ACS OMEGA

Open Access

This article is licensed under CC-BY-NC-ND 4.0

http://pubs.acs.org/journal/acsodf

Article

Coal Classification Based on Reflection Spectroscopy and the IAT-TELM Algorithm

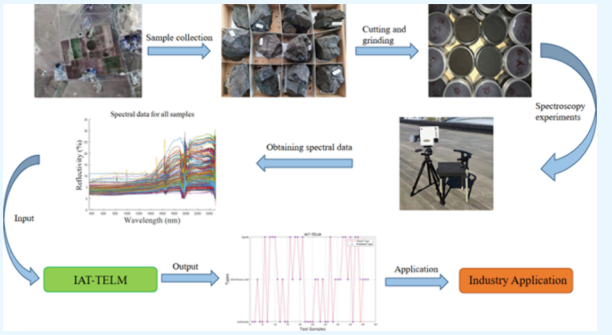
Boyan Li, Dong Xiao,* Hongfei Xie, Jie Huang, and Zelin Yan

Cite This: *ACS Omega* 2023, 8, 35232−35241

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: As a principal energy globally, coal's quality and variety critically influence the effectiveness of industrial processes. Different coal types cater to specific industrial requirements due to their unique attributes. Traditional methods for coal classification, typically relying on manual examination and chemical assays, lack efficiency and fail to offer consistent accuracy. Addressing these challenges, this work introduces an algorithm based on the reflectance spectrum of coal and machine learning. This method approach facilitates the rapid and accurate classification of coal types through the analysis of coal spectral data. First, the reflection spectra of three types of coal, namely, bituminous coal, anthracite, and lignite, were collected and preprocessed. Second, a model utilizing two hidden layer extreme learning machine (TELM) and affine transformation function is introduced, which is called affine transformation function TELM (AT-TELM). AT-TELM introduces an affine transformation function on the basis of TELM, so that the hidden layer output satisfies the maximum entropy principle and improves the recognition performance of the model. Third, we improve AT-TELM by optimizing the weight matrix and bias of AT-TELM to address the issue of highly skewed distribution caused by randomly assigned weights and biases. The method is named the improved affine transformation function (IAT-TELM). The experimental findings demonstrate that IAT-TELM achieves a remarkable coal classification accuracy of 97.8%, offering a cost-effective, rapid, and precise method for coal classification.



1. INTRODUCTION

Coal, a prominent constituent of the world's major energy resources, holds an irreplaceable position in the international industry and economy. Representing about half of China's energy consumption, it serves as a critical energy pillar. Coal's unique properties such as calorific value, combustion attributes, and emission characteristics deeply influence thermal power plant operations, fuel selection, and environmental conservation efforts. Furthermore, coal stands as a vital raw material in heavy industries like steel, chemical, and building materials, where the coal type directly influences product quality, production efficiency, and environmental footprint.

Coal classification, centered around its composition, characteristics, and origin, incorporates anthracite, bituminous coal, and lignite.[1] Each exhibits distinct chemical properties, leading to differing industrial applications. Lignite, though high in moisture and impurities, generates a low calorific value when burned and is primarily used as power generation fuel. Bituminous coal, boasting a high calorific value, serves as an energy source in heavy industries and also in coke production. Anthracite, primarily utilized for heating and high-quality coke production, possesses unique properties.

Existing classification methodologies encompass manual experience methods, the weighing method, and the chemical analysis method, each carrying their disadvantages, such as time intensity, high costs, and sample damage. Hence, the need for a more efficient coal classification method is self-evident.[2]

Spectral analysis technology, widely employed and developed across multiple fields, offers a noninvasive means to analyze coal properties.[3−8] Investigating the interactions between coal and various light wavelengths allows us to deduce its physical and chemical properties, enabling a more precise classification. Additionally, spectral information can facilitate the regression analysis of coal composition. For instance, Liu et al. integrated laser-induced breakdown spectroscopy (LIBS) with chemical analysis to decipher coal types and combustion efficiency, indicating a promising future for LIBS in coal analysis.[9] Similarly, Yao et al. assessed the impact of coal volatiles on plasma and spectral signals via LIBS, facilitating a comparative study of coal and char's primary components.[10] Le et al.

Received: July 12, 2023
Accepted: August 25, 2023
Published: September 16, 2023

© 2023 The Authors. Published by American Chemical Society

obtained the spectral characteristics of coal through the spectrometer and then used the extreme learning machine (ELM) to classify it. By integrating remote sensing principles, this approach allows us to capture a comprehensive distribution image of the coal mines.[11] Xie et al. explored the Raman spectra of various coals, analyzing their spectral characteristics, and employed high-temperature Raman spectroscopy to study the transformation of coal structure throughout the pyrolysis process.[12]

The machine learning landscape has been witnessing significant transformation in recent years, particularly in the context of its ability to manage and scrutinize voluminous and complex data arrays. Therefore, many coal classification methods combined with machine learning have been proposed. Zhang et al. measured the coal's ash content samples by LIBS and then used the support vector machine (SVM) to identify coal. Partial least-squares (PLS) regression is then used to build an accurate model to improve the precision of ash and volatile matter determination in coal.[13] Coal gangue, primarily discarded waste in coal mining regions, has the potential for utilization to mitigate environmental contamination while simultaneously augmenting the economic profitability of coal mines. Song et al. analyzed the thermal infrared spectrum of coal gangue using deep learning methods.[14] Jayaganthan et al. utilized epoxy micronano composite specimens to coat four distinct coal variants. By applying LIBS, they employed the logistic regression method (LR) for classification of these samples, resulting in significantly higher accuracy.[15]

Spectral data, characterized by high dimensionality, strong correlations, and noise interference, necessitate careful preprocessing. At present, principal component analysis (PCA) is a commonly used method for data dimension reduction, but improper usage can impair model accuracy. Deep learning has found extensive applications in diverse domains.[16−18] Xiao et al. used the local receptive field (LRF) to extract key features from spectral data and combines them with ELM.[19] Le et al. converted one-dimensional spectral data into two-dimensional format and extracted the characteristics of data through a convolutional neural network (CNN).[20]

Among many algorithms, the coal classification method combined with the ELM has achieved good results. ELM, introduced by Huang et al., is a single hidden layer feedforward neural network (SLFN).[21] Since ELM demonstrates excellent performance in both classification and regression tasks and has good generalization ability, many scholars have applied and improved ELM.[22−24] The ELM algorithm performs well in processing the spectral data. Mao et al. used multilayer ELM to classify coal by the spectrum of coal and obtained an accurate and fast classification model.[25] Le et al. classified the collected sandstone, shale, coal gangue, and other coal samples according to the content of fixed carbon and proposed an incremental multilayer learning machine algorithm combined with remote sensing images to identify coal quality, which has achieved good results.[26]

ELM is an effective algorithm for coal classification tasks. However, since ELM's weights and biases are randomly assigned, it will lead to extreme conditions in the output distribution of the hidden layer, resulting in feature loss. In this paper, we combined the affine transformation function and two-hidden-layer ELM, which is named affine transformation function TELM (AT-TELM), to modify the output of the initial hidden layer satisfying the maximum entropy principle to improve this phenomenon by estimating the affine parameters s and t. Furthermore, an innovative method for optimizing the weight and bias by the affine parameter of the first hidden layer is proposed and named improved affine transformation function TELM (IAT-TELM). The preprocessed data set, when predicted through the model, yields high classification precision.

