

Research Article

Characterization and Identification of Coal and Carbonaceous Shale Using Visible and Near-Infrared Reflectance Spectroscopy

En Yang , Shirong Ge, and Shibo Wang 

School of Mechatronic Engineering, China University of Mining and Technology, Xuzhou 221116, China

Correspondence should be addressed to Shibo Wang; wangshb@cumt.edu.cn

Received 31 July 2018; Revised 17 October 2018; Accepted 22 October 2018; Published 6 November 2018

Academic Editor: Vincenza Crupi

Copyright © 2018 En Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Because of the high organic carbon concentration in carbonaceous shale, a large proportion of carbonaceous shales are often misclassified into coals using visible and near-infrared (VIS-NIR) reflectance spectroscopy in the field of coal-gangue identification of hyperspectral remote sensing of coal mine. In order to study spectral characterization of coal and carbonaceous shale, three bituminite samples and three carbonaceous shales were collected from a coal mine of China, and their spectral reflectance curves were obtained by a field spectrometer in the wavelength range of 350–2500 nm. Only one carbonaceous shale could be easily identified from the three bituminite samples according to obvious absorption valleys near 1400 nm, 1900 nm, and 2200 nm of its reflectance curve while the other two carbonaceous shales have similar reflectance curves to the three bituminite samples. The effect of carbon concentration on reflectance curve was simulated by the mixed powder of ultralow ash bituminite and clay in 0.5 mm grain size under various mixing ratios. It was found that absorption valleys near 1400 nm, 1900 nm, and 2200 nm of the mixed powder become not obvious when the bituminite content is more than 30%. In order to establish an effective identification method of coal and carbonaceous shale, 250 other samples collected from the same coal mine were divided into 150 training samples and 100 prediction samples. Principal component analysis (PCA) and Gauss radial basis kernel principal component analysis (GRB-KPCA) were employed to extract principal components (PCs) of continuum removed (CR) spectra of the training samples in eight selected wavelength regions which are related to the main mineral and organic compositions. Two support vector machine- (SVM-) based models PCA-SVM and GRB-KPCA-SVM were established. The results showed that the GRB-KPCA-SVM model had better identification accuracies of 94% and 92% for powder and nature block prediction samples, respectively.

1. Introduction

Coal is the main energy source in China and still the irreplaceable main energy in the foreseeable future. Carbonaceous shale which is sandwiched in coal seam and was formed during coal formation is mainly composed of minerals and organic carbon. During the coal production, lots of carbonaceous shales which are the main part of gangues are often cut together with coals and transported to the ground. These gangues which occupy a large amount of land on the ground can cause damages to the environment, and the ones which are mixed in coals increase the ineffective transports [1, 2].

Due to the huge coal demand and coal production capacity requirement, lots of advanced spectral technologies

have been employed in coal mine investigation, coal-gangue identification, and coal property determination [3–6]. Reflectance spectrum is easily acquired using cheap instrument and suitable for online analysis in the visible and near-infrared (VIS-NIR) region [7]. However, in coal mine areas, most carbonaceous shales have similar black color to coals. Meanwhile, due to the high carbon concentration in carbonaceous shale, there is no obviously different spectral absorption feature between coals and many carbonaceous shales, and they may have similar spectral characterization in the VIS-NIR region. Therefore, a large proportion of carbonaceous shales are often misclassified into coals using VIS-NIR reflectance spectroscopy in the field of coal-gangue identification of hyperspectral remote sensing of coal mine [8–10].

Under the same measurement conditions, the difference between reflectance spectra of coal and carbonaceous shale mainly depends on the material compositions. Coal which is mainly composed of complicated organic matters including aliphatic structures, aromatic structures and oxygen-containing functional group structures often contains many minerals while carbonaceous shale which is mainly composed of clay minerals and quartz often contains many organic matters [11]. The spectral properties of coal and carbonaceous shale had been studied by some scholars in the VIS-NIR region [12, 13]. However, the effect of carbon concentration on the VIS-NIR reflectance spectra of coal and carbonaceous shale had seldom been investigated.

Some scholars have focused on various experiments, algorithms, and models to improve the spectral identification accuracies of coals and gangues including a large number of carbonaceous shales in coal mine areas, such as Mao's normalized difference coal index (NDCI) model [8], Le's improved distribution maps model, and Song's classification method based on combination of reflectance spectra in the VIS-NIR region and spectral emissivity in the thermal infrared (TIR) region [9, 10]. However, all the above-mentioned scholars have not accurately discriminated coals and carbonaceous shales by the reflectance spectra in the VIS-NIR region, and carbonaceous shale is the main type of gangue that reduces the classification accuracies of coals and gangues. For instance, in Song's experiment, based on the reflectance spectra in the VIS-NIR region, all carbonaceous shales were misclassified into coals in all samples including 12 coals and 9 carbonaceous shales [10]. Therefore, an effective identification method based on the reflectance spectra of coal and carbonaceous shale in the VIS-NIR region is very necessary to be studied.

In this paper, spectral reflectance characterization (350–2500 nm) of typical bituminite and carbonaceous shale samples of China was studied. The effect of carbon concentration on spectral reflectance curves of coal and carbonaceous shale in the VIS-NIR region was investigated through an experiment. An effective coal-carbonaceous shale identification method based on VIS-NIR reflectance spectroscopy was established.

2. Materials and Methods

2.1. Spectra and Compositions Acquisition of Typical Samples. Figure 1 shows six typical samples including three bituminite samples and three carbonaceous shales collected from Malan coal mine area in Shanxi province of China. It is obvious that the appearances of carbonaceous shales and bituminite samples in Figure 1 are similar.

Reflection spectra of rough or mat solid surfaces of remote samples are closer to reflection spectra of powder samples in larger grain size (0.25–1.6 mm) than smaller grain size to determine the compositions of the surfaces [14]. For powder samples of sedimentary rock, larger particle size (0.25–1.2 mm) displays lower reflectance, but shows more obvious spectral features in their spectral bands [12]. The ancillary files of sedimentary rocks in ASTER spectral library version 2.0 of JPL pointed out that

reflectance spectra of fresh (rough) solid surfaces of sedimentary or metamorphic rocks could be approximately simulated by spectra of the very coarse particulate (0.5–1.5 mm) samples [15]. Some scholars had used reflectance spectra of powdered carbonate rock (a type of sedimentary rock) samples with grain size fractions between 0.125 mm and 0.5 mm and known chemical compositions to simulate reflectance spectra of fresh solid surfaces of block carbonate rock samples [16]. Coal and carbonaceous shale belong to sedimentary rock. Therefore, in this study, the six typical samples were crushed, air-dried, and sieved to achieve size fraction of 0.5 mm. These powder samples are divided into three fractions which were used for spectra measurement, X-ray fluorescence (XRF) analysis, and proximate analysis, respectively.

The sample powder in 0.5 mm grain size used for spectra measurement was placed in a culture dish, and the powder surface was smoothed with the culture dish cover. Then the powder surface was measured by an ASD FieldSpec field spectrometer (Analytical Spectral Devices Inc., USA) in a dark laboratory. The spectral range is 350–2500 nm, and the spectral resolution is 3 nm for 350–1000 nm, 8.5 nm for 1000–1800 nm, and 6.5 nm for 1800–2500 nm. As is shown in Figure 2, each sample was horizontal, and the fiber optic cable with field of view (FOV) of 25° of the spectrometer was 15 cm vertical above the surface of the sample. With the distance of 15 cm to the sample, the recorded spectrum was an average spectral reflectance of the surface area of 35 cm², which equals a circle of 6.5 cm in diameter. A halogen lamp at a distance of 15 cm above the surface of the sample was the light source, and the incidence angle was 45°. Ten spectral reflectance curves of each sample were recorded in the first direction, then the sample was rotated 90° in the horizontal plane and measured to record the second 10 curves and so on, obtaining 40 curves of the sample in all four directions. Mean of the 40 curves of each sample was calculated, and a total of 6 spectral curves of the samples in Figure 1 were obtained with a spectral dimension of 2151 in the wavelength range of 350–2500 nm.

