Coal Classification Method Based on Improved Local Receptive Field-Based Extreme Learning Machine Algorithm and Visible–Infrared Spectroscopy

Dong Xiao,* Hongzong Li, and Xiaoyu Sun

ABSTRACT: In the process of using coal, if the type of coal cannot be accurately determined, it will have a significant impact on production efficiency, environmental pollution, and economic loss. At present, the traditional classification method of coal mainly relies on technician’s experience. This requires a lot of manpower and time, and it is difficult to automate. This paper mainly studies the application of visible-infrared spectroscopy and machine learning methods in coal mine identification and analysis to provide guidance for coal mining and production. This paper explores a fast and high-precision method for coal identification. In this paper, for the characteristics of high dimensionality, strong correlation, and large redundancy of spectral data, the local receptive field (LRF) is used to extract the advanced features of spectral data, which is combined with the extreme learning machine (ELM). We improved the coyote optimization algorithm (COA). The improved coyote optimization algorithm (I-COA) and local receptive field-based extreme learning machine (ELM-LRF) are used to optimize the structure and training parameters of the extreme learning machine network. The experimental results show that the coal classification model based on the network and visible–infrared spectroscopy can effectively identify the coal types through the spectral data. Compared with convolutional neural networks (CNN algorithm) and principal component analysis (PCA algorithm), LRF can extract the spectral characteristics of coal more effectively.

1. INTRODUCTION

Coal is the main source of energy in the world. With the development of society and industry, the analytical quality of coal plays a decisive role in production efficiency and environmental pollution. In the process of coal mining and production, the classification of coal has important guiding significance for production planning and resource estimation. However, in the current mining process, the traditional coal classification method mainly relies on artificial experience classification. This requires a lot of manpower and material resources, and it is difficult to achieve automation. Therefore, how to distinguish different types of coal quickly, accurately, and in real time is an important research on the mining and use of coal.

According to the degree of coalification, coal is mainly divided into three categories, namely, anthracite, bituminous coal, and lignite. Different coal types have different usage methods in different occasions. For example, the type and quality of coal used in thermal power plants is an important basic basis for boiler design and production process. Therefore, in the process of coal mining, distinguishing mining by coal type is the conventional production method of coal mining. There are two common traditional coal classification methods: one is to use manual classification method, although the classification speed is relatively fast, but the classification accuracy is low, and it is difficult to achieve automation; the second is to use chemical analysis methods, although the recognition accuracy is high. This method has the disadvantages of high cost and long detection time. Accurate and efficient determination of the type of coal is of great research significance for reducing classification costs and improving classification efficiency. Therefore, how to quickly and accurately determine the type of coal is an urgent problem to be solved by modern mineral processing technology.

In the previous research on coal classification, Freidina1 proposed the principle of faceting method and mass method.

Received: June 25, 2020
Accepted: September 2, 2020
Published: October 2, 2020

© 2020 American Chemical Society

http://pubs.acs.org/journal/acsona
and introduced the classification method of fossil coal from the perspective of the useful quality of coal. Dong, Li, and others classified iron ore based on BP algorithm and iron ore spectrum and achieved good results. Wang adopted the gray fixed weight clustering evaluation model and calculated the objective weight of each index using the projection pursuit method so as to obtain the clustering coefficient of coal samples belonging to each gray class, namely, anthracite, bituminous coal, and lignite, to solve the problem of coal type identification. Wang proposed a rapid coal classification method based on near-infrared spectroscopy. Wang first used the discriminant analysis function of statistical software SPSS to screen out the main indicators for discriminating coal types and then used variables to establish a Bayesian stepwise linear discriminant function to identify modeling samples and test samples. Pandit proposed to use a classic unsupervised clustering technology for coal classification, namely, "K-means clustering" and a nonlinear clustering form based on artificial intelligence (AI), namely, "self-organization" mapping (SOM). Carlsen used fast pyrolysis gas chromatography/mass spectrometry combined with statistical methods such as principal component analysis (PCA) and hierarchical cluster analysis (HCA) to classify coal.

With the development of spectroscopy, compared with chemical methods, it has the advantages of fast detection speed, economy, and convenience. Therefore, spectral analysis technology has been used in many fields such as rock identification, composition analysis, and food qualification. Since the factors that affect the spectrum of coal mines are mainly the oxides of carbon, sulfur, iron, etc., the spectral data generally contains a lot of irrelevant information for coal type identification, which makes the spectral data of coal mines have high dimensions and redundancy. The margin is large. Therefore, before analyzing the spectrum, effective feature extraction must be done.

In recent years, machine learning and big data analysis have attracted great attention from researchers in different disciplines. To our knowledge, the success of machine learning depends on three key factors: a powerful computing environment, rich dynamic data, and efficient learning algorithms. More importantly, in big data applications, higher requirements are placed on efficient machine learning algorithms.

Most traditional neural network training methods, such as the BP algorithm, involve a large number of gradient descent search steps, which have the problems of slow convergence, local minimum, and serious human intervention. Extreme learning machine (ELM) aims to overcome these shortcomings and limitations faced by traditional learning theories and technologies. ELM provides an efficient and unified learning framework for "generalized" single hidden layer feedforward neural networks (SLFN), including but not limited to Sigmoid networks, RBF networks, threshold networks, trigonometric networks, and fuzzy inference, fully complex neural networks, high-order networks, Ridge polynomial networks, wavelet networks, Fourier series, etc.

