	1.12	0.492
12	23.05	9.54	66.34	23.23	1.07	0.281
13	15.28	9.72	74.15	25.8	0.85	0.661
14	15.51	10.79	72.18	25.39	1.53	0.439
15	8.33	9.15	81.85	28.15	0.67	0.551
16	8.98	37.05	42.51	21.78	11.45	0.091
17	9.72	37.96	41.16	21.6	11.16	0.066
18	3.94	42.37	43.76	23.62	8.93	0.408
19	8.1	39.64	41	21.96	11.2	0.103
20	11.08	40.15	43.25	21.67	11.29	0.201

2.1.2. The determination of the proximate analysis of coal

After testing the spectra of the coal samples, we used the chemical method to determine the proximate analysis of coal. These indexes are the main standards to judge the coal quality. The chemical method measures the moisture, ash, volatile matter, fixed carbon, sulphur contents and the low heating value in coal through complex experiments and the standard apparatus. For example, taking the volatile matter, we isolated the air and heated the coal samples at a high temperature so that the organic substance in the coal would have a pyrolytic reaction. Some of the substances turned into gas and escaped, while the rest stayed as solid bodies. Volatile matter refers to the escaping substances (excluding moisture). According to the laboratory

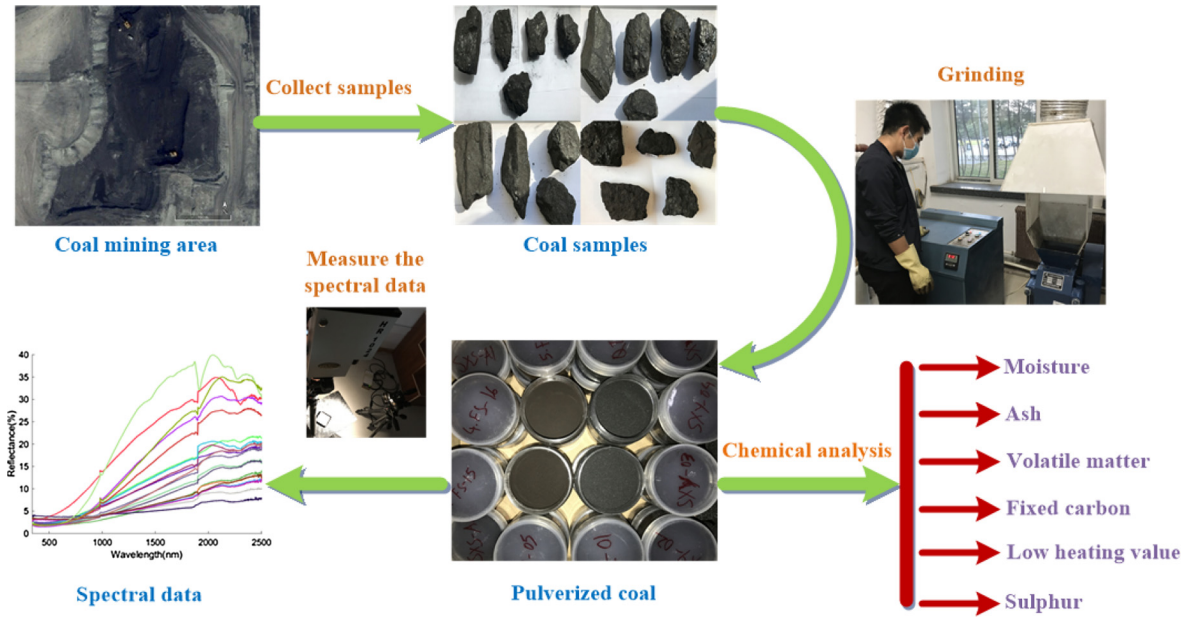

Fig. 1. Data collection experiment.

report, the proximate analysis results of some of the coal samples are shown in Table 1.

2.2. The model construction

2.2.1. Convolutional neural network

The convolutional neural network's input data are in matrix form. The hidden layers include alternating convolutional and sampling layers, and the last output layer is a one-dimensional network. If layer l is a convolutional layer, then the feature mapping of j is

$$x_j^l = f\left(\sum_{x_i^{n-1}\in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right) \tag{1}$$

Here, M_j is the collection of input data; f is nonlinear function such as the Sigmoid function, the Rectified Linear Units function and or the Softplus function; k_{ij}^l is the convolutional kernel of datum i in layer $l-1$ and datum j in layer l; and b_j^l is the bias.

If layer l is a sampling layer, then the feature mapping of j is

$$x_j^l = f(\omega_j^l DO(x_j^{l-1}) + b_j^l) \tag{2}$$

Here, $DO(\cdot)$ is the sampling function, with may include different sampling function schemes such as random sampling, averaging and maximizing; ω_j^l is the weight; and b_j^l is the bias.