2. THEORY AND METHODS

2.1. Collection of Spectral Data. The research samples for this experiment consist of three types of coal collected from various mining areas, namely, lignite, anthracite, and bituminous coal, with quantities outlined in Table 1. We label anthracite as 1,

Table 1. Number of the Coal Samples

types	label	number
anthracite	1	71
bituminous coal	2	80
lignite	3	58

bituminous coal as 2, and lignite as 3. We employed the SVC HR-1024 spectrometer for capturing spectral data of the coal within a range of 350−2500 nm. Prior to spectral measurement, the coal samples were washed, cut, and ground to powder. Outdoor experiments were conducted under clear, sunlit skies with a scan time of 1 s, a probe distance of 480 mm from the sample, and minimal interference from the experimenters. The resulting raw spectral curves, as depicted in the Figure 1, have high dimensions and require preprocessing before experimentation.

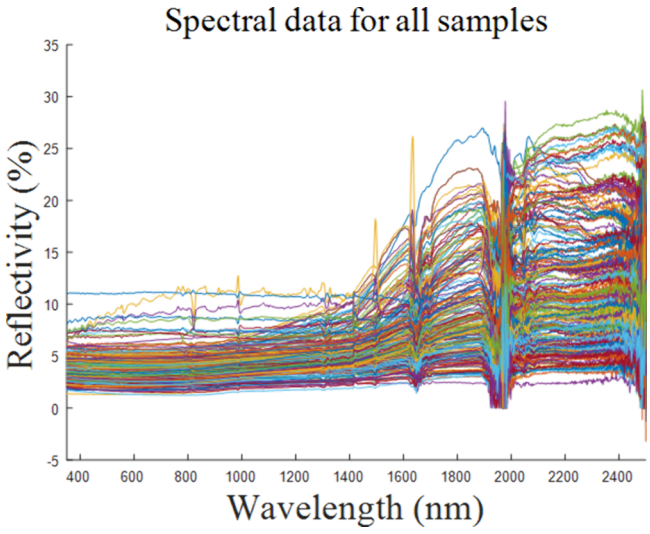


Figure 1. Raw spectral data of coal samples.

2.2. Mahalanobis Distance. The Mahalanobis distance (MD), indicating the covariance separation of data, provides a metric for data resemblance. This measure accounts for the interrelation among attributes and remains unaffected by dimensionality, such that the distance between any two points is not tied to the original data's unit of measurement. Additionally, it effectively mitigates the disturbance caused by variables' correlation.

Let a data set contain m samples, and the mean value of the data set is $u_i = (u_1, u_2, ..., u_n)^T$, where any sample is $x_i = (x_1, x_2, ..., x_n)^T$; the MD of the sample is

https://doi.org/10.1021/acsomega.3c04999
ACS Omega 2023, 8, 35232−35241

$$\mathrm{MD} = \sqrt{(x_i - u_i)^T M^{-1} (x_i - u_i)} \quad (1)$$

where M is the covariance matrix and n is the dimension of data.

In the data set chosen for this study, outlier samples have a larger MD. The steps of using MD to determine sample outliers include:

- Spectral dimensionality reduction using PCA
- Calculation of the average spectrum
- Centralization
- Calculation of MD
- Selection of outliers by threshold.

The MD of the data set is shown in Figure 2.

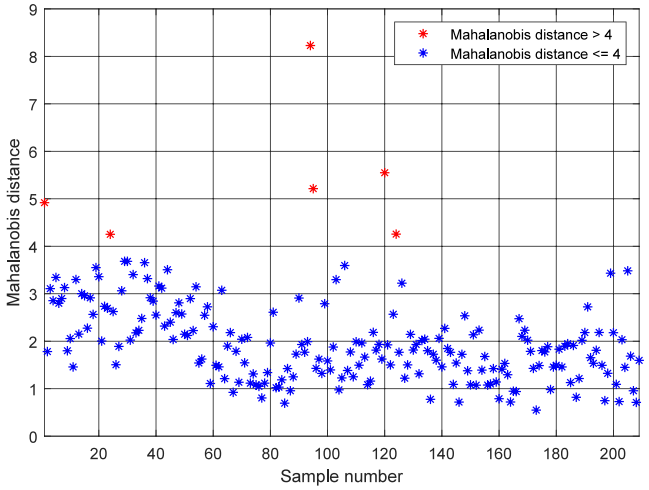


Figure 2. Mahalanobis distance of reflectance spectra of the coal samples in this paper.

According to the MD of the data set, the threshold is set to 4 to effectively eliminate outliers in order to improve the prediction accuracy of the model and the removed outlier samples are shown in Table 2.

2.3. PCA. The obtained spectral data dimension is 973, and the data correlation is strong. When the raw data are used directly as input, the model structure will be complicated and the training time will be greatly prolonged. Therefore, preprocessing the data to reduce the data dimensionality while preserving its essential information as much as possible is the premise of the experiment. In our study, PCA is employed for data set dimensionality reduction.

The PCA procedure is as follows:

1. Centralize the $m \times n$ data set matrix X:

$$x_i' = x_i - \frac{1}{m}\sum_{j=1}^{m} x_j \quad (2)$$

where m signifies the dimensionality of each sample while n indicates the total count of samples.

2. Calculate the covariance matrix of matrix X:

$$\mathrm{cov}(X) = \frac{1}{m-1} XX^T \quad (3)$$

3. Determine the eigenvalues λ_i and eigenvectors β_i of the covariance matrix.
4. Arrange the eigenvalues in descending order to get $\lambda = (\lambda_1', \lambda_2', \ldots, \lambda_n')$, at the same time, the eigenvectors is arranged in the order of λ_i to get $\beta_i = (\beta_1', \beta_2', \ldots, \beta_n')$.
5. Compute each principal component's individual contribution rate and the cumulative contribution rate of numerous principal components.

$$\theta_i = \frac{\lambda_i'}{\sum_{j=1}^{m} \lambda_i'} \quad (4)$$

$$\Theta = \sum_{i=1}^{d} \theta_i \quad (5)$$

where θ_i is the principal component contribution rate and Θ is the cumulative principal component contribution rate.

6. According to the set threshold, the first d feature vectors are taken to reduce the data from n dimension to d dimension.

$$W = (\beta_1, \beta_2, \ldots, \beta_d) \quad (6)$$

$$X_{\mathrm{new}} = XW \quad (7)$$

where X_{new} is the data matrix after dimension reduction.

