Optics & Laser Technology 181 (2025) 111734

Contents lists available at ScienceDirect

Optics and Laser Technology

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

Full length article

Research on accurate analysis of coal quality using NIRS-XRF fusion spectroscopy in complex coal type scenarios

Jiaxuan Li [a,b], Rui Gao [a,b], Yan Zhang [c], Lei Zhang [a,b,*], Lei Dong [a,b], Weiguang Ma [a,b], Wangbao Yin [a,b,*], Suotang Jia [a,b]

[a] State Key Laboratory of Quantum Optics and Quantum Optics Devices, Institute of Laser Spectroscopy, Shanxi University, Taiyuan 030006, China
[b] Collaborative Innovation Center of Extreme Optics, Shanxi University, Taiyuan 030006, China
[c] School of Optoelectronic Engineering, Xi'an Technological University, Xian 710021, China

ARTICLE INFO

Keywords:
Near-infrared spectroscopy (NIRS)
X-ray fluorescence (XRF)
t-Distributed Stochastic Neighbor Embedding (t-SNE)
Coal quality analysis

ABSTRACT

Coking coal in coal chemical enterprises presents a challenge due to its diverse types, wide sources, and variable quality, influenced by varying degrees of metamorphism and physicochemical properties. These differences not only impact coal quality but also hinder accurate analysis. Rapid and precise coal quality detection amidst diverse coal types is crucial for maintaining stable production and ensuring coke quality. This study employs near-infrared spectroscopy (NIRS) and X-ray fluorescence spectroscopy (XRF) fusion spectroscopy analysis, along with Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE), for data dimensionality reduction and visualization to devise a coal sample classification strategy. Subsequently, Support Vector Machine (SVM) is employed for automatic coal sample classification based on this strategy. Finally, Partial Least Squares Regression (PLSR) is used to establish regression models and evaluate their performance in predicting coal quality. Results show that classified regression model achieves R^2 values of 0.9987, 0.9955, and 0.9997 for ash content, volatile matter, and sulfur content, with corresponding root mean square error for prediction ($RMSE_P$) of 0.31 %, 1.34 %, and 0.05 %, and the mean absolute relative error for prediction ($MARD_P$) of 2.48 %, 3.58 %, and 3.57 %, respectively. Compared to the unclassified model, there is a significant enhancement in prediction accuracy. The classification and modeling method proposed herein effectively improve the accuracy of coal quality analysis in complex coal type scenarios, crucial for industries like coal chemical engineering to enhance production efficiency and optimize coal resource utilization.

1. Introduction

Coal is the most abundant fossil fuel in the world, mainly used in industries such as coking, thermal power generation, metallurgy, and cement production [1]. Rapid and accurate detection of coal quality can increase the efficiency of coal utilization, optimize industrial production processes, and reduce emissions of carbon dioxide and other pollutants. Although the traditional chemical analysis method can accurately detect the coal indexes, their procedures are quite cumbersome and typically require several hours or even longer, which cannot meet the demand of rapid analysis in the industrial field [2]. Therefore, it has a broad market prospect and important application value to develop the coal quality detection technology that meets the requirements of coal chemical enterprises in terms of speed, precision and accuracy.

At present, the techniques for rapid analysis of coal quality include prompt gamma neutron activation analysis (PGNAA), laser-induced breakdown spectroscopy (LIBS), near-infrared spectroscopy (NIRS) and X-ray fluorescence spectroscopy (XRF). The application of PGNAA is limited due to its radioactive hazards [3,4]. LIBS has been used for coal quality detection in industrial field, but its measurement repeatability is limited by factors such as Rayleigh-Taylor instability, pulse energy fluctuations, and susceptibility of the plasma to external interference [5–7]. NIRS can rapidly analyze organic functional groups in coal [8–10], but it cannot analyze ash-forming elements. XRF can stably analyze ash-forming elements in coal [11–13], but it is powerless to analyze the elements with atomic numbers lower than 11 (e.g., the key elements of C and H in coal). In summary, it is difficult to analyze coal quality using a single spectroscopy technique to meet the actual needs of

* Corresponding authors at: State Key Laboratory of Quantum Optics and Quantum Optics Devices, Institute of Laser Spectroscopy, Shanxi University, Taiyuan 030006, China.
E-mail addresses: k1226@sxu.edu.cn (L. Zhang), ywb65@sxu.edu.cn (W. Yin).

https://doi.org/10.1016/j.optlastec.2024.111734
Received 18 June 2024; Received in revised form 16 August 2024; Accepted 29 August 2024
Available online 31 August 2024

industrial production.

In our previous work, we have verified the feasibility of NIRS-XRF fusion spectroscopy for coal quality analysis and developed a prototype and applied it in a power plant [14]. Compared with a single spectral technique, two complementary spectral techniques can provide more comprehensive and accurate information about a sample, and better predictions can be obtained by using fusion spectroscopy [15]. In this study, we further extend this technology to coking enterprises. Coking enterprises usually use gas coal, fat coal, coking coal and lean coal to produce coking raw materials, compared with power plants, the types of coal used by coking enterprises are more complex, and different types of coal have significant differences in degree of metamorphic, physicochemical properties, which make it difficult to do an accurate analysis of coal quality of multiple coal types. In order to improve the accuracy of NIRS-XRF fusion spectroscopy for coal quality analysis in the case of complex coal type scenarios, this study intends to seek the optimal coal classification strategy and automatic classification algorithm, then establish the analysis model corresponding to each type of coal, and finally evaluate the performance of coal quality analysis.