Through XRF by an S8 Tiger spectrometer (Bruker Inc., Germany) and proximate analysis by a VMF10/6 volatile muffle furnace (Carbolite Inc., United Kingdom), an AAF12/18 ash muffle furnace (Carbolite Inc., United Kingdom), and a BS124S electronic analytical balance (Sartorius Inc., Germany), the contents of major mineral elements, volatile, and ash of these six samples were acquired. Categories (coal or rock) of these samples were determined by their ash yields according to ISO 11760 that ash yield of coal is less than or equal to 50% comparing with rock.

2.2. Spectra Acquisition of Bituminite and Clay Mixtures. Ultralow ash bituminite powder (2% ash yield, 8% volatile) with a grain size of 0.5 mm was mixed evenly with clay powder (62% kaolinite, 19% quartz, 10% hydromuscovite, 4% montmorillonite, 3% limonite, and 1.2% illite) with a grain size of 0.5 mm under various mixing ratios to simulate the effect of carbon concentration on reflectance



FIGURE 1: Six typical bituminite (a–c) and carbonaceous shale (d–f) samples of China.

curve. The weight mixing ratios of bituminite to clay were 0:100, 5:95, 10:90, . . . , 95:5, and 100:0 varying at the interval of 5%, and a total of 21 mixed powder samples were made. Reflectance spectra of these 21 samples were obtained under the same condition as the sample in Figure 2.

2.3. Spectra and Categories Acquisition of Modeling Samples. Two hundred other bituminite and carbonaceous shale samples collected from Malan coal mine were randomly divided into 150 training samples and 50 prediction samples in order to establish an effective identification model of coal

and carbonaceous shale. All these 200 modeling samples were crushed, air-dried, and sieved to achieve a size fraction of 0.5 mm. Each powder sample was divided into two fractions which were used for spectra measurement and proximate analysis, respectively. The training set included 56 bituminite samples and 94 carbonaceous shales, and the prediction set included 22 bituminite samples and 28 carbonaceous shales according to ISO 11760 and the ash yields acquired through proximate analysis of all these 200 samples. Reflectance spectra of these 200 powder samples in 0.5 mm grain size were obtained under the same condition as the sample in Figure 2. The preprocessing method of

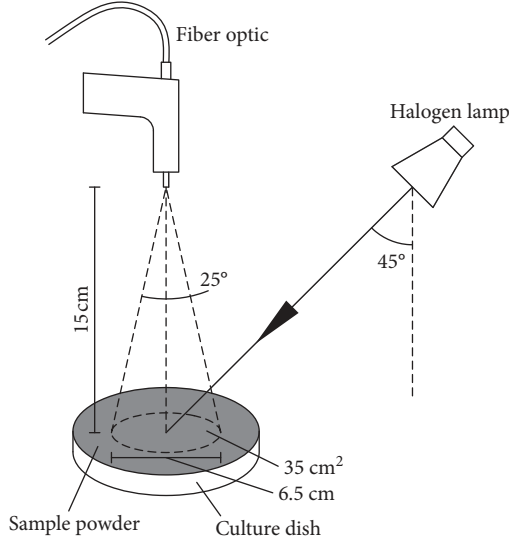


FIGURE 2: Reflectance spectra acquisition of powder sample in 0.5 mm grain size using a field spectrometer and a halogen lamp light source.

continuum removal (CR) was employed to remove background noise of the reflectance curves and isolate particular absorption features [17].

2.4. Spectra and Categories Acquisition of Block Samples. Fifty extra nature block samples collected from the same coal mine as above 200 samples were air-dried and measured using the same field spectrometer and halogen lamp in the same dark laboratory. Reflectance spectra of these 50 block samples were used to verify the practicability of the identification model established by the training samples discussed above. As is shown in Figure 3, each block sample was horizontal, and the fiber optic cable of the spectrometer was vertical above the top solid surface of the sample. The recorded spectrum was the average spectral reflectance of the circular area of about 35 cm². As the sample in Figure 2, mean of 40 reflectance curves in all four directions of each sample was acquired in the wavelength range of 350–2500 nm. The spectral reflectance curves of top solid surfaces of the 50 block samples were preprocessed by CR.

After reflectance spectra acquisition, the small sample pieces were cut from the measured circular area of the top solid surface of each sample and crushed into powder for proximate analysis. The 50 block samples included 19 bituminite samples and 31 carbonaceous shales according to ISO 11760 and the ash yields acquired through proximate analysis of these 50 cut samples.

2.5. Kernel Principal Component Analysis. Principal component analysis (PCA) is used for spectral feature extraction, and good results can be achieved when dealing with linear problems [18–20]. However, this method has some limitations when dealing with many spectral analyses, which are often not simple linear problems. Kernel principal component analysis (KPCA) is based on the PCA [21–23]. With

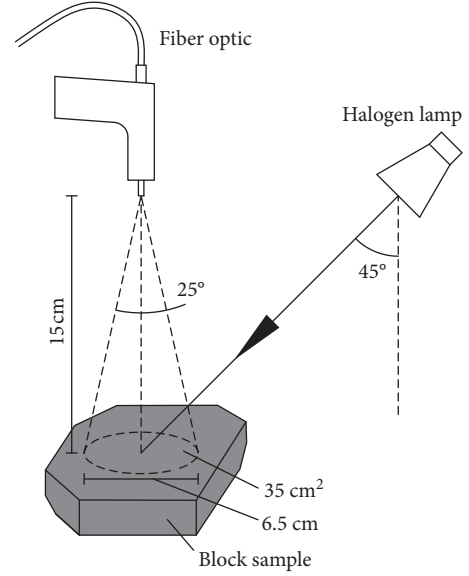


FIGURE 3: Reflectance spectra acquisition of the top solid surface of each block sample using a field spectrometer and a halogen lamp light source.

the kernel function introduced, the data are mapped into a high-dimensional space, and then, the nonlinear features are extracted by the PCA method, thus the unsatisfactory results of PCA under nonlinear data distribution are improved.

The procedures of KPCA and PCA are similar, but in KPCA, the kernel function is used instead of the original data [22]. For the linearly inseparable spectral data set $\mathbf{X}_{l \times p}$ ($\mathbf{X}_{l \times p} \in \mathbf{R}^p$) of the samples, where $\mathbf{X}_{l \times p} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_l]^T$, $\mathbf{x}_i = [\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \dots, \mathbf{x}_{i,p}]$, $i = 1, 2, \dots, l$, l is the number of samples and p is the spectral dimension, and they could be mapped into a higher dimension space \mathbf{F} through nonlinear transformation Φ .

$$\begin{cases} \Phi : \mathbf{R}^p \longrightarrow \mathbf{F} \\ \mathbf{x}_i \longrightarrow \Phi(\mathbf{x}_i) \end{cases} \quad (1)$$

$\Phi(\mathbf{X}_{l \times p})$ is defined as

$$\Phi(\mathbf{X}_{l \times p}) = [\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \dots, \Phi(\mathbf{x}_l)]^T, \quad (2)$$

and kernel function matrix is defined as

$$\begin{aligned} \mathbf{K} &= K(\mathbf{X}_{l \times p}) = \Phi(\mathbf{X}_{l \times p})\Phi(\mathbf{X}_{l \times p})^T \\ &= \begin{bmatrix} k(\mathbf{x}_1, \mathbf{x}_1) & \cdots & k(\mathbf{x}_1, \mathbf{x}_l) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_l, \mathbf{x}_1) & \cdots & k(\mathbf{x}_l, \mathbf{x}_l) \end{bmatrix}, \end{aligned} \quad (3)$$

where kernel function $k(\mathbf{x}_i, \mathbf{x}_j) = \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_j)$, $i, j = 1, 2, 3 \dots l$. The covariance matrix \mathbf{C} in \mathbf{F} is

$$\mathbf{C} = \frac{1}{l} \Phi(\mathbf{X}_{l \times p})^T \Phi(\mathbf{X}_{l \times p}) = \frac{1}{l} \sum_{i=1}^l \Phi(\mathbf{x}_i)^T \Phi(\mathbf{x}_i). \quad (4)$$

