

A novel algorithm combined X-ray fluorescence and Neural Network (XRF-NN) for coal ash content prediction: Algorithm design and performance evaluation

Chuanzhen Wang¹, Jintao Lv¹, Xinyi Wang¹, Andile Khumalo¹, Anghong Yu¹

¹ Anhui Engineering Research Center for Coal Clean Processing and Carbon Reduction, College of Material Science and Engineering, Anhui University of Science and Technology, Huainan 232001, China

Corresponding author: faxofking@cumt.edu.cn (Chuanzhen Wang)

Abstract: This study investigated a precise algorithm combining X-ray Fluorescence and Neural Network (XRF-NN) for predicting ash content. The 261 sets of XRF tests show that the 34 elements in the chosen Guqiao coal can be categorized as major, secondary, and tiny elements, whose cumulative ratios were ~95%, 4-5%, and <1%, respectively. Referring to the machine learning theory, the construction strategy of the Element-Ash dataset was determined viz. value determination → standardization → division → optimization of generalization ability. Then, the hyperparameters optimization displays that the XRF-NN model with 34 inputs and 1 output was suitable to predict coal ash content, where the activation function, loss function, and optimizer were ReLU, MAE, and Momentum, respectively. After iterative training, the new XRF-NN model provides the precise prediction of coal ash content with absolute errors between -2.0% and 2%. Moreover, the prediction accuracy rose from 57.69% to 100%, as the expected relative error increased from 1% to 5%. Furthermore, the comparisons between different prediction methods reveal that the minimum MRE of 1.22% can be obtained by XRF-NN with total elements, which was only half of those given by the conventional Multiple Linear Regression and Partial Least Squares Regression. Besides, the XRF-NN model presents the root mean squared error of 0.797%, a mean absolute error of 0.625%, and the coefficient of determination of 0.999, which were significantly superior to those calculated by Dual-Energy Gamma-ray Transmission, Ploy, RFR, XGBoost, and DNN model. The results of this study suggest the excellent performance of the new XRF-NN model in predicting ash content.

Keywords: ash content prediction, X-ray fluorescence, neural network, machine learning, mineral processing

1. Introduction

As the world's largest and most widely distributed primary source, coal plays a vital role in the world's primary energy (Zhou et al., 2021). According to the latest "Coal 2023" report published by the International Energy Agency (<https://www.iea.org/reports/coal-2023>), global coal consumption in 2023 exceeded a new record of more than 8 billion tons. At present, CO₂ generated from coal utilization is the main source of carbon emissions, accounting for 40% of the total global carbon emissions (Liu et al., 2023). Thus, the clean processing and efficient utilization of coal play important roles in the sustainable development of the economy and society. Ash content is the core evaluation index of coal quality, which is not only closely related to the clean utilization of coal but also the key to intelligent regulation in the coal selection process (Hower et al., 2022). Recently, the intelligent regulation method for coal preparation based on ash content has developed rapidly. That is the production process in a coal parathion plant can be optimized and regulated by using ash detection combined with process control technology, such as the intelligent control of heavy medium separation and flotation (Yang et al., 2013; Lu et al., 2019). The existing industrial practice shows that these techniques can effectively improve the efficiency of coal preparation, promote the development of clean and intelligent coal, and have broad application prospects. However, the industrialization promotion is highly limited by the unsatisfactory performance improvement of accuracy and real-time coal ash detection.

Therefore, the rapid and accurate detection of ash content is a technological bottleneck for the development of clean and intelligent coal washing. Recently, the ash content can be measured by either the offline method or the online technique in coal preparation plants. Firstly, the offline slow-ashing method, referring to Solid mineral fuels determination of ash (ISO 1171:2010), provides high-accuracy ash measurements. However, due to the complexity of the operation, more than 2 hours are expended for each test, leading to significant time delays in production process control (Begum et al., 2020; Bai et al., 2021). Secondly, online detection methods have significantly advanced to provide timely ash content based on element detection. Typically, atomic radiation-based radioactivity detection technologies, such as Dual-Energy Gamma-ray Transmission (DEGT) (Khoonthiwong et al., 2022; Chen et al., 2023) or Prompt Gamma-ray Neutron Activation Analysis (PGNAA) (Borsaru et al., 2001; Gaft et al., 2008), have emerged for coal ash content by detecting the elements Si, Fe, Al, Ca, Mg and so on. However, these methods are limited in the coal preparation industry due to the dangers associated with radiation sources. Afterward, the non- or less-radioactive detection techniques, for instance, the Natural Gamma-Ray (NGR) (Zhang et al., 2018) elements analysis or the Laser-Induced Breakdown Spectroscopy (LIBS) (Ctvrtnickova et al., 2010; Rajavelu et al., 2021), have been gradually developed to provide a secure testing environment. However, because of the high susceptibility of NGR to environmental interference and the high cost of LIBS, their industrial applications are limited. Thirdly, the machine vision method (Wang et al., 2020) is also adopted with the rapid development of image recognition to predict ash content, which shows good development prospects. Yet, the technology is still in the experimental research stage.

Moreover, since ash is composed of coal residue after complete combustion, the value of ash content is inevitably closely related to the elemental composition of coal. Therefore, the technique to accurately measure the composition of various elements in coal is significant for ash content prediction. X-ray fluorescence (XRF) spectroscopy, as a non-radioactive detection technique, utilizes secondary X-rays (X-ray fluorescence) to analyze the composition elements of a sample. In the XRF test, the content of elements is determined by the spectral characteristics of different secondary X-rays, allowing for accurate measurements of almost all elements (Rebiere et al., 2019). At present, XRF has been widely used in mineral composition analysis. Yang et al., (2019) quantified the alkali species and other components in the coal ash for evaluating the ash slagging and fouling tendency. Meanwhile, Li et al. (2020) et al. proposed an XRF-assisted LIBS analysis method, which can effectively analyze combustible organic light elements as well as highly stable inorganic ash-forming elements. Later, Shen et al. (2022) et al., used Energy-Dispersive X-ray Fluorescence (EDXRF) spectroscopy for the quantitative analysis of Cr, As, and Se in bituminous coal. Recently, contemporaneous works by our group (Wang et al., 2023) and Wen et al. (2023) reveal the well-size potential of XRF in predicting ash content.

Generally, these prior studies indicate that XRF relative to the other ash detection methods can provide a safer environment, lower costs, and faster detection. However, a model for establishing the precise mapping from elements to ash content is urgently needed. Thus, a combination of the XRF technique and neural network algorithm has been developed, and the objectives of the current study are as follows:

- (1) Elaborate on the details of slow ash detection, XRF test, and the neural network principle for coal ash detection.
- (2) Comprehensively analyze the characteristics of coal quality, including ash changes and element distribution, and construct the element-ash dataset for neural network algorithms based on machine learning ideas.
- (3) Systematically optimize the hyperparameters, including neurons, hidden layers, activation function, loss function, and optimizer, to create the precise XRF-NN algorithm for ash content prediction.
- (4) Objectively evaluate the performance of this XRF-NN model based on ash prediction results, comparative analysis, and its potential for industrial application.

2. Materials and methods

2.1. Materials

In this research, coal samples were selected from the Guqiao coal mine in Anhui Province, China. To ensure complete representation, the sampling period spanned the entire year of 2021. More than 20

samples were collected every month, resulting in a total of 261 samples. The samples included raw coal, clean coal, middlings, and gangue from the industrial production process. Subsequently, all the samples were prepared by sieving, dividing, drying, and grinding for the subsequent detection of ash content and X-ray fluorescence. The sampling and preparation of coal were conducted following the method for manual sampling of commercial coal (Chinese standard GB 475-2008) and the method for the preparation of coal samples (Chinese standard GB 474-2008), respectively, which were based on the Hard coal and coke - Manual sampling (ISO 18282:2006) standard. For the slow ash detection procedure, the proximate analysis of coal (Chinese standard GB 212-2008) based on the Solid mineral fuels - Determination of ash (ISO 1171-1997) was employed. The procedure was carried out as follows:

- Step 1.** Weigh the sample. The coal samples were weighed using an electronic balance in a dish, as shown in Fig. 1(a). The weight m_0 (g) of each sample before combustion was set to 1 g with an accuracy of 0.0002 g. Each sample was evenly spread in the ashtray to ensure full combustion.
- Step 2.** Burn. The dish with the sample was placed into the muffle furnace with a temperature of 85°C. The furnace temperature was then slowly raised to 500°C within 35-40 min and kept at this temperature for 30 min. Subsequently, the temperature was raised to 815±10°C and maintained for about 1 hour to ensure that the coal sample was fully burned.
- Step 3.** Cool and weigh. The ashtray was taken out from the furnace and allowed to cool in the air for 5 min, then moved into a dryer to cool to room temperature. The dried sample, shown in Fig. 1(b), was weighed as the coal sample after combustion, denoted as m_1 (g).
- Step 4.** Repeatability test and ash calculation. To ensure accuracy, Steps 1 to 3 were repeated three times for each sample, and the ash content was considered as the average value of the three repeated tests. The ash content A_d (%) was calculated using Eq. 1.

$$A_d = \frac{m_1}{m_0} \times 100\% \quad (1)$$

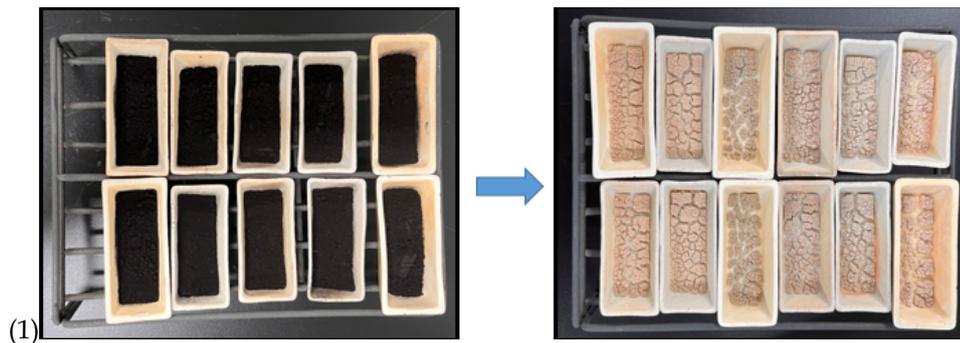


Fig. 1. Coal sample (a) before and (b) after combustion

2.2. Methods

2.2.1. X-ray fluorescence (XRF) analysis

The preparation process of the coal sample in this investigation is shown in Fig. 2, which includes (a) tablet press, (b) sample tablet, and (c) XRF spectrometer (Type: Bruker AXS S8-TIGER, Germany). In detail, 100 g of coal sample was ground in a mortar until its size was less than 0.074 mm. Then, the surface of the press mold with an outer and inner diameter of 40 mm and 30 mm, respectively, was cleaned with anhydrous alcohol. Afterward, the coal sample was fed into the press, as shown in Fig. 2(a), and pressed for 20 seconds under a pressure of 20 tons to obtain a coal tablet, as exhibited in Fig. 2(b). Boric acid, indicated by the white circle in Fig. 2(b), was used to fix the coal tablet with a thickness of 10~15 mm. Finally, the coal sample tablet could be tested using the S8-TIGER type XRF spectrometer, as shown in Fig. 2(c), at the China University of Mining and Technology located in Xuzhou, China.