The cumulative contribution rate is set to 99%, so that it can not only make the information loss little but also reduce the dimension of the data. As shown in Figure 3, upon considering seven principal components, the cumulative contribution rate initially attains 99%. Therefore, selecting seven-dimensional data as the input of the model can greatly reduce the amount of calculation and enhance the model's precision and generalizability.

2.4. ELM. ELM is a single-hidden-layer feedforward neural network. Other widely used SLFNs usually use the gradient descent method. Disadvantages of the gradient descent method are that (1) multiple iterations are required, resulting in long calculation time and slow model training, and (2) due to the existence of the learning rate, the algorithm is easily affected by the learning rate and is prone to falling into the local optimal solution. ELM determines the parameters by using the least-squares solution. This sidesteps the issues of sluggish training pace, learning rate sensitivity, and propensity to get trapped in local optimum solutions typically associated with gradient descent methods.

ELM comprises the input layer, hidden layer, and output layer. In the input layer, the input matrix $X_{m\times q}$ is represented by

$$X_{m\times q} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1q} \\ x_{21} & x_{22} & \cdots & x_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mq} \end{bmatrix} \quad (8)$$

Table 2. Removed Outlier Samples

serial number of removed outlier sample serial number	type of removed outlier sample serial number
1,24,94,95,120,124	lignite,lignite,anthracite,anthracite,anthracite,anthracite

https://doi.org/10.1021/acsomega.3c04999
ACS Omega 2023, 8, 35232−35241

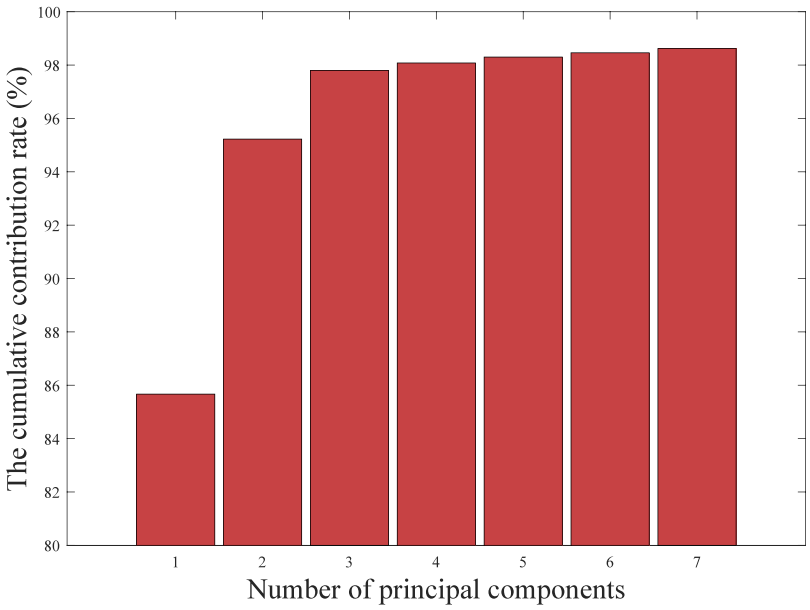


Figure 3. Cumulative contribution rate of different numbers of principal components

where m represents the sample dimension and q denotes the sample count. The label matrix is $Y_{1q} = [y_{11} \quad y_{12} \quad \cdots \quad y_{1q}]^T$. X and Y form a set of samples.

Then, the hidden layer of the ELM is introduced. There is a weight matrix and a bias between the first two layers of ELM. The interval of the weight is $(-1,1)$, which is randomly given at the beginning of the algorithm and obeys a uniform distribution, while the offset interval is $(0,1)$, which is also randomly given and obeys a uniform distribution. The weight matrix W and bias B are shown as follows:

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ w_{21} & w_{22} & \cdots & w_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ w_{l1} & w_{l2} & \cdots & w_{lm} \end{bmatrix} \tag{9}$$

$$B = [b_1 \quad b_2 \quad b_3 \quad \cdots \quad b_m]^T \tag{10}$$

where l denotes the count of hidden layer's nodes.

Under the selected activation function $f(x)$, the output matrix H of the hidden layer can be expressed as

$$H = f(WX + B) \tag{11}$$

An output weight matrix β is present between the hidden layer and the output layer. The output matrix $T = [t_1 \quad t_2 \quad \cdots \quad t_q]$ can be obtained by output weight and hidden layer output.

$$T' = H\beta \tag{12}$$

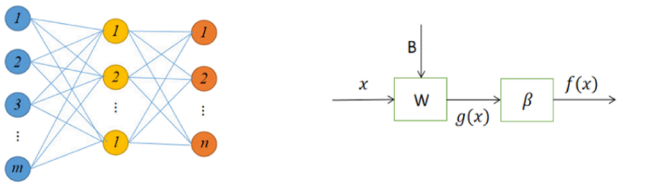
To enhance model accuracy, the output matrix should be closer to the true value, since both W and B are randomly given at the beginning of the algorithm; the loss function is

$$E = \min_{\beta} \left\| H\beta - Y \right\| \tag{13}$$

Then, we obtain the least-squares solution of β by using the Moore−Penrose generalized inverse method:

$$\beta = H^{+}Y' \tag{14}$$

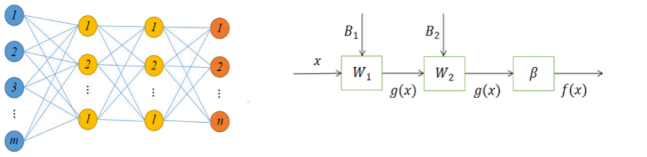
where H^{+} is the Moore−Penrose generalized inverse matrix of H. The structure of ELM is shown in Figure 4a, and the workflow of ELM is shown in Figure 4b.



(a) The structure of ELM (b) The workflow of ELM

Figure 4. Structure and workflow of ELM.