2. Experiments

2.1. Samples

The experiment collected 225 coal samples with a particle size of 0.2 mm provided by the Coal Preparation Plant of Yangguang Coking Group in Shanxi Province within one week, and we used the 180 samples from the first five days as the calibration set and the 45 samples from the last two days as the prediction set. According to the Chinese national standard "GB/T 5751-2009 Classification of Chinese Coals," these samples can be classified into four types: gas coal, fat coal, coking coal, and lean coal. The corresponding coal types and coal quality indicators are listed in Table 1. All coal quality indicators were tested by the employees of the enterprise according to the Chinese national standard "GB/T 212-2008 Methods for Chemical Analysis of Coal".

2.2. Spectral acquisition and preprocessing

In this experiment, we used the previously developed NIRS-XRF coal quality rapid analyzer for the acquisition of NIRS spectra and XRF spectra. The NIRS-XRF coal quality rapid analyzer consisted of a NIRS module, an XRF module, a sample delivery module, and a control module. The NIRS module utilized a Fourier transform infrared spectrometer (Hamamatsu, C15511-01), which operates within the wavelength range of 1100–2500 nm, with a spectral resolution of 5.7 nm. The wavelength range of the light source (AvaLight-HAL-S Mini, Avantes) used is 360–2500 nm. The XRF module was based on energy dispersive XRF technology, which is composed of an X-ray tube (VF-50J, VARIAN), a high-voltage power supply (MNX50P50, SPELLMAN), a silicon drift detector (VIAMP, KETEK), a sealed chamber, a beryllium window, and a collimator. The control module controlled the sample delivery module to pass sample through the NIRS Module and the XRF Module in turn during the measurement. To ensure the accuracy and repeatability of the experiment, the experimental parameters of the NIRS module and the XRF module were set in detail, and the specific setting methods and parameters are described in reference [14]. The experiment was conducted at a temperature of 22–25 °C and a humidity of 40 %–50 %.

After NIRS spectra and XRF spectra were collected, we first employed the Savitzky-Golay (SG) smoothing and standard normal variation (SNV) to preprocess the two spectra separately, aiming to eliminate high-frequency noise and scattering effects, the window size for smoothing was set to 5 and the polynomial order was 3. Then we performed a low-level fusion of the two spectra by concatenating them end-to-end. Due to the large differences in the intensity of the two spectra, we scaled the fused spectrum to fall within the range of (−1, 1), and the scaling method used is given by the following equation:

$$y = \frac{(y_{max} - y_{min})^*(x - x_{min})}{x_{max} - x_{min}} + y_{min} \quad (1)$$

where $y_{\max}$ and $y_{\min}$ are the upper and lower bounds of the interval, x is the variable to be scaled, and $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the variable, respectively, y represents the variable after scaling. The fused spectrum after preprocessing contains a total of 3739 variables [14]. All preprocessing methods and the algorithms mentioned later were performed on MATLAB.

2.3. Data dimensionality reduction and visualization algorithms

Spectral data usually have a large number of features or variables, which complicates data analysis and processing. Dimensionality reduction algorithms can reduce the dimensionality of spectral data, simplify the data structure and improve the analysis efficiency. At the same time, since high-dimensional data are difficult to be represented intuitively on graphs, dimensionality reduction can map the data into two- or three-dimensional space, which is easy to visualize and observe. We plan to use dimensionality reduction algorithms to map the spectral data of coal samples into a three-dimensional space, aiming to deeply explore the differences and relationships between complex coal types and thereby determine the optimal classification strategy for coal samples. Two commonly used algorithms are as follows.

2.3.1. Principal Component analysis (PCA)

Principal Component Analysis (PCA) is a commonly used method for data dimension reduction, as well as an unsupervised algorithm that does not rely on any information about category labels [16]. The main idea of PCA is to transform the original data into a set of representations that are linearly independent in each dimension through a linear transformation. This transformation can be used to extract the main feature components of the data. In other words, PCA can project high-dimensional data onto a lower-dimensional space while preserving as much information as possible from the original data and ensuring the linear independence among the features in the new dimensions.

2.3.2. t-Distributed Stochastic Neighbor Embedding (t-SNE)

t-Distributed Stochastic Neighbor Embedding (t-SNE) is a machine learning algorithm for dimensionality reduction, first proposed by Maaten and Hinton in 2008 [17]. t-SNE starts by computing the similarity between data points in the high-dimensional space and assigns a probability distribution to each data point, representing its similarity to other data points. This probability distribution can be viewed as a representation of the "neighborhood relationships." Then, in the low-dimensional space, t-SNE finds a new position for each data point and computes the similarity between them. Finally, t-SNE employs an optimization method known as Kullback-Leibler Divergence (KL

Table 1
The statistical results of 225 samples.

Type	Ash (%)	Volatile (%)	Sulfur (%)	Sample size of calibration set	Sample size of prediction set
Gas coal	8.21–11.64	39.02–42.28	0.46–0.55	9	3
Fat coal	7.06–14.23	28.72–34.40	0.39–1.73	43	15
Coking coal	10.01–11.46	21.18–27.77	0.68–3.90	74	16
Lean coal	7.72–11.70	12.56–18.26	0.37–2.38	54	11

Divergence) to measure the difference between these two probability distributions, and uses algorithms like gradient descent to minimize this difference. In this way, t-SNE can preserve the similarity relationships between data points in the high-dimensional space as much as possible in the low-dimensional space, which improves the visualization result.

2.4. Classification algorithms

2.4.1. Back propagation neural network (BPNN)

The backpropagation neural network (BPNN) is one of the most widely used neural network models today [18,19]. The BPNN is composed of an input layer, hidden layers, and an output layer, with neurons in each layer only connected to neurons in adjacent layers. The neurons in the input layer are responsible for receiving external information and passing it to the hidden layer. The hidden layer transforms and processes the received information before passing it to the output layer. The output layer compares the neural network's output with the desired output from the external environment and calculates the error. When the error between the actual output and the expected output exceeds a preset threshold, the backpropagation process begins. The error propagates back through the layers using the gradient descent method, and the weights of each layer are adjusted to minimize the output error. Through continuous forward propagation of information and backward propagation of errors, the weights of each layer are constantly adjusted, which constitutes the learning and training process of the neural network.