If layer l is the output layer, then the output vector is

$$x^l = f(\beta^l v^{l-1} + b^l) \tag{3}$$

v^{l-1} is the output vector of the last layer; β^l is the weight; and b^l is the bias.

2.2.2. Extreme learning machine

For any different N samples $(\mathbf{x}_i, \mathbf{t}_i)$ in which $\mathbf{x}_i = [x_{i1}, x_{i2}, \cdots, x_{in}]^T \in \mathbf{R}^n, \mathbf{t}_i = [t_{i1}, t_{i2}, \cdots, t_{im}]^T \in \mathbf{R}^m$, the single hidden layer feed-forward neural network of the standard hidden nerve cells at L is

$$\sum_{i=1}^{L} \beta_i g(\mathbf{w}_i \cdot \mathbf{x}_j + b_i) = \mathbf{t}_j, \quad j = 1, \cdots, L; \quad b_i, \beta_i \in \mathbf{R} \tag{4}$$

Here, $\mathbf{w}_i = [w_{i1}, w_{i2}, \cdots, w_{in}]^T \in \mathbf{R}^n$ stands for the input weight between the nerve cell in the input layer and nerve cell i in the hidden layer; β_i is the output weight; b_i is the bias; $g(x)$ is activation function; and $\mathbf{w}_i \cdot \mathbf{x}_j$ stands for the inner product of $\mathbf{w}_i$and $\mathbf{x}_j$.

Then, the ELM function can be expressed as

$$\mathbf{H}\beta = \mathbf{T} \tag{5}$$

Here,

$$\mathbf{H}(\mathbf{w}_1, \cdots, \mathbf{w}_L, b_1, \cdots, b_L, \mathbf{x}_1, \cdots, \mathbf{x}_N) = \begin{bmatrix} g(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_1 + b_L) \\ \vdots & & \vdots \\ g(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \cdots & g(\mathbf{w}_L \cdot \mathbf{x}_N + b_L) \end{bmatrix}_{N\times L} \tag{6}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L\times m} \qquad \mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_L^T \end{bmatrix}_{N\times m} \tag{7}$$

$\mathbf{H}$ is the matrix output by the hidden layer of neural network; $\mathbf{T}$ is the expected output; β is output weight.

The method developed by Professor Huang [16] randomly chooses the input weights and hidden layer bias. Training this network resembles the process that obtains the least squares solution $\hat{\beta}$ of the linear system $\mathbf{H}\beta = \mathbf{T}$.

$$\text{Minimize: } \|\mathbf{H}\beta - \mathbf{T}\| \tag{8}$$

Finally, we calculate that the minimum of the least squares solution of the linear system is

$$\hat{\beta} = \mathbf{H}^+\mathbf{T} \tag{9}$$

Here, $\mathbf{H}^+$ is the Moore-Penrose generalized inverse matrix of $\mathbf{H}$. The minimum of the least squares solution of $\mathbf{H}\beta = \mathbf{T}$ is unique.

2.2.3. Artificial bee colony algorithm

The artificial bee colony algorithm (ABC) [22] was proposed in 2007, and it imitates the behaviours of foraging bees. Bees in the algorithm can be categorized into three different types, including the employed bees, the onlooker bees and scout bees. The locations of flower sources represent a set of feasible solution D to the optimization of the problem. The nectar volume represents the fitness of the optimization of the problem. Every location of a food source corresponds with one bee.

Assuming that the number of food sources is SN, the fitness value of the food source number $i\,(i = 1, 2, ..., SN)$ is fit_i. Then, the employed bees search for food sources according to equation (10).

$$v_{ij} = u_{ij} + \varphi_{ij}(u_{ij} - u_{kj}) \tag{10}$$

In the equation, v_{ij} is the new food source; u_{ij} is the current food source; φ_{ij} is the random value between $[-1, 1]$; k is the random number between $[1, SN]$; and $k \# i, \quad 1 \leqslant i \leqslant SN, \quad 1 \leqslant j \leqslant d$, d is the dimension of feasible solution D. If the fitness value of v_{ij} is greater than fitness value of u_{ij}, the employed bees will substitute v_{ij} for u_{ij}; otherwise, u_{ij} is kept.

After all the employed bees finish the search, they will go back to the beehive to share the information regarding food sources with the onlooker bees. The onlooker bees choose the food source number $i\,(i = 1, 2, ..., SN)$ according to the roulette algorithm. After that, the onlooker bees search the areas near the chosen food sources for new food sources, during which time they are calculating and trying retain the food sources with the best fitness.

If food source u_{ij} is recycled due to limited time without the updated location, the scout bees search for new food sources according to Eq. (11).

$$u_{ij} = u_{\min,j} + rand(0, 1)(u_{\max,j} - u_{\min,j}) \tag{11}$$

In this equation, $u_{\min,j}$ and $u_{\max,j}$ are the respective upper and lower limits of element j in the feasible solution D.