2.5. Two-Hidden-Layer ELM. Building on the structure of ELM, the TELM algorithm introduces an additional hidden layer.[27] The structure of TELM is shown in Figure 5a, and the



(a) The structure of TELM (b) The workflow of TELM

Figure 5. Structure and workflow of TELM.

workflow of TELM is shown in Figure 5b. Typically, both hidden layers have an equal number of nodes. TELM optimizes the algorithm by calculating the weight matrix and bias between the first hidden layer and the second hidden layer. Let H be the output of the first hidden layer. Regarding two hidden layers as one, the output weight matrix can be expressed as

$$\beta = H^{+}Y \tag{15}$$

The expected output of the second hidden layer H_1 can be expressed as

https://doi.org/10.1021/acsomega.3c04999
ACS Omega 2023, 8, 35232−35241

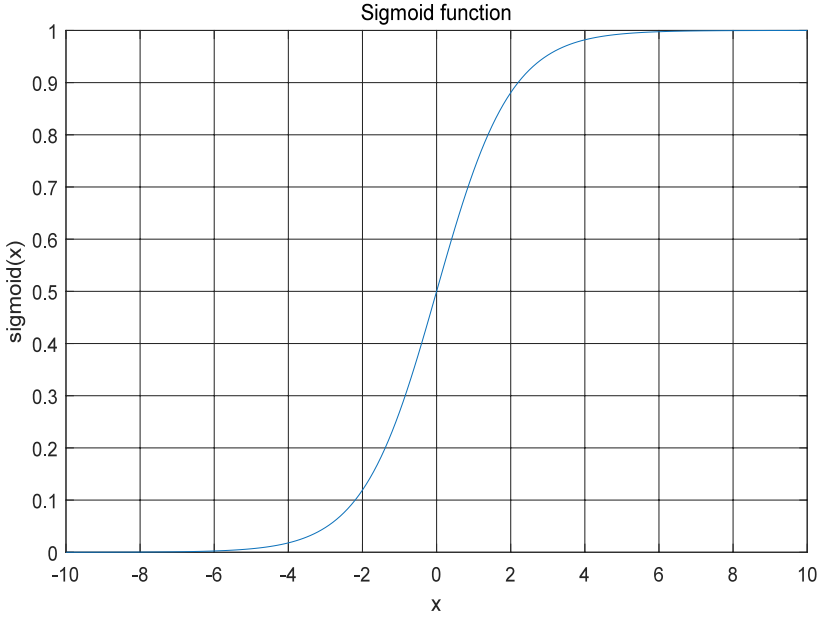


Figure 6. Output of the sigmoid function.

$$H_1 = Y\beta^{+} \tag{16}$$

where W_1 and B_1 are the weight and bias matrices of the second hidden layer, respectively. In order to calculate the weight and bias of the second hidden layer, we use $W_E = [B_1 \quad W_1]$ and $H_E = [\mathbf{1} \quad H]^T$ to express the weights and bias of the second hidden layer as

$$W_E = g^{-1}(H_1){H_E}^{+} \tag{17}$$

where $\mathbf{1}$ is a q-dimensional column vector. Then, the output of the second hidden layer H_2 can be calculated as

$$H_2 = g(W_E H_E) \tag{18}$$

Therefore, the weight matrix β_{new} is finally determined by H_2 and T as

$$\beta_{\text{new}} = {H_2}^{+}Y \tag{19}$$

2.6. Maximum Entropy Principle. The concept of entropy first appeared in physics to describe the degree of disorder of the system. In information theory, entropy is used to describe the unpredictability of information content.[28] The maximum entropy principle is a criterion for model learning, advocating that the optimal model among all potential models is the one with the greatest entropy. If a random variable X has k states and its probability distribution is $P(X = x_i) = P_i$, then the entropy of the random variable is

$$H = \sum_i P_i \log_2 \frac{1}{P_i} \tag{20}$$

In order to maximize the entropy, the issue can be solved utilizing the method of Lagrange multipliers:

$$\mathcal{L} = \sum_{i=1}^{k} P_i \log_2 \frac{1}{P_i} + \mu\left(\sum_{i=1}^{k} P_i - 1\right) \tag{21}$$

Cao et al. pointed out that in the hidden layer output of ELM, the distribution with maximum entropy is uniformly distributed.[29] Let the hidden layer input matrix $V = WX + B$, so the expectation of V can be calculated as

$$E(v_i) = l \times E(w_i)E(x_i) + E(b_i) \tag{22}$$

Let W and B satisfy the uniform distribution of $(-1,1)$. Then, the variance of V can be calculated as

$$E(v_i) = l \times \mathrm{var}(w_i)E({x_i}^2) + \mathrm{var}(b_i) \tag{23}$$

When the hidden layer output has the maximum entropy, the output of each node conforms to a uniform distribution. However, in the first hidden layer of TELM, since both W and B are randomly given and obey the uniform distribution on $(-1,1)$ and $(0,1)$, the output is difficult to obey the uniform distribution; that is, the output is difficult to satisfy the maximum entropy principle. In order to make the output as much as possible in line with the maximum entropy principle, the model is optimized. We will add an affine transformation function to the hidden layer input to adjust the distribution of the output after the activation function.

2.7. Improved Affine Transformation TELM. *2.7.1. Affine Transformation TELM.* In this section, the functions we use in the TELM algorithm are all sigmoid functions. The output of the sigmoid function is shown in Figure 6.

As shown in Figure 6, there is a saturation interval at both ends of the sigmoid function. To enhance the effectiveness of the model, we let the expected output range of the network be $[0.2,0.8]$, avoiding the saturation interval of the sigmoid function. Let H be the output of the first hidden layer in TELM, and the activation function is $g(x)$, then $H = g(WX + B)$. All elements in the H matrix are sorted from small to large, and the corresponding elements in H are supplanted by ordinal numbers to obtain the matrix O_1,

$$O_1 = \begin{bmatrix} O_{11} & O_{12} & \cdots & O_{1q} \\ O_{21} & O_{22} & \cdots & O_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ O_{l1} & O_{l2} & \cdots & O_{lq} \end{bmatrix} \tag{24}$$

https://doi.org/10.1021/acsomega.3c04999
ACS Omega 2023, 8, 35232−35241

where O_{ij} is the ordinal number of H_{ij} in H from small to large. Since the expected output range is [0.2,0.8], the expected output matrix subject to uniform distribution is

$$O = \frac{(0.8 - 0.2)}{l^*q - 1} * (O_1 - I) \tag{25}$$

where I is a matrix of $l \times q$ composed of 1.

To approximate the hidden layer's distribution as closely as possible to the uniform distribution, we introduce an affine transformation function in the activation function to make H become

$$H = g[s(WX + B) + T_0] \tag{26}$$

where s is the stretching coefficient and T_0 is the translation coefficient matrix whose all elements are t. Thus, s and t are determined by

$$\text{error} = \|g[s(WX + B) + T_0] - O\|_F \tag{27}$$

Next, we need to find the correct parameters s and t by the gradient descent method to minimize the error. When the parameters of all nodes are s and t, the partial derivatives of the loss function *error* for s and t are

$$\frac{\partial \text{error}}{\partial s} = \sum_{j=1}^{q} \sum_{i=1}^{l} [g(s \times V_{ij} + t) - O_{ij}] \times \frac{\partial g}{\partial s} \tag{28}$$

$$\frac{\partial \text{error}}{\partial t} = \sum_{j=1}^{q} \sum_{i=1}^{l} [g(s \times V_{ij} + t) - O_{ij}] \times \frac{\partial g}{\partial t} \tag{29}$$

where $V = WX + B$. Then, s and t are transformed as follows in each iteration:

$$s(\kappa + 1) = s(\kappa) - \eta \frac{\partial \text{error}}{\partial s(\kappa)} \tag{30}$$

$$t(\kappa + 1) = t(\kappa) - \eta \frac{\partial \text{error}}{\partial t(\kappa)} \tag{31}$$

where η is the learning rate.