2.4.2. Support Vector Machine (SVM)

Support Vector Machine (SVM) is a supervised learning model widely used in the fields of classification, regression analysis and anomaly detection [20,21]. Its basic idea is to binary classify data by finding an optimal hyperplane that maximally separates samples of different classes and ensures the classifier's ability to generalize to new samples. It can be extended to multi-class nonlinear classifiers by employing kernel functions and techniques such as one-to-one and one-to-coefficient.

2.5. Regression algorithms

Partial least squares regression (PLSR) provides a regression modeling method for multiple dependent variables on multiple independent variables, which is very suitable for regression modeling when the number of samples is less than the number of independent variables [22]. In this study, two different prediction models were established based on PLSR, one using all samples for regression analysis, and the other classifying coal samples first according to the optimal coal sample classification strategy before establishing a regression model. Considering that the amount of gas coal is too small to establish a representative model, gas coal is not used in both PLSR models. All PLSR models in this paper were optimized by 5-fold cross-validation.

2.6. Evaluation indexes

In this study, the confusion matrix is utilized to demonstrate the relationship between the model's prediction results and the actual labels. The classification performance of the model is evaluated using accuracy. The formula for calculating accuracy is as follows:

$$accuracy = \frac{\sum_{i=1}^{K} TP_i}{\sum_{i=1}^{K} (TP_i + FN_i)} \tag{2}$$

where K is the number of types, TP, FN represent True Positives, and False Negatives, and the subscript i denotes the category to which it belongs.

The evaluation indexes for the regression model in this study include the correlation coefficient of the fitted curve (R^2) of the calibration set, the root mean square error for prediction ($RMSE_P$) and the mean absolute relative error for prediction ($MARD_P$), and the corresponding formulas are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{m} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{m} (y_i - \bar{y}_i)^2} \tag{3}$$

$$RMSE_p = \sqrt{\frac{\sum_{i}^{n} (y_i - \hat{y}_i)^2}{n}} \tag{4}$$

$$MARD_P = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{y_i} \times 100\% \tag{5}$$

where y_i is the reference value of the sample, $\hat{y}_i$ is the prediction value of the sample, $\bar{y}_i$ is the average value of the reference value of all the samples, m is the number of samples in the calibration set and n is the number of samples in the prediction set. the closer R^2 is to 1, the better the linearity of the calibration set model is; the closer $RMSE_P$ and $MARD_P$ are to 0, the closer the predicted value is to the true value, reflecting the better accuracy of the prediction model.

3. Results and discussion

3.1. Evaluation of unclassified regression models

Fig. 1 presents the analysis results of regression modeling using all coal samples without classification. The R^2 values of the unclassified model for ash, volatile, and sulfur are 0.9848, 0.9938, and 0.9968, respectively, while the $RMSE_P$ values are 0.56 %, 2.00 %, and 0.08 %, and the $MARD_P$ values are 4.12 %, 6.09 %, and 8.52 %. Judging from the distribution of samples in the prediction set, the prediction accuracy of this model for coal quality is not particularly satisfactory. Combined with Table 1, it can be seen that the types of coal purchased by this enterprise are quite diverse, with a wide range of values for various indicators. Different types of coal exhibit significant differences in physical and chemical properties, and this diversity significantly reduces the prediction performance of the model. Therefore, in the face of complex and diverse coal types, it is not recommended to use an unclassified model to directly analyze coal quality.

3.2. Visualization of coal classification

Both PCA and t-SNE algorithms can be used for dimensionality reduction and visualization of high-dimensional spectral data. The 180 preprocessed calibration set sample spectral data were dimensionally reduced using PCA with the confidence interval set to 95 %, and the spectral data variables were reduced from the initial 3739 variables to 89. The first three principal components are selected for PCA spectral visualization, and then t-SNE is used to further map the PCA dimensionality reduction data into the 3-dimensional space for t-SNE spectral visualization. Fig. 2 demonstrates the visualization effect of these two algorithms, it can be seen that the clustering results of coal samples after dimensionality reduction using the t-SNE is more obvious, and its visualization effect is better than the PCA. This is because only the first three principal components of PCA cannot fully demonstrate the difference between different coal types. In subsequent studies, the t-SNE will continue to be used to explore how to classify coal samples. However, unlike PCA, t-SNE does not provide an explicit mapping function that can be directly applied to new data. PCA can map new data to a low-dimensional space through a linear transformation matrix, whereas t-SNE finds the optimal low-dimensional representation on the training dataset through a complex optimization process, which cannot be directly applied to new data [17], so the PCA dimensionality reduced data will still be used as the input variables in the classification