Table 1 presents the details of the S8-TIGER type XRF spectrometer adopted in this study. From this table, it can be seen that the testing instrument is equipped with different types of spectroscopic crystals, such as LiF200, PET, and XS series artificial multilayer films. The maximum power of the device can reach 4 Kw. The test elements range from Be to U, corresponding to atomic numbers 4 to 92, and the element ratio ranges from 0 to 100%. However, in this XRF spectrometer, the elements with atomic

numbers less than 10 (such as H, C, N O, etc.) were regarded as combustible elements and were calculated as CO₂. The other ash-forming elements (e.g., Si, Al, etc.) were given as their corresponding oxides (see Table 1 for details). This means that the S8-TIGER spectrometer can accurately obtain the content values of almost all elements in coal samples except for hydrogen.

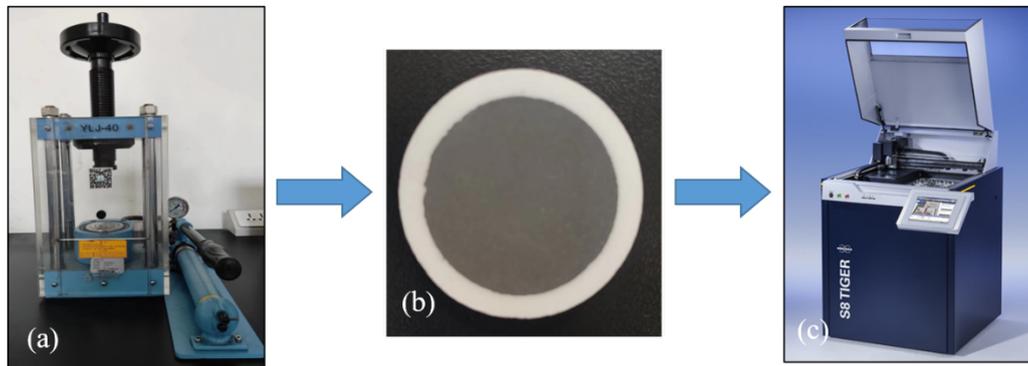


Fig. 2. Schematic diagram of sample preparation for fluorescence test: (a) tablet press (b) coal tablet (c) XRF spectrometer (Type S8-TIGER)

Table 1. Parameters of S8-TIGER type XRF spectrometer

Structure/Performance	Parameters/indicators
X-ray tubes	Rhodium (Rh), 75 μm Beryllium windows
Collimator	Autocollimator converter
Spectroscopic crystals	LiF 200, PET, XS series artificial multilayer film
Detector	Proportional flow gas counter and scintillation counter with MCA technology
Maximum power/current	4KW, 170mA
Accelerating voltage	0-60KV
Analyze element ranges	Be-U
Elemental quantitative range	0 to 100%

According to the principle of quantitative XRF spectroscopy, the S8-TIGER type spectrometer converts the measured characteristic XRF spectral intensities into element content determined by Eq. 2 (Vincze et al., 2002).

$$I_i = I_0 C_i S_i Abs_i \quad (2)$$

In Eq. 1, I_i is the detected characteristic X-ray intensity of element I, I_0 is the intensity of the primary X-ray beam, C_i is the concentration of element I within the particle, S_i the elemental sensitivity value of element i , and Abs_i is a term describing the self-absorption of both incident beam and emitted characteristic X-ray line within the particle. Additionally, due to the different excitation potentials and chemical properties of each element, the test conditions should be optimized based on the specific sample and test purpose. All of the aforementioned conditions selection, parameter settings, and concentration analysis were realized using the S8-TIGER software packages SPECTRA plus and QUANT-EXPRESS.

2.2.2. X-ray fluorescence (XRF) analysis

With the development of modern data analysis and machine learning algorithms, the neural network model provides new technical means to handle large data arrays and complex computational models with multiple factors. It has demonstrated superior performance in effectively solving multivariate nonlinear regression problems with fast computational speed, high adaptability, and fault tolerance. Therefore, it has been widely used in various industries, especially in the coal industry. For instance, Bai et al. (2021) explored a coal density range recognition method for multiple images with different light sources and different incidence angles based on the AlexNet convolutional neural network

algorithm. Cao et al. (2022) proposed a deep neural network FNet, motivated by knowledge and data fusion, for coal blast safety prediction in coal mine safety production. Temeng et al. (2020) developed a novel class of advanced artificial neural networks known as brain-inspired emotional neural networks (BI-ENN) to predict Aop arising from blasting during mining operations.

2.2.2.1. Structure of network

The neural network algorithm was utilized in this study to build a relationship model between the elements and ash content. Fig. 3 shows the schematic diagram of the neural network, which consists of an input layer, a hidden layer(s), and an output layer. The adjacent layers are connected by weights and biases. The input layer and output layer are responsible for the input feature values and output label values, respectively. The number of hidden layer(s) typically reflects the complexity of the neural network model and should be set reasonably. The universal approximation theorem (Cybenko et al., 1989; Hornik, 1991) of neural networks states that a multilayer feedforward network containing enough hidden layer neurons can approximate any predetermined continuous function with any precision. Meanwhile, Goodfellow et al. (2016) considered that a feedforward network with a single layer is sufficient to represent any function, but the layer may need to be infeasibly large and might fail to learn and generalize correctly. Building upon these insights, Zhang et al. (2021) insisted that simple depth two neural networks already have perfect finite sample expressivity as long as the number of parameters exceeds the number of data points. Thus, based on these previous studies, a double hidden layer neural network was selected for model construction.

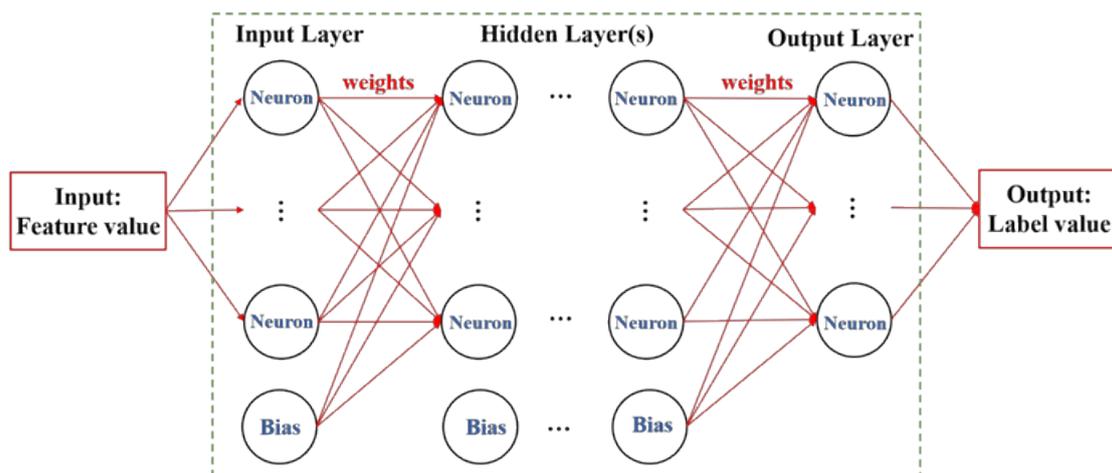


Fig. 3. Schematic diagram of neural network

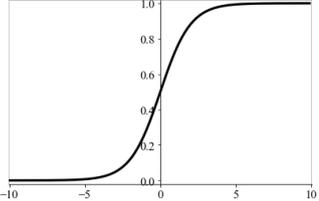
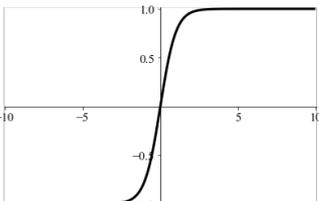
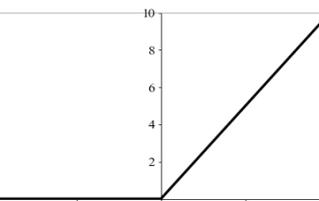
Additionally, in Fig. 3, the number of neurons in different layers was considered. The input layer and the output layer contain neurons equal to the number of features and labels, respectively. Namely, the number of neurons in the former was 34, and the number of neurons in the latter was 1. Moreover, determining the number of neurons in the hidden layers is a crucial part. Some related studies (Liu et al., 2019; Tran et al., 2020) showed that using too many or too few hidden neurons during the training process may lead to overfitting or underfitting in the prediction process. To determine a reasonable number of hidden layer neurons, previous studies (Sheela et al., 2013) provided relevant empirical formulas for reference. However, it is essential to obtain the optimal number of hidden layer neurons based on the actual situation and relevant data, which can be approximated through a series of experiments. The optimization process of hidden layer neurons is described in Section 3.2 in detail.

2.2.2.2. Activation function

As an essential part of the whole architecture of the neural network, the introduction of activation functions has greatly improved the model's ability to handle nonlinear relationships and made it possible to approximate arbitrarily complex functions. Moreover, activation functions have different degrees of impact on enhancing the model's robustness, alleviating the gradient vanishing issue, and

accelerating the convergence of the model. In this study, the most common activation functions were selected for the neural networks, including Sigmoid, Tanh, and ReLU. Table 2 presents the curves and formulas of the aforementioned activation functions.