Because the element count in the output matrix of the data in reality is too large, the complexity of the gradient descent method and the calculation time is increased. Therefore, the matrix O and H can be downsampled to reduce the complexity of the calculation. We arrange the elements in V,O from small to large into two vector and then sample every p element to obtain $v = [v_1' \ v_2' \ \cdots \ v_k']$and $o = [o_1' \ o_2' \ \cdots \ o_k']$to represent the vectors obtained after downsampling, where k is the number of elements obtained by downsampling $l \times q$ elements and p is the sampling interval. Then, eqs 28 and 29 can be expressed as

$$\frac{\partial \text{error}}{\partial s} = \sum_{j=1}^{l \times q} [g(s \times v_j + t) - o_j] \times \frac{\partial g}{\partial s} \tag{32}$$

$$\frac{\partial \text{error}}{\partial t} = \sum_{j=1}^{l \times q} [g(s \times v_j + t) - o_j] \times \frac{\partial g}{\partial t} \tag{33}$$

After the parameters s and t are determined, the activation function with an affine transformation function is used in the first hidden layer of the model.

2.7.2. Improved Affine Transformation TELM. Although the TELM algorithm can effectively achieve the classification task, the weights and biases of the first hidden layer are set randomly. The random distribution of weights and biases will lead to the fact that in some cases, even if the parameters s and t of the affine transformation function are effectively estimated, the gap between the output matrix and the expected output matrix conforming to the maximum entropy principle is still large. Therefore, we need to find a way to set the weight and bias so that the output matrix can better conform to the maximum entropy principle after the stretching and translation of s and t.

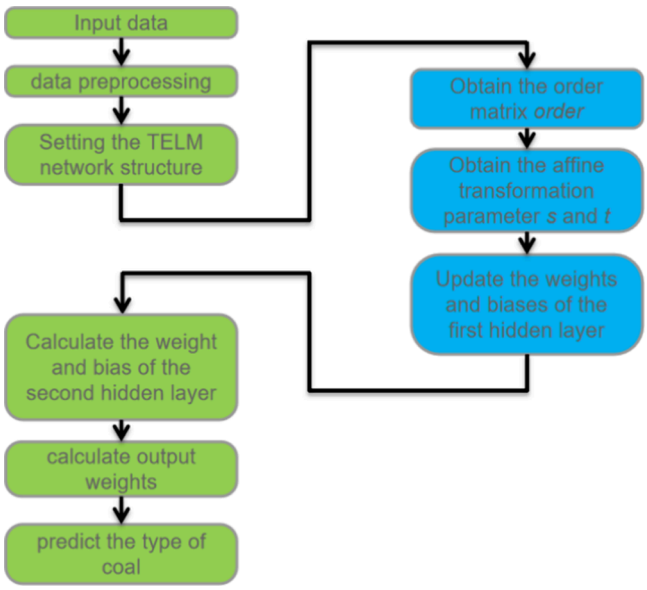


Figure 7. Flow of the IAT-TELM algorithm.

The emphasis of the IAT-TELM algorithm is to reconstruct the weight matrix and the bias matrix by estimating the relationship between the output and the expected output after estimating parameters s and t.

The output is $H = g[s(WX + B) + T_0]$, and the expected output determined by the maximum entropy principle is O. Let the optimized weights and biases be W_0 and B_0, then the input of the first hidden layer can be expressed as

$$[B_0 \ \ W_0] = \frac{g^{-1}(O) - T_0}{s} \boldsymbol{X}^+ \tag{34}$$

where $\mathbf{1}$ is a q-dimensional column vector and

$$\boldsymbol{X} = [\mathbf{1} \ \ X]^T$$

The optimized W_0 and B_0 are used as the weights and biases of the first hidden layer to classify the targets. The AT-TELM flow is shown in Figure 7.

https://doi.org/10.1021/acsomega.3c04999
ACS Omega 2023, 8, 35232−35241

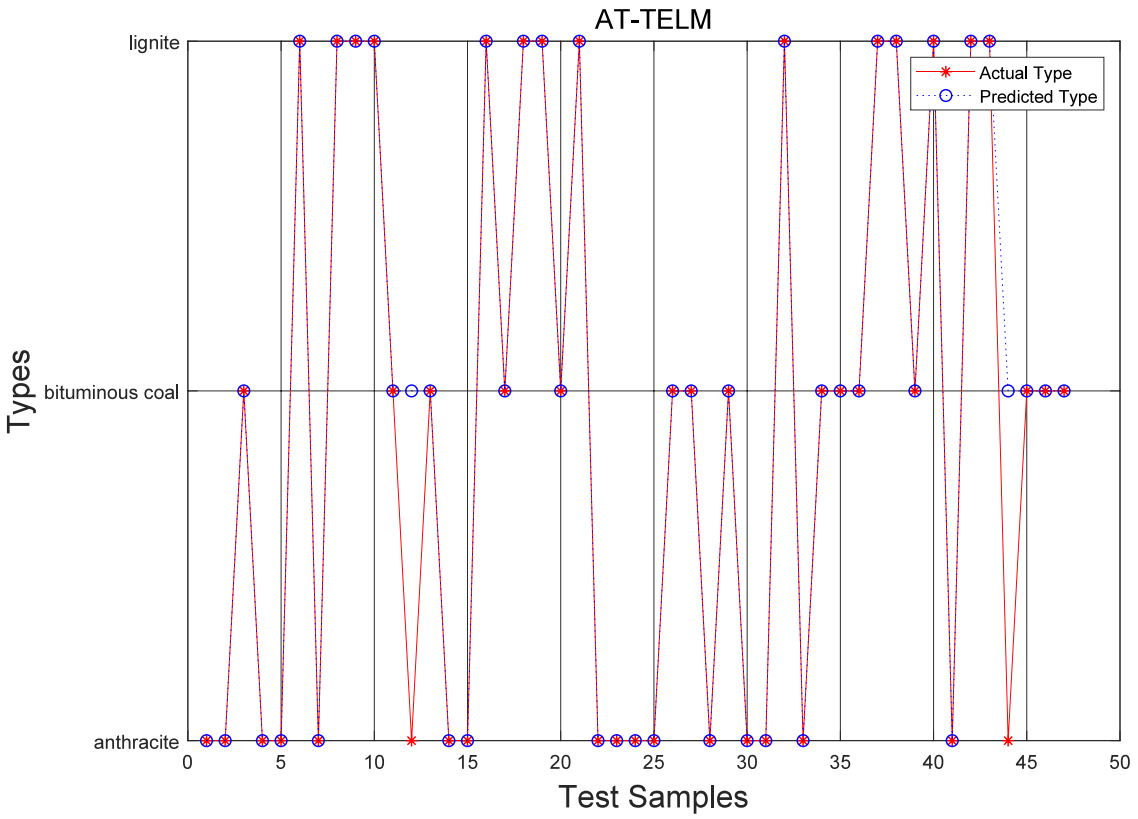


Figure 8. Classification results of the AT-TELM model.