2.2.4. A combination of CNN, ABC and ELM

The CNN extracts the features of the objects through convolutional and sampling layers and uses the gradient descent to train the network to find the minimum global error. Therefore, the CNN can be used to extract features that provide greater benefits to the prediction analysis. However, its perceptron in the last layer is not a good predictor.

The ELM is a feedforward neural network with single hidden layer. Its, advantages include fast training speed, high generalization and high classification accuracy. However, the ELM performs with high accuracy only with ample training data. In addition, in the ELM network, the weights and biases have random values. This leads to the consequence that part of the weights and biases don not reach the optimal stage. Therefore, we apply the new optimal ABC algorithm that can seek the optimal value for the ELM network.

According to these situations, we combined the CNN, ABC and ELM algorithms (CNN-AELM) to overcome their disadvantages and adopt their advantages. Fig. 2 shows the network structure when the CNN-AELM algorithm is applied in coal analysis and prediction. The specific process is as follows:

(1) Input the spectral data into the CNN. For the network structure of the CNN, choose 2 convolutional layers (C1, C2) and 2 sampling layers (S1, S2). Choose Sigmoid as the activation function for the convolutional layers and mean sampling function for the sampling

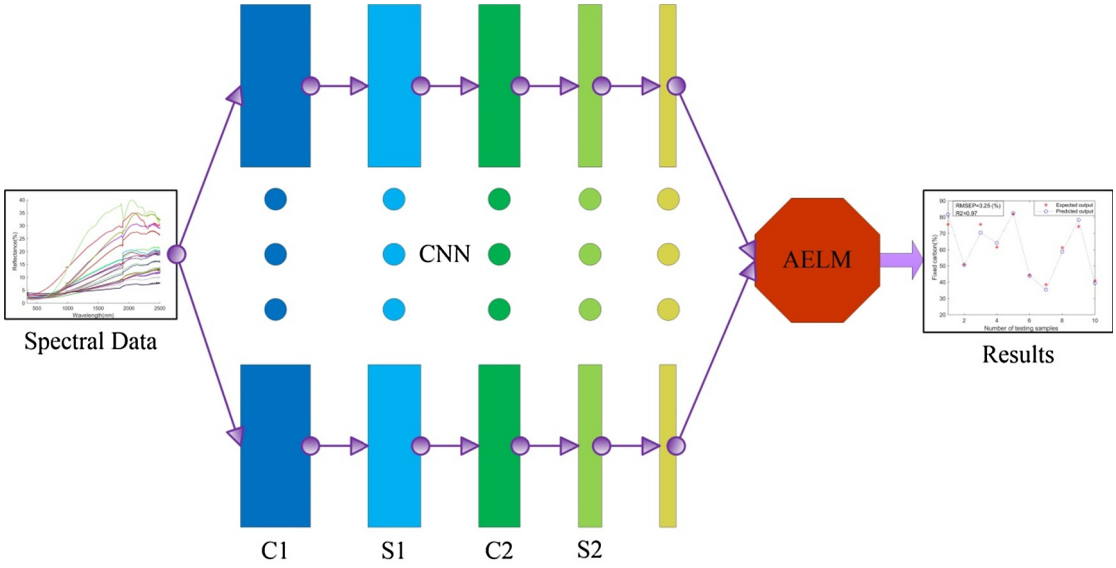

Fig. 2. The CNN-AELM proximate analysis model of coal.

layers. Every spectral datum has 1024 characteristics. However, after the feature extraction process through the CNN, the number of characteristic data that are outputted is 300.

(2) Input the characteristic data into the AELM to train and optimize the algorithm. Set the number of food sources in the ABC algorithm as 100. Additionally, set the number of employed, onlooker and scout bees as 50 each. Then, the feasible solution is d = R * NE + NE. Here, NE is the number of nodes in the hidden layers in the ELM algorithm, and R is the characteristic quantity of every inputted datum. The activation function of the ELM algorithm is the Sin function, and the number of nodes in the hidden layers is 100.

(3) Apply the prediction model into the test set and evaluate the effectiveness of the method.

3. Results and discussion

For the proximate analysis of coal, the samples in this study belong to different coal samples from different coal mines in China. Among the 100 samples total samples, we chose 90 samples as the training and calibration set and 10 samples as the testing set. To examine the effectiveness of this method, the evaluation criteria of the model's performance we applied were the coefficient of determination of prediction (R2p) and cross validation (R2c), and the root-mean-square error of prediction (RMSEP) and cross validation (RMSECV) [23,24]. R2p, R2c, RMSEP and RMSECV are defined in the following equations:

$$R2p = 1 - \frac{\sum_{i=1}^{N_{test}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N_{test}} (y_i - \bar{y})^2} \quad (12)$$

$$R2c = 1 - \frac{\sum_{i=1}^{N_{cv}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N_{cv}} (y_i - \bar{y})^2} \quad (13)$$

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{N_{test}} (y_i - \hat{y}_i)^2}{N_{test}}} \quad (14)$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^{N_{cv}} (y_i - \hat{y}_i)^2}{N_{cv}}} \quad (15)$$

Here, N_{test} and N_{cv} are the number of samples in the testing set and cross validation set, respectively; y_i is the true value of the samples; $\bar{y}$ is the mean value of the true values; and $\hat{y}_i$ is the predicted value calculated through the model.

