International Core Journal of Engineering Volume 12 Issue 4, 2026
ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

Hyperspectral Response Characteristics and Quantitative Prediction of Moisture Content in Different Iron Ore Samples

Xinyao Li[1, †], Wanxin Li[2, †], Bai Yang[3, *]

[1] School of Civil Engineering, North China University of Science and Technology, Tangshan, Hebei 063299, China

[2] School of Science, North China University of Science and Technology, Tangshan, Hebei 063299, China

[3] School of Mining Engineering, North China University of Science and Technology, Tangshan, Hebei 063299, China

[†]These authors contributed equally to this work.

Abstract

The moisture content of iron ore is a key parameter affecting beneficiation efficiency, transport costs and subsequent smelting processes. Although the traditional drying and weighing method yields accurate results, it has drawbacks such as being time-consuming and causing sample damage. Accordingly, this study investigates methods for predicting the moisture content of iron ore using hyperspectral technology, using a total of nine representative iron ore samples (H1–H9) from Anshan-type magnetite in China, limonite in Australia and hematite in Brazil, covering the sub-types of limonite, magnetite and hematite. Under conditions where grade and particle size were kept consistent, different moisture content gradients were established, and spectral data in the 350–2400 nm wavelength range were acquired using a FieldSpec 4 field spectrometer. Outliers were removed using Pearson correlation analysis, followed by a comparison of 11 classical pre-processing methods and their 1–3-step cascaded combinations. The top 15 feature bands were selected using analysis of variance (ANOVA). Based on this, a mean fusion model combining partial least squares regression (PLSR) and random forest (RF) was constructed. The results indicate that moisture content is generally negatively correlated with spectral reflectance, with distinct absorption features appearing near 1440 nm and 1920 nm; the optimal pre-processing method combinations vary across different mineral samples; and the characteristic bands are primarily concentrated near wavelengths of 360 nm, 750 nm, 1140 nm, 1402 nm, 1880 nm and 2490 nm. The PLSR-RF fusion models for the H1, H2, H3, H6, H7 and H8 all achieved an R^2 of over 0.94 for the training set, over 0.83 for the test set, and an overall R^2 of over 0.91, with the H8 sample exhibiting the highest R^2 across the entire dataset (0.9747). Compared with existing studies, this method demonstrates a significant advantage in predictive accuracy, providing a viable solution for the rapid, non-destructive detection of moisture content in iron ore.

Keywords

Hyperspectral Technology; Moisture Content; Spectral Pre-Processing; Hybrid Predictive Model; Iron Ore.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

1. Introduction

The moisture content of iron ore is one of the key parameters affecting beneficiation efficiency, transport costs and subsequent smelting processes[1]. Traditional methods for determining moisture content typically employ the oven-drying and weighing method[2]; although this yields accurate results, it suffers from drawbacks such as being time-consuming, destructive to the sample, and unsuitable for online detection. Due to its rapid, non-destructive and in-situ detection capabilities, hyperspectral technology has demonstrated promising prospects for the quantitative analysis of mineral moisture in recent years[3]. However, spectral data from iron ore is susceptible to interference from factors such as particle scattering, baseline drift and instrument noise, and the faint absorption features related to moisture content are often masked in the raw spectra[4]. Furthermore, the spectral response mechanisms of moisture vary across different ore types-including limonite, magnetite and hematite[5]-as well as under different grades and particle size conditions[6]; consequently, a single pre-processing method or linear model struggles to meet the universal prediction requirements for diverse ore samples[7]. To address these issues, this study examines nine typical iron ore samples to systematically investigate the noise reduction and feature enhancement effects of various pre-processing methods and their cascaded combinations[8]. By combining analysis of variance (ANOVA) to screen for characteristic spectral bands[9] and integrating both a partial least squares (PLS) linear model[10] and a random forest non-linear model[11], the study aims to establish a high-precision, highly robust method for predicting iron ore moisture content, thereby providing technical support for rapid on-site detection.

2. Experiments and Methods

(1) Sample collection and processing

This study selected a total of nine representative iron ore samples (H1–H9), comprising Anshan-type magnetite from China, limonite from Australia and hematite from Brazil, as experimental materials.

Figure 1. Photographs of iron ore samples
(H2, H8, H6, H4, H9, H1, H5, H7, H3)

Throughout the experiments, both ore grade and particle size were kept constant: a systematic dilution method was employed, in which high-grade raw ore was uniformly mixed with high-purity silica in a predetermined mass ratio to control the grade; following comminution in a ball mill, the ore was classified into specific target particle sizes using standard sieves, in accordance with laboratory sieve specifications and particle sizes applicable to actual production[12]. The dried iron ore powder was sealed and stored in black, airtight bags.

Prior to the experiment, water was slowly added to the iron ore powder and thoroughly mixed; saturation measurements were taken once all free water had completely evaporated from the sample surface. Based on these results, different moisture content gradients were established for each ore sample to simulate varying humidity conditions, thereby investigating how the spectral characteristics of the iron ore samples vary with moisture content.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

Table 1. Parameters for each sample

Sample	Taste (%)	Granularity (mm)	Span (%)	Category	Levels
H1	63	35-60	0-48	Hematite	24
H2	60	100-150	0-50	Magnetite	26
H3	63	35-60	0-50	Hematite	30
H4	59	35-60	0-59	Limonite	34
H5	65	35-60	0-50	Limonite	29
H6	63	35-60	0-50	Limonite	25
H7	65	35-60	0-50	Hematite	29
H8	67	35-60	0-50	Limonite	29
H9	55	100-150	0-50	Magnetite	25

(2) Spectroscopic analysis of sample moisture content

Spectral testing of the moisture content in iron ore samples was carried out using a portable field spectrometer (FieldSpec 4) manufactured by Analytical Spectral Devices (USA), with a spectral range of 350–2500 nm. The experiments were conducted in a dark, light-shielded environment to minimise interference from external light sources. The prepared mineral powder samples were thoroughly shaken and then placed in a black light-shielding box. The moisture content of the samples was determined at regular intervals using the evaporation method to obtain samples with different moisture content levels. During the spectral measurement process, the spectrometer's lens was positioned at a 90° angle to the sample, and 20 spectral curves were collected repeatedly, with the average value taken as the raw data.