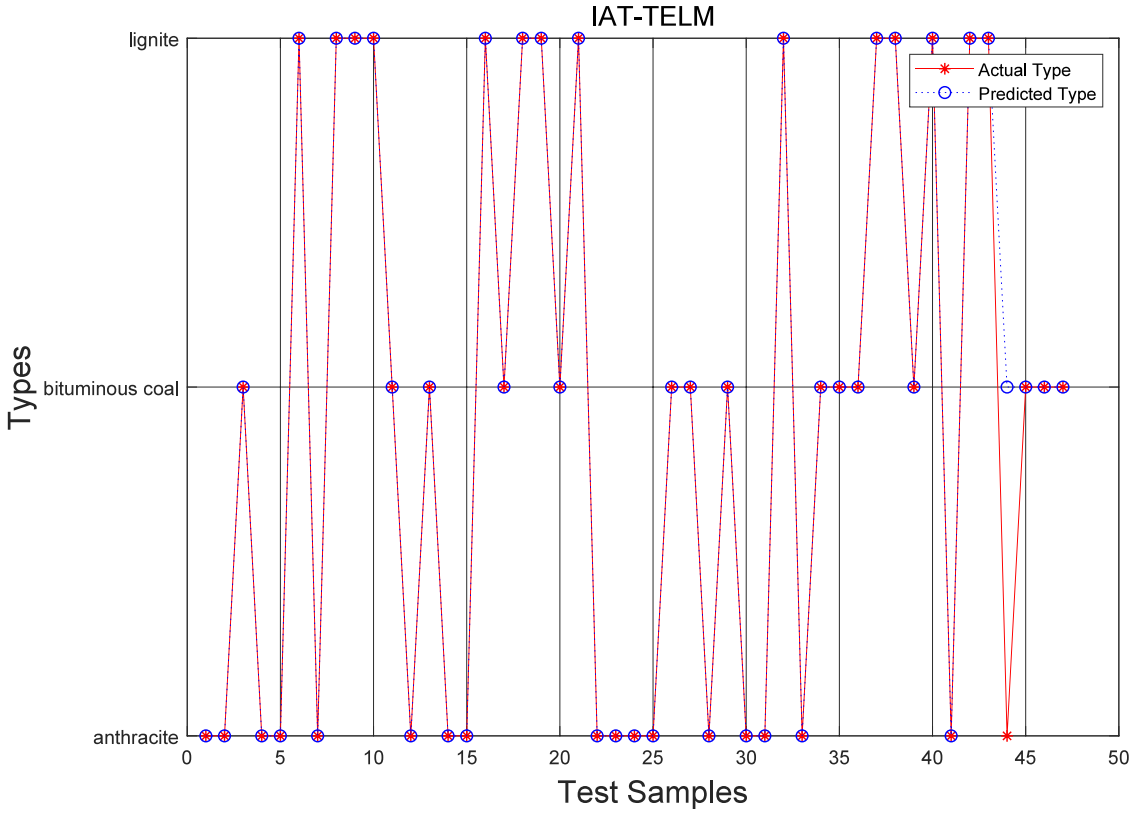


Figure 9. Classification results of the IAT-TELM model.

https://doi.org/10.1021/acsomega.3c04999
ACS Omega 2023, 8, 35232−35241

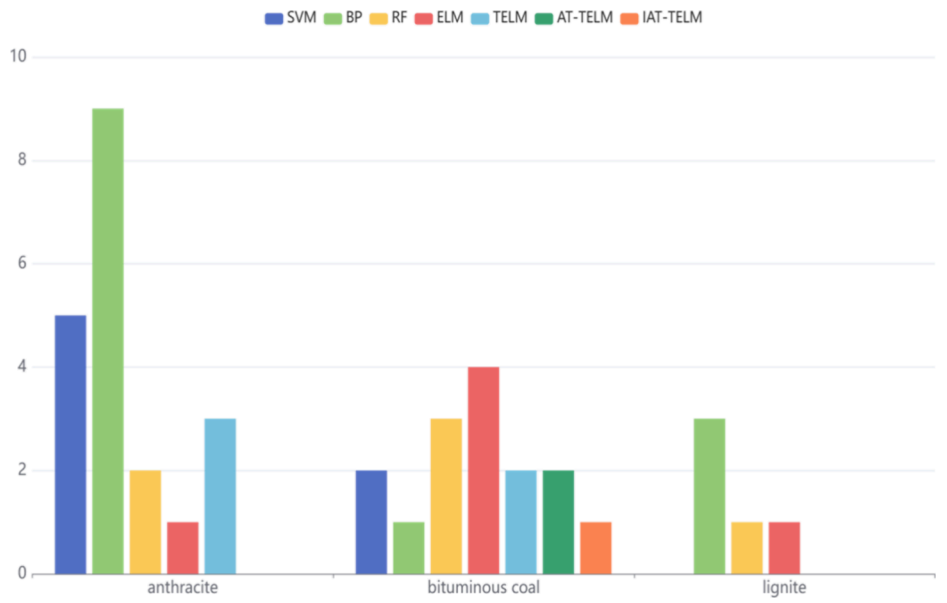


Figure 10. Number of prediction errors in coal classification for different models.

Algorithm: IAT-TELM

Input: Training data set $\{(x_j, y_j)\}_{j=1}^{q}$, the nodes' count l, activation function $g(\cdot)$.

Output: The predicting results $T = [t_1 \quad t_2 \quad \cdots \quad t_q]$.

1 Stochastically create the weight matrix W and the bias matrix B of the first hidden layer;

2 Calculate the output H and get the order matrix O_1 and downsample these matrix;

3 **While** $(\kappa < K)$ **do**

4 Calculate the loss function by (27) and the gradients by (28) and (29);

5 $s(\kappa + 1) = s(\kappa) - \eta \frac{\partial erro}{\partial s(\kappa)}$;

6 $t(\kappa + 1) = t(\kappa) - \eta \frac{\partial error}{\partial t(\kappa)}$;

7 $\kappa = \kappa + 1$;

8 **End**

9 Obtain the optimized weight matrix W_0 and bias matrix B_0 of the first hidden layer by (34);

10 Calculate the weight matrix W_1, bias matrix B_1 and the output weight matrix β of the second hidden layer by using (17) and (19);

11 Predict the type of the coal

3. RESULTS AND DISCUSSION

All of the experiments in this paper are carried out in the Windows 10 system. The coal recognition model is established, and the experimental results are visualized by using MATLAB R2018a.