The value range of R2p and R2c are between [0, 1]. The closer R2p and R2c are to 1 and the smaller the value of RMSEP and RMSECV are, the better the model performs.

3.1. The proximate analysis of coal based on the CNN-AELM method

3.1.1. Moisture detection

In the real burning process, moisture cannot burn or release heat. Rather, the vaporization of moisture will absorb some heat, and the existence of moisture can influence the characteristics of coal, such as the burning, the heating value and the formation of ash. Thus, the detection of moisture is undoubtedly an important item in the proximate analysis of coal.

The moisture predicted results of coal is shows in Fig. 3. The moisture value range in the testing set varies from 2.4% to 12%. We can

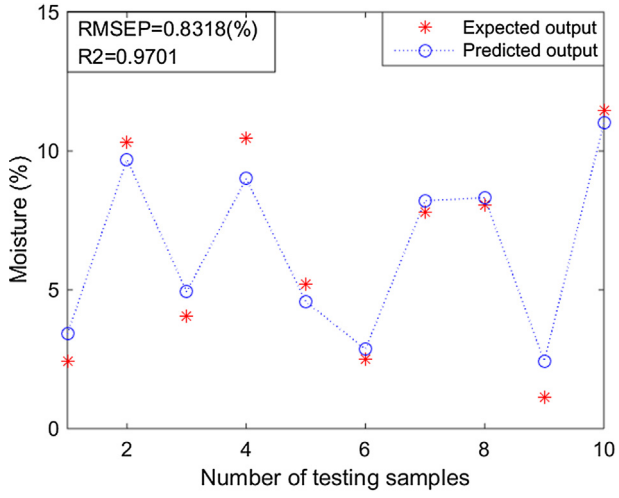

Fig. 3. The moisture predicted results of coal.

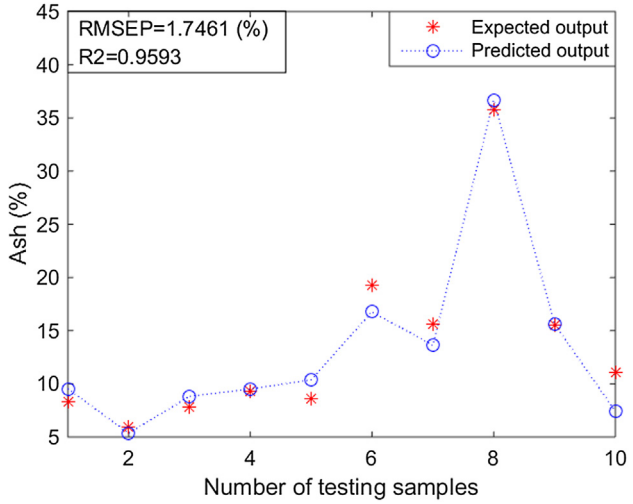

Fig. 4. The ash predicted results of coal.

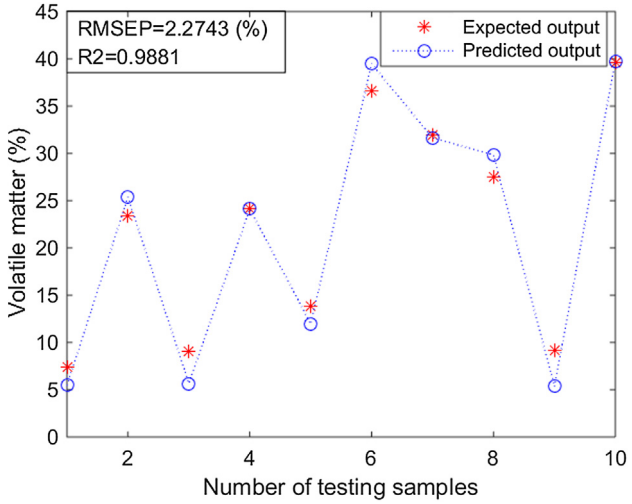

Fig. 5. The volatile matter predicted results of coal.

see that through the calculation of CNN-AELM model, we can well predict the moisture in coal. The CNN can well extract the spectral features of water. The R2p of 0.9701, the R2c of 0.9117, the RMSECV of 1.3057% and the RMSEP of 0.8318% denote the accuracy and reliability of the model.