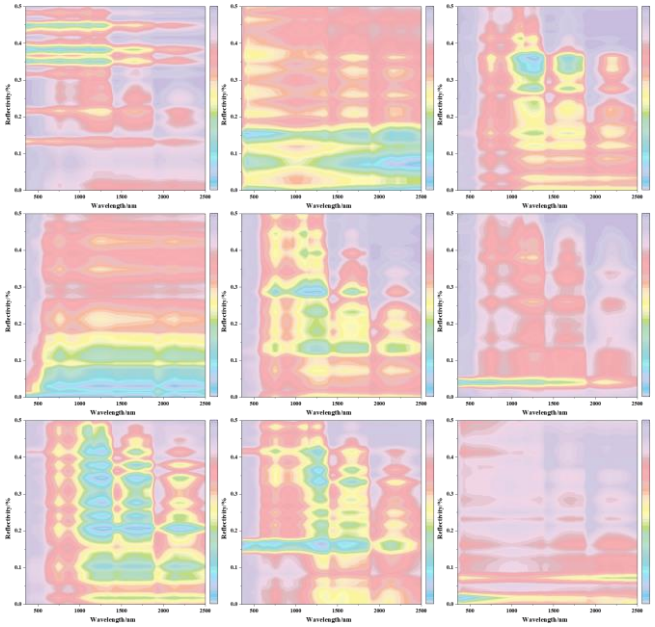

Figure 2. Initial results of spectral testing (H1–H9)

As the data in the 2401–2500 nm wavelength band exhibited significant noise and low stability[13], only spectral data from the 350–2400 nm wavelength band was selected for data analysis and modelling.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

(3) Pearson correlation analysis

Under controlled moisture conditions, some experimental data may exhibit spectral curve anomalies due to uneven sample preparation, such as excessively thick local water films, agglomeration of fine particles, or instrument errors[14]. Such data can interfere with model training and the analysis of interference mechanisms, leading to biased results. Consequently, this study employs Pearson's correlation coefficient to analyse the correlation between spectral curves of similar iron ores at different moisture contents[15], excluding experiments that fall outside the specified range. This purifies the dataset, enhances the accuracy of spectral analysis and modelling, and ensures the reliability and scientific validity of subsequent research conclusions.

$$\rho_{x,y} = \frac{\sum_{i=1}^{n}(x_i - x)\cdot(y_i - y)}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{1}$$

where n is the number of spectral data samples; x_i is the spectral reflectance of the sample with moisture content class i in band x; y_i is the spectral reflectance of the sample with moisture content class i in band y.

(4) Spectral data pre-processing

During spectral acquisition, factors such as the instrument's condition and ambient temperature may introduce noise, thereby affecting the stability and accuracy of the model. Consequently, it is usually necessary to pre-process the spectral data prior to model construction. This paper employs 11 classical pre-processing methods[16], allowing for cascaded combinations of 1 to 3 steps, including deviation standardisation, mean centring, Z-score standardisation, Pareto scaling, moving average smoothing, Savitzky-Golay convolution smoothing, multicollinearity correction, standard normal transformation, first-order derivative, detrending, and envelope removal, to eliminate noise and improve data quality.

Method 1 (Standardised deviation): Due to the significant variation in reflectance amplitude within hyperspectral data of iron ore, with reflectance values spanning a wide range across different wavelength bands and samples, there is a marked issue of dimensionality interference. This can cause the model to over-emphasise high-amplitude bands during training, thereby neglecting moisture-related features in low-amplitude bands and reducing the model's predictive accuracy. Therefore, a standardised data processing method based on standard deviation can be employed to map reflectance uniformly to a fixed range, thereby eliminating differences in dimensionality and amplitude.

$$x_{\mathrm{MMN}} = \frac{x_{i,k} - \min(x_i)}{\max(x_i) - \min(x_i) + \varepsilon} \tag{2}$$

In the equation, x_i,k represents the raw reflectance of the i-th sample in the k-th band; the numerator is the difference between the current reflectance and the minimum reflectance of that spectral line, whilst the denominator is the difference between the maximum and minimum reflectances of that spectral line; the last is a small constant that prevents the denominator from becoming zero; after normalisation, the reflectance is mapped to the interval [0,1], thereby eliminating dimensionality issues.

Method 2 (Mean centralisation): As raw spectral data are susceptible to factors such as instrument baseline shifts and overall fluctuations in illumination, an overall baseline shift may occur, causing the spectral curve to shift upwards or downwards as a whole. This masks the subtle spectral changes caused by moisture content, making it impossible to accurately extract features related to moisture

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

content. Therefore, mean-centring can be employed to eliminate the overall baseline shift on a sample-by-sample basis, thereby highlighting the relative spectral changes.

$$x_{\text{centered}} = x_{i,k} - \frac{1}{m}\sum_{i=1}^{m} x_{i,k} \tag{3}$$

After centring, the mean of each spectrum is set to 0, making it easier to detect subtle fluctuations caused by moisture content.

Method 3 (Z-score standardisation): As the mean reflectance values and levels of variability across different spectral bands in the raw spectral data vary considerably, bands with high variance tend to dominate during model training, obscuring moisture-related features in bands with low variance and resulting in poor model generalisation. Therefore, Z-score standardisation based on global statistical features can be employed to ensure that data from each band follows an approximate standard normal distribution, thereby balancing the weights of the various bands.

$$x_{\text{z-score}} = \frac{x_{i,k} - \overline{x}_k}{\sigma_i} \tag{4}$$

Among which,

$$\sigma_i = \sqrt{\frac{\sum_{i=1}^{m}(x_{i,k} - \overline{x}_k)^2}{m}} \tag{5}$$

Following standardisation, the mean value for data across all bands is 0 and the standard deviation is 1, effectively balancing the weights of the different bands and avoiding interference from bands with high variance that are not relevant.

Method 4 (Pareto scaling): As the spectral features resulting from the moisture content in certain types of iron ore are often rather weak, Pareto scaling can be employed to standardise the data by taking the square root; this method suppresses high-variance bands more gently, thereby balancing the needs of high-variance bands with those of low-variance, weak features.

$$x_{\text{pareto}} = \frac{x_{i,k} - \overline{x}_k}{\sqrt{\sigma_i}} \tag{6}$$

Method 5 (Moving average smoothing): During on-site experimental analysis, spectral data are susceptible to influences such as instrumental electronic noise and ambient stray light, resulting in high-frequency random fluctuations. This causes the spectral curves to appear rough, obscuring the characteristic absorption peaks associated with moisture content and thereby affecting the model's accuracy in identifying moisture content. Therefore, a moving average smoothing technique can be employed to rapidly suppress high-frequency noise by calculating local averages, thereby smoothing the spectral curves.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

$$x_{k,\text{smooth}} = \overline{x}_k = \frac{1}{2w+1}\sum_{i=-w}^{+w} x_{k+i} \tag{7}$$

In the equation, 2W+1 represents the width of the smoothing window, which contains an odd number of wavelength points; k is the central wavelength point within the window; and the wavelength points are replaced by the average of the measurements taken at w points before and after k.

Method 6 (Savitzky-Golay Convolutional smoothing): As the characteristic absorption peaks corresponding to the moisture content of certain types of iron ore are crucial for detection, Savitzky-Golay convolution smoothing can be employed to achieve peak-preserving smoothing through local polynomial fitting. This approach suppresses noise whilst fully preserving the spectral peak shapes and absorption band information, thereby preventing distortion of the characteristic features.

$$\sigma_i = \sqrt{\frac{\sum_{i=1}^{m}(x_{i,k} - \overline{x}_k)^2}{m}} \tag{8}$$

Among which,

$$H = \sum_{i=-w}^{+w} h_i \tag{9}$$

In the equation, H is the normalisation factor, and h_i is the smoothing coefficient; multiplying the measured value by the smoothing coefficient serves to minimise the impact of smoothing on useful information.