Table 2. Three common activation functions

Activation function	Function curve	Function & derivative
Sigmoid		$\varphi(x) = \frac{1}{1+e^{-x}} \quad (3)$ $\varphi'(x) = \varphi(x)[1 - \varphi(x)] \quad (4)$
Tanh		$\varphi(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \quad (5)$ $\varphi'(x) = [1 - \varphi(x)]^2 \quad (6)$
ReLU		$\varphi(x) = \max(0, x) \quad (7)$ $\varphi'(x) = \begin{cases} 1, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (8)$

Sigmoid is a function that compresses the input feature data to the range of 0 to 1. It allows the data amplitude to be maintained without significant variations in the neural network. However, when the activation values are around 0 and 1, the gradient is almost 0, which may cause the phenomenon of premature gradient disappearance and result in a slow convergence rate (Glorot et al., 2010). Moreover, sigmoid outputs are not zero-centered, which can affect the operation of gradient descent and lead to slower convergence of the neural network (Goodfellow et al., 2016). The Tanh function scales the values of neurons to the range of -1 to 1 (Parkes et al., 1996). It effectively solves the non-zero-centered issue, but the problem of gradient disappearance still exists. The ReLU function propagates identically for the positive part and outputs zero for the negative part. Its main advantage is that it alleviates the vanishing gradient problem of the traditional activation functions (Zhou et al., 2021). However, ReLU is non-negative, and its mean activation degree is much greater than zero, which may not be conducive to the convergence of the network model (Clevert et al., 2015). In conclusion, each activation function has its advantages and disadvantages, and the best one to utilize depends on the performance of the model.

2.2.2.3. Training process of neural network

The training process of the model can be divided into two parts: forward transmission of information and backpropagation of error. The forward transmission process involves the input layer receiving the input signal, passing it to the hidden layer for information processing and transformation, and finally transmitting it to the output layer to obtain prediction results. When the predicted result significantly differs from the actual result, the model enters the backpropagation stage to correct the error. The error between the predicted value and the true value of the neural network model is often measured by a loss function. Generally, a smaller loss value indicates a closer match between the predicted value and the true value, leading to a better prediction effect of the model. The Mean Squared Error (MSE) and Mean Absolute Error (MAE) are used as the loss functions, as shown in Eqs. 9 and 10:

$$J_{MSE}(\mathcal{Y}, \hat{\mathcal{Y}}) = \frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2 \quad (9)$$

$$J_{MAE}(\mathcal{Y}, \hat{\mathcal{Y}}) = \frac{1}{m} \sum_{i=1}^m |y_i - \hat{y}_i| \quad (10)$$

where m is the total number of samples, y_i represents the true value of the i th sample and \hat{y}_i is regarded as the predicted value of the corresponding sample. Precisely, in this study, y_i is the actual ash content of the sample measured by the slow ash test, while \hat{y}_i is considered as the approximate ash predicted by the neural network model.

To find the global optimal solution and minimize the loss function, the model parameters are usually optimized using the gradient descent method, as shown in Eqs. 11 and 12:

$$\nabla J(\theta)_{\theta=\theta_j} = \frac{\partial J(\theta)}{\partial \theta_j} \quad (11)$$

$$\theta_j \leftarrow \theta_j - \eta \cdot \nabla J(\theta)_{\theta=\theta_j} \quad (12)$$

where θ denotes the model parameter vector, including the bias term θ_0 and feature weights $\theta_1, \theta_2, \dots, \theta_n$; $J(\theta)$ represents the loss function, and η is the learning rate, which adjusts the size of each iterative update. The learning rate is a crucial hyperparameter in network optimization and typically ranges from 0.0 to 1.0.

The neural network training process involves continuously optimizing the model parameters through a large number of iterative updates to approximate the real data. Eq. 12 briefly illustrates the process of updating the model parameter vector in one iteration using the gradient descent method. However, traditional gradient descent may converge to a local minimum and end the iteration prematurely in flatter regions as the gradient approaches zero. To address this, more robust optimizers, such as Momentum, AdaGrad, RMSprop, and Adam, have been proposed, which can adaptively adjust the learning rate. In this study, these optimizers were considered for network model training. More details about these optimizers can be found in reference (Bhakta et al., 2023). The orthogonal experimental approach was used to select the best combination of activation function, loss function, and optimizer, as these parameters interact with one another. The specific procedure is described in Section 3.2.

3. Results and discussion

As mentioned, the combination of X-ray Fluorescence detection and the Neural Network (XRF-NN) algorithm was used to detect the ash content of coal. In this section, the dataset was first constructed based on machine learning principles. Then, the optimization of hyperparameters for the XRF-NN model was investigated. Afterward, the prediction performance was systematically evaluated. The details are described as follows.

3.1. Dataset construction

3.1.1. Coal quality characteristics

As mentioned in Section 2.1, slow ash tests were conducted for Guqiao coal samples, where each sample was tested three times. Fig. 4 shows the distribution of ash content consisting of raw coal \circ , clean coal \circ , middlings \circ , and gangue \circ .

In Fig. 4, the horizontal coordinate was the number of each sample, which corresponds one-to-one with the sample numbers of the original dataset submitted with the paper. The left vertical coordinate indicates the ash content value of the sample, while the right vertical coordinate represents the cumulative number of samples within the corresponding ash range. For example, the accumulated number 38 (see the red dashed circle in Fig. 4) indicates the number of coal samples with an ash content of 45~50%.

In Fig. 4, it is clear that the ash content with a large span ranged from 20% to 90%. Actually, the larger the span, the higher the performance requirements for the stability and accuracy of the prediction model. In addition, a gap of ash content 55~80% was observed. This can be attributed to the actual industrial production in the Guqiao Mine. On one side, the ash content of raw coal mined from Guqiao Mine is approximately 50%; on the other side, the clean coal is used as different types of steam coal, so

its ash content is considered between 20% to 35%; the ash content of middlings and gangue were 35% and 80%, respectively.

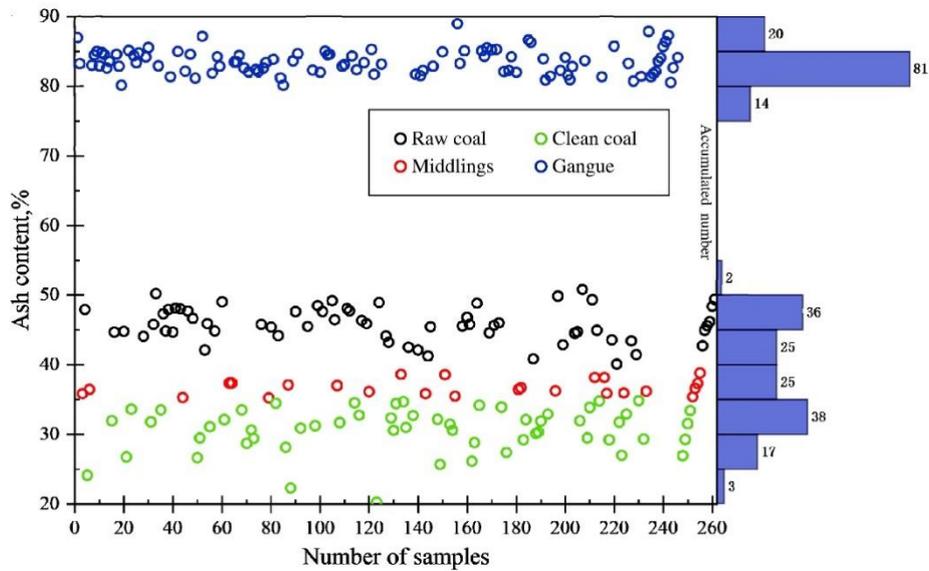


Fig. 4. Distribution of average ash content for Guqiao Coal Mining: raw coal \circ , clean coal \circ , middlings \circ , and gangue \circ

Furthermore, the 261 samples were sequentially detected by the Bruker AXS S8-TIGER XRF spectrometer. Due to article length limitations, several representative samples were chosen to describe the fluorescence characteristics (more data can be downloaded from the dataset submitted with the paper). Fig. 5 shows the Ka XRF spectra of coals with different ash contents: (a) 25.72%, (b) 45.45%, (c) 50.84%, and (d) 76.38%, corresponding to clean coal, middlings, raw coal, and gangue, respectively. In these Fig.s, the horizontal coordinate indicates the energy of electron leaping between different orbits of radiation, measured in KeV (Kiloelectron Volts). Elements exhibit identical positions of electron leap energy. For instance, the ash content increased from Fig. 5 (a) to (d), yet the electron leap energy of element Al remains at approximately 1.49 KeV (see the red circular dashed box). The vertical coordinate, KCps (kilocycles per second), represents the dimensionless count and stands for the statistical number corresponding to where the element produces an electron leap under XRF irradiation. The higher the peak, the higher the element content, and vice versa. Moreover, it is clear from Fig. 5 that the coal samples with different ash contents display analogous elemental composition, including Al, Si, S, Cl, K, etc. However, the content of each element varies with the ash content, indicating that the metal element content increases with the rising ash content. For example, in Fig. 5(a) to (d), the KCps of element Al increases from around 250 to 420, signifying that the content of Al in each coal sample is consistently increasing. It is worth noting that the inherent peak of Rh (shown in the red dashed box) can be observed in these Fig.s, which is attributed to the Rh-ray tube used in the aforementioned S8-TIGER spectrometer. However, these Rh peaks can be effectively filtered out to eliminate any impact on XRF test results.

Based on the previous XRF spectra, the percentages of elements in all coal samples were calculated using Eq. 2 (see details in Section 2.2). Fig. 6 shows the percentage of element content in the four different ash samples mentioned before. The element ratio of all 261 samples in this study can be observed in the submitted dataset. It is noted that the "CO₂" represents all combustible elements including H, C, N, and O (see Section 2.2 for details), while the remaining Si, Al to Pb indicate the ash-forming elements. It is clear from Fig. 6 that as the ash content increased, the percentage of combustible elements (CO₂) significantly decreased from 65.8% to 15.9%, while the ash-forming elements showed the opposite trend. This reveals that the ash content of coal is directly proportional to the content of ash-forming elements and inversely proportional to the combustible elements. Additionally, in terms of content percentage, H to Al was the major element with a cumulative ratio of \sim 95%, S to Mg was the secondary element with a cumulative ratio of 4-5%, and P to Pb was the tiny element with the cumulative ratio less than 1%.