3.1.2. Ash detection

The ash content of coal refers to the remnants after the coal is incompletely burned under the required conditions. It mainly includesAl_2O_3, CaO, SiO_2, MgO, Fe_2O_3 and some oxides of some rare elements. The ash and heating value are linearly related. When the ash increases, the heating value decreases, and the density of coal also increases. Therefore, ash is also an important item in regard to the proximate analysis of coal.

Fig. 4 shows the predicted results of ash content. The ash content value in the testing set varies from 5.9% to 35.7%. We can see that the CNN-AELM model can accurately predict the ash content in coal. The error in the testing set was mainly caused by the sixth and tenth samples. The reason for the large error is the intensive training. There was a lack of ash content values that were close to these predicted values. The R2p of 0.9593, the R2c of 0.9611, the RMSECV of 1.4364% and the RMSEP of 1.7461% show accuracy and stability of the model.

3.1.3. Volatile matter detection

The volatile matter of coal is the mass defect after coal is heated in the absence of air, which performs the moisture correction under the specified conditions. The volatile matter of coal is greatly concerned with the metamorphic grade of coal. As the coalification degree increases, the volatile matter decreases. For example, the volatile matter in lignite is usually over 38%, in bituminous coal it is over 10%, and in anthracite it is less than 10%. The volatile matter is the main factor that classifies coal.

Fig. 5 shows the predicted results of the volatile matter of coal. The volatile matter value in the testing set varies from 7.4% to 39.7%. We can see that the error is mainly caused in the third, sixth and ninth samples, while the rest of the samples are all aligned with the prediction. This shows that the CNN-AELM model can well predict the volatile matter values of coal. The R2p of 0.9881, R2c of 0.9841, the RMSECV of 1.6677% and the RMSEP of 2.2743% show the good predictive performance and stability of the model.

3.1.4. Fixed carbon detection

After we detected the volatile matter of coal samples, the coke button remained. After we excluded the ash content in the coke button, the residue is called the fixed carbon. The combustible solid in coal is the main component in coal that generates heat. The fixed carbon in coal varies according to different coal qualities. More fixed carbon content results in higher heat generated by the coal, which indicates and the better coal quality.

Fig. 6 shows the predicted results of the fixed carbon in coal. The fixed carbon value in the testing set varies from 38.5% to 82%. We observe that the predicted values are all very close to the expected values, which shows that the CNN can effectively extract the fixed spectral features in coal, and the CNN-AELM model can effectively predict the fixed carbon values of coal. The R2p of 0.9677, the R2c of 0.9734, the RMSECV of 2.398% and the RMSEP of 3.2570% show the good predictive performance and stability of the model.

3.1.5. Low heating value detection

The heating value of coal is the heat generated by the complete combustion of the unit mass of coal. According to the actual situation, we can divide the heating value of coal into high heating value and low heating value. The low heating value of coal refers to the heat value after the high heating value deducts all the latent heat of vaporization. Since the low heating value is close to the actual heating value of the boilers' operations, it has become the calculation basis for the designing and testing of boilers.

Fig. 7 shows the predicted results of the low heating value of coal. The low heating value in the testing set varies from 19.8 J/g to 27.6 J/g.

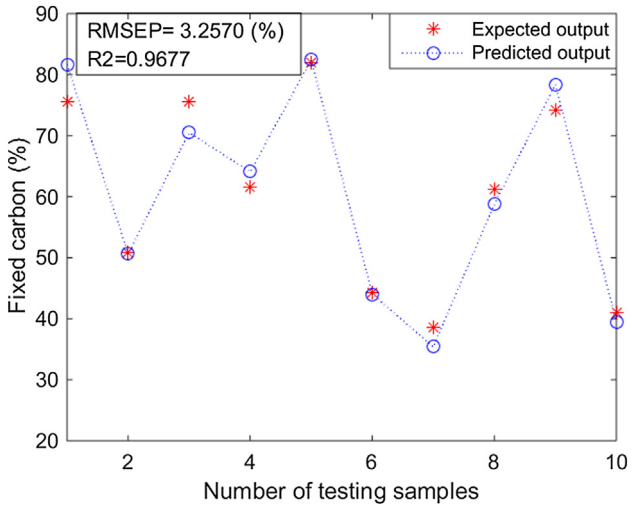

Fig. 6. The fixed carbon predicted results of coal.