Method 7 (Multipath correction): As it is common for ore samples to exhibit uneven particle sizes and irregular distributions during iron ore production and crushing, this can cause scattering effects in the spectral signals, leading to spurious variations in spectral reflectance. These variations overlap with those caused by moisture content, making it impossible to distinguish between them accurately and thereby reducing the accuracy of model predictions. Therefore, multivariate scattering correction can be employed to specifically eliminate the interference caused by particle scattering and restore the true spectral information.

$$x_{\text{MSC}} = \frac{(x_{i,k} - b_0)}{b} \tag{10}$$

In the equation, b and b_0 represent the relative offset coefficient and the translation variable, respectively, obtained by performing simple linear regression on the spectra of each sample against the average spectrum.

Method 8 (Transformations of standard normal variables): During on-site analysis, factors such as uneven illumination and variations in the surface smoothness of mineral samples-in addition to particle scattering-can cause spectral baseline drift and overall amplitude shifts. This prevents direct comparison of spectra from different samples, thereby affecting the generalisability of the model. Therefore, standardisation can be performed on a sample-by-sample basis using a standard normal transformation, which simultaneously eliminates scattering and baseline drift, thereby ensuring the comparability of the spectral data.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

$$x_{\mathrm{SNV}} = \frac{x - \overline{x}}{\sqrt{\frac{\sum_{k=1}^{n}(x_k - \overline{x})^2}{n-1}}} \tag{11}$$

Among which,

$$\overline{x} = \frac{\sum_{k=1}^{n} x_k}{n} \tag{12}$$

Method 9 (First derivative): In the raw spectra of iron ore, absorption peaks caused by moisture content are often masked by the overall spectral trend, and the presence of linear baseline drift renders these features indistinct, preventing the model from accurately capturing moisture-related information. Therefore, the first derivative can be employed to effectively remove the linear baseline drift, highlight the rate of change in spectral reflectance, and enhance the distinctiveness of the absorption peaks.

$$X'_{i,j} = \frac{\mathrm{d}X_{i,j}}{\mathrm{d}\lambda} \approx \frac{X_{i,j+1} - X_{i,j-1}}{\lambda_{j+1} - \lambda_{j-1}} \tag{13}$$

Method 10 (Trend-neutral processing): As the magnetite spectrum may exhibit not only linear baseline drift but also slow non-linear baseline drift-caused by factors such as instrument ageing or gradual changes in light intensity-such drift can distort spectral characteristics and affect the accuracy of model predictions. Therefore, trend removal can be employed to specifically eliminate linear or non-linear baseline drift, thereby restoring the true fluctuations in the spectrum.

$$x_{\mathrm{DT}} = x - \hat{x} \tag{14}$$

Among which,

$$\hat{x} = b_0 + b_1 x_m + b_2 x_m^2 \tag{15}$$

In the equation, x_m represents the absorbance value in the original spectrum.

Method 11 (Envelope removal): As the spectral reflectance of iron ore is influenced by various factors, such as mineral composition and surface condition, the overall spectral trend may mask the absorption troughs corresponding to moisture content, making it impossible to accurately extract spectral parameters related to moisture content. Therefore, envelope subtraction can be employed to normalise the depth of the absorption peaks, thereby highlighting the relative absorption characteristics and clearly revealing the absorption troughs corresponding to moisture content.

$$X^*_{i,j} = \frac{X_{i,j}}{H_{i,j} + \varepsilon} \tag{16}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

In the equation, H_i,j represents the interpolation of the spectral envelope at the jth band; the raw reflectance is divided by the envelope to normalise the spectral peaks to 1, making the absorption troughs corresponding to moisture content more pronounced. This eliminates interference from the overall spectral trend, facilitating the extraction of moisture-related features and enabling the analysis of different batches of mineral samples collected in the field.

(5) Selection of characteristic bands

In order to effectively reduce the dimensionality of hyperspectral data, minimise the interference caused by multicollinearity between bands, highlight the spectral information closely related to the moisture content of iron ore, and improve the efficiency and predictive accuracy of subsequent modelling, the variance of each band is calculated for the pre-processed spectral data following spectral pre-processing[17]:

$$Var(j) = \frac{1}{n-1}\sum_{i=1}^{n}(X_{i,j}^{*} - \overline{X_{j}^{*}})^{2} \tag{17}$$

In the equation, the mean is the average of all samples in the jth band after pre-processing; Var(j) reflects the band's ability to distinguish between different samples.

The magnitude of the variance can be used to quantify the information content of different spectral bands: a larger variance indicates that the reflectance differences between iron ore samples with varying moisture contents are more pronounced in that band, and that it contains richer, useful information regarding moisture content; conversely, a smaller variance suggests that the band contains redundant information and is less sensitive to changes in moisture content.

Based on this, the bands are ranked in descending order of variance, and the top 15 bands with the highest variance are selected as feature bands. These are then combined to form a low-dimensional feature spectral matrix for use in subsequent model construction.

(6) Development and Accuracy Assessment of a Moisture Content Prediction Model

Common modelling methods are categorised into linear and non-linear regression. In this study, to address the modelling requirements for hyperspectral data and sample moisture content, partial least squares regression and random forest were selected as the core algorithms. A multi-model fusion strategy was developed to enhance prediction accuracy and robustness.

As a classic method of linear multiple analysis, partial least squares regression effectively addresses the issues of high dimensionality and multicollinearity in spectral data. By projecting variables to uncover covariance structures, it is suitable for scenarios where the spectral dimension exceeds the sample size, enabling the precise extraction of linear relationships between spectral data and moisture content.

$$Y = X\beta + \sigma \tag{18}$$

In particular, following the extraction of two principal components, the final prediction formula is:

$$\hat{y}_{PLSR} = b_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_{15} x_{15} \tag{19}$$

In the equation, X represents the spectral feature matrix; Y represents the true moisture content vector of the sample; beta_1 represents the regression coefficient for the i-th feature band; x_i represents the spectral reflectance of the i-th feature band following optimal data processing.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

Random Forests are non-linear ensemble learning algorithms that do not require a predefined linear relationship; they can adaptively capture the non-linear mapping between spectral data and moisture content. They are highly resistant to noise and outliers, can prevent overfitting, and offer fast fitting speeds without the need for complex pre-processing, making them well-suited to the requirements of this study. In this study, the Random Forest consists of an ensemble of 20 decision trees, each with a maximum depth of 6.

$$\hat{y}_{\mathrm{RF}} = \frac{1}{T}\sum_{t=1}^{T} f_t(\mathbf{X}) \tag{20}$$

In the formula, T represents the number of decision trees; and f(X) represents the prediction result of the tth decision tree.