In biomedical analysis, chemical process, system modeling, power system, motion recognition, hyperspectral image, and many other different applications, it provides competitive accuracy with extremely high efficiency. In almost all ELM implementations implemented in the past few years, hidden nodes are fully connected to input nodes. Fully connected ELM has good generalization performance in many applications and achieves high efficiency. However, in applications such as spectral analysis and image processing, there may be strong local correlations, so it is reasonable to expect that the corresponding neural network has local connections instead of full connections in order to learn local correlations. As we all know, there is a local receptive field in the retina module of the biological learning system, which helps to consider the local correlation of the input image. The ELM theory proves that hidden nodes can be randomly generated according to any continuous probability distribution. Naturally, in some applications, hidden nodes can be generated according to some continuous probability distributions, which are denser around certain input nodes and sparser in farther places. Therefore, the ELM theory is actually valid for local receptive fields.

This paper explores a new coal classification model based on spectral technology and deep learning. In view of the characteristics of high dimensionality, strong correlation, and high redundancy of spectral data, this paper proposes to combine the local receptive field with ELM to solve the problem of coal classification, and to further improve the classification accuracy, the coyote optimization algorithm is studied. An improved coyote algorithm is proposed, and the improved coyote algorithm is used to improve and optimize the structure and training parameters of the model. Finally, different experiments verify the effectiveness of the proposed method.
2. EXPERIMENTAL SECTION

2.1. ELM-LRF Algorithm. Although different types of local receptive fields and combined nodes can be used in ELM, for ease of implementation, we use a simple step probability function as the sampling distribution and square root/square root pooling structure to form the combined node. According to the ELM-LRF theory,21 the receptive field of each hidden node will consist of input nodes within a predetermined distance from the center. In addition, the input weights are simply shared to different hidden nodes to directly perform the convolution operation, and it is easy to implement.

In this way, we have constructed a specific case for the general ELM-LRF (Figure 1):

1) The random convolution node in the Feature Map in Figure 1 is a locally connected hidden node.

2) The node in the Pooling Map in Figure 1 is an example of a combined node.

In order to obtain a complete representation of the input, K different input weights are used to generate K different feature maps. Figure 2 describes the implementation network used by all K maps in this paper.

In Figure 2, the hidden layer consists of random convolution nodes. The input weights of the same feature map are shared among different feature maps. The input weights are randomly generated and then orthogonalized as follows:

1) Randomly generate the initial weight matrix $\hat{A}_{\text{init}}$, the feature map is $(d - r + 1) \times (d - r + 1)$ under the input size $d \times d$ and receptive field $r \times r$.

2) The orthogonalized initial weight matrix $\hat{A}_{\text{init}}$ uses singular value decomposition (SVD), and $A$, $a_k$ is the orthogonal basis of $\hat{A}_{\text{init}}$.

Orthogonalization allows the network to extract a more complete feature set than non-orthogonal features, thereby further improving the generalization performance of the network.

The input weight to the $k$th feature map is $a_k \in \mathbb{R}^{r \times r}$ and it corresponds to $\hat{a}_k \in \mathbb{R}^{r \times r}$. The $k$th feature mapping convolution node $(i, j)$, $i, j = 1, \ldots, (d - r + 1)$ can be calculated as

$$c_{i,j,k}(x) = \sum_{m=1}^{r} \sum_{n=1}^{r} x_{i+m-1,j+n-1} a_{m,n,k}$$

$i, j = 1, \ldots, (d - r + 1)$ (1)

A square/square root pool structure is used to form a combined node. The pool size $e$ is the distance between the center and the edge of the pool area, as shown in Figure 2. The pool map is the same size as the feature map, which is $(d - r + 1) \times (d - r + 1)$. $c_{i,j,k}$ and $h_{p,q,k}$ represent the node $(i, j)$ in the $k$th feature map and the combined node $(p, q)$ in the $k$th pool map, respectively:

$$h_{p,q,k} = \sqrt{\sum_{i=(p-e)}^{p+e} \sum_{j=(q-e)}^{q+e} c_{i,j,k}^2}$$

$p, q = 1, \ldots, (d - r + 1)$

If $(i, j)$ exceed the limit: $c_{i,j,k} = 0$ (2)

For an input sample $x$, the value $h_{p,q,k}$ of the combined node is calculated by the cascade of eqs 3 and 4:

$$c_{i,j,k}(x) = \sum_{m=1}^{r} \sum_{n=1}^{r} x_{i+m-1,j+n-1} a_{m,n,k}$$

$p, q = 1, \ldots, (d - r + 1)$

If $(i, j)$ exceeds the limit: $c_{i,j,k} = 0$ (3)

$$h_{p,q,k} = \sqrt{\sum_{i=(p-e)}^{p+e} \sum_{j=(q-e)}^{q+e} c_{i,j,k}^2}$$

$p, q = 1, \ldots, (d - r + 1)$

If $(i, j)$ exceeds the limit: $c_{i,j,k} = 0$ (4)
Simply concatenate the values of all combined nodes into a row vector and put the rows of N input samples together to get the combined layer matrix \( H \in \mathbb{R}^{N \times K \cdot (d - r + 1)^2} \):