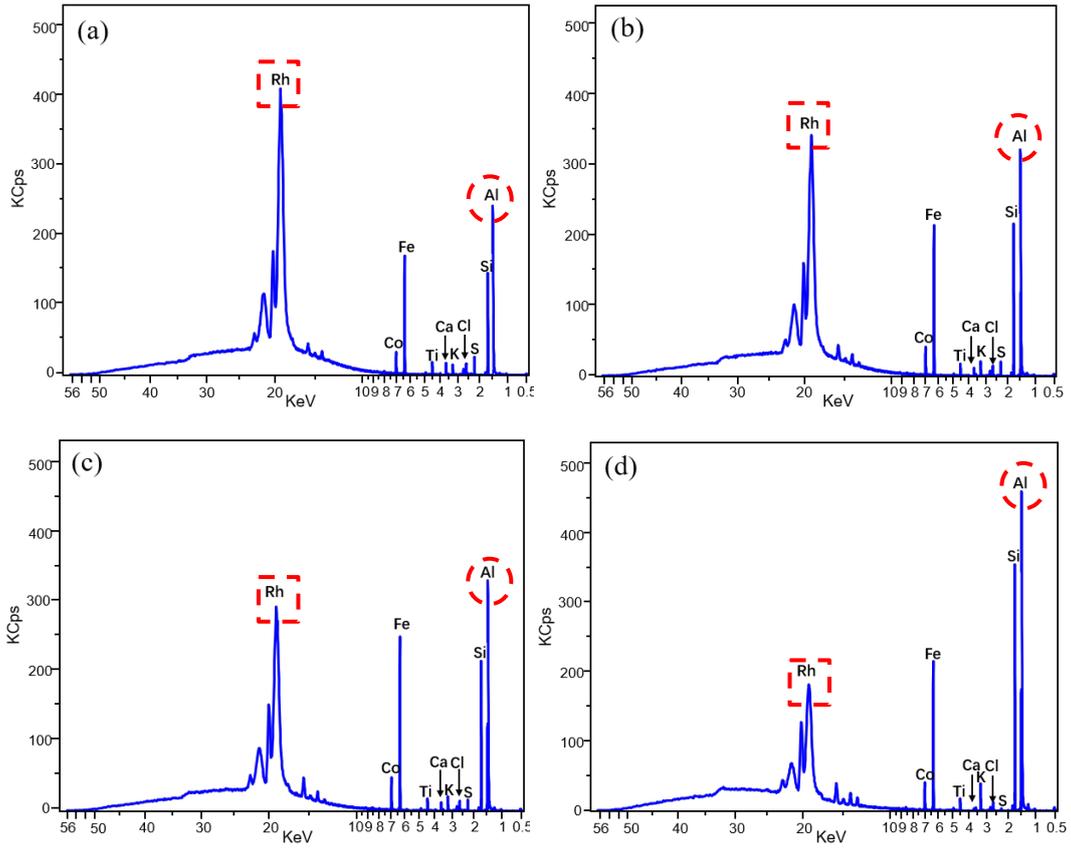


Fig. 5. XRF spectra of coals with different ash contents: (a) 25.72% (clean coal), (b) 45.45% (middling), (c) 50.84% (raw coal), and (d) 85.09% (gangue)

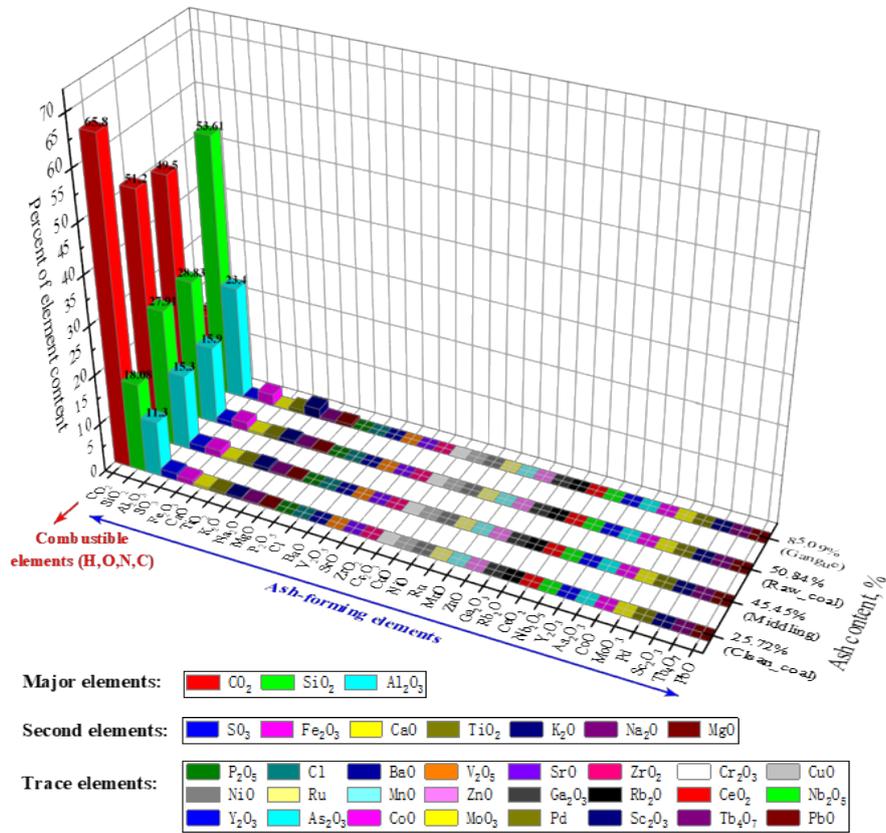


Fig. 6. Distribution of element content with different ash content in coal

3.1.2. Construction strategy of element-ash dataset

Based on these elements and ash content, the Element-Ash dataset was constructed referring to the supervised machine learning theory (Ouadah et al., 2022) step by step.

Firstly, the determination of dataset values. In machine learning, the input of the network model consists of feature parameters, while the output consists of label parameters. In this research, the ash content needed to be accurately predicted using the limited known information about the coal. Therefore, the 34 elements measured by the S8 TIGER wavelength-dispersive XRF spectrometer were used as the feature values, and the ash content obtained from the slow ash test was considered as the label value.

Secondly, the standardization. The combination of element proportion feature and ash content label was adopted to compose the original E-A dataset shown in Fig. 7(a). As can be seen from Fig. 7(a), the E-A dataset is a matrix with a size of 261×35. Here, 261 rows denote the number of coal samples in the present study. 35 columns stand for the sample properties, where column 1~34 and column 35 mean the element proportion feature and ash content label, respectively. Hence, the size of the E-A dataset can flexibly change based on the number of test samples and the number of detection elements. This feature was applied to the subsequent comparative exploration (see details in Section 3.3). Moreover, considering the influence of different magnitude orders on ash content (in Fig. 6), the content of each element x was standardized by the min-max normalization, given in Eq. 13, to normalized x' with the value range 0~1.

Furthermore, as observed in Fig. 6, the proportion of different elements in coal varies from around 60% to approximately 0.001%, with a difference of four orders of magnitude. To consider the influence of different magnitude orders on the ash content, the content of each element x was standardized using the min-max normalization, given by Eq. 13, to obtain the normalized value x' with a range of 0 to 1:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (13)$$

where x_{max} and x_{min} represent the maximum and minimum value of x in all the XRF analyses, respectively. Thus, the original E-A dataset can be converted to the standard E-A dataset shown in Fig. 7(b).

Thirdly, the division. The E-A dataset was divided into two parts, as indicated by the red dashed line in Fig. 8. Specifically, 90% of the entire dataset (235×35) was evenly extracted as the training set, while the remaining 10% (26×35) served as the test set. The training set was used for iterative training to build the neural network prediction model, mapping element proportions to ash content for coal. On the other hand, the test set was used to independently evaluate the effectiveness of the established model.

Finally, the optimization of generalization ability. In this study, 10-fold cross-validation was accepted to avoid overfitting the model. This method was first proposed by Mervyn Stone (Stone, 1974) and Seymour Geisser (Geisser, 1975). This algorithm is characterized by cutting data samples into smaller subsets, which is usually adopted in machine learning to improve the generalization performance of the model (Lu et al., 2019). Specifically, the dataset was initially split into 10 parts, as shown by the blue dashed line in Fig. 8. Next, nine of these parts were used for training, and the remaining part was used for validation. Then, this process was repeated for each part, yielding corresponding loss and accuracy results (details provided in Section 3.2). Lastly, the average of these 10 results, as shown by the purple dashed line in Fig. 8, was used to estimate the model's accuracy.

3.2. Hyperparameters optimization for XRF-NN model

As mentioned earlier, hyperparameters play a crucial role in constructing the XRF-NN model, and optimizations have been conducted for these hyperparameters. Fig. 9 illustrates the neural network prediction with different numbers of neurons in the hidden layers. In this case, the activation function, loss function, and optimizer were ReLU, MAE, and Momentum, respectively. The horizontal axis "14-7" indicates that the number of neurons in the first and second hidden layers were 14 and 7, respectively. The vertical axis represents the loss calculated using Eq. 10. It is evident from Fig. 9 that as the number of neurons in the hidden layers increases, the loss fluctuates between 0.6 and 1. Specifically, it first significantly decreases to a valley of ~0.7, then gradually increases to a peak of ~0.78, before decreasing

again. The minimum loss occurs at the combination of 30-15, indicating that this configuration is suitable for the hidden layers.

	CO2	SiO2	Al2O3	SO3	Fe2O3	CaO	TiO2	K2O	Na2O	MgO	...	Nb2O5	Y2O3	As2O3	CoO	MoO3	Pd	Sc2O3	Tb4O7	PbO	Ad	
0	14.3	53.88	25.28	0.238	2.500	0.103	0.907	1.670	0.508	0.472	...	0.002	0.0033	0.000	0.000	0.000	0.0	0.0	0.0	0.0	0.0	86.987453
1	17.9	51.42	24.19	0.383	2.221	0.305	0.822	1.450	0.635	0.477	...	0.000	0.0000	0.002	0.001	0.000	0.0	0.0	0.0	0.0	0.0	83.220241
2	61.1	21.19	12.76	1.400	1.222	0.558	0.530	0.546	0.320	0.216	...	0.000	0.0000	0.000	0.000	0.000	0.0	0.0	0.0	0.0	0.0	35.849932
3	48.2	30.06	16.39	1.140	1.553	0.513	0.637	0.738	0.326	0.278	...	0.001	0.0030	0.000	0.000	0.000	0.0	0.0	0.0	0.0	0.0	47.966322
4	65.7	18.17	11.40	1.560	1.069	0.517	0.523	0.422	0.280	0.190	...	0.000	0.0000	0.000	0.001	0.006	0.0	0.0	0.0	0.0	0.0	24.141610
...
256	53.1	26.30	15.36	1.260	1.458	0.415	0.586	0.727	0.363	0.249	...	0.000	0.0000	0.000	0.000	0.000	0.0	0.0	0.0	0.0	0.0	44.949360
257	51.5	27.65	15.29	1.300	1.615	0.414	0.579	0.833	0.344	0.302	...	0.000	0.0000	0.000	0.000	0.000	0.0	0.0	0.0	0.0	0.0	45.602562
258	51.6	27.76	15.57	1.230	1.382	0.470	0.582	0.664	0.381	0.256	...	0.000	0.0000	0.000	0.000	0.000	0.0	0.0	0.0	0.0	0.0	46.219135
259	49.1	29.40	16.03	1.300	1.573	0.365	0.596	0.855	0.310	0.296	...	0.001	0.0030	0.000	0.001	0.000	0.0	0.0	0.0	0.0	0.0	48.396152
260	48.8	29.78	16.28	1.120	1.448	0.434	0.594	0.746	0.369	0.280	...	0.000	0.0000	0.000	0.000	0.000	0.0	0.0	0.0	0.0	0.0	49.438834