However, individual models have their limitations: partial least squares regression lacks accuracy in non-linear predictions, whilst random forests perform poorly in generalising to small samples. This study combines the prediction results of mean-averaged partial least squares regression and random forests to complement their respective strengths, thereby enhancing the model's stability and generalisation ability[18]. The final predicted value of the fusion model is:

$$\hat{y}_{\mathrm{fusion}} = \frac{1}{2}\left(\hat{y}_{\mathrm{PLSR}} + \hat{y}_{\mathrm{RF}}\right) \tag{21}$$

In this paper, the samples are randomly divided into a training set and a test set in a 7:3 ratio for each iteration, and cross-validation is performed accordingly. Model performance is evaluated using the coefficient of determination for both the training set and the test set; the closer the coefficient of determination is to 1, the smaller the root mean square error, and the greater the model's stability and predictive power.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{22}$$

In the equation, y_i represents the true moisture content of the i-th sample.

3. Results and Discussion

(1) Spectral curve analysis of iron ore with varying moisture content

A comparison of the trends in the reflectance spectra of iron ores under different moisture content conditions reveals that, whilst the spectral trends vary across different iron ores, there is an overall negative correlation: reflectance is higher when the moisture content is 0%; as the moisture content increases, reflectance decreases. However, changes in moisture content have little effect on the overall shape and trend of the spectral curves. Distinct spectral absorption features appear near 1440 nm and 1920 nm; these are associated with the vibration of the –OH bond in water molecules, and the intensity of the absorption peaks increases with rising moisture content[19].

It can thus be seen that changes in moisture content significantly affect the reflectance and the depth of the absorption peak of the iron ore; the higher the moisture content, the lower the reflectance and the deeper the absorption peak. This indicates that spectral absorption characteristics can be used to characterise the moisture content of magnetite samples.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

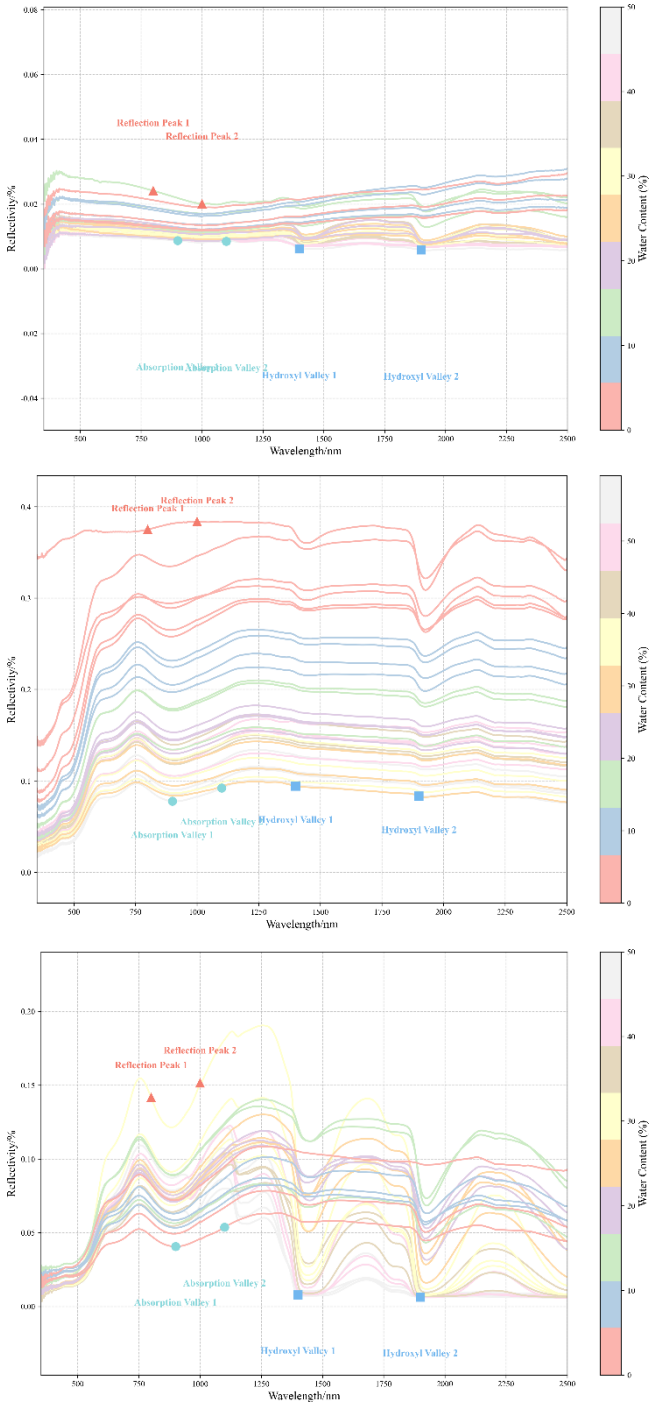

Figure 3. Partial spectral curves (H3, H4, H1)

As shown in Figure 4, correlation and clustering analyses of the spectral curves at different moisture contents revealed that, with the exception of the H4 mineral sample-which exhibited a high positive correlation and relatively consistent spectral characteristics across all moisture conditions-the spectral curves of the remaining mineral samples displayed significant differences across different moisture content ranges. The primary reason for this lies in the significant differences in surface moisture content, interparticle scattering properties, and water–mineral interactions between low and high moisture content states. These differences lead to marked alterations in the shape of spectral reflectance, the intensity of absorption peaks, and the overall trend of the curves[20], ultimately making it difficult for spectral characteristics to remain highly consistent across different moisture content ranges.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