1. If \( N \leq K \cdot (d - r + 1)^2 \)
   \[
   \beta = H^T \left( \frac{1}{C} + HH^T \right)^{-1} T
   \]  
   \( (5) \)

2. If \( N > K \cdot (d - r + 1)^2 \)
   \[
   \beta = \left( \frac{1}{C} + H^T H \right)^{-1} H^T T
   \]  
   \( (6) \)

2.2. ICELM-LRF. The coyote optimization algorithm (COA) is a population-based algorithm.\(^\text{22} \) It is inspired by canines that are divided into two groups: group intelligence and evolutionary heuristics, and it is inspired by the behavior of coyotes.\(^\text{23,24} \) Unlike the grey wolf optimizer (GWO),\(^\text{25} \) the COA has different algorithmic structure settings, and it does not pay attention to the social hierarchy and domination rules of these animals, even if alpha is used as a group leader (as previously described). In addition, the COA focuses on the social structure and exchange of experiences of wolves, rather than hunting prey as in GWO.

The social condition \( \text{soc} \) (set of decision variables) of \( t \)th instant \( pt \) is written as

\[
\text{soc}^{P,t} = \vec{x} = (x_1, x_2, \ldots, x_D)
\]  
\( (7) \)

It means the adaptation of the wolf to the environment (the cost of the objective function) \( \text{fit}^{P,t} \in \mathbb{R} \).

The first step of the COA is to initialize the global coyote population. It is achieved by assigning random values in the search space of the \( t \)th coyote of the \( P \)th package in the \( j \)th dimension, as follows:

\[
\text{soc}^{P,t} = l_b^j + r_j(ub^j - l_b^j)
\]  
\( (8) \)

where \( l_b^j \) and \( ub^j \) respectively denote the lower and upper bounds of the \( j \)th decision variable, \( D \) is the dimension of the search space, and \( r_j \) is the real random number generated in the range \([0,1]\) using uniform probability. After that, assess the adaptation of the coyotes under their current social conditions:

\[
\text{fit}^{P,t} = f(\text{soc}^{P,t})
\]  
\( (9) \)

Initially, coyotes were randomly assigned to wolves, but sometimes they would leave the wolves:

\[
P_i = 0.005 \cdot N_c^2
\]  
\( (10) \)

Considering the minimization problem, the \( \alpha \) of \( P \)th package at instant \( t \)th is defined as

\[
\text{alpha}^{P,t} = \{ \text{soc}^{P,t}_{\text{arg min}(\text{soc}^{P,t})} \}
\]  
\( (11) \)

The COA links all the information from the coyotes and calculates it as the cultural tendency of the ethnic group:

\[
\text{cult}^{P,t} = \begin{cases} 
    O_{P,t}^{P,t+1/2,j}, & N_i \text{ is odd} \\
    O_{P,t}^{P,t+j} + O_{P,t}^{P,t+1,j}, & \text{otherwise} 
\end{cases}
\]  
\( (12) \)

In the equation, \( O_{P,t}^{P,t+1/2,j} \) represents the ranking social status of all \( Pt \) wolves in the range \([1,D]\) for each \( j \) at the instant of time \( t \). In other words, the cultural orientation of the ethnic group is calculated as the average social status of all coyotes from a specific ethnic group.

Considering the two main biological events, birth and death, COA calculates the age (in years) of the coyote and expresses it as the age \( \text{age}^{P,t} \) in \( \mathbb{N} \). The birth of the new coyote is written by the combination of the social conditions of two parents (randomly selected) plus environmental impact, like this

\[
p_{\text{birth}}^{P,t} = \begin{cases} 
    \text{soc}^{P,t}_{\text{arg min}(\text{soc}^{P,t})}, & \text{rnd}^j < P_i \text{ or } j = j_1 \\
    \text{soc}^{P,t}_{\text{arg max}(\text{soc}^{P,t})}, & \text{rnd}^j \geq P_i + P_a \text{ or } j = j_2 \\
    R_j, & \text{otherwise}
\end{cases}
\]  
\( (13) \)

where \( r_1 \) and \( r_2 \) are random coyotes from package \( p \)th, \( j_1 \) and \( j_2 \) are the two random dimensions of the problem, \( P_i \) is the scattering probability, \( P_a \) is the correlation probability, \( R_j \) is the random number within the bounds of the \( j \)th decision variable, and \( \text{rnd}^j \) is the random number in \([0,1]\) generated by uniform probability. Diversity and probability of association guide the cultural diversity of wolves. In the initial version of COA, \( P_i \) and \( P_a \) are defined as

\[
P_i = 1/D
\]  
\( (14) \)

\[
P_a = (1 - P_i)/2
\]  
\( (15) \)

where \( P_i \) has the same effect on both parents.