261 rows × 35 columns

(a) Original E-A dataset

	CO2	SiO2	Al2O3	SO3	Fe2O3	CaO	TiO2	K2O	Na2O	MgO	...	Nb2O5	Y2O3	As2O3	CoO	MoO3	Pd	Sc2O3	Tb4O7	PbO	Ad	
0	0.0729	0.9125	0.9861	0.0390	0.2478	0.0379	0.9422	0.6854	0.6164	0.5052	...	0.6667	0.8049	0.0000	0.0000	0.0000	0.0	0.0	0.0	0.0	0.0	86.987453
1	0.1300	0.8579	0.9139	0.1177	0.2030	0.1913	0.7602	0.5865	0.8980	0.5126	...	0.0000	0.0000	0.2597	0.3333	0.0000	0.0	0.0	0.0	0.0	0.0	83.220241
2	0.8158	0.1879	0.1564	0.6692	0.0426	0.3834	0.1349	0.1802	0.1996	0.1270	...	0.0000	0.0000	0.0000	0.0000	0.0000	0.0	0.0	0.0	0.0	0.0	35.849932
3	0.6110	0.3845	0.3970	0.5282	0.0957	0.3492	0.3640	0.2665	0.2129	0.2186	...	0.3333	0.7317	0.0000	0.0000	0.0000	0.0	0.0	0.0	0.0	0.0	47.966322
4	0.8889	0.1210	0.0663	0.7560	0.0180	0.3523	0.1199	0.1245	0.1109	0.0886	...	0.0000	0.0000	0.0000	0.3333	0.7595	0.0	0.0	0.0	0.0	0.0	24.141610
...
256	0.6888	0.3012	0.3287	0.5933	0.0805	0.2748	0.2548	0.2616	0.2949	0.1758	...	0.0000	0.0000	0.0000	0.0000	0.0000	0.0	0.0	0.0	0.0	0.0	44.949360
257	0.6634	0.3311	0.3241	0.6150	0.1057	0.2741	0.2398	0.3092	0.2528	0.2541	...	0.0000	0.0000	0.0000	0.0000	0.0000	0.0	0.0	0.0	0.0	0.0	45.602562
258	0.6650	0.3336	0.3426	0.5770	0.0683	0.3166	0.2463	0.2333	0.3348	0.1861	...	0.0000	0.0000	0.0000	0.0000	0.0000	0.0	0.0	0.0	0.0	0.0	46.219135
259	0.6253	0.3699	0.3731	0.6150	0.0989	0.2368	0.2762	0.3191	0.1774	0.2452	...	0.3333	0.7317	0.0000	0.3333	0.0000	0.0	0.0	0.0	0.0	0.0	48.396152
260	0.6206	0.3783	0.3897	0.5174	0.0789	0.2892	0.2719	0.2701	0.3082	0.2216	...	0.0000	0.0000	0.0000	0.0000	0.0000	0.0	0.0	0.0	0.0	0.0	49.438834

261 rows × 35 columns

(b) Standard E-A dataset

Fig. 7. Structure diagram for the Element-Ash dataset (E-A) dataset

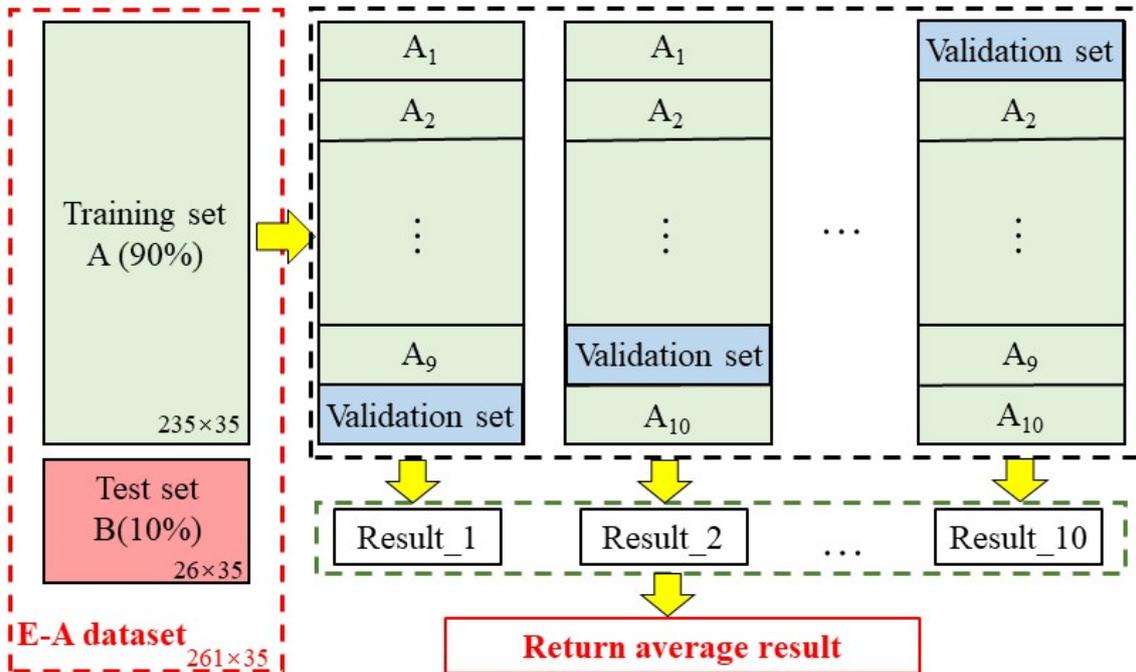


Fig. 8. Schematic diagram of construction strategy for Element-Ash (E-A) dataset

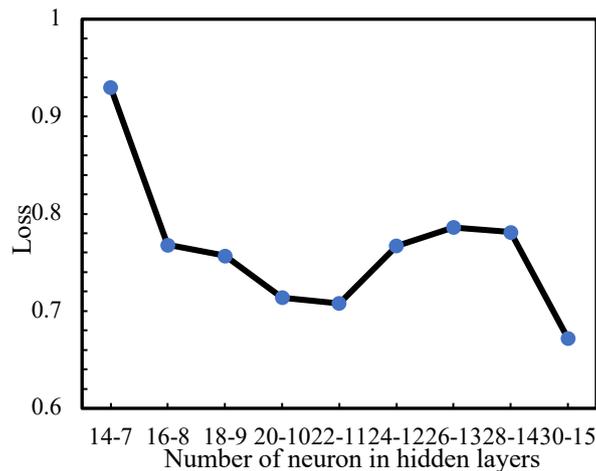


Fig. 9. Evaluation of different numbers of neurons in hidden layers

In this study, an orthogonal design table L16(34) was used to optimize the permutation of three factors: activation function, loss function, and optimizer. Table 3 presents the 16 sets of orthogonal experiments conducted in this study. The activation function (factor A) includes Sigmoid (level 1), Tanh (level 2), and ReLU (level 3); the loss function (factor B) consists of mean squared error (MSE, level 1) and mean absolute error (MAE, level 2); the optimizer (factor C) includes Momentum (level 1), AdaGrad (level 2), RMSprop (level 3), and Adam (level 4). The training loss (indicator M) and the test loss (indicator N) were selected as the evaluation indexes in the orthogonal experiment. The training loss reflects the fitting ability of the neural network model, while the test loss represents the generalization ability.

In Table 3, K_i represents the sum of the results of the experiments with level i ($i=1,2,3,4$ in this study) in any column. For example, considering the activation function (factor A), K_1 is the sum of the experimental results corresponding to Sigmoid (the 1st, 2nd, 3rd, 4th, and 14th experiments). Under the condition of the training loss (indicator M), $K_1 = 1.62 + 10.74 + 43.40 + 0.88 + 1.51 = 58.15$. Similarly, for factors B and C, K_1 is 196.28 and 4.18, respectively. K_i indicates the arithmetic mean of the experimental results with level i in any factor, which is the ratio of K_i to the occurrence number of each level in any column. For factor A, $k_1 = 58.15/5 = 11.63$ (indicator M), while the k_1 values corresponding to factors B and C are 28.04 and 1.05, respectively. R represents the extreme difference, which is the maximum value of K_i minus the minimum value of a factor. For factor A, its corresponding extreme difference is $R = 30.97 - 0.85 = 30.12$ (indicator M).

Normally, the larger the extreme difference, the greater the effect of the change in the column of factors on the experimental index. In this experiment, since $R_C > R_A > R_B$, the primary and secondary factors are as follows: C (optimizer), A (activation function), and B (loss function). This demonstrates that among these three factors, the selection of the optimizer has the greatest effect on the loss value, while the choice of the loss function has the least effect.

The loss function was used as the performance evaluation in this study, the smaller loss reflected the better performance. Thus, the level corresponding to the smallest value of k_i in each factor should be selected first. Due to the existence of the following results for both of the indicators M, N:

- Factor C: $k_1 > k_4 > k_3 > k_2$
- Factor A: $k_3 > k_2 > k_1$
- Factor B: $k_2 > k_1$

Therefore, C1A3B2 was accepted, i.e., the optimizer, activation function, and loss function were Momentum, ReLU, and MAE, respectively, in this study.

Table 4 exhibits the key hyperparameters of neural networks including not only the studied parameters but also the initial learning rate of 0.005 and batch size of 4. Note that the learning rate was adaptively adjusted during the training process. Further, TensorFlow2.0 software was adopted in this study for the neural network model, while all the training ran on the workstation conFig.d with 64 GB memory, Intel(R) Core (TM) i9-9900K CPU, and 16 processors.