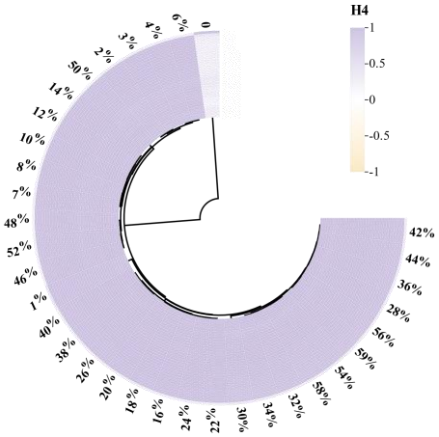

Figure 4(a). Correlation analysis of H4 spectral curves

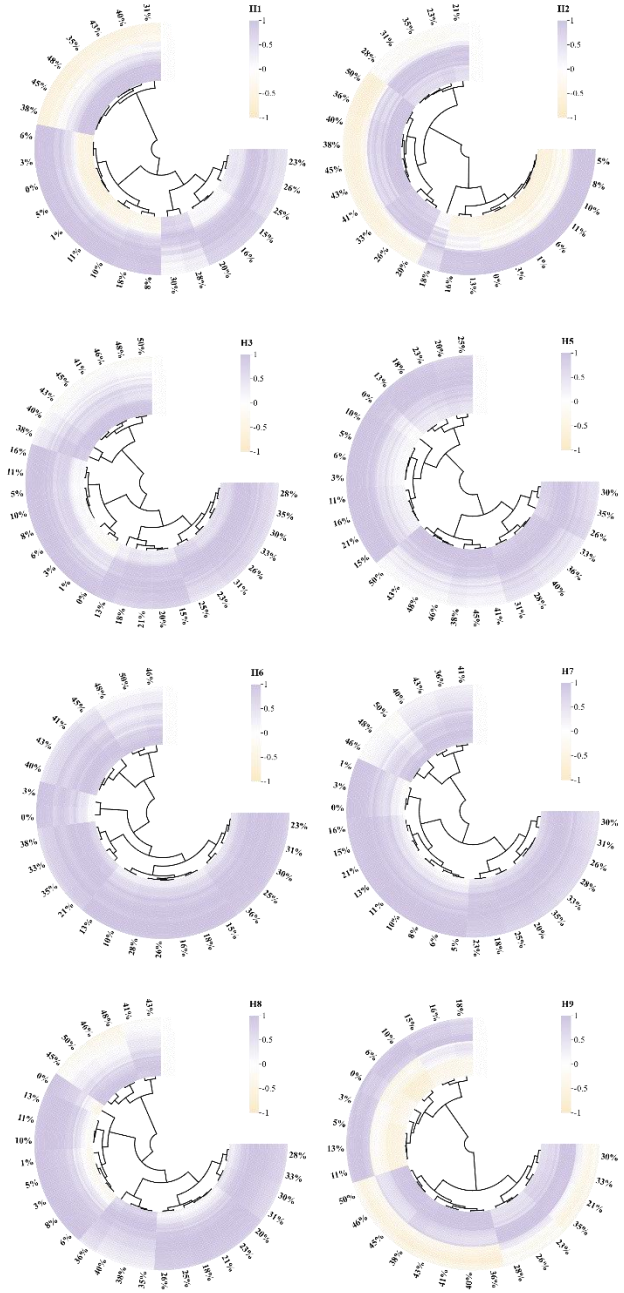

Figure 4(b). Correlation analysis of spectral curves (H1, H2, H3, H5, H6, H7, H8, H9)

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

An analysis of reflectance variations and correlations was conducted on spectral curves under different moisture content conditions[21]. The results indicate that, under identical grade and grain size conditions, the characteristics of reflectance variations in different mineral samples caused by changes in moisture content, as well as the degree of linear correlation between these variations and target properties, vary with wavelength. The H1 mineral sample is most sensitive to moisture content in the short-wavelength band, with reflectance decreasing rapidly as moisture content increases; in the long-wavelength band, it is tolerant to moisture, and reflectance remains consistently low. For the H2 mineral sample, reflectance increases as moisture content rises, with the mid-wavelength band being the core region where reflectance is enhanced. In the H3 mineral sample, the long-wavelength band is most sensitive to moisture content; as moisture content increases, reflectance expands from localised high reflectance to widespread high reflectance. The mid-to-long-wave bands of the H4 mineral sample exhibit strong tolerance to moisture content; reflectance declines only gradually with increasing moisture, and the band structure remains stable. The mid-wave band of the H5 mineral sample is a moisture-sensitive region; reflectance changes abruptly from continuous high reflectance to low reflectance as moisture content increases. The short-wave band of the H6 mineral sample is extremely sensitive to moisture content; reflectance breaks down rapidly and disappears as moisture increases. The long-wave band reflectance of the H7 mineral sample remains stable despite fluctuations in moisture content, indicating that the mineral structure is highly resistant to moisture interference. In the H8 mineral sample, the reflectance in the mid-to-long-wave bands decreases continuously with increasing moisture content, with warm clusters evolving from dense to sparse and eventually disappearing. The H9 mineral sample exhibits extremely low reflectance across all wavelength bands and is highly sensitive to moisture; it is the type with the lowest reflectance and the most significant moisture influence among the nine mineral samples. Furthermore, the optical response of moisture in different mineral samples exhibits synergistic effects in certain wavelength bands.

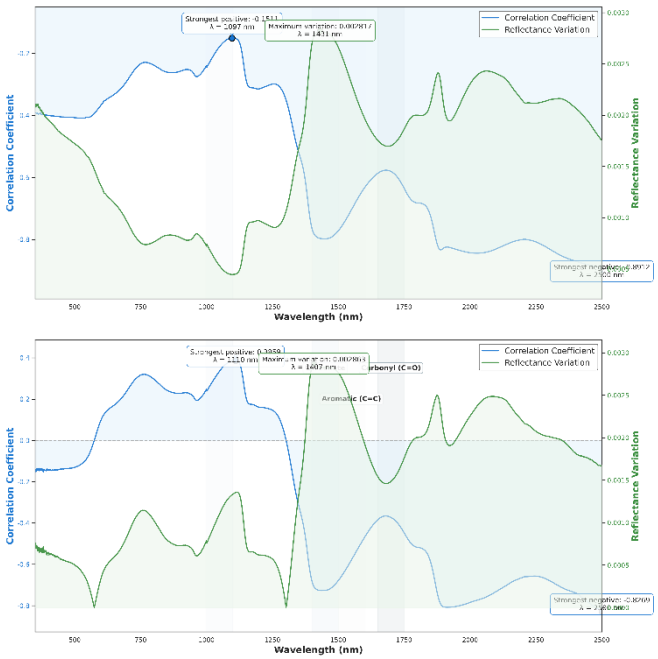

Figure 5. Correlation and amplitude curves for selected mineral samples (H6, H8)

(2) Analysis of the results of spectral data pre-processing

By comparing 11 data pre-processing methods and their combinations, and ranking them according to the coefficient of determination on the test set, the optimal data pre-processing methods for the nine mineral samples were ultimately identified, as shown in Table 2.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

Among these, the optimal combination for H1 is 'Z-score standardisation + multivariate scatter correction' (R^2 = 0.9463). The optimal combination for H9 is 'standardised deviation + moving average smoothing + envelope removal' (R^2 = 0.7851).

The average R^2 for nine types of iron ore samples, following various data processing methods, can reach 0.9 or higher.

Table 2. Optimal data pre-processing methods

Sample	Index of Methods	R^2
H1	(4)+(7)	0.946299
H2	(6)+(7)+(11)	0.897548
H3	(1)+(2)+(9)	0.934194
H4	(1)+(3)+(10)	0.879406
H5	(5)+(8)+(9)	0.935969
H6	(4)+(5)+(8)	0.879686
H7	(2)+(5)+(8)	0.937898
H8	(4)+(7)+(10)	0.911362
H9	(1)+(5)+(11)	0.785135

The spectral curves of various iron ore samples following multiple data transformations are shown in Figure 6. These transformed curves can be classified into three main types based on their smoothness, feature prominence and amplitude of fluctuation: the transformed spectral curves of samples H1, H5 and H7 generally exhibit high smoothness and low noise; the curves for samples H3, H4 and H9 exhibit prominent feature information and significant amplitude fluctuations; whilst the curves for samples H2, H6 and H8 offer the combined advantage of noise suppression and effective feature retention. The aforementioned data transformation methods can effectively enhance the spectral features of interest and suppress random noise interference, thereby laying a more representative and stable spectral data foundation for subsequent quantitative and qualitative modelling.