Eigenvector \mathbf{V} and eigenvalue λ of \mathbf{C} satisfy the relationship $\mathbf{C}\mathbf{V} = \lambda \mathbf{V}$ and $\mathbf{a} = (\alpha_1, \alpha_2, \dots, \alpha_l)$ exists, making

$$\mathbf{V} = \sum_{i=1}^l \alpha_i \Phi(\mathbf{x}_i). \quad (5)$$

Based on Equation (5), the relationship $\mathbf{K}\boldsymbol{\alpha} = l\lambda \boldsymbol{\alpha}$ can be concluded and $\boldsymbol{\alpha}$ can be calculated. Based on Equation (5) and $\boldsymbol{\alpha}$, the nonlinear features of any inputted spectral data \mathbf{x} in the higher dimension space \mathbf{F} can be expressed by the following equation:

$$(\mathbf{V}, \Phi(\mathbf{x})) = \sum_{i=1}^l \alpha_i k(\mathbf{x}_i, \mathbf{x}). \quad (6)$$

The above process is derived when the following equation,

$$\sum_{i=1}^l \Phi(\mathbf{x}_i) = 0, \quad (7)$$

is assumed, but in actual spectral data, K' is used instead of \mathbf{K} by

$$K' = \mathbf{K} - \mathbf{L}\mathbf{K} - \mathbf{K}\mathbf{L} + \mathbf{L}\mathbf{K}\mathbf{L}, \quad (8)$$

where

$$\mathbf{L} = \left(\frac{1}{l}\right)_{l \times l}. \quad (9)$$

In this study, Gauss radial basis (GRB) kernel function $k(\mathbf{x}_i, \mathbf{x}_j)$ is chosen and expressed by

$$k(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}\right), \quad (10)$$

where \mathbf{x}_i and \mathbf{x}_j are the spectral data of the i -th and j -th sample and σ is the width parameter.

2.6. Support Vector Machine. Support vector machine (SVM) is a machine-learning algorithm based on statistic learning theory (SLT) established by Vapnik et al. in 1990s, in which structural risk minimization (SRM) criterion is used to reduce the upper bound of the model generalization error while minimizing the sample error in order to improve the generalization ability of the model [24]. SVM can solve the problem of pattern recognition in small samples and nonlinear and high-dimensional data space [25, 26]. When dealing with nonlinear problems, the support vector classification (SVC) maps data in original space into high-dimensional space and constructs an optimal classification plane in the high-dimensional space so that the distances of all samples from the plane are maximized. Radial basis function (RBF) is chosen as the kernel function of SVM in this study.

2.7. The Proposed Procedure for Establishment of Coal-Carbonaceous Shale Prediction Models. Principal components (PCs) were obtained through PCA and KPCA of CR spectra of the training samples in eight selected wavelength regions which are related to the main mineral and organic compositions. With the PCs combined with the categories

(coal and rock) as input parameters, the SVM model was trained.

The optimal width parameter σ in formula (10) of Gauss radial basis kernel principal component analysis (GRB-KPCA) was determined by cross validation. For each width parameter σ from 1 to 100 in step length of 1, the maximum average prediction accuracy (PA) of validation set of the training set was obtained through 5-fold cross validation in SVM. At the same time, based on the maximum average PA, the optimal penalty factor c and RBF variance parameter γ of SVM for each σ were obtained. Then the optimal width parameter σ was determined by the maximum value of all maximum average PAs of all width parameters σ . GRB-KPCA-SVM model was established according to the abovementioned procedure. Figure 4 shows flowcharts of procedures of the two prediction models: PCA-SVM and GRB-KPCA-SVM.

Prediction accuracies (PAs) of categories of the 50 powder and 50 nature block prediction samples were employed to assess the prediction models. Establishment of prediction models and assessment of the models were all performed in Matlab R2013b software (MathWorks Inc., USA).

3. Results and Discussion

3.1. Spectral Reflectance Curves and Compositions of Typical Samples. Spectral reflectance curves of the six representative samples in Figure 1 in the wavelength range of 350–2500 nm are shown in Figure 5. Eight spectral feature bands are marked using No. 1–8 in Figure 5, and interactive processes of the related major compositions are shown in Table 1. In Figure 5, absorption valleys of 1400 nm (5) (5 is the spectral feature band number, and so on), 1900 nm (6), and 2200 nm (7) bands of the reflectance curve of carbonaceous shale (f) are more obvious than those of the other five samples. Carbonaceous shale (f) could be easily identified from the three bituminite samples according to the above three obvious absorption valleys, while the other two carbonaceous shales have similar reflectance curves to the three bituminite samples. The reflectance curves of carbonaceous shale (d) and bituminite (c) have similar obvious absorption valleys of 700 nm (2) and 870 nm (3) bands. The frequent absorption valleys of 400–550 nm (1) and 1000–1100 nm (4) bands appear in all reflectance curves, and the typical strong absorption bands from 2300 nm to 2500 nm (8) resulting from organic matters appear in all reflectance curves, too. There is another obvious different feature between reflectance curve of carbonaceous shale (f) and that of the other five samples that the overall shape of the reflectance curve of carbonaceous shale (f) is straight while the other five are obviously concave.

Table 2 shows the contents of major mineral elements (SiO_2 , Al_2O_3 , and Fe_2O_3) through XRF analysis and the contents of volatile and ash through proximate analysis of these six samples in Figure 1. Matrix of XRF refers to C, H, O, and N in organic matters and matrix of carbonaceous shale (d) and (e) (47.80% and 41.25%) is much more than that of carbonaceous shale (f) (20.24%). This may be the

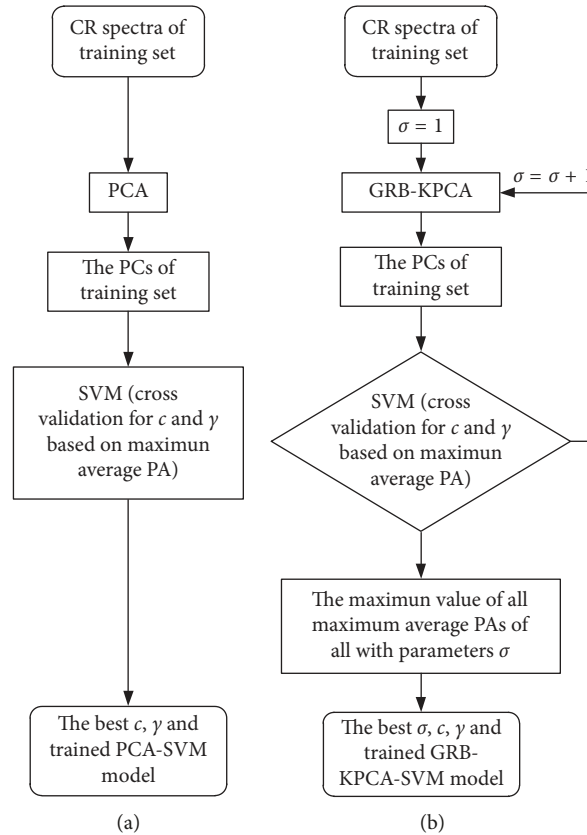


FIGURE 4: Flowcharts of procedures of PCA-SVM (a) and GRB-KPCA-SVM (b) models.