In order to maintain the stability of the population, COA synchronizes the birth and death of wolves, as described in Alg.1, where \( \omega \) and \( \phi \) respectively represent the coyote group that is more adaptable to the environment than puppies (that is, the solution that presents the worst objective function cost Program group) and the number of coyotes in this group. Note that two or more coyotes may be of similar age (on line 4). In this case, the coyotes with poor adaptability will die.

\begin{algorithm}[H]
\caption{Birth and death of a population}
1. Calculate \( \omega \) and \( \phi \) (\( \omega \) refers to the individuals weaker than the offspring, and \( \phi \) refers to the number of individuals weaker than the offspring)
2. if \( \phi = 1 \) then
3. The offspring survives and the only coyote in \( \omega \) dies
4. if \( \phi > 1 \) then
5. The offspring survived and died in the oldest coyote at \( \omega \)
6. else
7. Offspring
8. end if
\end{algorithm}

In order to express the cultural interaction within the ethnic group, COA assumes that the coyotes are affected by alpha \( \alpha \) and ethnic influence \( \beta \). The first refers to a cultural difference, from a random coyote pack \( cr_1 \) to an alpha coyote, while the second is a cultural difference, from a random coyote \( cr_2 \) to a culturally inclined pack. The random coyote is selected by the uniform distribution of probability, \( \delta_1 \) and \( \delta_2 \) are written as

\[
\beta_i = \alpha^{P,t} - \text{soc}^{P,t}_{ci}
\]  
\( (16) \)
\[ \beta_2 = cull^{h,t} - soc^{c,t} \]  

(17)

Therefore, the new social conditions of the coyotes are updated using the following equations using alpha and group influence: selection by uniform probability distribution, \( \delta_1 \) and \( \delta_2 \) are written as

\[
\text{new}_{soc}^{c,t} = soc^{c,t} + r_1 \beta_1 + r_2 \beta_2
\]  

(18)

where \( r_1 \) and \( r_2 \) are the weights of alpha and pack, respectively. Initially, \( r_1 \) and \( r_2 \) were defined as random numbers in the range \([0,1]\) generated with uniform probability. Then, assess the new social situation:

\[
\text{new}_{fit}^{c,t} = f(\text{new}_{soc}^{c,t})
\]  

(19)

The cognitive ability of the wolf determines whether the new social conditions are more suitable to maintain it than the old social conditions, which means

\[
\text{soc}^{c,t+1} = \begin{cases} 
\text{new}_{soc}^{c,t}, & \text{new}_{fit}^{c,t} < \text{fit}^{c,t} \\
\text{soc}^{c,t}, & \text{otherwise} 
\end{cases}
\]  

(20)

Finally, choose the social conditions of the coyote that can best adapt to the environment as the overall solution to the problem. The pseudocode for COA is described in Algorithm 2, where \( N_c \) can be set to the first guess in the range \([5,10]\), and \( N_p \) can be subsequently adjusted to define the overall size of the algorithm.

Algorithm 2 COA pseudocode

1. initialize population \( N_c \) and coyote \( N_p \) (Equation 8)
2. check the adaptability of the coyote (Equation 9)
3. while Stop condition is not met do
4. for each population \( p \) do
5. define the alpha coyote (Equation 11)
6. calculate the social tendency of the population (Equation 12)
7. for Coyote \( c \) of each population \( p \) do
8. update social conditions (Equation 18)
9. evaluate new social conditions (Equation 19)
10. adaptation (Equation 20)
11. end for
12. survival and death (Equation 13) (Algorithm 1)
13. end for
14. transfer between populations (Equation 10)
15. update the age of the coyotes
16. end while
17. choose the most suitable coyote

With the rapid development of intelligent algorithms, many optimization problems have become more complicated. Inspired by the theory of swarm intelligence, the advantages of the artificial coyote algorithm (COA) have attracted the attention of researchers all over the world and achieved most research results. In most cases, the COA is very suitable for regression optimization problems.

Usually, the performance of an algorithm is usually measured by two indicators: local search capability and global convergence capability. Local search refers to the ability to infinitely approach the optimal solution, while global convergence ability refers to the ability to find the approximate location of the global optimal solution. Local search ability and global search ability are indispensable in optimization algorithms.

However, a problem with the COA is the lack of global search capabilities. Although its strong optimization capabilities can allow group diversity and expand the search range, it also means that the convergence rate needs to be slow. Therefore, the basic COA search performance is strong but the development performance is weak.

In the COA, for the birth and death of a population, the oldest individual dies at \( \varphi > 1 \), but when there are multiple individuals of the same age in \( \omega \) at the same time, the algorithm does not give a strategy.

In order to improve the global search ability of the COA, at \( \varphi > 1 \), a random factor \( \gamma \) is added to the oldest coyote of \( \omega \), and the \( r\)th individual is randomly selected among the oldest individuals to die. The improved algorithm is as follows.