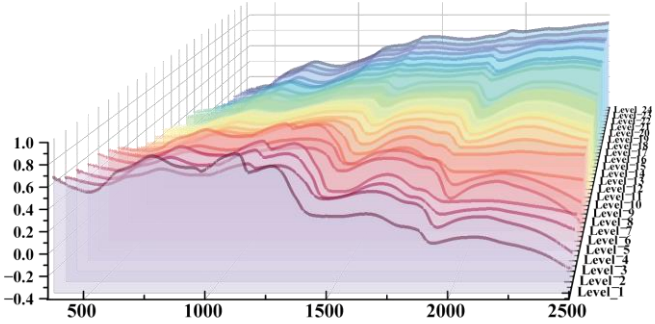

Figure 6(a). Results of H1 spectral changes

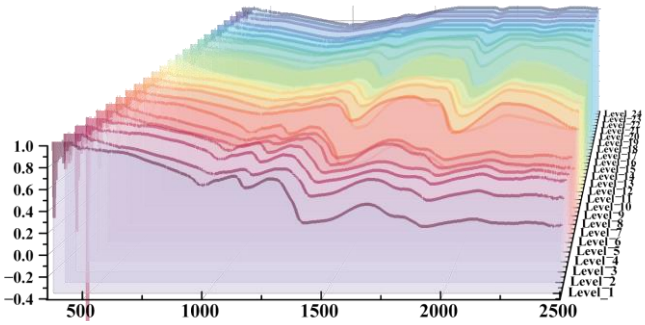

Figure 6(b). Results of H2 spectral changes

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

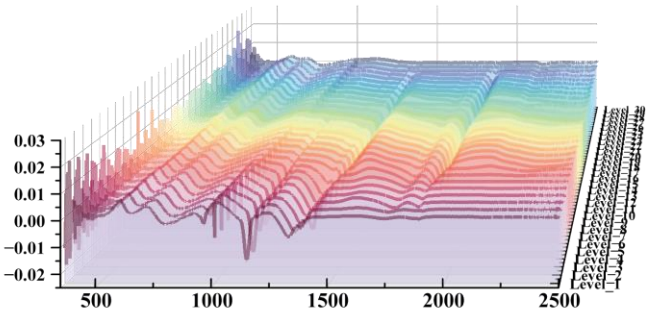

Figure 6(c). Results of H3 spectral changes

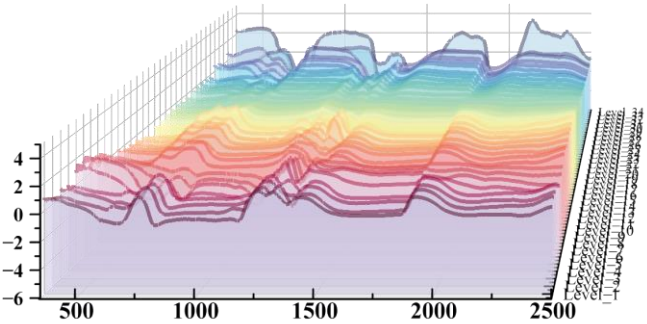

Figure 6(d). Results of H4 spectral changes

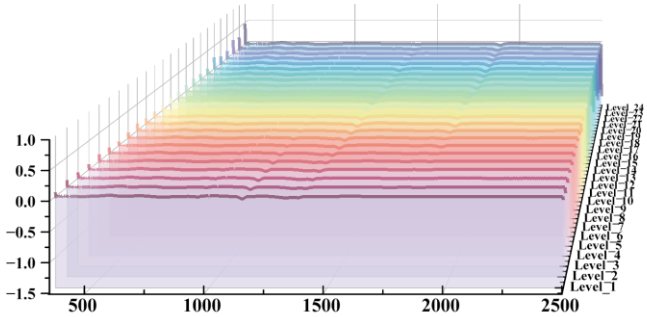

Figure 6(e). Results of H5 spectral changes

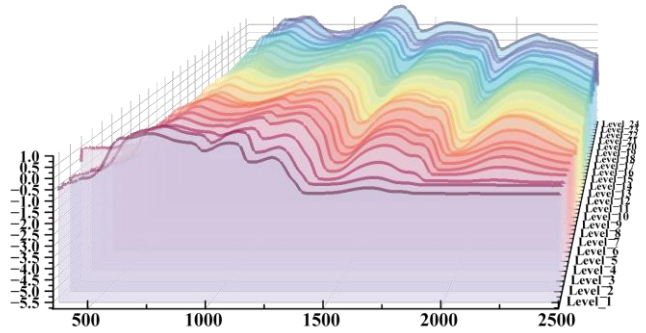

Figure 6(f). Results of H6 spectral changes

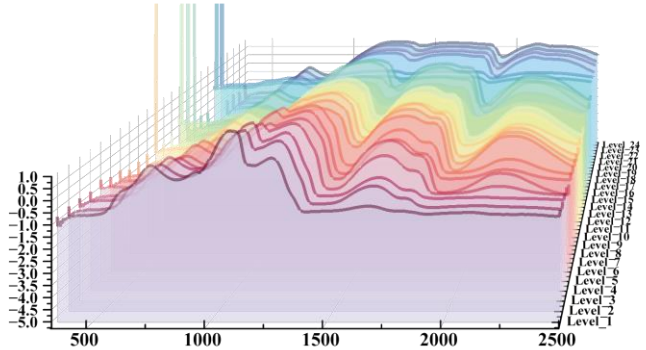

Figure 6(g). Results of H7 spectral changes

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

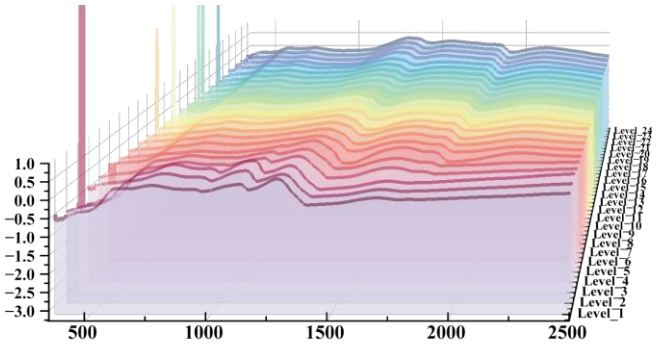

Figure 6(h). Results of H8 spectral changes

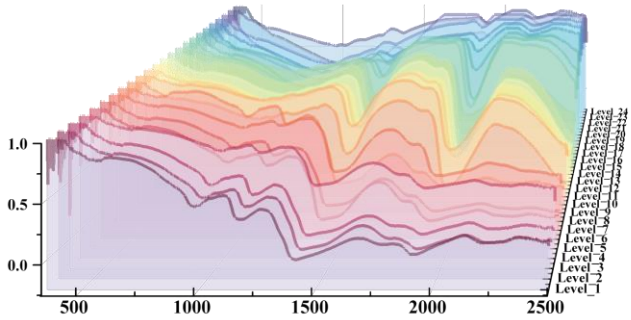

Figure 6(i). Results of H9 spectral changes

(3) Results and Analysis of Feature Band Screening

The distribution of characteristic bands corresponding to the nine mineral samples following feature screening is shown in Figure 7. The characteristic bands of samples H2, H3, H5, H6, H7 and H8 are concentrated around 360 nm, whilst those of samples H1, H2, H5 and H9 are clustered around 2490 nm; Furthermore, characteristic bands of different mineral samples are also densely distributed near wavelengths such as 750 nm, 1140 nm, 1402 nm and 1880 nm. By effectively eliminating redundant data through feature screening, data quality can be significantly improved, laying a more reliable data foundation for subsequent in-depth analysis.