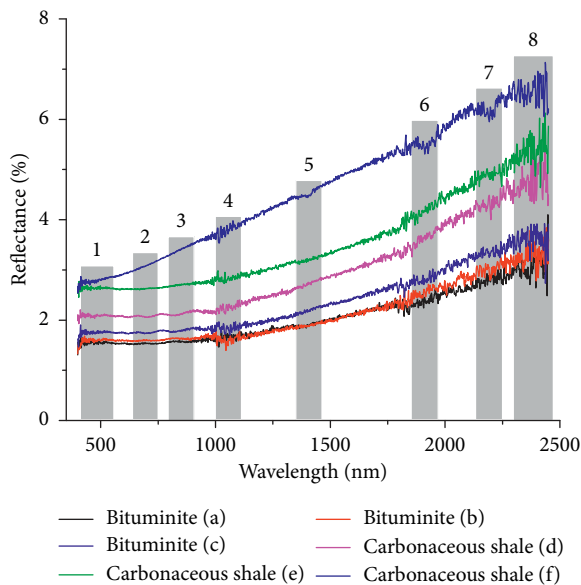


FIGURE 5: Spectral reflectance curves of the six representative samples in Figure 1 in the wavelength range of 350–2500 nm.

main reason why reflectance curves of carbonaceous shale (d) and (e) are similar to that of the three bituminite samples [12, 31, 32]. The obvious absorption valleys of 1400 nm, 1900 nm, and 2200 nm bands of the reflectance curve of carbonaceous shale (f) depends on the low matrix content

(20.24%) and clay and quartz minerals from which the high ash yield (78.94%) originates [27, 28]. The top two highest ferriferous mineral element contents (3.39% and 3.14% Fe_2O_3) in Table 2 are the main cause of obvious absorption valleys of 700 nm and 870 nm bands of carbonaceous shale (d) and bituminite (c) [27, 28].

According to the main compositions in Table 2 and the major related interactive processes in Table 1, the eight selected wavelength regions in Table 1 can reflect the main composition information of a bituminite or carbonaceous shale sample. From the waveforms of reflectance curves in Figure 5 and composition contents in Table 2, it can easily be deduced that matrix content and ash yield which are correlated to organic matters and minerals of a sample are the key factors to determine the waveform of reflectance curve of the sample. In other words, the effect of carbon concentration on the VIS-NIR reflectance spectra of coal and carbonaceous shale should be systemically investigated.

3.2. Spectral Reflectance Characterization of Bituminite and Clay Mixtures. Figure 6 shows spectral reflectance curves of the 21 mixed powder samples of dry ultralow ash bituminite and clay in 0.5 mm grain size in the wavelength range of 350–2500 nm. The bituminite content increases from 0% to 100% at the interval of 5%. The eight spectral feature bands of Table 1 are also marked in Figure 6. The pure clay powder has the highest spectral reflectance curve. However, spectral reflectance curve of the mixed powder of 5% bituminite

TABLE 1: Spectral regions related to the major compositions [27–30].

No.	Absorption band (nm)	Major related interactive process	Selected wavelength region (nm)
1	400, 430, 450, 490, 510, 520, 550	Fe ²⁺ and Fe ³⁺ in ferriferous minerals	400–550
2	700	Fe ³⁺ in ferriferous minerals	650–750
3	870	Fe ³⁺ in ferriferous minerals	820–920
4	1000–1100	Fe ²⁺ in ferriferous minerals	1000–1100
5	1400	OH ⁻ vibrations in clay minerals	1350–1450
6	1900	H ₂ O vibrations in milky quartz	1850–1950
7	2200	Al–OH vibrations in clay minerals	2150–2250
8	2300–2500	Overtones and combinations of organic absorption groups	2300–2500

TABLE 2: X-ray fluorescence (XRF) and proximate analysis of samples in Figure 1.

Sample	XRF analysis (wt. %)				Proximate analysis (wt. %)	
	SiO ₂	Al ₂ O ₃	Fe ₂ O ₃	Matrix	Volatile	Ash
Bituminite (a)	8.63	5.01	0.87	82.60	13.74	16.05
Bituminite (b)	10.61	6.17	0.81	79.12	16.14	18.22
Bituminite (c)	10.68	8.40	3.39	73.94	21.57	23.95
Carbonaceous shale (d)	26.46	19.98	3.14	47.80	7.86	51.80
Carbonaceous shale (e)	32.13	21.20	1.04	41.25	4.31	57.55
Carbonaceous shale (f)	45.65	30.14	0.92	20.24	1.51	78.94

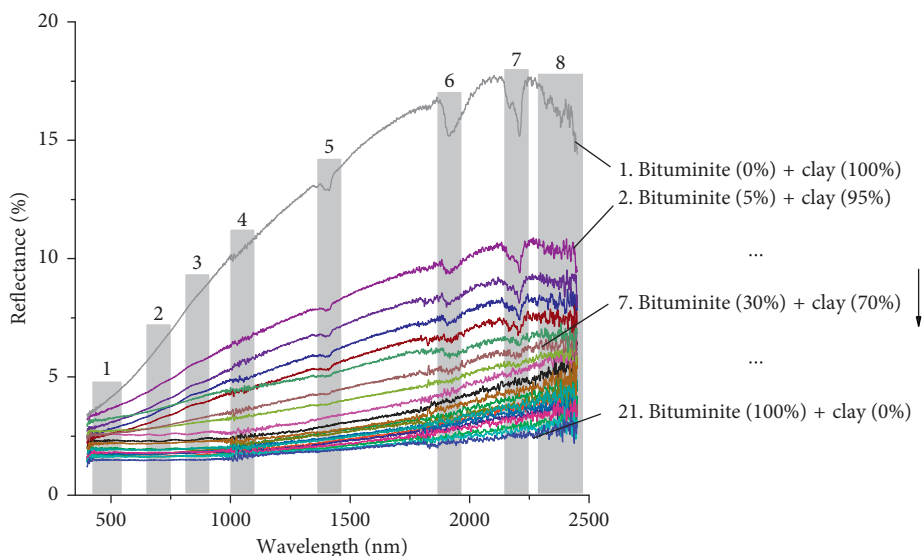


FIGURE 6: Spectral reflectance curves of 21 mixed powder samples of dry bituminite and clay in 0.5 mm grain size under different mixing ratios in the wavelength range of 350–2500 nm.

content dramatically falls comparing with the curve of pure clay powder especially in the near-infrared (NIR) region (780–2500 nm). When bituminite content is more than 5%, with increasing bituminite content, spectral reflectance curve of the mixed powder falls at a smaller interval. With increasing bituminite content, the depths of absorption valleys of 1400 nm, 1900 nm, and 2200 nm bands of the reflectance curve all decrease. When the bituminite content is more than 30%, absorption valleys of 1400 nm, 1900 nm, and 2200 nm bands of the reflectance curve become not obvious, and the overall shape of the reflectance curve changes from convex to concave. The overall shapes of the

curves are more dense, similar, and horizontal when the bituminite contents are more than 40%. The pure bituminite powder has the lowest spectral reflectance curve, and the overall shape of the curve is close to the horizontal line.

The experimental results of mixed powder of bituminite and clay explain the spectral reflectance curves of the six representative samples in Figure 5. It could be deduced that it is difficult to identify carbonaceous shale from bituminite according to the waveform of spectral reflectance curve of the sample in the VIS-NIR region when carbon concentration of the carbonaceous shale is more than about 30%. However, the material composition information contained

in the reflectance spectra in the eight selected wavelength regions in Table 1 might provide reference for identifying carbonaceous shale with high carbon concentration from bituminite. And an effective identification method based on the reflectance spectra of coal and carbonaceous shale in the VIS-NIR region should be studied.

3.3. Spectra of Modeling Samples and Establishment of Identification Models. In order to establish an effective identification model of coal and carbonaceous shale, spectral reflectance curves of 200 powder samples in 0.5 mm grain size including 150 training samples and 50 prediction samples in the wavelength range of 350–2500 nm were obtained and shown in Figure 7(a). Figure 7(b) shows the 200 spectral curves preprocessed by CR. It is obvious that less reflectance curves (12 samples) with overall convex shapes could be easily identified from others according to the CR method in Figure 7(b). But it is difficult to identify most of the samples because of many overlappings, crossovers, and similarities among these spectral curves.