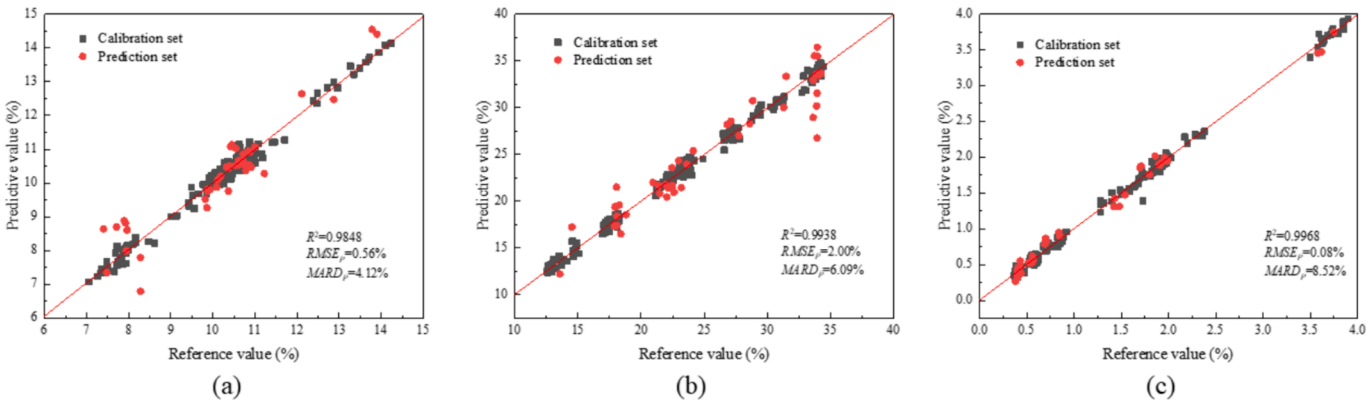

Fig. 1. Unclassified model analysis results, (a) ash, (b) volatiles, (c) sulfur.

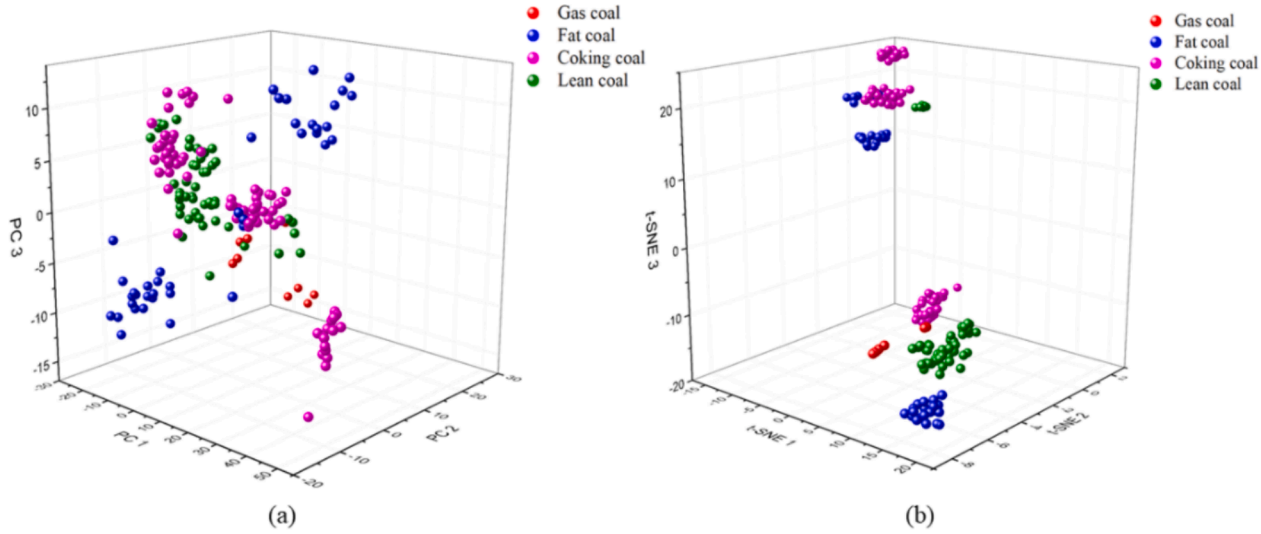

Fig. 2. Visualization of coal classification results using (a) PCA and (b) t-SNE.

modeling.

It can also be seen from Fig. 2 that the classification method according to gas coal, fat coal, coking coal, and lean coal does not distinguish the coal samples visually. Therefore, we further refine the classification according to the spectral data after visualization, and the optimized classification results are shown in Fig. 3. As can be seen in Fig. 3(a), the coal samples with different sulfur content show a good clustering result, we differentiated the samples in the calibration set according to their sulfur content, naming samples with sulfur content below 1.0 % as low-sulfur coals, those between 1.0 % and 3.0 % as medium-sulfur coals, and those greater than 3.0 % as high-sulfur coals. The medium-sulfur and the low-sulfur coals can be further classified. For the medium-sulfur coal, we can see from Fig. 2(b) that the medium-sulfur coal consists of fat coal, coking coal and a small amount of lean coal, in which the coking coal and some fat coals cluster together making it difficult to separate, and the amount of lean coal is too small to be suitable for establishing the modeling as a class alone. From Fig. 3(b), we can see that if we consider the ash of medium-sulfur coal, it can be divided into two obvious classes, one with relatively low ash (average ash of 10.33 %) called medium-sulfur low-ash coal, and the other is medium-sulfur high-ash coal (average ash of 13.27 %). In addition, even though they are both medium-sulfur coals, there are still slight differences in sulfur content between them, and the sulfur content of medium-sulfur low-ash coal (average sulfur content of 1.93 %) is slightly higher than medium-sulfur high-ash coals (average sulfur content of 1.47 %). For the low-sulfur coal, we find that the clustering results are effective according to the classification strategy of gas coal, fat coal, coking coal, and lean coal, as shown in Fig. 3(c).

In summary, different types of coal samples are well separated using t-SNE, the sulfur content has a great influence on the differentiation of different coal types, and when the sulfur content is similar, coal samples with similar ash content or volatile matter will be clustered together. We think that the best classification strategy is to divide the coal samples into seven types, which are low-sulfur gas (LS-G) coal, low-sulfur fat (LS-F) coal, low-sulfur coking (LS-C) coal, low-sulfur lean (LS-L) coal, medium-sulfur low-ash (MS-LA) coal, medium-sulfur high-ash coal (MS-HA) and high-sulfur (HS) coal. After classifying into seven types, we randomly selected a sample from each type of coal to plot its NIRS spectrum and XRF spectrum, Typical NIRS spectra and XRF energy spectra of seven types of coal are shown in Fig. 4(a) and (b), respectively, and we can see that the classification strategy we used matches well with the rules presented by the coal spectra. The statistical results for seven types of coal samples are shown in Table 2.