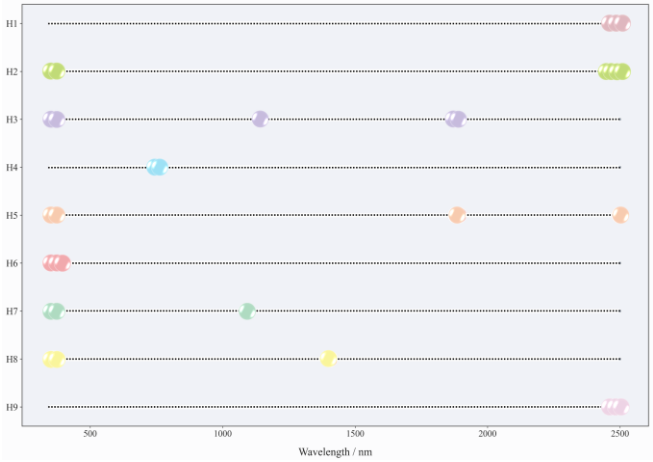

Figure 7. Results of spectral feature band screening (H1–H9)

(4) Results and Analysis of the Moisture Content Prediction Model

Based on the selected feature bands, a fusion model combining a Random Forest model and a Partial Least Squares (PLS) regression model was employed to predict the moisture content of iron ore.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

Model performance was evaluated using the coefficient of determination (R^2) from the test set. The results indicated that the optimal model for predicting the H1, H2, H3, H6, H7, H8 and H9 ore samples was the fusion model, whilst the optimal model for predicting the H4 and H5 ore samples was the PLS regression model. The average coefficients of determination (R^2) for the nine categories of ore samples in the training set, test set and full dataset were 0.9226, 0.8899 and 0.9143, respectively. The measured and predicted values for the optimal spectral transformation models of ore samples are shown in Figure 9.

When comparing the nine ore samples, the H8 sample performed best on the training set, with an R^2 of 0.9865, followed by the H1 sample, with an R^2 of 0.9826; Ore sample H1 performed best on the test set, with an R^2 of 0.9532, followed by ore sample H1, with an R^2 of 0.9511.

Across all samples, the ore samples with the highest model accuracy were H8, H1 and H3, with R^2 values of 0.9747, 0.9746 and 0.9654 respectively.

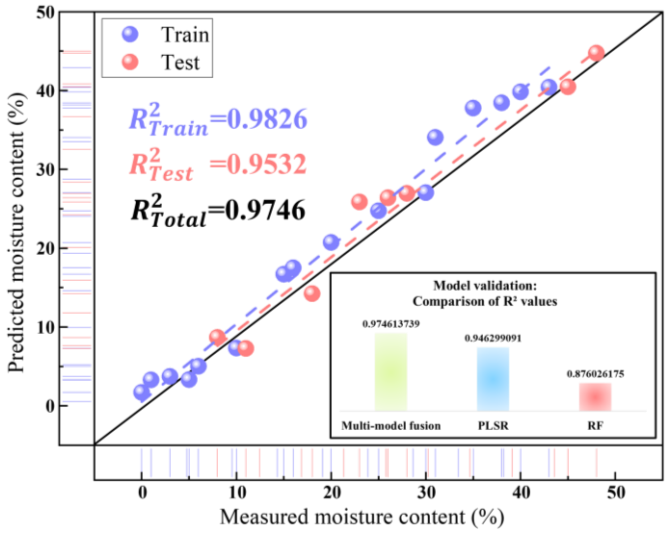

Figure 8(a). Comparison of the optimal model's predicted moisture content for H1 with the model itself

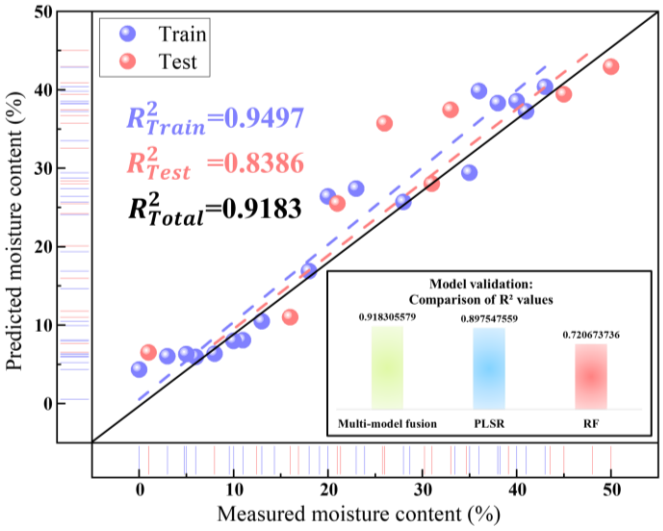

Figure 8(b). Comparison of the optimal model's predicted moisture content for H2 with the model itself

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

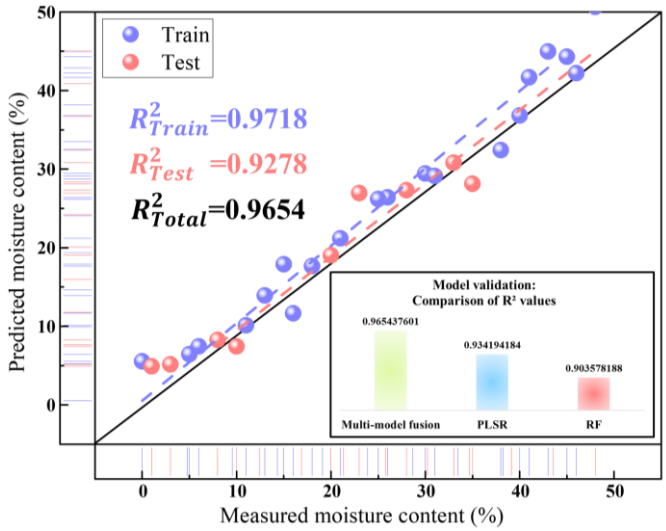

Figure 8(c). Comparison of the optimal model's predicted moisture content for H3 with the model itself