Figure 7(b) indicates that the absorption features are amplified. CR spectra of the training set in the eight selected wavelength regions in Table 1 were used to establish identification models. PCA was first employed to eliminate irrelevant information and reduce dimensionality of preprocessed spectra in the eight regions. Table 3 lists cumulative contribution rates of the first ten PCs of CR spectra of the training set in the eight regions through PCA. The cumulative contribution rate of the first three PCs is 90.0403%, and the 3D spatial distribution of the first three PCs of the 150 training samples through PCA is exhibited in Figure 8(a). Figure 8(a) indicates that a large part of rock (carbonaceous shale) balls are mixed in coal balls.

The first ten PCs of the training set through PCA combining with the categories (0 stands for coal and 1 stands for rock) were taken as input variables to train the SVM model. 5-fold cross validation was employed to determine the optimal penalty parameter c and radial basis function variance parameter γ . The search ranges of c and γ were both from 2^{-10} to 2^{10} , and step length of the exponent was 0.5. The process of determining the optimal parameter c and γ is shown in Figure 8(b). With $c = 2^{1.5}$ and $\gamma = 2^{-2}$ as optimal parameters, the maximum average PA of categories of the validation set is 90.67%.

KPCA which has more abilities to deal with the nonlinear data distribution might get a better result when employed to reduce dimensionality of preprocessed spectra of the training set in the eight regions. When the width parameter σ in Formula (10) of GRB-KPCA increases from 1 to 100 in step length of 1, Figure 9 shows the variation trend of the maximum average PA of categories of the validation set through GRB-KPCA-SVM. As is shown in Figure 9, the optimal width parameter σ is 63 for GRB-KPCA-SVM. Table 3 lists the cumulative contribution rates of the first ten PCs of CR spectra of the training set in the eight regions through GRB-KPCA ($\sigma = 63$). The cumulative contribution rate of the first three PCs

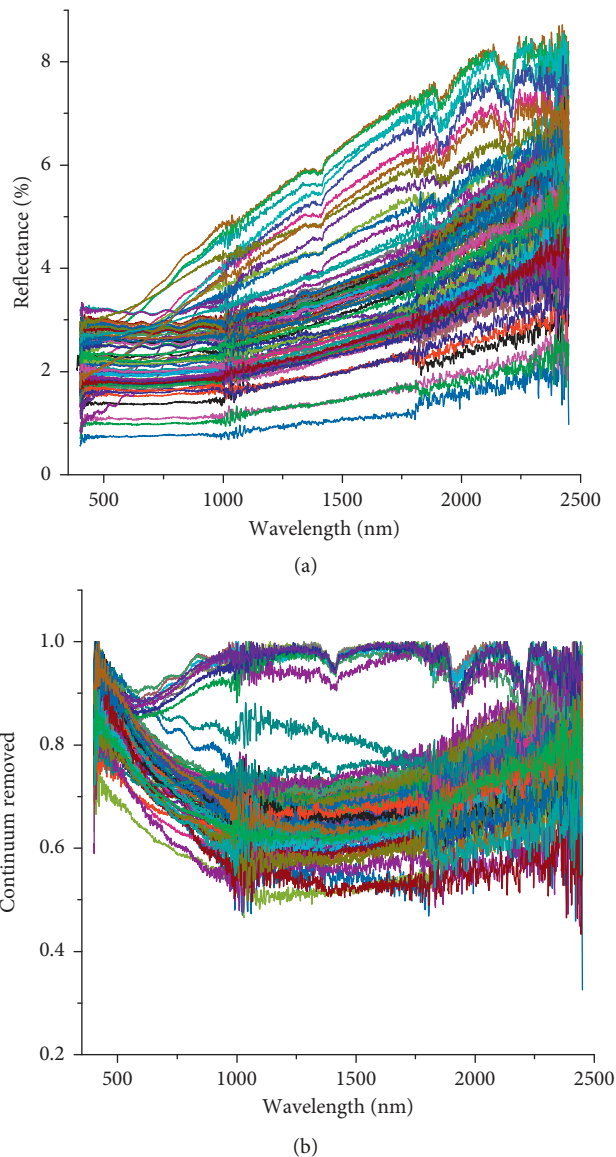


FIGURE 7: Spectral reflectance curves (a) and continuum-removed spectral curves (b) of 200 modeling samples in the wavelength range of 350–2500 nm.

TABLE 3: Cumulative contribution rates of the first ten PCs of continuum-removed spectra of the 150 training samples in the eight selected wavelength range in Table 1.

PC	PCA (%)	GRB-KPCA ($\sigma = 63$) (%)
PC1	50.0150	62.1243
PC2	82.8114	92.8212
PC3	90.0403	98.1378
PC4	93.9476	98.2543
PC5	96.6978	98.3647
PC6	97.0984	98.7834
PC7	97.2688	99.0579
PC8	98.3690	99.1521
PC9	98.6629	99.3829
PC10	98.9098	99.9251

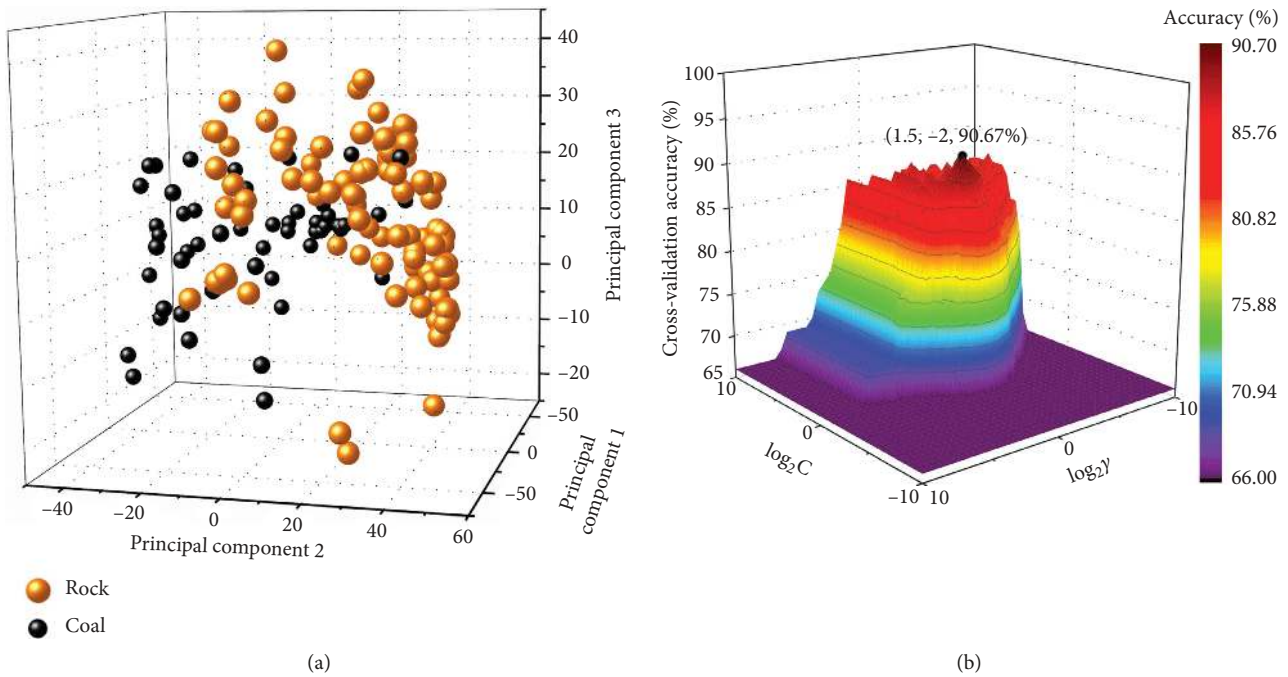


FIGURE 8: (a) 3D spatial distribution of the first three PCs of CR spectra of the 150 training samples in the eight selected regions in Table 1 through PCA; (b) process of searching for the optimal penalty parameter c and the radial basis function variance parameter γ of PCA-SVM model through cross validation with the first ten PCs as input variables.