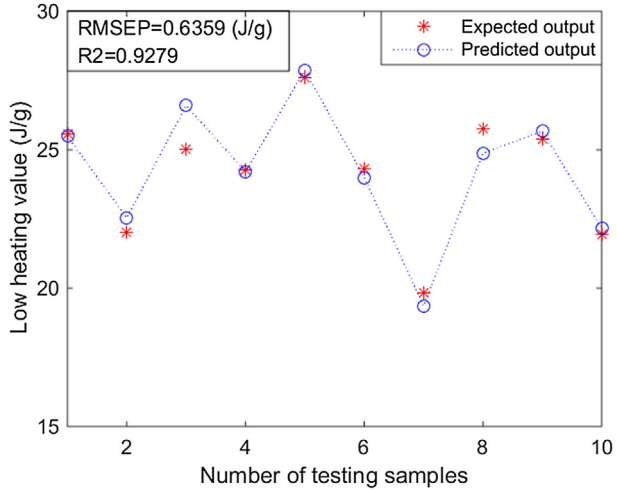

Fig. 7. The predicted results of the low heating value of coal.

We can see that apart from the tiny prediction error in the third and eighth coal samples, the CNN-AELM model has accurately predicted the low calorific values for the rest of the coal samples. The R2p of 0.9279, the R2c of 0.8994, the RMSECV of 0.7568 J/g and the RMSEP of 0.6359 J/g show the good predictive performance of the model.

3.1.6. Sulphur detection

Sulphur is one of the most harmful elements contained in coal. Coal contains a lot of sulphur. Therefore, when it combusts, a lot of SO_2 is generated. The SO_2 emitted by combusting coal takes up 90% of its total emission load, which severely harms the natural environment and erodes boilers. The sulphur in coal will depreciate the quality of coke, which will directly influence the quality of iron and steel and increase the amount of slag discharged by blast furnaces. Therefore, before using coal, it is indispensable to determine the amount of sulphur in coal.

Fig. 8 shows the predicted results of the amount of sulphur in coal. The sulphur amount in the testing set varies from 0.06% to 0.71%. We can see that CNN-AELM model can accurately predict the amount of sulphur in coal. The error in the testing set was mainly caused by the sixth and tenth samples. The R2p of 0.918, the R2c of 0.9537, the RMSECV of 0.048% and the RMSEP of 0.0529% show the accuracy and stability of the model.

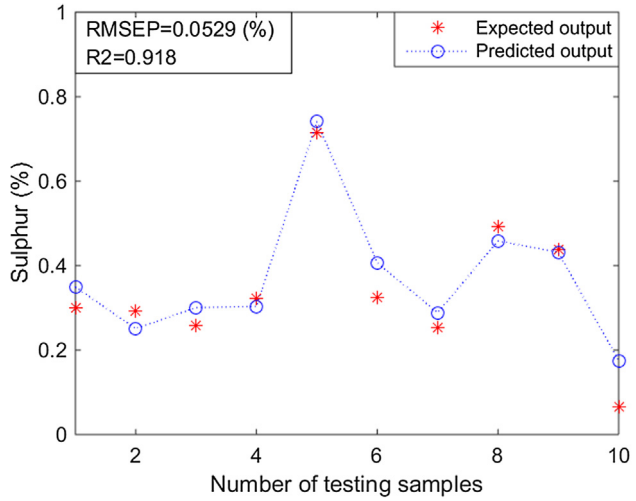

Fig. 8. The predicted results of the amount of sulphur in coal.

3.2. A comparison among all kinds of methods

The successive projections algorithm (SPA) is an important method to extract features through wave length variations [25]. The advantage of this method lies in the fact that it can extract several characteristic wavelengths in the whole spectral range and eliminate the redundant information. Therefore, it can reduce the complexity of the mode, reduce the calculation amount and increase the speed and efficiency of the model. This algorithm is usually used in the screening of spectral characteristic wavelengths [26].

The support vector machine (SVM) is a machine learning method developed by Vapnik [27], and is a general-purpose learning method based on statistical theories. It can effectively solve the problems such as finite samples, high dimensions and nonlinearity. In recent years, it has been widely applied in many classification and model construction fields [28–30].

In this study, we compared the predicted results of the CNN-ELM, CNN-AELM, CNN-SVM, SPA-ELM, SPA-AELM and SPA-SVM methods. The predicted results are shows in Table 2.

According to Table 2, we can see that for all the models that applied the CNN to extract features, the R2p values are above 0.9. The predictions of various indexes made through the CNN-AELM model is overall better than that of the CNN-ELM and CNN-SVM models, especially when predicting volatile matter in which the R2p value of the CNN-AELM model reached 0.99. Relatively speaking, the models that applied the SPA to extract features had poorer model performance. The result demonstrates that data that had gone through the CNN feature extraction process can allow models to more accurately predict the indexes of coal, and the model's performance can be better overall than the counterpart models with SPA feature extraction. It also proves that the artificial bee colony algorithm can effectively optimize the ELM network. The improved ELM model has better performance than before.

Table 3 shows the comparisons of the investment costs and the time consumed between the chemical analysis method and the CNN-AELM network model analysis, which have been respectively applied to 100 coal samples. We can see that the chemical method entails a large amount of money and time. Moreover, the chemical analysis method requires the purchase of experimental equipment, and some chemical experimental devices can cost more than $300,000 USD. Although it has high accuracy, it requires huge costs and a rather long time. At the same time, the costs of the method discussed in this research are below $30,000 USD, including the costs of the spectrometer and computers. Relatively speaking, this method has the advantages of lower costs, higher analysis accuracy and shorter time.