3.3. *Evaluation of classification models*

After determining the classification strategy for the coal samples, we compared the classification performance of BPNN and SVM. Using the data after PCA dimensionality reduction as input to the classification model. For BPNN, a three-layer neural network structure was employed, namely the input layer – hidden layer – output layer. By comparing the accuracy of the prediction set, the optimal number of hidden layer

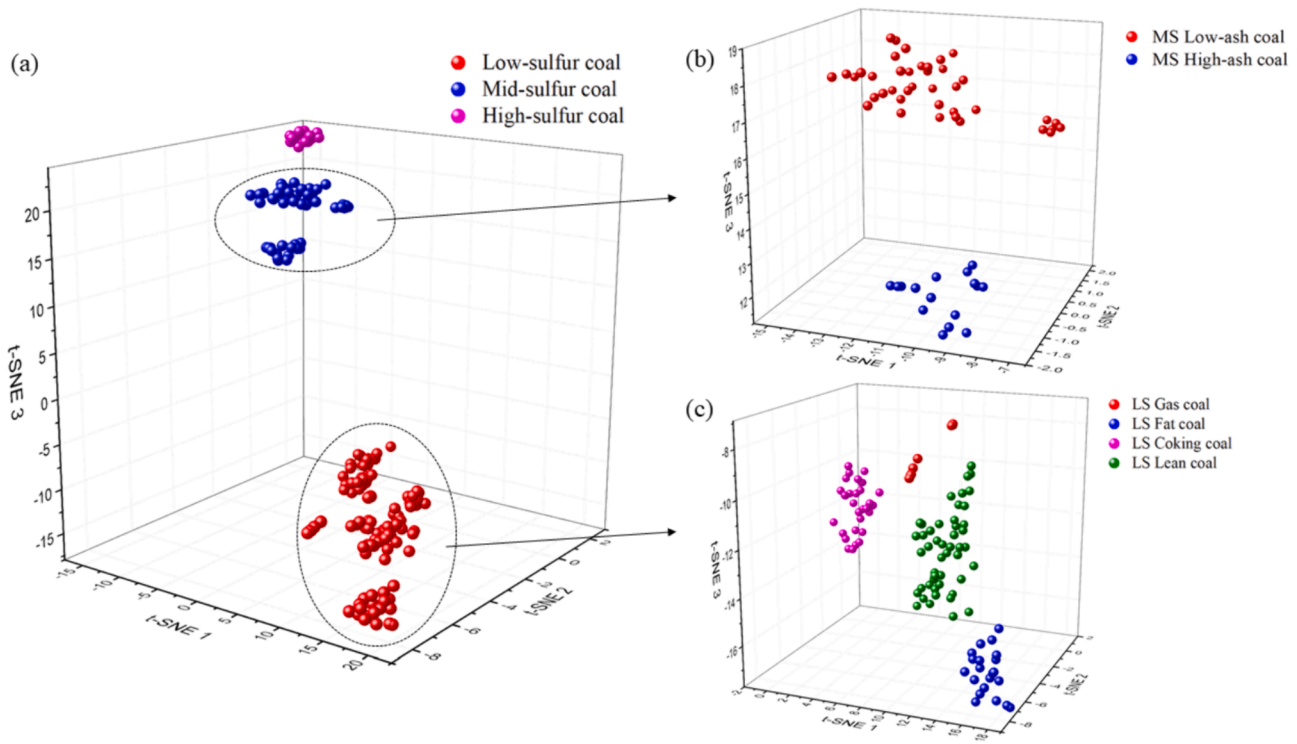

Fig. 3. Visualization of coal classification results, (a) classification by sulfur, (b) further classification of medium-sulfur coal, (c) further classification of low-sulfur coal.

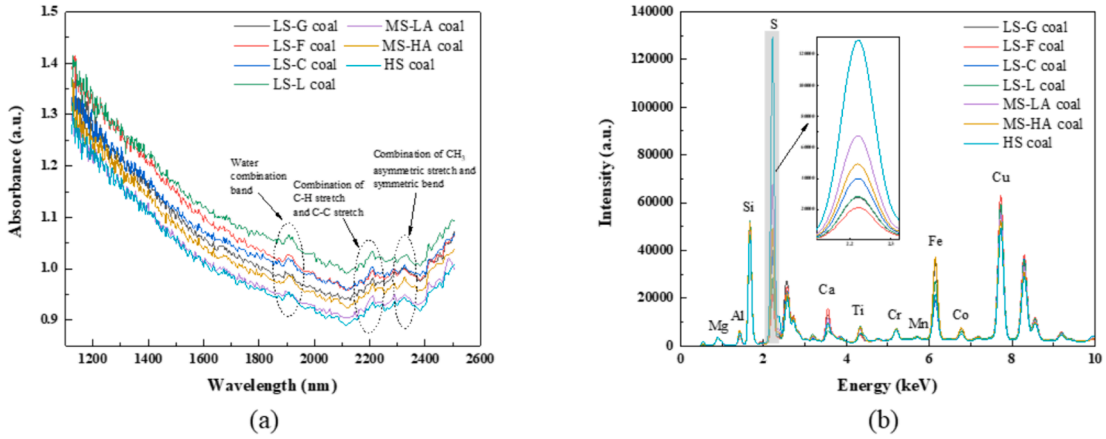

Fig. 4. Spectra of sample selected from each type, (a) NIRS, (b) XRF.

Table 2
Statistical results for seven types of coal samples.