Table 3. Orthogonal design table for parameter selection

Num	Activation function (Factor A)	Loss function (Factor B)	Optimizer (Factor C)	Training loss (Indicator M)	Test loss (Indicator N)
1	Sigmoid (level 1)	MSE (level 1)	Momentum (level 1)	1.62	2.16
2	Sigmoid	MAE (level 2)	AdaGrad (level 2)	10.74	10.79
3	Sigmoid	MSE	RMSprop (level 3)	43.40	44.49
4	Sigmoid	MAE	Adam (level 4)	0.88	1.85
5	Tanh (level 2)	MSE	AdaGrad	133.71	138.77
6	Tanh	MAE	Momentum	0.76	1.34
7	Tanh	MSE	Adam	14.99	18.33
8	Tanh	MAE	RMSprop	0.79	1.84
9	ReLU (level 3)	MSE	RMSprop	1.08	2.21
10	ReLU	MAE	Adam	0.66	1.06
11	ReLU	MSE	Momentum	0.99	1.83
12	ReLU	MAE	AdaGrad	1.08	0.97
13	ReLU	MSE	Adam	0.49	1.85
14	Sigmoid	MAE	RMSprop	1.51	2.06
15	Tanh	MAE	AdaGrad	4.58	4.80
16	ReLU	MAE	Momentum	0.81	1.05
Training loss (M)	K_1	58.15	196.28	4.18	
	K_2	154.83	21.81	150.11	
	K_3	5.11		46.78	
	K_4			17.02	
	k_1	11.63	28.04	1.05	
	k_2	30.97	2.42	37.53	
	k_3	0.85		11.70	
	k_4			4.26	
	R	30.12	25.62	36.48	
Test loss (N)	K_1	61.35	209.64	6.38	
	K_2	165.08	25.76	155.33	
	K_3	8.97		50.60	
	K_4			23.09	
	k_1	12.27	29.95	1.60	
	k_2	33.02	2.86	38.83	
	k_3	1.50		12.65	
	k_4			5.77	
	R	31.52	27.09	37.23	

Table 4. Key hyperparameters of neural network

Num	Hyperparameter	Value
1	Number of hidden layers	2
2	Number of neurons from input layer to output layer	34-30-15-1
3	Activation function	ReLU
4	Loss function	MAE
5	Optimizer	Momentum
6	Initial learning rate	0.005
7	Batch size	4

3.3. Performance evaluation of XRF-NN

3.3.1. Prediction performance of XRF-NN model

Following the aforementioned dataset and hyperparameters optimization, the XRF-NN model was trained. Fig. 10 displays the loss distribution of the training set, validation set, and test set. It is evident from Fig. 10 that with the increasing number of training iterations, the loss of all the sets first rapidly decreases from a higher value (<100 iterations), then gradually stabilizes (100-165 iterations), and eventually converges (>165 iterations) to ~ 0.6 (see the black circle in Fig. 10). This indicates that the XRF-NN model has been trained to a stable state. It should be pointed out here that attention should be paid to avoiding overfitting of the model.

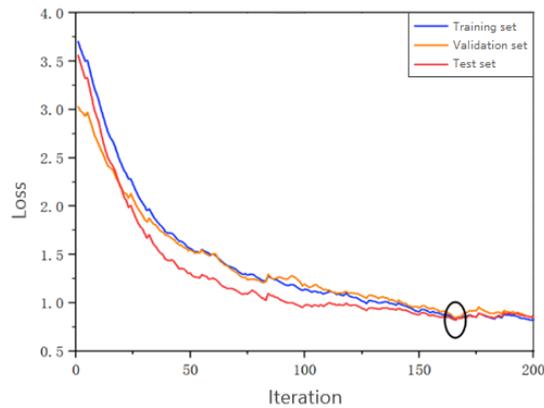


Fig. 10. Iteration process of XRF-NN model

The test set was utilized to evaluate the prediction performance of the trained XRF-NN model, as shown in Fig. 11(a) for the comparison of predicted and actual ash content, and Fig. 11(b) for the distribution of absolute error. In Fig. 11(a), the blue squares represent the actual values, while the red triangles indicate the predicted values and the error bars reveal the allowed relative error range of 5%. From Fig. 11(a), it is evident that all relative errors between the predicted and actual values are within the allowable error range of 5%. Furthermore, Fig. 11(b) illustrates the corresponding distribution of absolute errors. As seen from Fig. 11(b), all absolute errors of prediction fell within the range of -2.0% to 2.0%. More specifically, 54% of the predicted data is within the absolute error of $\pm 0.5\%$, while approximately 80% of the prediction data has an absolute error within 1%. These results indicate that the XRF-NN model can accurately predict the ash content.

Furthermore, in industrial applications, the predicted ash content does not need to be completely consistent with the actual value, as long as their relative error is within the expected error. Thus, the expected relative error (s) was considered as the threshold to control the prediction accuracy. This means that as the relative error calculated exceeds s (i.e. $> s$), the prediction calculated by the XRF-NN model was invalid, and vice versa. Namely, the lower the threshold s , the higher the performance of the

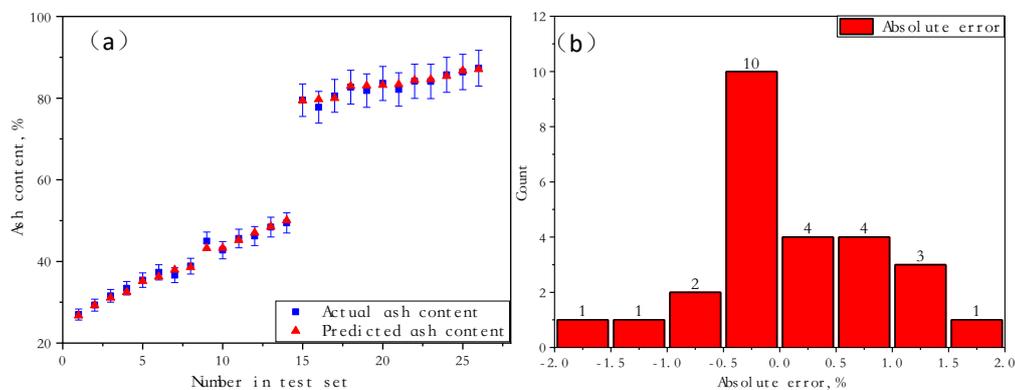


Fig. 11. Predictive performance evaluation: (a) comparison of predicted and actual ash content and (b) distribution of absolute error

XRF-NN model. Here, the prediction accuracy (Acc) was defined in Eq. 16 to count the proportion of predicted results within the expected error. n is the number of rows in the test data.

$$Acc = \frac{1}{n} \sum_{i=1}^n Count(Actual, Predicted) \times 100\% \quad (16)$$

$$Count(Actual, Predicted) = \begin{cases} 1 & \left| \frac{Actual - Predicted}{Actual} \right| \leq s \\ 1 & \left| \frac{Actual - Predicted}{Actual} \right| > s \end{cases} \quad (17)$$

Table 5 presents the prediction accuracy of the XRF-NN model under different expected relative errors on the test set. It can be observed that Acc continuously decreased with the decreasing expected relative error (s). Specifically, when s was higher than 4%, Acc was maintained at 100%. As s gradually reduced to 2%, Acc began to decline but remained at a high level (> 80%). However, when s was less than 1%, Acc significantly fell to about 57%. These changes indicate that although reducing the s was beneficial for improving the prediction performance of the XRF-NN model, the Acc was reduced simultaneously. Therefore, choosing a reasonable expected relative error is necessary for the application of the XRF-NN model.

Table 5. Prediction accuracy of neural network models with different accuracy evaluation functions

Expected relative error (s), %	Prediction Accuracy (Acc), %
<=5	100.00
<=4	100.00
<=3	92.31
<=2	80.77
<=1	57.69

3.3.2. Comparisons between different prediction methods

To further analyze the performance of the XRF-NN model, the comparison between different prediction methods was explored. In Table 6, Col. 1 shows the actual ash in the test data. Col. 2~4 shows the ash content predicted by the conventional Multiple Linear Regression (MLR) (Korkmaz, 2019), Partial Least Squares regression (PLS) (Xia, 2017) and XRF-NN, respectively. These three models were built based on the aforesaid training set, which includes all elements detected by the XRF test. Col. 5~7 display the ash calculated by XRF-NN based on main, secondary, and tiny elements (see Fig. 6 for details), respectively. It is noticed that Col. 4 can be regarded as the ash calculated by XRF-NN based on total elements namely that is equal to main, secondary, and tiny elements together. Col. 8~13 exhibit the relative error of ash corresponding to different methods.

From Col. 1~4 and Col. 8~10 in Table 6, the ash contents predicted by MLR, PLS, and XRF-NN models show similar trends, wherein the predicted values increased with the increasing actual ash content. However, XRF-NN displays the predicted results closer to the true values than MLR and PLS. Especially for lower ash content, the former gave higher prediction accuracy. For instance, when the actual ash content was 26.97%, its relative error predicted by XRF-NN was -0.97, which was significantly lower than that predicted by MLR (6.32%) and PLS (7.69%). Furthermore, the maximum relative errors of MLR and PLS were 9.11 and 7.69, respectively, while the maximum relative error of the XRF-NN model was 3.48. In addition, the Mean Relative Error (MRE) of XRF-NN was 1.22, which was nearly half that of MLR and PLS. These results indicate that the XRF-NN model exhibits better performance for ash content prediction than MLR and PLS models.

From Col. 1, Col. 4~7, and Col. 10~13, it is obvious that the predicted ash content of coal varied significantly with different elements. In detail, when the ash content was predicted by the XRF-NN model based on total, major, secondary, and trace elements, their MREs were 1.22%, 1.70%, 2.78%, and 7.26%, respectively. Among them, the minimum error corresponding to total elements is only 71.76%, 43.88%, and 16.80% of the other three cases. This means that the higher the element content used for modeling, the more conducive it is to improving model prediction accuracy. Compared to Col. 8~13, it can be seen that XRF-NN with all elements detected by the XRF test can provide the minimum average relative error.