3.1. Experimental Results. This study utilizes a total of 209 coal specimens encompassing three categories. After the outliers were removed, the remaining 203 samples are used to train and test the model. The training set is composed of 156 samples, while the test set includes 47 specimens. The accuracy of the two models proposed in this paper is shown in Figures 8 and 9. In terms of classification accuracy, the accuracy of AT-TELM reaches 95.7% and the accuracy of IAT-TELM reaches 97.8%. AT-TELM incorrectly classifies two anthracites into bituminous coal, and IAT-TELM incorrectly classifies one anthracite into bituminous coal. Both methods can completely identify bituminous coal and lignite correctly and have a high recognition rate for anthracite. Among them, IAT-TELM is better than AT-TELM in accuracy. The accuracy of this method can meet the application of a coal mine in the process of mining and production.

3.2. Algorithm Comparison. In order to reflect the accuracy and effectiveness of the proposed method, we compare the AT-TELM and IAT-TELM algorithms with several other algorithms commonly used in spectrum-based coal classification applications, including SVM, backpropagation algorithm (BP), random forest algorithm (RF), ELM, and TELM. By comparison with ELM, SVM, BP, and RF, the effectiveness of TELM in the field of spectrum-based coal classification is proved. By comparison with TELM, the effectiveness of the proposed method in the improved ELM algorithm is demonstrated and it can prove that the affine transformation function improves the performance of TELM. Table 3 shows the accuracy of various algorithms for spectral-based coal classification on the data set that is used in this paper. From the table, we can see that IAT-TELM achieves the best accuracy of 97.8% among the compared models. Comparing with TELM, it can be seen that IAT-TELM has achieved an improvement in accuracy, which proves the effectiveness of affine transformation function.

Table 3. Comparison of Classification Results of Different Algorithms

algorithms	accuracy
SVM	85.2%
RF	87.2%
BP	70.2%
ELM	85.1%
TELM	89.4%
AF-TELM	95.7%
IAF-TELM	97.8%

https://doi.org/10.1021/acsomega.3c04999
ACS Omega 2023, 8, 35232−35241

The classification errors of each model on different types of coal are shown in Figure 10.

It can be seen from the figure that the various models cited in this section have more classification errors on anthracite and the least classification errors on lignite. Among them, BP has the largest number of classification errors in anthracite and lignite classification and ELM has the largest number of classification errors in bituminous coal. The proposed method had no errors in the categories of lignite and anthracite. It can be seen from the above charts that compared with the methods used in this section, IAT-TELM has better performance in the classification of lignite, bituminous coal, and anthracite. In addition, compared with the traditional manual inspection and chemical inspection methods, we can see from Table 4 that the coal classification using machine learning has not only the advantages of improved accuracy but also lower inspection cost.[30]

Table 4. Comparison of Different Methods for Coal Classification

methods	time (h)	accuracy (%)	costs (USD)
manual experience	2	60	50
weighing method	3	70	50
chemical analysis	300	100	1000
IAT-TELM	8	97.8	90

4. CONCLUSIONS

The classification of coal is a prerequisite for the application of coal in industry. In this paper, a method combining machine learning and reflection spectrum is proposed for the identification of lignite, bituminous coal, and anthracite coal and its application. In the data processing stage, we use the MD method to eliminate outliers in the data set and PCA to reduce the dimension of the spectral data. In the experiment, it is proved that the IAT-TELM model has the best performance and high accuracy for coal classification tasks in these comparison algorithm. This method is a coal classification method with low cost, short time consumption, and high accuracy, which can meet industrial needs. Furthermore, due to the good generalization performance of ELM, the research field of this method can be extended to the reflection spectra of other important resources such as soil and iron ore.

AUTHOR INFORMATION

Corresponding Author

Dong Xiao − *School of Information Science and Engineering and Liaoning Key Laboratory of Intelligent Diagnosis and Safety for Metallurgical Industry, Northeastern University, Shenyang 110819, China*; orcid.org/0000-0002-0401-6654; Email: xiaodong@ise.neu.edu.cn

Authors

Boyan Li − *School of Information Science and Engineering and Liaoning Key Laboratory of Intelligent Diagnosis and Safety for Metallurgical Industry, Northeastern University, Shenyang 110819, China*

Hongfei Xie − *School of Information Science and Engineering and Liaoning Key Laboratory of Intelligent Diagnosis and Safety for Metallurgical Industry, Northeastern University, Shenyang 110819, China*

Jie Huang − *School of Information Science and Engineering and Liaoning Key Laboratory of Intelligent Diagnosis and Safety for Metallurgical Industry, Northeastern University, Shenyang 110819, China*

Zelin Yan − *School of Information Science and Engineering and Liaoning Key Laboratory of Intelligent Diagnosis and Safety for Metallurgical Industry, Northeastern University, Shenyang 110819, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c04999

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported in part by the National Key Research and Development Program of China under Grant 2020AAA0109200; in part by the Liaoning Revitalization Talents Program under Grant XLYC2008020; in part by the National Natural Science Foundation of China under Grant 52074064; in part by the Natural Science Foundation of Science and Technology Department of Liaoning Province under Grant 2021-BS-054; and in part by the Fundamental Research Funds for the Central Universities of China under Grants N2204006, N2104026, and N2304006.

REFERENCES

(1) Chen, P. Study on Integrated Classification System for Chinese Coal. *Fuel Process. Technol.* **2000**, *62*, 77−87.

(2) Zou, L.; Yu, X.; Li, M.; Lei, M.; Yu, H. Nondestructive Identification of Coal and Gangue via Near-Infrared Spectroscopy Based on Improved Broad Learning. *IEEE Trans. Instrum. Meas.* **2020**, *69* (10), 8043−8052.

(3) Xiao, D.; Xie, H.; Fu, Y.; Li, F. Mine Reclamation Based on Remote Sensing Information and Error Compensation Extreme Learning Machine. *Spectrosc. Lett.* **2021**, *54* (2), 151−164.

(4) Huang, G.; Yuan, L.-M.; Shi, W.; Chen, X.; Chen, X. Using One-Class Autoencoder for Adulteration Detection of Milk Powder by Infrared Spectrum. *Food Chem.* **2022**, *372*, No. 131219.

(5) Zhang, D.; Liang, P.; Chen, W.; Tang, Z.; Li, C.; Xiao, K.; Jin, S.; Ni, D.; Yu, Z. Rapid field trace detection of pesticide residue in food based on surface-enhanced Raman spectroscopy. *Microchim. Acta* **2021**, *188* (11), 1−28, DOI: 10.1007/s00604-021-05025-3.

(6) Jiang, S.; He, H.; Ma, H.; Chen, F.; Xu, B.; Liu, H.; Zhu, M.; Kang, Z.; Zhao, S. Quick Assessment of Chicken Spoilage Based on Hyperspectral NIR Spectra Combined with Partial Least Squares Regression. *Int. J. Agric. Biol. Eng.* **2021**, *14* (1), 243−250.