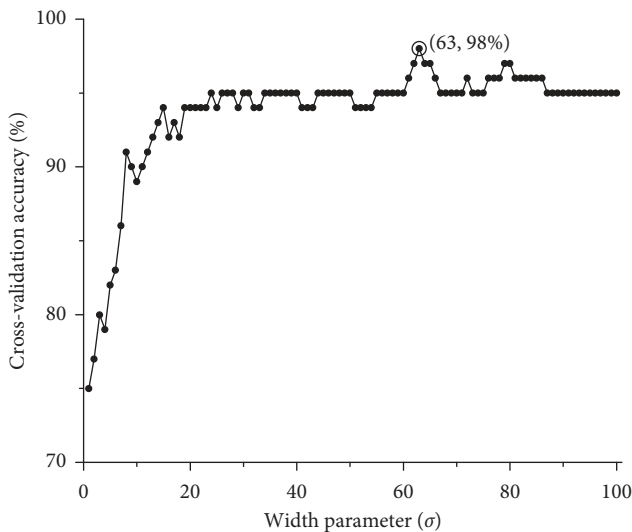


FIGURE 9: Process of determining the optimal parameter σ of GRB-KPCA: the variation trend of the maximum average PA of categories of the validation set through GRB-KPCA-SVM with the change of σ .

is 98.1378%, and the 3D spatial distribution of the first three PCs of the 150 training samples through GRB-KPCA ($\sigma = 63$) is exhibited in Figure 10(a). Figure 10(a) indicates that the 3D spatial cluster of coal and rock (carbonaceous shale) balls is better than that of PCA, and the mixed part of balls decreases.

When the width parameter σ in Formula (10) of GRB-KPCA is 63, the first ten PCs of the training set through

GRB-KPCA combining with the categories (0 stands for coal and 1 stands for rock) were taken as input variables to train the SVM model. 5-fold cross validation was employed to determine the optimal penalty parameter c and radial basis function variance parameter γ . The search ranges of c and γ were also both from 2^{-10} to 2^{10} and step length of the exponent was also 0.5. The process of determining the optimal parameter c and γ is shown in Figure 10(b). The optimal parameters are $c = 2^4$ and $\gamma = 2^0$ based on GRB-KPCA-SVM, and the maximum average PA of categories of the validation set is 98%.

From the results of cross validation of the validation set, the dimension reduction method using KPCA is better than that of PCA for CR spectra of the training set in the eight selected wavelength regions in Table 1. Two coal-carbonaceous shale prediction models, PCA-SVM and GRB-KPCA-SVM, were established according to above processes.

3.4. Identification of Samples by the Prediction Models

3.4.1. Identification of Powder Prediction Samples. CR spectra of the 50 prediction samples in 0.5 mm grain size in the eight selected wavelength regions in Table 1 were input into the above PCA-SVM ($c = 2^{1.5}$, and $\gamma = 2^{-2}$) and GRB-KPCA-SVM ($\sigma = 63$, $c = 2^4$, and $\gamma = 2^0$) models, and the results predicted by the two models are shown in Figure 11.

3.4.2. Spectra and Identification of Block Samples. Spectral reflectance curves and CR curves of top solid surfaces of the 50 extra block samples in the wavelength

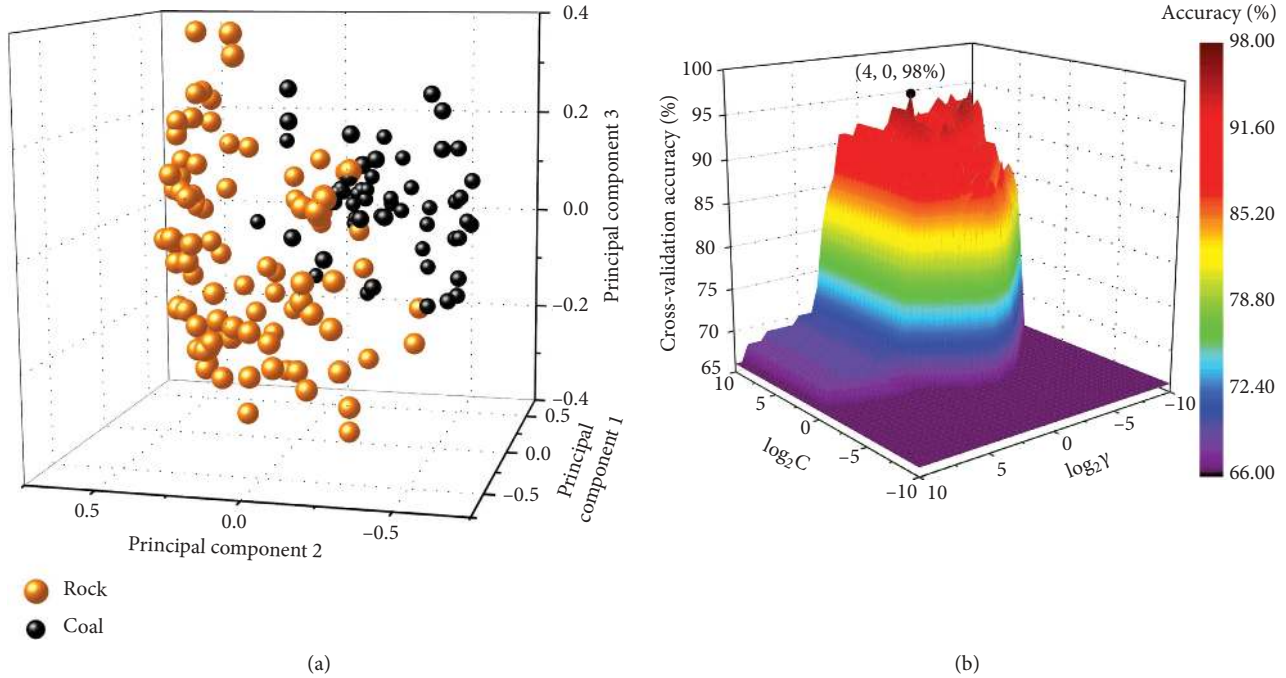


FIGURE 10: (a) 3D spatial distribution of the first three PCs of CR spectra of the 150 training samples in the eight regions in Table 1 through GRB-KPCA ($\sigma = 63$); (b) process of searching for the optimal penalty parameter c and radial basis function variance parameter γ of GRB-KPCA-SVM model through cross validation with the first ten PCs as input variables.

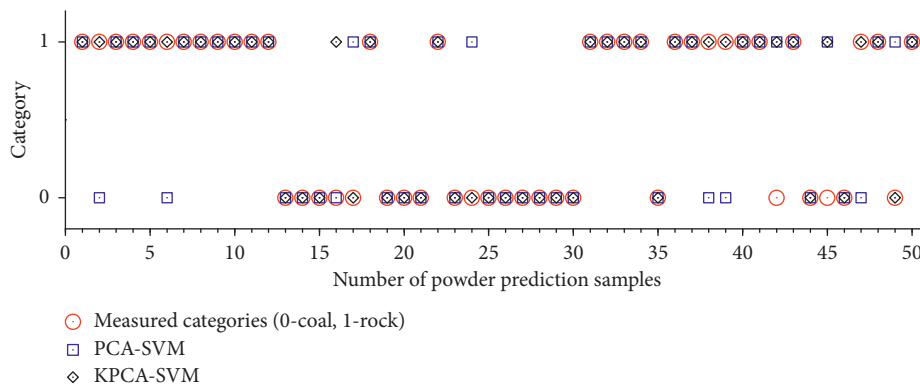


FIGURE 11: Prediction results of 50 powder prediction samples by PCA-SVM and GRB-KPCA-SVM models.

range of 350–2500 nm are shown in Figure 12(a) and Figure 12(b), respectively. Figure 12 indicates that spectral curves of block samples are more similar to spectral curves of powder samples in 0.5 mm grain size in Figure 7. Likewise, it is difficult to identify most of these block samples using the observed spectral curves in Figure 12.