Algorithm 3 Improved birth and death of a population

1. Calculate \( \omega \) and \( \varphi \) (\( \omega \) refers to the individuals weaker than the offspring, and \( \varphi \) refers to the number of individuals weaker than the offspring)
2. if \( \varphi = 1 \) then
3. The offspring survived and the only coyote died in \( \omega \)
4. if \( \varphi > 1 \) then
5. The offspring survived and died in the oldest \( \omega \) coyote \( \gamma \)
6. else
7. Offspring
8. end if

The second point is that in the COA, there are problems in optimizing the ELM parameters in order to express the cultural interaction within the ethnic group. The evaluation of the alpha influence \( \delta_1 = alpha^{h,t} - soc^{c,t} \) and the ethnic influence \( \delta_2 = cull^{h,t} - soc^{c,t} \) cannot reflect the influence of the group on the individual. The alpha effect \( \delta_1 = alpha^{h,t} - soc^{c,t} \) is improved to

\[
\beta_1 = alpha^{h,t} - soc^{c,t}
\]  

(21)

Ethnic influence \( \delta_2 = cull^{h,t} - soc^{c,t} \) is improved to

\[
\beta_2 = cull^{h,t} - soc^{c,t}
\]  

(22)

In the mutual influence of the ethnic groups, it will also be affected by a part of the environment. In order to increase the global search ability of the algorithm, for the new social conditions \( \text{new}_{soc}^{c,t} = \text{soc}^{c,t} + r_1 \cdot \delta_1 + r_2 \cdot \delta_2 \) of the coyotes, the impact of adding the environment on the ethnic group is improved as follows:

\[
\text{new}_{soc}^{c,t} = (\beta + (1 - \beta) \times r_3)(\text{soc}^{c,t} + r_1 \beta_1 + r_2 \beta_2) + (1 - \beta - (1 - \beta) \times r_3)(lb + r_3(ab - lb))
\]  

(23)

where \( \beta \) is the reference value, which is a fixed value, and \( r_3 \) is a \( 0-1 \) random value. This not only reflects the internal impact on individuals but also reflects the environmental impact on individuals.

Returning to the ELM-LRF network, although the weights and deviation vectors are randomly given, the ELM-LRF network can approximate any continuous function. However,
in practical applications, such a random parameter method has a significant impact on the performance of the model because the random parameter method does not guarantee the optimal state of the model output. With the development of swarm intelligence algorithms, more and more scholars use swarm intelligence algorithms to optimize artificial intelligence systems, especially on artificial neural network algorithms. Swarm intelligence optimization algorithms have good random search capabilities and fast convergence speed. The generalization performance of neural networks is strong, so they can be combined to give play to their own advantages.

With the rapid development of swarm intelligence theory, it has the advantages of good random exploration ability and fast convergence speed. In this section, we will discuss the combination of I-COA and ELM-LRF network to form a more complete coal composition analysis model. This method is called ICELM-LRF algorithm. The specific algorithm flow is shown in Algorithm 4:

Algorithm 4 Pseudocode of I-COA

1. Initialize population $N_p$ and coyote $N_c$ (Equation 8)
2. Check the adaptability of the coyote (Equation 9)
3. While stop condition is not met do
   4. For each population $p$ do
      5. Define the alpha coyote (Equation 11)
      6. Calculate the social tendency of the population (Equation 12)
      7. For each coyote $c$ of each population $p$ do
         8. Update social conditions (Equation 23)
         9. Evaluate new social conditions (Equation 19)
         10. Adaptation (Equation 20)
      11. End for
   12. Survival and death (Equation 13) (Algorithm 3)
   13. End for
   14. Transfer between populations (Equation 10)
   15. Update the age of the coyotes
4. End while
5. Choose the most suitable coyote

In this paper, the improved COA is named I-COA. The I-COA is used to optimize the parameters of the ELM-LRF algorithm. First, the ELM-LRF is initialized, the I-COA is initialized, and the parameters are determined, including the number of wolves and the size of the wolves, the maximum number of iterations, etc.; then check the adaptability of the coyote, get the alpha coyote of the population $i$, and get the social tendency of the population $i$; call ELM-LRF to calculate the new social conditions of the population, update the social conditions of the population $i$; probability of transition between populations; call ELM-LRF to calculate the fitness value of progeny to obtain the survival and death conditions of the population, update the age of the coyote; until the stopping conditions are met, output the ELM optimal parameters and optimal values.

2.3. Acquisition and Processing of Spectral Data. This article collected samples collected from the Shenhua Zibao Rixile, Jiajin, Zhijin, and Yimin mining areas. Four kinds of spectrum are anthracite, bituminous coal, lignite, and gangue. It includes 315 samples, of which 282 coal samples are used as the training set and 33 coal samples are used as the test set. The anthracite label is 1, the bituminous coal label is 2, the lignite label is 3, and the gangue label is 4. The detailed data attributes of coal types are shown in Table 1.

| Table 1. Description of Data Set |
|----------------------------------|
| coal type | training set | test set |
|-----------|--------------|----------|
| 1: anthracite | 98 | 8 |
| 2: bituminous coal | 50 | 8 |
| 3: lignite | 74 | 6 |
| 4: gangue | 60 | 11 |
| sum | 282 | 33 |

This subject is based on the interdisciplinary technology of spectroscopy and computer science to analyze, identify, and monitor coal mines. We collected a large number of coal and non-coal samples in several coal mining areas. Use a Spectra Vista’s SVC HR-1024 ground spectrometer as an experimental instrument to measure the spectral information of the sample. The spectral range of the instrument: 350–2500 nm. Built-in memory: 500 scans (scan). Weight: 3 kg. Number of channels: 1024. The spectral resolution (FWHM $\leq 8.5$ nm) is in the range of 1000–1850 nm. Minimum integration time: 1 ms.