4. Conclusion

In this study, we proposed a coal proximate analysis model based on a combination of visible-infrared spectroscopy and deep neural networks. We first collected the spectral data of 100 samples of different types and applied the deep learning CNN and ELM algorithms to construct a coal analysis model. To further improve accuracy, we incorporated the artificial bee colony algorithm to optimize the CNN-ELM network model. The result illustrated that the CNN-AELM network can well predict the industrial indexes of coal. Compared with the SPA feature extraction method, the CNN can better extract the features of spectral data. Up until now, the CNN has been widely applied to image processing and voice recognition systems and has been rarely used in other fields. In this research, we discovered that the CNN can well extract the features of spectral data. The CNN and ELM network can compensate for each other's respective disadvantages when combined and can construct a better analysis model. Compared with traditional coal analysis, the method that combines visible-infrared spectroscopy with deep neural networks has unparalleled advantages and important value for practical applications in terms of efficiency, speed and accuracy.

Table 2
The comparisons between different coal proximate analysis models.

Feature extraction methods	Predict algorithms	Ash (%)		Volatile matter (%)		Fixed carbon (%)		Low heating value (J/g)		Moisture (%)		Sulphur (%)	
		R2p	RMSEP	R2p	RMSEP	R2p	RMSEP	R2p	RMSEP	R2p	RMSEP	R2p	RMSEP
CNN	ELM	0.86	7.52	0.88	4.80	0.94	4.11	0.89	1.07	0.95	0.95	0.88	0.18
	AELM	**0.96**	**1.75**	**0.99**	**2.27**	**0.97**	**3.26**	**0.93**	**0.64**	**0.97**	**0.83**	**0.92**	**0.053**
	SVM	0.92	3.41	0.94	3.42	0.86	6.30	0.86	1.23	0.91	1.23	0.85	0.17
SPA	ELM	0.82	8.94	0.81	8.48	0.87	9.23	0.83	0.90	0.95	0.96	0.71	0.23
	AELM	0.88	2.90	0.93	3.41	0.91	5.07	0.87	1.07	0.96	0.76	0.88	0.16
	SVM	0.80	6.74	0.87	4.58	0.92	5.83	0.75	1.74	0.89	1.30	0.77	0.14

Table 3
The comparisons between different analysis methods.

Analysis methods	Time (h)	Costs (USD)
Chemical analysis	240	$5000
This work (CNN-AELM model)	10	$100

Conflict of interest

The authors declared that there is no conflict of interest.

Acknowledgment

This research work is supported by National Natural Science Foundation of China (Grant Nos. 41371437, 61203214, 61473072, 61773105 and 61374147). Fundamental Research Funds for the Central Universities (Grant Nos. N150402001 and N160404008). National Key Research and Development Plan (2016YFC0801602). National Twelfth Five-Year Plan for Science and Technology Support (2015BAB15B01) PR China