Table 6. Prediction error comparison between different methods: (a) different models and (b) different elements

No.	Actual ash content, %	Predicted Ash content, %						Relative Error, %					
		(a) all elements with different models			(b) XRF-NN model with different elements			(a) all elements with different models			(b) XRF-NN model with different elements		
		MLR	PLS	XRF-NN	Main	Secondary	Tiny	MLR	PLS	XRF-NN	Main	Secondary	Tiny
1	2	3	4	5	6	7	8	9	10	11	12	13	
1	26.97	28.67	29.04	26.71	29.44	30.11	30.60	6.32	7.69	-0.97	9.17	11.65	13.48
2	29.31	31.98	31.42	29.20	30.59	31.82	36.29	9.11	7.21	-0.36	4.39	8.59	23.84
3	31.57	32.94	32.68	31.08	31.08	31.21	33.78	4.34	3.51	-1.54	-1.54	-1.12	7.01
4	33.43	34.48	33.85	32.48	32.02	34.98	34.21	3.16	1.27	-2.84	-4.21	4.66	2.35
5	35.43	36.60	36.36	35.19	34.50	36.43	36.66	3.30	2.63	-0.66	-2.60	2.83	3.49
6	37.34	37.37	36.92	36.26	36.57	37.19	32.33	0.09	-1.11	-2.90	-0.18	1.53	-11.75
7	36.63	37.72	38.61	37.91	36.63	35.83	38.85	2.97	5.40	3.48	-1.91	-4.03	4.05
8	38.84	38.97	39.12	38.55	38.05	38.96	37.57	0.33	0.70	-0.75	-2.03	0.31	-3.28
9	44.95	43.65	43.55	43.25	43.15	44.53	40.62	-2.88	-3.12	-3.78	0.94	4.18	-4.96
10	42.75	42.13	42.64	43.26	43.03	46.26	47.05	-1.43	-0.26	1.21	-4.27	2.91	4.67
11	45.60	45.29	45.49	45.20	45.84	46.69	53.36	-0.70	-0.24	-0.89	0.52	2.39	17.02
12	46.22	47.25	47.12	47.02	45.63	44.64	55.76	2.23	1.95	1.73	-1.28	-3.41	20.64
13	48.40	47.47	47.22	48.48	48.49	48.10	52.99	-1.92	-2.42	0.17	0.20	-0.62	9.48
14	49.44	48.65	48.22	50.04	48.86	48.00	59.81	-1.59	-2.47	1.22	-1.18	-2.91	20.98
15	79.50	79.27	79.67	79.41	79.15	79.43	80.49	-0.29	0.21	-0.11	1.80	2.16	3.53
16	77.75	77.15	76.99	79.72	77.81	79.08	82.06	-0.77	-0.97	2.53	-2.13	-0.54	3.22
17	80.55	80.16	79.92	80.00	81.05	82.51	74.08	-0.48	-0.78	-0.68	0.62	2.43	-8.03
18	82.69	84.55	83.64	82.98	82.03	82.15	84.93	2.25	1.15	0.35	0.26	0.40	3.80
19	81.83	82.50	82.90	83.03	82.29	85.17	83.19	0.82	1.32	1.48	0.21	3.71	1.30
20	83.59	83.59	83.92	83.18	82.89	80.62	84.62	0.00	0.40	-0.49	0.24	-2.50	2.33
21	82.12	84.07	83.93	83.36	83.51	83.08	80.34	2.38	2.20	1.51	-0.10	-0.61	-3.88
22	84.12	85.26	84.81	84.42	83.32	84.28	76.39	1.36	0.83	0.36	-0.91	0.23	-9.16
23	84.09	84.42	85.21	84.59	83.42	83.42	83.55	0.40	1.34	0.59	-0.82	-0.83	-0.67
24	85.70	85.63	86.41	85.37	85.13	82.58	86.99	-0.08	0.82	-0.38	-0.66	-3.65	1.50
25	86.40	87.16	87.36	86.71	84.71	87.18	88.87	0.87	1.10	0.36	-1.97	0.90	2.85
26	87.33	91.20	91.08	87.05	87.24	84.58	85.96	4.43	4.29	-0.33	-0.11	-3.15	-1.57
Mean relative error (MRE), %								2.10	2.13	1.22	1.70	2.78	7.26

3.3.3. Evaluation of application potential

As previously mentioned in the introduction, the DEGT technique is widely used in current industrial applications to detect the ash content of coal. In our previous study, we investigated the relational model between the metal elements and ash content based on this method, which has been successfully applied in the LinHuan coal preparation plant (HuaiBei, China) (Yu et al., 2022). Here, the comparison between XRF-NN and the gamma-ray method was considered. Fig. 13 illustrates the distribution of relative errors between the actual and predicted ash content, where the red squares and blue dots represent the values obtained by the gamma-ray method and the XRF-NN model, respectively. The black dashed line represents a relative error of 0%, indicating the coincidence of actual and predicted ash content. It can be observed from Fig. 12 that the relative error predicted by the gamma-ray method is within $\pm 5\%$, while that calculated by the XRF-NN model is within $\pm 4\%$. Additionally, the MRE of 2.29% obtained by the gamma-ray method is about twice as much as that (1.22% in Table 6) predicted by the XRF-NN model. These trends indicate that the XRF-NN model provides predicted values closer to the actual ash content compared to the gamma-ray method. In other words, the XRF-NN model shows better industrial application potential than the dual-energy gamma-ray method.

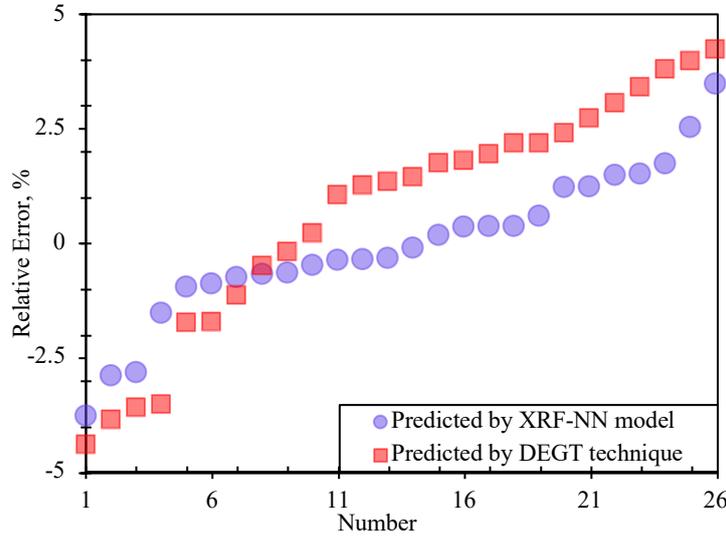


Fig. 12. Distribution of the relative error predicted by the XRF-NN model and DEGT technique

Although, Wen et al. (2023) also explored the coal ash content based on the XRF test. In their study, multiple Polynomial regression (Ploy), Random Forest Regression (RFR), Extreme Gradient Boosting (XGBoost), and Deep Neural Network regression (DNN) were adopted to predict ash content. The model's performances were evaluated by root mean squared error (RMSE), mean absolute error (MAR), and coefficient of determination (R^2). The equations for these indexes are given as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \bar{y}_i| \quad (14)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (15)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (16)$$

where n is the number of test samples, y_i is the actual value, \bar{y}_i is the value predicted by each model, \bar{y} and is the mean of observed values.

Fig. 13 exhibits the comparison of $RMSE$, MAR , and R^2 between the XRF-NN model and Wen's study. Here, the indexes of XRF-NN were obtained based on Col. 1 and Col. 4 in Table 6, while those of Poly, RFR, XGBoost, and DNN were from Wen's study (Wen et al., 2023). Note that the higher the MAE , MAR , and the lower R^2 , the better the model's performance. It is apparent from Fig. 13 that the $RMSE$, MAR , and R^2 obtained by the XRF-NN model were 0.797%, 0.625%, and 0.999, respectively, which are significantly better than those obtained by the other four models in the literature. This result can be

attributed to the fact that in Wen's research, the nine most important elements (Al, S, Si, Fe, Ca, Ti, K, Sr, and Zr) were supposed to have the greatest contribution to model performance. This view is similar to the classic X-ray modeling for ash content fitting based on the main elements. However, in the present study, as many elements as possible (up to 34 kinds) in coal were used for modelling, and 10-fold cross-validation (see Section 3.1.2 for details) was adopted for E-A dataset optimization.

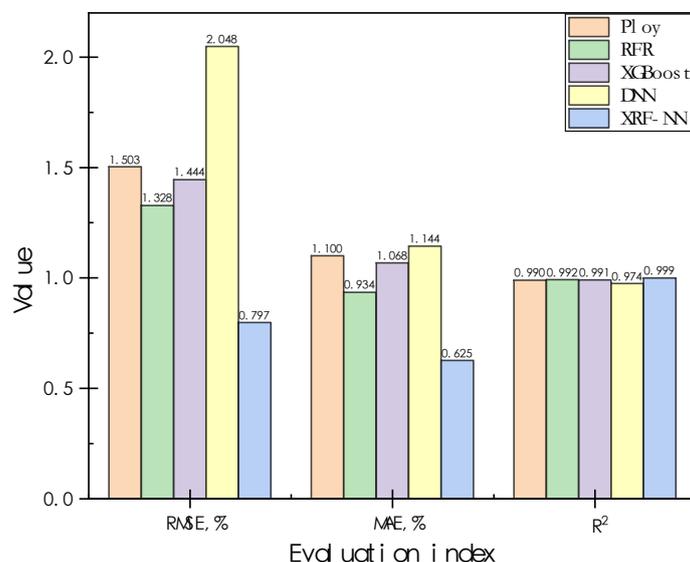


Fig. 13. Comparison of RMSE, MAE, and R² between the XRF-NN model and Wen's study (Ploy, RFR, XGBoost, and DNN model)

4. Conclusions

In the present research, the combination of X-ray Fluorescence detection and Neural Network algorithm (XRF-NN) were developed aiming to accurately predict the coal ash content. The key features of this new approach include dataset construction, hyperparameter optimization, and performance evaluation. Extensive experiments were carried out using the XRF-NN model with different ash content samples from the Guqiao coal mine in China. Based on the results obtained, the following conclusions can be drawn:

- (1) The XRF test revealed that the selected Guqiao coal consisted of combustible and ash-forming elements. It was discovered that the ash content was directly proportional to the proportion of ash-forming elements and inversely proportional to the combustible elements. Based on the percentage of element content, the elements in coal can be categorized that H to Al was the major element with a cumulative ratio of ~ 95%, S to Mg was the secondary element with a cumulative ratio of 4-5%, and P to Pb was the tiny elements with the cumulative ratio less than 1%.
- (2) The machine learning theory was used to build the element-ash (E-A) dataset step by step including the determination of dataset values, standardization, division, and optimization by the 10-fold cross-validation.
- (3) The orthogonal experiments for parameter optimization demonstrate that a dual hidden layer neural network with 34 inputs and 1 output can effectively establish the relationship between element content and ash content. ReLU, MAE, and momentum were chosen as the activation function, loss function, and optimizer, respectively.
- (4) After undergoing over 165 iterations, the XRF-NN model rapidly converged to a loss value of 0.6 on the training set. The prediction results on the test set display all absolute errors that fell within the range of -2.0% to 2.0%. Further, with the increased expected relative error from 1% to 5%, the prediction accuracy for ash content accordingly rose from 57.69% to 100%. However, as long as the expected relative error exceeded 3%, the prediction accuracy of the XRF-NN model was higher than 80%.
- (5) Compared to the classic MLR and PLS model, the MRE of ash content predicted by XRF-NN was half of the formers'. In addition, the effect of major, second, tiny, and total elements on prediction

accuracy shows that XRF-NN with a total of 34 elements can provide an MRE error of 1.22%. In addition, relative to the conventional DEGT technique and Wen's study (Ploy, RFR, XGBoost, and DNN), XRF-NN also displays significant improvement in predictive performance.