In order to further verify the practicability of the above two trained models, categories of the 50 extra nature block samples were predicted. CR spectra of the 50 block samples in the eight selected wavelength regions in Table 1 were input into the above PCA-SVM ($c = 2^{1.5}$, and $\gamma = 2^{-2}$) and GRB-KPCA-SVM ($\sigma = 63$, $c = 2^4$, and $\gamma = 2^0$) models. The identification results of categories of the 50 block samples by the two models are shown in Figure 13.

3.4.3. Performance of Prediction Models. Identification accuracies of categories of the 50 powder prediction samples and the 50 extra block samples are listed in Table 4. The 150 training samples were also predicted by the two models and are listed in Table 4.

According to the results in Table 4, it can be seen that the categories of coal and carbonaceous shale samples based on GRB-KPCA-SVM have higher PAs comparing with PAs of PCA-SVM. The PA of 92% for nature block samples is acceptable, and GRB-KPCA-SVM is a more practicable and ideal coal-carbonaceous shale identification method. And the GRB-KPCA-SVM model provides a valid reference for coal-carbonaceous shale identification using VIS-NIR reflectance spectroscopy.

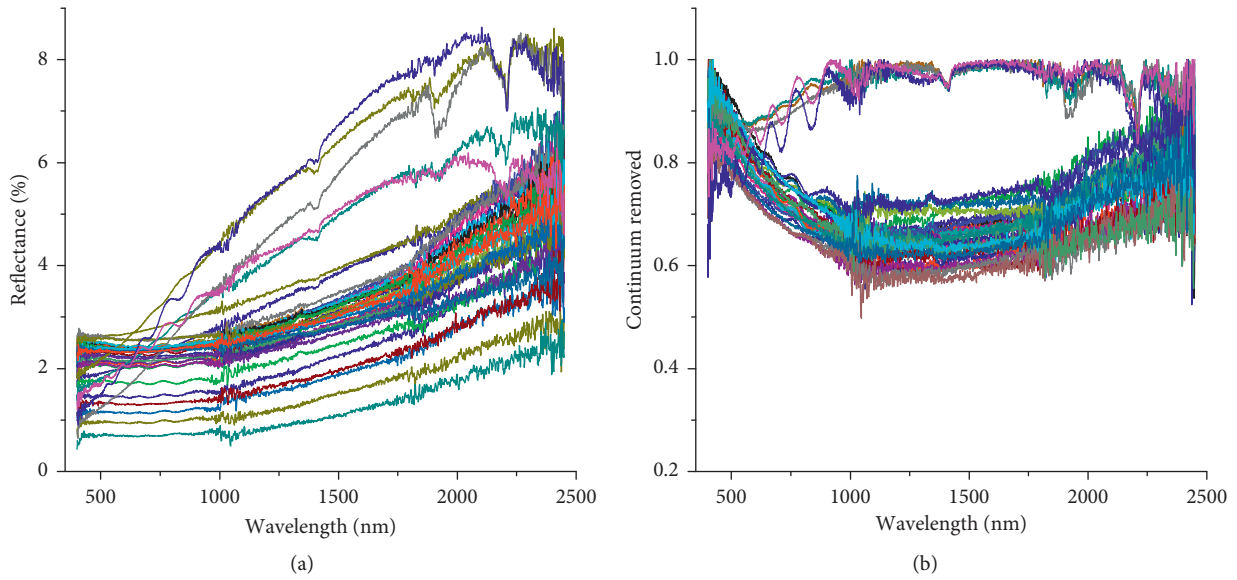


FIGURE 12: Spectral reflectance curves (a) and continuum-removed spectral curves (b) of 50 nature block samples in the wavelength range of 350–2500 nm.

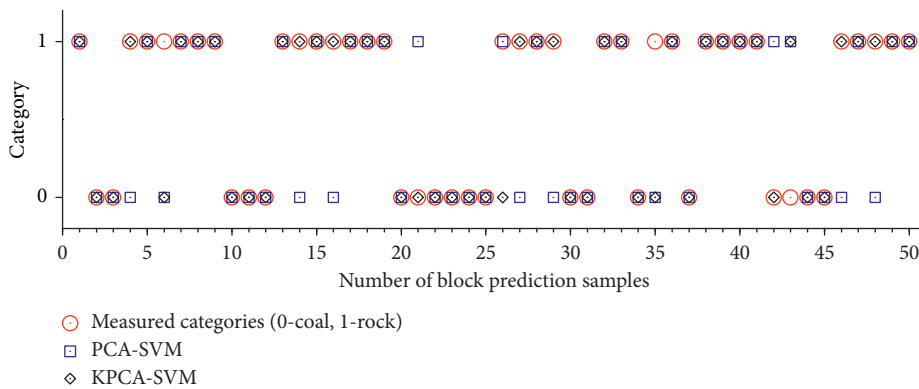


FIGURE 13: Prediction results of 50 nature block samples by PCA-SVM and GRB-KPCA-SVM models.

TABLE 4: Identification accuracies of categories of all bituminite and carbonaceous shale samples collected from Malan coal mine area in Shanxi province of China by the prediction models.

Model	150 training samples	50 prediction samples	50 block samples
PCA-SVM ($c = 2^{1.5}, \gamma = 2^{-2}$)	92%	80%	76%
GRB-KPCA ($\sigma = 63$)-SVM ($c = 2^4, \gamma = 2^0$)	100%	94%	92%

4. Conclusions

In order to solve the problem of misclassification between coal and carbonaceous shale using VIS-NIR reflectance spectroscopy in the field of coal-gangue identification of hyperspectral remote sensing of coal mine, spectral reflectance characterization of six typical bituminite, and carbonaceous shale samples of China in the wavelength range of 350–2500 nm were studied. In the six samples, two carbonaceous shales without obvious absorption valleys have similar reflectance curves to the three bituminite samples. Through simulation experiment of the effect of bituminite content on reflectance curve of mixed powder of ultralow ash bituminite and clay, it

was found that absorption valleys near 1400 nm, 1900 nm, and 2200 nm of the mixed powder become not obvious when bituminite content is more than 30%. It could be deduced that when carbon concentration of the carbonaceous shale is more than about 30%, it is difficult to identify carbonaceous shale from bituminite according to the waveform of spectral reflectance curve of the sample in the VIS-NIR region.

Eight wavelength regions including 400–550 nm, 650–750 nm, 820–920 nm, 1000–1100 nm, 1350–1450 nm, 1850–1950 nm, 2150–2250 nm, and 2300–2500 nm which are related to the main mineral and organic compositions of the samples were selected. CR spectra of 150 other coal and carbonaceous shale samples of China in the eight

selected wavelength regions were used to establish two coal-carbonaceous shale prediction models: PCA-SVM and GRB-KPCA-SVM. The two models were verified through predicting categories of 50 powder samples and 50 nature block samples. The GRB-KPCA-SVM model with prediction accuracies of 94% and 92% for powder and nature block prediction samples, respectively, is a more practicable and ideal coal-carbonaceous shale identification method. The basic principles and models in this paper provide references for coal-carbonaceous shale identification using VIS-NIR reflectance spectroscopy in coal mines.

Data Availability

The spectral data, XRF data, and proximate analysis data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

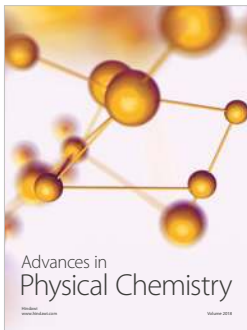
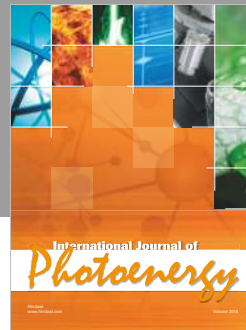
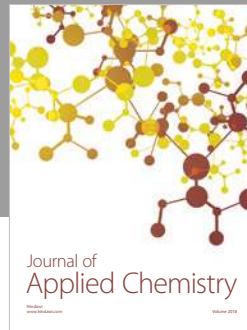
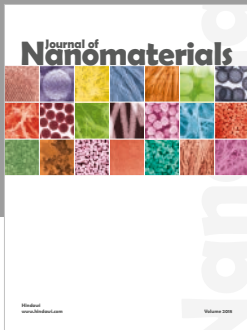
Acknowledgments

This work was supported by the Joint Funds of the National Natural Science Foundation of China (grant numbers U1610251 and U1510116), the National Key Research and Development Program of China (grant number 2018YFC0604503), and the Priority Academic Program Development (PAPD) of Jiangsu Higher Education Institutions.