The measurement spectrum experiment was carried out in different environments outdoors and indoors. First, clean the surface of the sample and then cut the coal sample into two parts: one for grinding and the other for the block. For outdoor experiments, using solar radiation as the light source, the experimental observation is selected between 10:00 and 14:00 under the sun. When the sun’s altitude angle is $60^\circ$, the sky is clear without clouds or few clouds, the scanning time is in seconds/time, and the probe of the spectrometer is 480 mm away from the sample surface. The surface of the coal sample should be kept as horizontal as possible, and the spectrometer lens and observation surface should be kept vertical. Experimenters are required to wear dark clothing and not allowed to move around in order to minimize the data error caused by the reflection of the clothing. Because the solar radiation value is different at different times, it is necessary to perform whiteboard measurement calibration every 10 min in the experiment to calibrate the equipment.

For indoor experiments, a halogen lamp is used as the light source, the scanning time is also 1 s/time, and the probe of the spectrometer is 480 mm away from the surface of the coal sample and perpendicular to the surface of the sample piece. The halogen lamp is 320 mm away from the coal sample and forms an angle of $45^\circ$ with the surface of the coal sample. In order to reduce the interference caused by the environment to the experiment and avoid the interference of other light sources, the spectrum measurement experiment was completed in a closed room. Experimenters are not allowed to walk around and wear dark clothing to reduce the interference of the environment with the spectral data. Figure 3 shows the sample collection and data collection process.

For the coal content, chemical analysis methods are used to determine the coal composition. These indicators are the main basis for judging coal quality. The chemical analysis method is to measure through complex experiments and standard instruments to obtain moisture, volatile matter, ash, fixed carbon, low calorific value, and sulfur content in coal. For example, the volatile matter is heating the coal under air under...
high temperature conditions, the organic matter in the coal will react, part of the mass will become gas escape, and the rest of the material will be left in solid form.

The SVC HR-1024 spectrometer has a total of 1024 bands, and the spectral range is 350–2500 nm.

Figure 4 shows photos of different samples, and Figure 5 shows the spectral curves of these samples. Each coal sample has 961 spectral characteristics. Spectral curves between different kinds of samples are mixed together, especially bituminous coal and anthracite, because the coal itself is black, most of the energy injection is absorbed, the reflection intensity is low, and the spectrum has little valid information. If manual methods are used, it is difficult to distinguish these coal types, especially anthracite, fat coal, and coking coal. Therefore, using machine learning methods for recognition can effectively improve the classification accuracy.

3. RESULTS AND DISCUSSION

3.1. Coal Classification Model Based on the ICELM-LRF Algorithm. The spectral data of coal has high dimensions, strong correlation, and great redundancy. If the unprocessed spectral data is directly used for modeling and identification, the model identification ability will be reduced. Therefore, the ICELM-LRF algorithm established in this paper is mainly to solve the problems of hyperspectral data and at the same time improve the accuracy of coal classification. Based on this idea, the parameters of the ICELM-LRF coal classification model designed in this paper are as follows: For the LRF network, this paper uses ReLU as the activation function and the average sampling function. The kernel size of the convolutional layer is $4 \times 4$, the number of output features of the convolutional layer is 2, and the sampling function size of the sampling layer is $3 \times 3$. I-COA sets the maximum number of iterations to 100, the number of populations to 5, and the number of individuals in each population to 3. Use the Sigmoid function as the activation function of the ELM algorithm.

In order to evaluate the performance of the ICELM-LRF model, this article establishes three models based on LRF, COA, and ELM algorithms and compares them, including LRF combined with basic ELM (ELM-LRF model), LRF combined with basic COA and ELM (LRF-ICELM model), and LRF combined to improve COA and ELM (ICELM-LRF model). In this paper, different experiments were carried out on indoor and outdoor spectra. The comparison results are shown in Figure 6, Figure 7, Figure 8, Figure 9, and Table 2.

After the algorithm ran 30 times, we take the best one. Figure 6 is the distribution diagram of the prediction results of the test and training sets of the ELM-LRF model. It can be seen that, due to the lack of optimization of the ELM
structure and training parameters, the classification results of the test and training sets have a large deviation. In particular, there is a large deviation in the classification results of the first type of anthracite. Therefore, the ELM-LRF model has limited ability to identify the type of coal. According to the results shown in Figure 9 and Table 2, the classification accuracy of the ELM-LRF model in the spectral experiment test set is 75.76%. The classification accuracy of the ICELM-LRF model in the spectral experiment test set is 100%. The training time of CELM-LRF and ICELM-LRF models is about 50 s, and the training time is basically the same.

From the above analysis, it can be seen that the ICELM-LRF model has good performance for coal classification, and using the LRF network to extract features for data in the ELM algorithm can achieve highly accurate classification. Using the I-COA to optimize the ELM-LRF network can also improve the performance of the classification model. Compared with the original COA, the improved COA has a higher accuracy and stability. This provides a fast and efficient identification method for coal type identification technology.