Type	Ash (%)	Volatile (%)	Sulfur (%)	Sample size of calibration set	Sample size of prediction set
LS-G	8.21–11.64	39.02–42.28	0.46–0.56	9	3
LS-F	7.06–7.98	33.17–34.40	0.39–0.47	21	9
LS-C	10.32–11.18	22.21–23.98	0.68–0.92	31	5
LS-L	9.00–11.70	12.56–18.22	0.37–0.62	48	11
MS-LA	7.72–11.46	17.40–33.54	1.68–2.38	40	10
MS-HA	12.38–14.23	28.72–31.33	1.28–1.65	16	4
HS	10.01–10.76	26.51–27.77	3.50–3.90	15	3

neurons was determined to be 16. Further, the transfer function of the hidden layer of BPNN was tansig, the transfer function of the output layer was purelin, the learning rate was 0.1, the maximum number of training iterations was 100, and the target error was 10^{-4}. In the SVM, the radial basis function (RBF) kernel was chosen, a grid search combined with 5-fold CV was used to optimize the hyperparameters of SVM, determine the appropriate parameters by comparing the accuracy of the prediction set, and the penalty coefficient C and the width parameter γ were determined to be 0.5743 and 0.0039, respectively.

Fig. 5 and Fig. 6 show the confusion matrices of the training and test

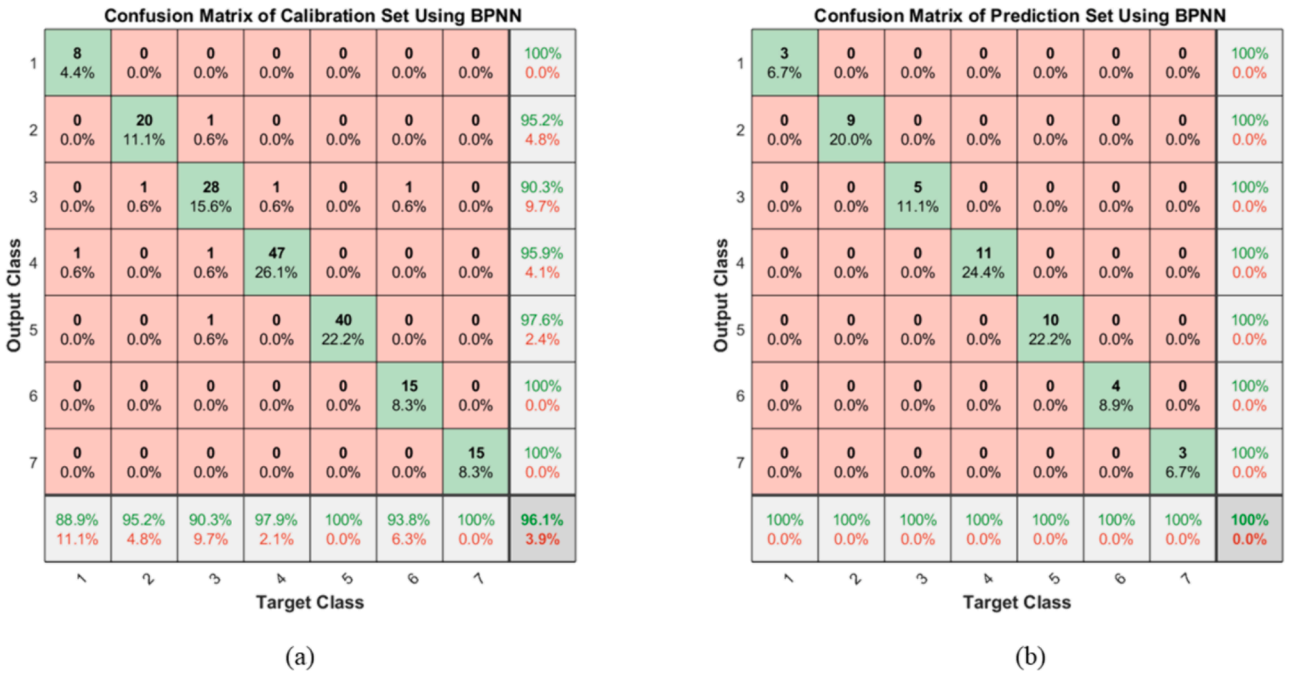

Fig. 5. Confusion matrix for classification by BPNN, (a) calibration set, (b) prediction set.

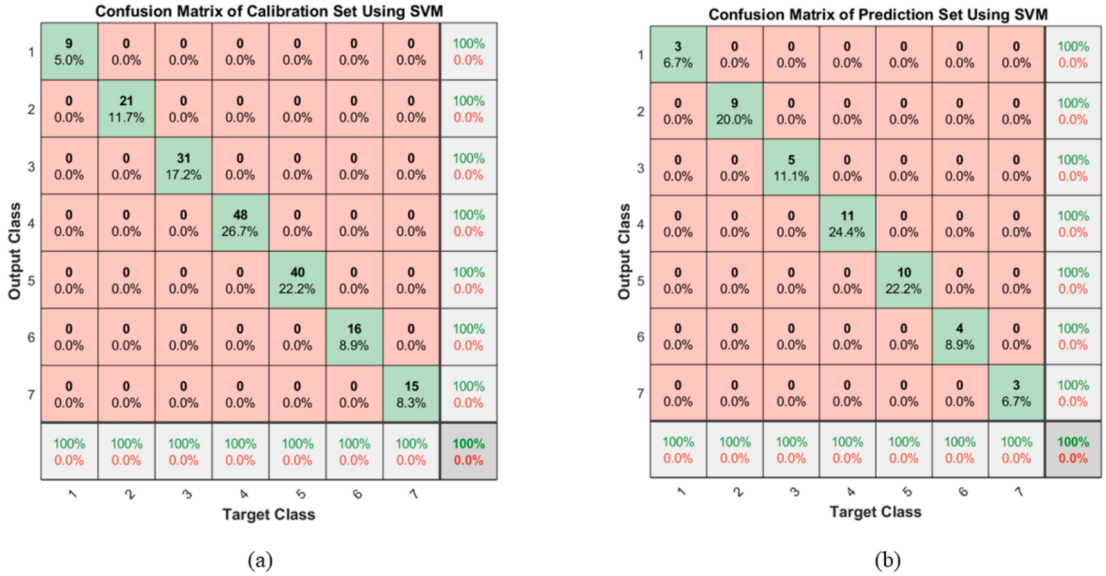

Fig. 6. Confusion matrix for classification by SVM, (a) calibration set, (b) prediction set.