(7) Xiao, D.; Huang, J.; Li, J.; Fu, Y.; Mao, Y.; Li, Z.; Bao, N. Inversion study of soil organic matter content based on reflectance spectroscopy and the improved hybrid extreme learning machine. *Infrared Phys. Technol.* **2023**, *128*, No. 104488, DOI: 10.1016/j.infrared.2022.104488.

(8) Xiao, D.; Yan, Z.; Li, J.; Fu, Y.; Li, Z. Rapid proximate analysis of coal based on reflectance spectroscopy and deep learning. *Spectrochim. Acta Part A* **2022**, *287* (Pt 2), No. 122042, DOI: 10.1016/j.saa.2022.122042.

(9) Liu, K.; He, C.; Zhu, C.; Chen, J.; Zhan, K.; Li, X. A review of laser-induced breakdown spectroscopy for coal analysis. *TrAC Trends Anal. Chem* **2021**, *143*, No. 116357, DOI: 10.1016/j.trac.2021.116357.

(10) Yao, S.; Zhao, J.; Wang, Z.; Yoshihiro, D.; Lu, Z.; Lu, J. Analysis of spectral properties for coal with different volatile contents by laser-induced breakdown spectroscopy. *Spectrochim. Acta Part B* **2018**, *149*, 249−255.

(11) Le, B. T.; Xiao, D.; Okello, D.; He, D.; Xu, J.; Doan, T. T. Coal exploration technology based on visible-infrared spectra and remote sensing data[J]. *Spectrosc. Lett.* **2017**, *50* (8), 440−450, DOI: 10.1080/00387010.2017.1354889.

(12) Xie, Y.; You, J.; Lu, L.; Wang, M.; Wang, J. Raman Spectroscopic Study of Coal Samples during Heating. *Appl. Sci.* **2019**, *9* (21), 4699 DOI: 10.3390/app9214699.
(13) Zhang, W.; Zhuo, Z.; Lu, P.; Tang, J.; Tang, H.; Lu, J.; Xing, T.; Wang, Y. LIBS Analysis of the Ash Content, Volatile Matter, and Calorific Value in Coal by Partial Least Squares Regression Based on Ash Classification. *J. Anal. At. Spectrom.* **2020**, *35* (8), 1621−1631.
(14) Song, L.; Yu, Y.; Yan, Z.; Xiao, D.; Sun, Y.; Zhang, X.; Li, X.; Cheng, B.; Gao, H.; Bai, D. Rapid Analysis of Composition of Coal Gangue Based on Deep Learning and Thermal Infrared Spectroscopy. *Sustainability* **2022**, *14* (23), 16210 DOI: 10.3390/su142316210.
(15) Jayaganthan, S.; Babu, M. S.; Vasa, N. J.; Sarathi, R.; Imai, T. Classification of coal deposited epoxy micro-nanocomposites by adopting machine learning techniques to LIBS analysis. *J. Phys. Commun.* **2021**, *5* (10), 105006 DOI: 10.1088/2399-6528/ac2b5d.
(16) Wang, L.; Li, J.; Li, T.; Liu, H.; Wang, Y. Method Superior to Traditional Spectral Identification: FT-NIR Two-Dimensional Correlation Spectroscopy Combined with Deep Learning to Identify the Shelf Life of Fresh Phlebopus Portentosus. *ACS Omega* **2021**, *6* (30), 19665−19674.
(17) Urban, C. J.; Gates, K. M. Deep Learning: A Primer for Psychologists. *Psychol. Methods* **2021**, *26* (6), 743−773.
(18) O'Byrne, C.; Abbas, A.; Korot, E.; Keane, P. A. Automated Deep Learning in Ophthalmology: AI That Can Build AI. *Curr. Opin. Ophthalmol.* **2021**, *32* (5), 406−412.
(19) Xiao, D.; Li, H.; Sun, X. Coal Classification Method Based on Improved Local Receptive Field-Based Extreme Learning Machine Algorithm and Visible-Infrared Spectroscopy. *ACS Omega* **2020**, *5* (40), 25772−25783, DOI: 10.1021/acsomega.0c03069.
(20) Le, B. T.; Xiao, D.; Mao, Y.; Song, L.; He, D.; Liu, S. Coal Classification Based on Visible, Near-Infrared Spectroscopy and CNN-ELM Algorithm. *Spectrosc. Spectral Anal.* **2018**, *38*, 2107−2112.
(21) Huang, G.-B.; Zhu, Q.-Y.; Siew, C.-K. Extreme Learning Machine: Theory and Applications. *Neurocomputing* **2006**, *70* (1−3), 489−501.
(22) Cao, J.; Lin, Z.; Huang, G. B. Self-adaptive evolutionary extreme learning machine. *Neural Process. Lett.* **2012**, *36* (3), 285−305.
(23) (23) DENG, W. Y.; ZHENG, Q. H.; CHEN, L.; XU, X. B. Research on extreme learning of neural networks. *Chin. J. Comput.* **2010**, *33* (2), 279−287.
(24) Gu, Y.; Liu, J.; Chen, Y.; Jiang, X.; Yu, H. TOSELM: timeliness online sequential extreme learning machine. *Neurocomputing* **2014**, *128* (27), 119−127.
(25) Mao, Y.; Le, B. T.; Xiao, D.; He, D.; Liu, C.; Jiang, L.; Yu, Z.; Yang, F.; Liu, X. Coal Classification Method Based on Visible-Infrared Spectroscopy and an Improved Multilayer Extreme Learning Machine. *Opt. Laser Technol.* **2019**, *114*, 10−15.
(26) Le, B. T.; Xiao, D.; Mao, Y.; He, D.; Xu, J.; Song, L. Coal quality exploration technology based on an incremental multilayer extreme learning machine and remote sensing images. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57* (7), 4192−4201.
(27) Ding, S.; Zhang, N.; Xu, X.; Guo, L.; Zhang, J. Deep extreme learning machine and its application in EEG classification. *Math. Probl. Eng.* **2015**, *2015*, 1−11, DOI: 10.1155/2015/129021.
(28) Jaynes, E. T. Information theory and statistical mechanics. II. *Phys. Rev.* **1957**, *108* (2), 171.
(29) Cao, J.; Zhang, K.; Yong, H.; Lai, X.; Chen, B.; Lin, Z. Extreme Learning Machine With Affine Transformation Inputs in an Activation Function. *IEEE Trans. Neural Networks Learn. Syst.* **2019**, *30* (7), 2093−2107.
(30) Xiao, D.; Yan, Z. Coal Identification Based on Reflection Spectroscopy and Deep Learning: Paving the Way for Efficient Coal Combustion and Pyrolysis. *ACS Omega.* **2022**, *7* (27), 23919−23928.

https://doi.org/10.1021/acsomega.3c04999
ACS Omega 2023, 8, 35232−35241