3.2. Comparison with Other Methods.
Principal component analysis combined with ELM has been widely used in many problems, and there have been many related studies in the field of spectroscopy. Jiang et al.28 used Fourier transform spectroscopy combined with multivariate analysis to determine the hardness and solid content of pears. They use principal components and independent components to extract valid information from the original spectral data. Then, use ELM to build a regression model. Zheng et al.29 applied ELM, PCA,30 and spectroscopy to food classification and counting and compared with other typical intelligent methods (including SVM, squared discriminant analysis, and back-propagation BP network). The accuracy is better than other classification methods. In addition, the execution speed of ELM is much faster than these algorithms. This shows that ELM is a promising method of food classification. Liu et al.31 combined laser-induced breakdown spectroscopy with machine learning methods to identify non-transgenic and transgenic corn and achieved 100% test accuracy.

CNN32 is a feature extraction method for deep learning. CNN mainly uses multiple hidden layers to extract data
features and uses the back-propagation method to find the optimal parameters of the network. CNN proved to be an efficient feature extraction method.

This article compares the LRF method, PCA method, and CNN method. Among them, PCA uses 10 principal components to replace the original data. For the recognition algorithm, in addition to using the CELM algorithm, this article also uses RBF-SVM for comparative analysis of the results. The experimental results are shown in Table 3 and Figure 10.

According to the results shown in Table 3 and Figure 10, it can be seen that the CELM-LRF algorithm has the highest recognition accuracy among all algorithms, reaching 100%, but the training takes a long time. For using CNN to extract features, the highest recognition precision is CNN-IPELM, and the test set recognition accuracy reaches 93.02%, while for PCA as feature extraction, the highest recognition accuracy is the PCA-IPELM model, and the test set recognition accuracy reaches 91.87%. The data after feature extraction using the LRF network enables the model to effectively identify the type of coal, and the accuracy rate is generally higher than the model using CNN and PCA. The results also verify the effectiveness of the improved coyote algorithm. The improved algorithm has a higher recognition accuracy than other typical algorithms.

Table 4 is a comparison of artificial experience, chemical testing methods, and the methods proposed in this article. This table is based on the identification costs of 206 coal samples. It can be seen that although the artificial experience method is very economical, the recognition accuracy is low. In addition to the need for expensive experimental drugs using chemical analysis methods, some chemical experiment measuring instruments cost 2 million or more. Therefore, although the chemical method is accurate, the cost is high, and it takes a long time. The cost of the method proposed in this paper only needs to be within 350,000, which is higher than the traditional artificial empirical method in recognition accuracy, less investment, and shorter time-consuming than
the chemical analysis method. This provides a fast, high-efficiency, and low-cost analysis tool for coal classification technology.

4. CONCLUSIONS
This paper presents a coal classification model with visible and infrared spectroscopy combined with deep learning network. First, the LRF network combined with ELM algorithm is used to construct a coal recognition model. To improve the recognition accuracy, this paper proposes an improved coyote algorithm to optimize the ELM-LRF model.
results show that the ICELM-LRF model can effectively identify the types of coal through spectral data. Compared with CNN algorithm and PCA algorithm, LRF can extract the spectral characteristics of coal more effectively. This provides an effective feature extraction method for spectroscopy. LRF and ELM are combined to give play to their respective advantages and complement the shortcomings between the two algorithms to build a more complete model. Compared with traditional methods, the combination of spectral technology and ICELM-LRF algorithm can greatly improve the economy, speed, accuracy, and other aspects and has a higher application value.

■ AUTHOR INFORMATION

Corresponding Author
Dong Xiao — College 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; Phone: +86-0-13840280858; Email: xiaodong@ise.neu.edu.cn; Fax: +86-024-8369525

Authors
Hongzong Li — College of Information Science and Engineering, Northeastern University, Shenyang 110819, China
Xiaoyu Sun — College of Resources and Civil Engineering, Northeastern University, Shenyang 110819, China

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.0c03069

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under grant 1674063, grant 52074064, grant 61203214, grant 61773105, and grant 61733003, in part by the Fundamental Research Funds for the Central Universities under grant N182401001; in part by the National Training Program of Innovation and Entrepreneurship for Undergraduates under grant 201910145206; in part by the Fundamental Research Funds for the Central Universities, China, under grant N180404012, grant N182608003, and grant N2001002; and in part by the Fundamental Research Funds for Liaoning Natural Science Foundation, China, under grant 2019-MS-120.