References

- [1] Y. Z. Sun, J. S. Fan, P. Qin, and H. Y. Niu, "Pollution extents of organic substances from a coal gangue dump of Jiulong Coal Mine, China," *Environmental Geochemistry and Health*, vol. 31, no. 1, pp. 81–89, 2009.
- [2] J. Fan, Y. Sun, X. Li et al., "Pollution of organic compounds and heavy metals in a coal gangue dump of the Gequan Coal Mine, China," *Chinese Journal of Geochemistry*, vol. 32, no. 3, pp. 241–247, 2013.
- [3] D. Xiao, B. T. Le, Y. Mao et al., "Research on coal exploration technology based on satellite remote sensing," *Journal of Sensors*, vol. 2016, Article ID 8214801, 9 pages, 2016.
- [4] Z. Song and C. Kuenzer, "Spectral reflectance (400–2500 nm) properties of coals, adjacent sediments, metamorphic and pyrometamorphic rocks in coal-fire areas: a case study of Wuda coalfield and its surrounding areas, northern China," *International Journal of Coal Geology*, vol. 171, pp. 142–152, 2017.
- [5] X. Wang, K. Hu, L. Zhang, X. Yu, and E. Ding, "Characterization and classification of coals and rocks using terahertz time-domain spectroscopy," *Journal of Infrared Millimeter and Terahertz Waves*, vol. 38, no. 2, pp. 1–13, 2017.
- [6] M. T. Bona and J. M. Andrés, "Reflection and transmission mid-infrared spectroscopy for rapid determination of coal properties by multivariate analysis," *Talanta*, vol. 74, no. 4, pp. 998–1007, 2008.
- [7] D. W. Kim, J. M. Lee, and J. S. Kim, "Application of near infrared diffuse reflectance spectroscopy for on-line measurement of coal properties," *Korean Journal of Chemical Engineering*, vol. 26, no. 2, pp. 489–495, 2009.
- [8] Y. Mao, B. Ma, S. Liu et al., "Study and validation of a remote sensing model for coal extraction based on reflectance spectrum features," *Canadian Journal of Remote Sensing*, vol. 40, no. 5, pp. 327–335, 2014.
- [9] B. T. Le, D. Xiao, D. Okello et al., "Coal exploration technology based on visible-infrared spectra and remote sensing data," *Spectroscopy Letters*, vol. 50, no. 8, pp. 440–450, 2017.
- [10] L. Song, S. Liu, M. Yu, Y. Mao, and L. Wu, "A classification method based on the combination of visible, near-infrared and thermal infrared spectrum for coal and gangue distinguishment," *Spectroscopy and Spectral Analysis*, vol. 37, no. 2, pp. 416–422, 2017.
- [11] A. K. Permana, C. R. Ward, Z. Li, and L. W. Gurba, "Distribution and origin of minerals in high-rank coals of the south Walker Creek area, Bowen basin, Australia," *International Journal of Coal Geology*, vol. 116, no. 9, pp. 185–207, 2013.
- [12] G. R. Hunt and J. W. Salisbury, "Visible and near-infrared spectra of minerals and rocks: XI. Sedimentary rocks," *Modern Geology*, vol. 5, pp. 211–217, 1976.
- [13] E. A. Cloutis, "Quantitative characterization of coal properties using bidirectional diffuse reflectance spectroscopy," *Fuel*, vol. 82, no. 18, pp. 2239–2254, 2003.
- [14] G. R. Hunt and R. K. Vincent, "The behavior of spectral features in the infrared emission from particulate surfaces of various grain sizes," *Journal of Geophysical Research*, vol. 73, no. 18, pp. 6039–6046, 1968.
- [15] A. M. Baldridge, S. J. Hook, C. I. Grove, and G. Rivera, "The ASTER spectral library version 2.0," *Remote Sensing of Environment*, vol. 113, no. 4, pp. 711–715, 2009.
- [16] N. Zaini, F. van der Meer, and H. van der Werff, "Determination of carbonate rock chemistry using laboratory-based hyperspectral imagery," *Remote sensing*, vol. 6, no. 5, pp. 4149–4172, 2014.
- [17] Q. Fang, H. Hong, L. Zhao et al., "Visible and near-infrared reflectance spectroscopy for investigating soil mineralogy: a review," *Journal of Spectroscopy*, vol. 2018, Article ID 3168974, 14 pages, 2018.
- [18] M. Blanco and A. Peguero, "Analysis of pharmaceuticals by NIR spectroscopy without a reference method," *Trends in Analytical Chemistry*, vol. 29, no. 10, pp. 1127–1136, 2010.
- [19] D. Barbin, G. Elmasry, D. Sun, and P. Allen, "Near-infrared hyperspectral imaging for grading and classification of pork," *Meat Science*, vol. 90, no. 1, pp. 259–268, 2012.
- [20] Y. Liu, X. Sun, and A. Ouyang, "Nondestructive measurement of soluble solid content of navel orange fruit by visible-NIR spectrometric technique with PLSR and PCA-BPNN," *LWT-Food Science and Technology*, vol. 43, no. 4, pp. 602–607, 2010.
- [21] J. Wu, J. Wang, and L. Liu, "Feature extraction via KPCA for classification of gait patterns," *Human Movement Science*, vol. 26, no. 3, pp. 393–411, 2007.
- [22] M. B. Richman and I. Adrianto, "Classification and regionalization through kernel principal component analysis," *Physics and Chemistry of the Earth*, vol. 35, no. 9, pp. 316–328, 2010.
- [23] G. Fu, F. Y. Shih, and H. Wang, "A kernel-based parametric method for conditional density estimation," *Pattern Recognition*, vol. 44, no. 2, pp. 284–294, 2011.

- [24] V. N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, Berlin, Germany, 1995.
- [25] T. Oommen, D. Misra, N. K. C. Twarakavi et al., "An objective analysis of support vector machine based classification for remote sensing," *Mathematical Geosciences*, vol. 40, no. 4, pp. 409–424, 2008.
- [26] C. Chang and C. Lin, "LIBSVM: a library for support vector machines," *ACM transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.
- [27] G. R. Hunt and J. W. Salisbury, "Visible and near-infrared spectra of minerals and rocks: I silicate minerals," *Modern Geology*, vol. 1, pp. 283–300, 1970.
- [28] G. R. Hunt, "Spectral signatures of particulate minerals in the visible and near infrared," *Geophysics*, vol. 42, no. 3, pp. 501–513, 1977.
- [29] J. M. Andrés and M. T. Bona, "Analysis of coal by diffuse reflectance near-infrared spectroscopy," *Analytica Chimica Acta*, vol. 535, no. 1, pp. 123–132, 2005.
- [30] X. Xie, Y. Zhao, P. Qiu et al., "Investigation of the relationship between infrared structure and pyrolysis reactivity of coals with different ranks," *Fuel*, vol. 216, pp. 521–530, 2018.
- [31] R. N. Clark, "Spectral properties of mixtures of montmorillonite and dark carbon grains: implications for remote sensing minerals containing chemically and physically adsorbed water," *Journal of Geophysical Research: Solid Earth*, vol. 88, no. B12, pp. 10635–10644, 1983.
- [32] R. N. Clark, "Spectroscopy of rocks and minerals, and principles of spectroscopy," *Manual of Remote Sensing*, vol. 3, pp. 3–58, 1999.



Hindawi

Submit your manuscripts at
www.hindawi.com