These conclusions demonstrate that the new XRF-NN model is more accurate and reliable in predicting ash content. The reduced error makes it a preferred choice for many applications where high precision is required. Additionally, the new method may also be faster and more efficient than traditional methods, making it a practical option for real-world application. In the future, industrial trials of this technology will be attempted.

Acknowledgments

This study was supported by the Youth Program of the National Natural Science Foundation of China (52004009); China Postdoctoral Science Foundation (2020M671837); Anhui Postdoctoral Foundation (2021A479); Anhui Province Coal Clean Processing and Carbon Reduction Engineering Research Center Foundation (CCCE-023001); The University Synergy Innovation Program of Anhui Province (GXXT-2023-104).

References

- BAI, F., FAN, M., YANG, H., DONG, L., 2021. *Rapid ash content determination method for coal particles through images captured under multiple ring light sources with various incident angles*. *Fuel*. 271, 117557.
- BEGUM, N., MAITI, A., CHAKRAVARTY, D., DAS, BS., 2020. *Reflectance spectroscopy based rapid determination of coal quality parameters*. *Fuel*. 280, 118676.
- BHAKTA, S., NANDI, U., SI, T., GHOSAL, SK., CHANGDAR, C., PAL, RK., 2023. *DiffMoment: an adaptive optimization technique for convolutional neural network*. *Appl. Intell.* 13, 16844-16858.
- BORSARU, M., BIGGS, M., NICHOLS, W., BOS, F., 2001. *The application of prompt-gamma neutron activation analysis to borehole logging for coal*. *Appl. Radiat. Isot.* 54, 335-343.
- CAO, A., LIU, Y., YANG, X., LI, S., LIU, Y., 2022. *FDNet: Knowledge and Data Fusion-Driven Deep Neural Network for Coal Burst Prediction*. *Sensors*. 22, 3088.
- CHEN, T-C., ILIYASU, AM., ALIZADEH, SM., SALAMA, AS., EFTEKHARI-ZADEH, E., HIROTA, K., 2023. *The use of artificial intelligence and time characteristics in the optimization of the structure of the volumetric percentage detection system independent of the scale value inside the pipe*. *Applied Artificial Intelligence*. 37, 2166225
- CHENYANG, Z., XIBO, L., YUEMIN, Z., XULIANG, Y., YANJIAO, L., LIANG, D., CHENLONG, D., ZHONGHAO, R., 2021. *Recent progress and potential challenges in coal upgrading via gravity dry separation technologies*. *Fuel*. 305, 121430.
- CTVRTNICKOVA, T., MATEO, MP., YAÑEZ, A., NICOLAS, G., 2010. *Laser Induced Breakdown Spectroscopy application for ash characterisation for a coal fired power plant*. *Spectrochim. Acta, Part B*. 8, 734-737
- CYBENKO, G., 1989. *Approximation by superpositions of a sigmoidal function*. *Mathematics of Control, Signals and Systems*. 4, 303-314
- DJORK-ARNE, C., THOMAS, U., SEPP, H., 2015. *Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)*. arXiv - CS - Machine Learning. 1511, 07289.
- FEIYAN, B., MINQIANG, F., HONGLI, Y., LIANPING, D., 2021. *Fast recognition using convolutional neural network for the coal particle density range based on images captured under multiple light sources*. *Int. J. Min. Sci. Technol.* 31, 1053-1061.
- GAFT, M., DVIR, E., MODIANO, H., SCHONE, U., 2008. *Laser Induced Breakdown Spectroscopy machine for online ash analyses in coal*. *Spectrochim. Acta, Part B*. 63, 1177-1182.
- GAO, R., LI, J., WANG, S., ZHANG, Y., ZHANG, L., YE, Z., ZHU, Z., YIN, W., JIA, S., 2023. *Ultra-repeatability measurement of calorific value of coal by NIRS-XRF*. *Anal. Methods*. 13, 1674-1680.
- GEISSER, S. 1975., *The Predictive Sample Reuse Method with Applications*. *J. Am. Stat. Assoc.* 70, 320-328.
- GOODFELLOW, I., BENGIO, Y., COURVILLE, A., 2016. *Deep Learning*. MIT Press.
- HORNIK, K. 1991., *Approximation capabilities of multilayer feedforward networks*. *Neural Networks*. 2, 251-257.
- HOWER, JC., FINKELMAN, RB., EBLE, CF., ARNOLD, BJ., 2022. *Understanding coal quality and the critical importance of comprehensive coal analyses*. *Int. J. Coal Geol.* 263, 104120.
- KHOONTHIWONG, C., SAENKAEW, P., CHANKOW, N., 2020. *Determination of the ash content of coal samples by nuclear techniques with bismuth germanate detectors*. *Int. J. Coal Prep. Util.* 4, 1-10.

- KORKMAZ, M. 2019., *A study over the Formulation of the Parameters 5 or Less Independent Variables of Multiple Linear Regression*. J. Funct. Spaces. 2019, 1-14.
- LIU, J., LIU, X., LIU, C., LE, BT., XIAO, D., 2019. *Random Search Enhancement of Incremental Regularized Multiple Hidden Layers ELM*. IEEE Access. 7, 36866-36878.
- LIU, Z., DENG, Z., DAVIS, S., CIAIS, P., 2023. *Monitoring global carbon emissions in 2022*. Nature Reviews Earth & Environment. 4, 205-206.
- LU, H-J., ZOU, N., JACOBS, R., AFFLERBACH, B., LU, X-G., MORGAN, D., 2019. *Error assessment and optimal cross-validation approaches in machine learning applied to impurity diffusion*. Comput. Mater. Sci. 169, 109075-109075.
- OUADAH., ABDELFFETTAH., ZEMMOUCHI, G., LEILA., SALHI., NEDJMA., 2022. *Selecting an appropriate supervised machine learning algorithm for predictive maintenance*. The International Journal of Advanced Manufacturing Technology. 119, 1-25.
- PARKES, EJ., DUFFY, BR., 1996. *An automated tanh-function method for finding solitary wave solutions to non-linear evolution equations*. Comput. Phys. Commun. 98, 288-300.
- REBIERE, H., KERMAÏDIC, A., GHYSELINCK, C., BRENIER, C., 2018. *Inorganic analysis of falsified medical products using X-ray fluorescence spectroscopy and chemometrics*. Talanta. 195, 490-496.
- SHEELA, KG., DEEPA, SN., 2013. *Review on Methods to Fix Number of Hidden Neurons in Neural Networks*. Math. Probl. Eng. 2013, 1-11.
- SHEN, L., QIAO, E., LIU, L., XUE, C., LIU, B., MIN, F-F., 2022. *Utilization of coal slime: Coal and kaolinite separation by classification, forward and reverse flotation method*. Physicochem. Probl. Miner. Process. 3, 58.
- STONE, M. 1974., *Cross-Validatory Choice and Assessment of Statistical Predictions*. Journal of the Royal Statistical Society: Series B (Methodological). 2, 111-133
- TEMENG, VA., ZIGGAH, YY., ARTHUR, CK., 2020. *A novel artificial intelligent model for predicting air overpressure using brain inspired emotional neural network*. Int. J. Min. Sci. Technol. 5, 683-689
- TRAN, TTK., LEE, T., KIM, JS., 2020. *Increasing Neurons or Deepening Layers in Forecasting Maximum Temperature Time Series? Atmosphere*. 11, 1072-1072.
- VINCZE., SOMOGYI., OSAN., VEKEMANS., TOROK., JANSSENS., ADAMS., 2002. *Quantitative trace element analysis of individual fly ash particles by means of X-ray microfluorescence*. Anal. Chem. 74, 1128-1135.
- WANG, C., LV, J., LIU, H., WANG, G., YU, A., 2023. *Research of GM(1,N) dynamic network based on the metabolic algorithm optimization for ash fitting*. Journal of China Coal Society. 12, 4549-4558
- WANG, C., LV, W., 2023. *Intelligent control of heavy media separation*. Int. J. Global Energy Issues. 45, 86-100.
- WANG, G., HE, T., KUANG, Y-L., LIN, Z., 2020. *Optimization of soft-sensing model for ash content prediction of flotation tailings by image features tailings based on GA-SVMR*. Physicochem. Probl. Miner. Process. 4, 56.
- WEN, Z., LIU, H., ZHOU, M., LIU, C., ZHOU, C., 2023. *Explainable machine learning rapid approach to evaluate coal ash content based on X-ray fluorescence*. Fuel. 332, 125991.
- YANG, M., XIE, Q., WANG, X., DONG, H., ZHANG, H., LI, C., 2018. *Lowering ash slagging and fouling tendency of high-alkali coal by hydrothermal pretreatment*. Int. J. Min. Sci. Technol. 29, 521-525.
- YANG, X., ALDRICH, C., 2013. *Optimizing control of coal flotation by neuro-immune algorithm*. Int. J. Min. Sci. Technol. 23, 407-413.
- YU, A., LIU, H., WANG, C., LV, J., WANG, F., HE, S., WANG, L., 2022. *Online Ash Content Monitor by Automatic Composition Identification and Dynamic Parameter Adjustment Method in Multicoal Preparation*. Processes. 10, 1432.
- ZHANG, C., BENGIO, S., HARDT, M., RECHT, B., VINYALS, O., 2021. *Understanding deep learning (still) requires rethinking generalization*. Commun. ACM. 64, 107-115.
- ZHANG, N., LIU, C., 2018. *Radiation characteristics of natural gamma-ray from coal and gangue for recognition in top coal caving*. Sci. Rep. 8, 190
- ZHOU, Y., LI, D., HUO, S., KUNG, S-Y., 2020. *Shape autotuning activation function*. Expert Syst. Appl. 171, 114534.