sets using the two algorithms for classification, where 1 to 7 represent low-sulfur gas coal, low-sulfur fat coal, low-sulfur coking coal, low-sulfur lean coal, medium-sulfur low-ash coal, medium-sulfur high-ash coal, and high-sulfur coal, respectively. It can be seen that the classification accuracy of these two algorithms for the training set is 96.1 % and 100 % respectively, while the classification accuracy for the test set is 100 % for both algorithms. The higher classification accuracies demonstrated by these two algorithms fully prove the rationality of the present coal sample classification strategy. Comprehensively, SVM shows better classification performance compared to BPNN.

3.4. Evaluation of regression models after classification

After dividing all the samples into seven classes and determining the appropriate classification algorithm, we establish regression model for each class individually. In order to have a direct comparison with subsection 3.1, instead of showing the performance of each of the seven models separately, we organize the results of the seven models together here. Fig. 7 demonstrates the coal quality analysis results of the regression model after classification for the samples of the calibration set and the prediction set. It can be seen that the R^2 of the model for ash, volatile matter, and sulfur are 0.9987, 0.9955, and 0.9997, respectively, the $RMSE_P$ is 0.31 %, 1.34 %, and 0.05 %, respectively, and the $MARD_P$ is 2.48 %, 3.58 %, and 3.57 %, respectively. The prediction accuracy of the model after classification is significantly improved compared to the unclassified model.

4. Conclusion

In this study, we utilized the NIRS-XRF fusion spectroscopy analysis

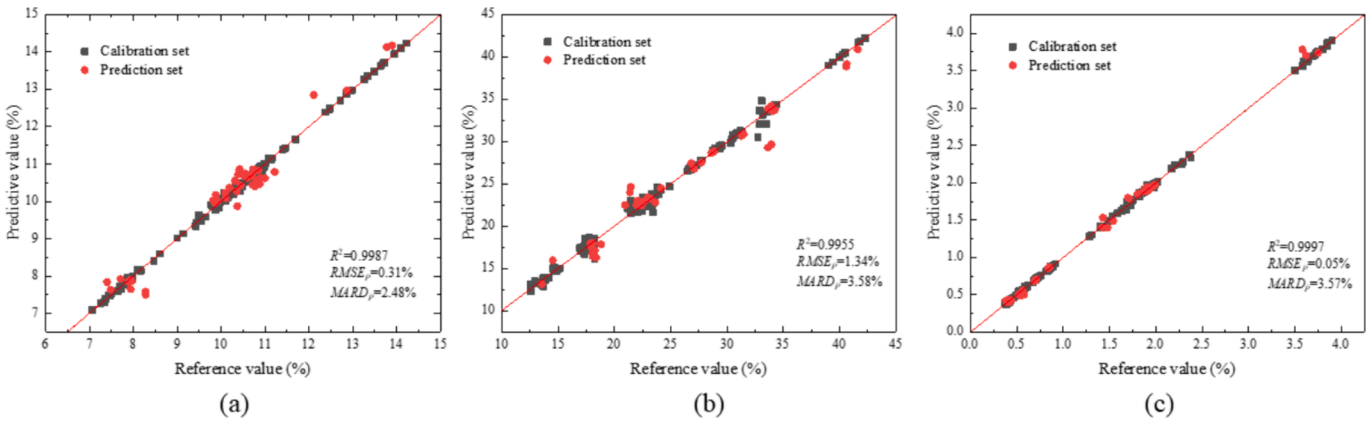

Fig. 7. Results of regression model after classification, (a) ash, (b) volatiles, (c) sulfur.

technique, combined with PCA and t-SNE algorithms through data visualization to carry out an investigation on the classification strategy, which provides a suitable classification basis for the classification of coal samples. In order to realize the automatic identification of coal types, we compared the classification performance of SVM and BPNN classification algorithms, and the results showed that the classification performance of SVM was better than that of BPNN. Meanwhile, the prediction performance of the unclassified model and the regression model after classification were compared. The R^2 values for ash, volatile matter, and sulfur in the unclassified model were 0.9848, 0.9938, and 0.9968, respectively, with $RMSE_P$ values of 0.56 %, 2.06 %, and 0.08 %, and $MARD_P$ values of 4.12 %, 6.09 %, and 8.52 %. In contrast, the R^2 values for the classified model were 0.9987, 0.9955, and 0.9997, with $RMSE_P$ values of 0.31 %, 1.34 %, and 0.05 %, and $MARD_P$ values of 2.48 %, 3.58 %, and 3.57 %, respectively. The proposed method significantly enhances prediction accuracy post-classification compared to the unclassified model. This approach effectively improves coal quality analysis accuracy in complex coal type scenarios, which is vital for enhancing production efficiency and achieving efficient coal resource utilization in coal chemical and related industries.