References

[1] J. Guo, W.M. Li, X.Y. Yao, Supply and demand analysis of global coal resource, China Coal 41 (2015) 125–129.
[2] H. Kaya-Celiker, P.K. Mallikarjunan, A. Kaaya, Mid-infrared spectroscopy for discrimination and classification of Aspergillus, spp. contamination in peanuts, Food Control 52 (2015) 103–111.
[3] Y.C. Mao, D. Xiao, J.P. Cheng, J.H. Jiang, B.T. Le, S.J. Liu, Research in magnesite grade classification based on near infrared spectroscopy and ELM algorithm, Spectrosc. Spect. Anal. 37 (2017) 89–94.
[4] B.T. Le, D. Xiao, D. Okello, D.K. He, J.L. Xu, T.T. Doan, Coal exploration technology based on visible-infrared spectra and remote sensing data, Spectr. Lett. 50 (2017) 440–450.
[5] R.R. Karna, G.M. Hettiarachchi, M. Newville, C. Sun, Q. Ma, Synchrotron-based X-Ray spectroscopy studies for redox-based remediation of lead, zinc, and cadmium in mine waste materials, J. Environ. Qual. 45 (2016) 1883–1892.
[6] D. Yang, D.D. He, A.X. Lu, D. Ren, J.H. Wang, Combination of spectral and textural information of hyperspectral imaging for the prediction of the moisture content and storage time of cooked beef, Infrared Phys. Technol. 83 (2017) 206–216.
[7] R.L. Nyenge, H.C. Swart, O.M. Ntwaeaborwa, Luminescent properties, intensity degradation and X-ray photoelectron spectroscopy analysis of CaS: Eu^{2+}, powder, Opt. Mater. 40 (2015) 68–75.
[8] J.M. Andres, M.T. Bona, Analysis of coal by diffuse reflectance near-infrared spectroscopy, Anal. Chim. Acta 535 (2005) 123–132.
[9] M.T. Bona, J.M. Andres, Coal analysis by diffuse reflectance near-infrared spectroscopy: hierarchical cluster and linear discriminant analysis, Talanta 72 (2007) 1423–1431.
[10] W.K. Dong, J.M. Lee, J.S. Kim, Application of near infrared diffuse reflectance spectroscopy for on-line measurement of coal properties, Korean J. Chem. Eng. 26 (2009) 489–495.
[11] Z. Li, P.M. Fredericks, C.R. Ward, L. Rintoul, Chemical functionalities of high and low sulfur Australian coals: a case study using micro attenuated total reflectance Fourier transform infrared (ATR–FTIR) spectrometry, Org. Geochem. 41 (2010) 554–558.
[12] Y.Y. Zhai, J.B. Chen, B.M. Wu, F. Xiao, Y.S. Wang, G. Wei, Study on a rapid coal ash determination method based on near-infrared diffuse reflectance spectroscopy, J. Yunnan Univ. 35 (2013) 214–218.
[13] Y. Lecun, L. Bottou, Y. Bengio, Gradient-based learning applied to document recognition, Proc. IEEE 86 (1998) 2278–2324.
[14] S. Turaga, J. Murray, V. Jain, Convolutional networks can learn to generate affinity graphs for image segmentation, Neural Comput. 22 (2010) 511–538.
[15] X. Niu, Y. Suen, A novel hybrid CNN–SVM classifier for recognizing handwritten digits, Pattern Recogn. 45 (2012) 1318–1325.
[16] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: theory and applications, Neurocomputing 70 (2006) 489–501.
[17] W. Li, C. Chen, H. Su, Local binary patterns and extreme learning machine for hyperspectral imagery classification, IEEE Trans. Geosci. Remote Sensing 53 (2015) 3681–3693.
[18] D. Xiao, B.T. Le, Y.C. Mao, J.H. Jiang, L. Song, S.J. Liu, Research on coal exploration technology based on satellite remote sensing, J. Sens. (2016) 1–9.
[19] A. Iosifidis, A. Tefas, I. Pitas, Dynamic action recognition based on dynemes and extreme learning machine, Pattern Recogn.Lett. 68 (2015) 1890–1898.
[20] X. Sun, J. Xu, C. Jiang, Extreme learning machine for multi-label classification, Entropy 18 (2016) 1–12.
[21] K. Cao, G. Wang, D. Han, J. Ning, X. Zhang, Classification of uncertain data streams based on extreme learning machine, Cogn. Comput. 7 (2015) 150–160.
[22] D. Karaboga, B. Gorkemli, C. Ozturk, N. Karaboga, A comprehensive survey: artificial bee colony (ABC) algorithm and applications, Artif. Intell. Rev. 42 (2014) 21–57.
[23] X. Tian, Q.Y. Wang, J.B. Li, F. Peng, W.Q. Huang, Non-destructive prediction of soluble solids content of pear based on fruit surface feature classification and multivariate regression analysis, Infrared Phys. Technol. 92 (2018) 336–344.
[24] X.D. Sun, M.X. Zhou, Y.Z. Sun, Variables selection for quantitative determination of cotton content in textile blends by near infrared spectroscopy, Infrared Phys. Technol. 77 (2016) 65–72.
[25] Y.Y. Fan, Z.J. Qiu, C. Jian, W.U. Xiang, H.E. Yong, Identification of varieties of dried red jujubes with near-infrared hyperspectral imaging, Spectrosc. Spect. Anal. 37 (2017) 836–840.
[26] S.K. Yan, H.H. Yang, B.C. Hu, C.C. Ren, Z.B. Liu, Variable selection method of NIR spectroscopy based on least angle regression and GA-PLS, Spectrosc. Spect. Anal. 37 (2017) 1733–1738.
[27] V. Vapnik, C. Cortes, Support vector networks, Mach. Learn. 20 (1995) 273–297.
[28] Y.S. Wang, M. Yang, G. Wei, R.F. Hu, Z.Y. Luo, G. Li, Improved PLS regression based on SVM classification for rapid analysis of coal properties by near-infrared reflectance spectroscopy, Sens. Actuator B-Chem. 193 (2014) 723–729.
[29] H. Zhou, Q. Tang, L. Yang, Y. Yan, G. Lu, K. Cen, Support vector machine based online coal identification through advanced flame monitoring, Fuel 117 (2014) 944–951.
[30] A.K. Patel, S. Chatterjee, A.K. Gorai, Development of machine vision-based ore classification model using support vector machine (SVM) algorithm, Arab. J. Geosci. 10 (2017) 1–16.