■ REFERENCES

(1) Freidina, E. V.; Botvinnik, A. A.; Dvornikova, A. N. Basic principles of coal classification by useful quality. J. Min. Sci. 2011, 47, 593–605.
(2) Xiao, D.; Li, H.; Jiang, G., Wuhan, Hubei, China; 2019; p 4.
(3) Wang, J. Application of grey clustering analysis based on projection pursuit in coal sample identification. Cem. eng. 2019, 24.
(4) Wang, Y. S.; Yang, M.; Luo, Z. Y.; Yang, W.; Li, G.; Hu, R. F. Rapid classification method of coal based on belief learning machine and near infrared spectroscopy. Guangxue Yu Guangfu Fenxi 2016, 36, 1685–1689.
(5) Wang, J. R.; Wen, H.; Zhao, Q. B. Application of Bayesian stepwise linear discriminant method based on SPSS in coal type identification. Xuanmei Jishu 2014.
(6) Pandit, Y. P.; Badhe, Y. P.; Sharma, B. K.; Tambe, S. S.; Kulkarni, B. D. Classification of Indian power coals using K-means clustering and Self Organizing Map neural network. Fuel 2011, 90, 339.
(7) Christiansen, J. V.; Feldthus, A.; Carlsten, L. Flash Pyrolysis of Coals. Temperature-dependent product distribution. J. Anal. Appl. Pyrolysis 1995, 32, 51.
(8) Rumelhart, D. E.; Hinton, G. E.; Williams, R. J. Learning representations by back-propagating errors. Nature 1986, 323, 533.
(9) Huang, G. B.; Lei, C.; Siew, C. K. Universal approximation using incremental constructive feedforward networks with random hidden nodes. IEEE Press Neural Networks 2006, 879.
(10) Huang, G.-B.; Chen, L. Convex incremental extreme learning machine. Neurocomputing 2007, 3056.
(11) Huang, G.-B. An Insight into Extreme Learning Machines: Random Neurons, Random Features and Kernels. Cognit. Comput. 2014, 6, 376.
(12) You, Z.-H.; Lei, Y.-K.; Zhu, L.; Xia, J.; Wang, B. Prediction of protein-protein interactions from amino acid sequences with ensemble extreme learning machines and principal component analysis. BMC Bioinf. 2013, 14, 1.
(13) Song, Y.; Crowcroft, J.; Zhang, J. Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine. J. Neurosci. Methods 2012, 210, 132.
(14) Zhang, Y.; Zhang, P. Optimization of nonlinear process based on sequential extreme learning machine. Chem. Eng. Sci. 2011, 66, 4702.
(15) Yang, Y.; Wang, Y.; Yuan, X. Bidirectional Extreme Learning Machine for Regression Problem and Its Learning Effectiveness. IEEE Trans. Neural Network Learn. Syst. 2012, 23, 1498.
(16) Zheng, Y.; Wang, J. Robust Model Predictive Control of Nonlinear Systems With Unmodeled Dynamics and Bounded Uncertainties Based on Neural Networks. IEEE Trans. Neural Networks Learn. Syst. 2014, 25, 457.
(17) Nizar, A. H.; Dong, Z. Y.; Wang, Y. Power utility nontechical loss analysis with extreme learning machine method. IEEE Trans. Power Syst. 2008, 23, 946.
(18) Minhas, R.; Mohammed, A. A.; Wu, Q. M. J. Incremental Learning in Human Action Recognition Based on Snappeets. IEEE Trans. Circuits Syst. Video Technol. 2012, 22, 1529.
(19) Zhou, Y.; Peng, J.; Chen, C. L. P. Extreme Learning Machine With Composite Kernels for Hyperspectral Image Classification. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 2015, 8, 2351.
(20) Huang, G. B.; Chen, L. Enhanced random search based incremental extreme learning machine. Neurocomputing 2008, 71, 3460.
(21) Huang, G. B.; Bai, Z.; Kasun, L. L. C.; Vong, C. M. Local Recceptive Fields Based Extreme Learning Machine. IEEE Comput. Intell. Mag. 2015, 10, 18.
(22) Pireezan, J.; Coelho, L. D. S. Coyote optimization algorithm: a new metaheuristic for global optimization problems. In 2018 IEEE Congress on Evolutionary Computation; 2018 1.
(23) Conner, M. M.; Ebinger, M. R.; Knowlton, F. F. Evaluating coyote management strategies using a spatially explicit, individual-based, socially structured population model. Ecol. Model. 2008, 219, 234.
(24) Pitt, W. C.; Box, P. W.; Knowlton, F. F. An individual-based model of canid populations: modelling territoriality and social structure. Ecol. Model. 2003, 166, 109.
(25) Mirjalali, S.; Mirjalali, S. M.; Lewis, A. Grey Wolf Optimizer. Adv. Eng. Software 2014, 69, 46.
(26) Silva, G. L. F.; Valente, T. L. A.; Silva, A. C.; De Paiva, A. C.; Gattass, M. Convolutional neural network-based PSO for lung nodule false positive reduction on CT images. Comput. Methods Programs Biomed. 2018, 162, 109.
(27) Melo, H.; Watada, J. Gaussian-PSO with fuzzy reasoning based on structural learning for training a Neural Network. Neurocomputing 2016, 172, 405.
(28) Jiang, H.; Zhu, W. Determination of Pear Internal Quality Attributes by Fourier Transform Near Infrared (FT-NIR) Spectroscopy and Multivariate Analysis. Food Anal. Methods 2013, 6, 569.
(29) Zheng, W.; Fu, X.; Ying, Y. Spectroscopy-based food classification with extreme learning machine. *Chemom. Intell. Lab. Syst.* 2014, 139, 42.

(30) Jolliffe, I. *Principal Component Analysis*; 2002.

(31) Liu, X.; Feng, X.; Liu, F.; Peng, J.; He, Y. Rapid Identification of Genetically Modified Maize Using Laser-Induced Breakdown Spectroscopy. *Food Bioprocess Technol.* 2018, 12, 347.

(32) Sahiner, B.; Chan, H.-P.; Petrick, N.; Wei, D.; Helvie, M. A.; Adler, D. D.; Goodsitt, M. M. Classification of mass and normal breast tissue: a convolution neural network classifier with spatial domain and texture images. *IEEE Trans. Med. Imaging* 1996, 15, 598.