CRediT authorship contribution statement

Jiaxuan Li: Writing – original draft, Methodology, Data curation. **Rui Gao:** Software, Investigation. **Yan Zhang:** Resources, Formal analysis. **Lei Zhang:** Visualization, Supervision. **Lei Dong:** Visualization, Supervision. **Weiguang Ma:** Resources. **Wangbao Yin:** Writing – original draft, Project administration. **Suotang Jia:** Supervision, Resources.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

National Natural Science Foundation of China (NSFC) (12374377, 61975103, 627010407); Changjiang Scholars and Innovative Research Team in University of Ministry of Education of China (IRT_17R70); National Energy R&D Center of Petroleum Refining Technology (RIPP, SINOPEC); 111 Project (D18001); Fund for Shanxi "1331KSC".

References

[1] K. Liu, C. He, C. Zhu, J. Chen, K. Zhan, X. Li, A review of laser-induced breakdown spectroscopy for coal analysis, TrAC Trends Anal. Chem. 143 (2021) 116357.
[2] S. Sheta, M.S. Afgan, Z. Hou, S.-C. Yao, L. Zhang, Z. Li, Z. Wang, Coal analysis by laser-induced breakdown spectroscopy: a tutorial review, J. Anal. At. Spectrom 34 (2019) 1047–1082.
[3] W.-B. Jia, D.-Q. Hei, A.-G. Xu, X.-W. Chen, A.-M. Li, Influence of sample weight in coal composition online analysis by PGNAA, Atomic Energy Science and Technology 45 (2011) 1011.
[4] J.-B. Yang, Z. Liu, K. Chang, R. Li, Research on the self-absorption corrections for PGNAA of large samples, Eur. Phys. J. Plus 132 (2017) 86.
[5] S. Yao, J. Mo, J. Zhao, Y. Li, X. Zhang, W. Lu, Z. Lu, Development of a rapid coal analyzer using laser-induced breakdown spectroscopy (LIBS), Appl. Spectrosc. 72 (2018) 1225–1233.
[6] W. Li, M. Dong, S. Lu, S. Li, L. Wei, J. Huang, J. Lu, Improved measurement of the calorific value of pulverized coal particle flow by laser-induced breakdown spectroscopy (LIBS), Anal. Methods 11 (2019) 4471–4480.
[7] Y.-T. Fu, W.-L. Gu, Z.-Y. Hou, S.A. Muhammed, T.-Q. Li, Y. Wang, Z. Wang, Mechanism of signal uncertainty generation for laser-induced breakdown spectroscopy, Front. Phys. 16 (2020) 22502.
[8] T.A. Lestander, R. Samuelsson, Prediction of resin and fatty acid content of biorefinery feedstock by on-line near-infrared (NIR) Spectroscopy, Energy Fuel 24 (2010) 5148–5152.
[9] W. Liu, B. Peng, X. Liu, F. Ren, L. Zhang, Intelligent proximate analysis of coal based on near-infrared spectroscopy, J. Appl. Spectrosc. 88 (2021) 645–652.
[10] L. Zou, J. Qiao, X. Yu, X. Chen, M. Lei, Intelligent proximate analysis of coal based on near infrared spectroscopy and multi-output deep learning, IEEE Trans. Artif. Intell. (2023).
[11] C.R. Ward, S.J. Kelloway, J. Vohra, D. French, D.R. Cohen, C.E. Marjo, I. E. Wainwright, In-situ inorganic analysis of coal seams using a hand-held field-portable XRF Analyser, Int. J. Coal Geol. 191 (2018) 172–188.
[12] Z. Yan, Z. XinLei, J. WenBao, S. Qing, L. YongSheng, H. DaQian, C. Da, Online X-ray fluorescence (XRF) analysis of heavy metals in pulverized coal on a conveyor belt, Appl. Spectrosc. 70 (2016) 272–278.
[13] X. Tian, L. Zhao, Determination of concentrations of Sr and Ba in coal and coal combustion by-products: a comparison between results by ICP-MS and XRF techniques, Talanta 266 (2024) 124919.
[14] R. Gao, S. Wang, J. Li, Z. Tian, Y. Zhang, L. Zhang, Z. Ye, Z. Zhu, W. Yin, S. Jia, Development and application of a rapid coal calorific value analyzer based on NIRS-XRF, J. Anal. At. Spectrom 38 (2023) 2046–2058.
[15] E. Borràs, J. Ferré, R. Boqué, M. Mestres, L. Aceña, O. Busto, Data fusion methodologies for food and beverage authentication and quality assessment – A review, Anal. Chim. Acta 891 (2015) 1–14.
[16] I.T. Jolliffe, J. Cadima, Principal component analysis: a review and recent developments, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 374 (2016) 20150202.
[17] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (2008).
[18] A. Hussain, H. Pu, D.-W. Sun, Measurements of lycopene contents in fruit: a review of recent developments in conventional and novel techniques, Crit. Rev. Food Sci. Nutr. 59 (2019) 758–769.
[19] A. Anani, S.O. Adewuyi, N. Risso, W. Nyaaba, Advancements in machine learning techniques for coal and gas outburst prediction in underground mines, Int. J. Coal Geol. 285 (2024) 104471.

[20] X. Yu, W. Guo, N. Wu, L. Zou, M. Lei, Rapid discrimination of coal geographical origin via near-infrared spectroscopy combined with machine learning algorithms, Infrared Phys. Technol. 105 (2020) 103180.
[21] Y. Wang, M. Yang, G. Wei, R. Hu, Z. Luo, G. Li, Improved PLS regression based on SVM classification for rapid analysis of coal properties by near-infrared reflectance spectroscopy, Sens. Actuators B 193 (2014) 723–729.
[22] J. Feng, Z. Wang, L. West, Z. Li, W. Ni, A PLS model based on dominant factor for coal analysis using laser-induced breakdown spectroscopy, Anal. Bioanal. Chem. 400 (2011) 3261–3271.