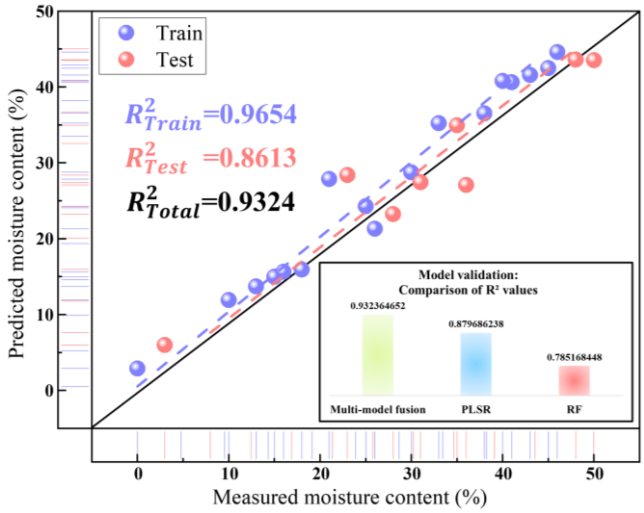

Figure 8(d). Comparison of the optimal model's predicted moisture content for H6 with the model itself

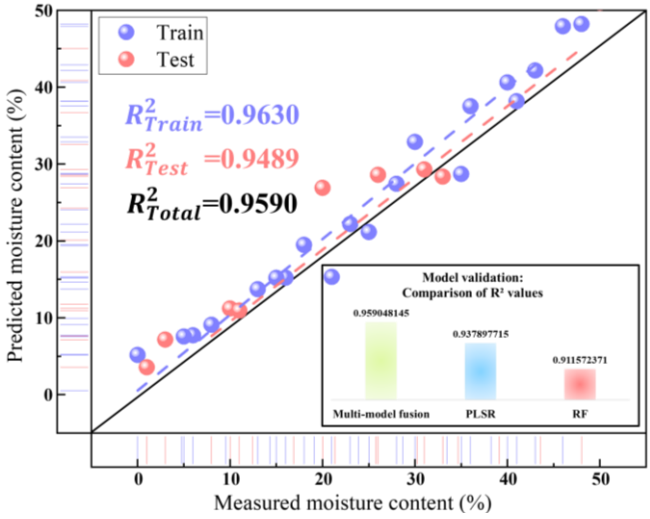

Figure 8(e). Comparison of the optimal model's predicted moisture content for H7 with the model itself

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

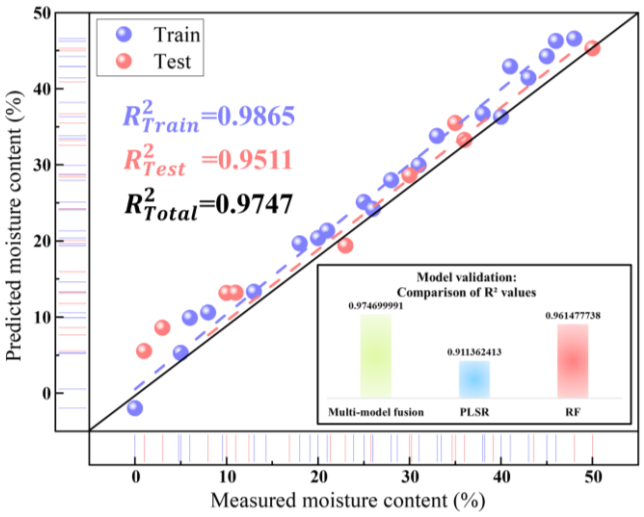

Figure 8(f). Comparison of the optimal model's predicted moisture content for H8 with the model itself

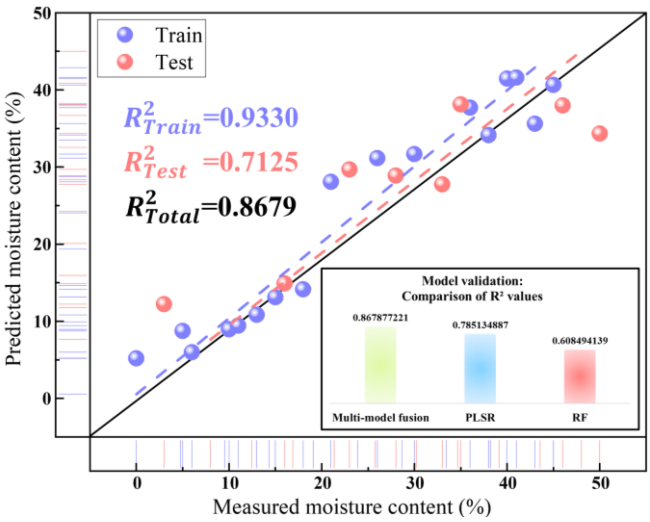

Figure 8(g). Comparison of the optimal model's predicted moisture content for H9 with the model itself

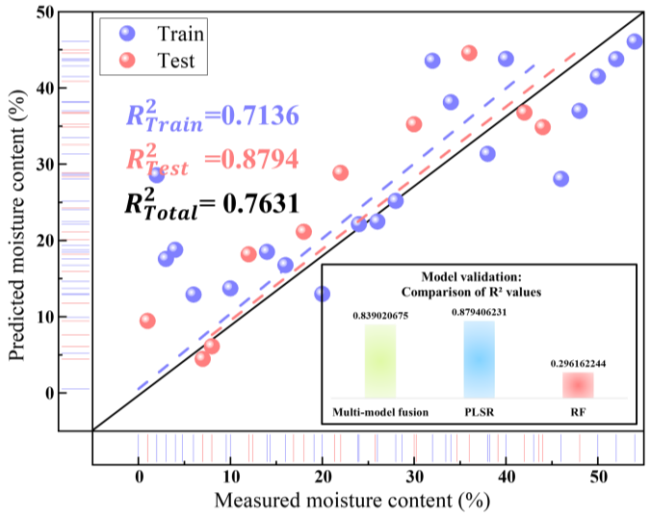

Figure 8(h). Comparison of the optimal model's predicted moisture content for H4 with the model itself

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

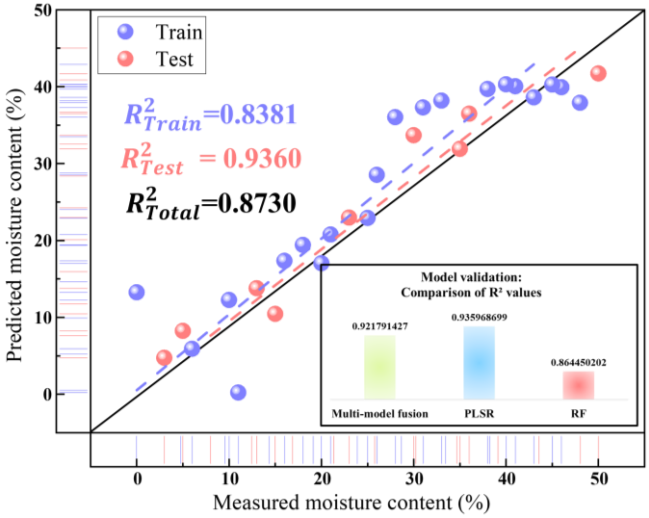

Figure 8(i). Comparison of the optimal model's predicted moisture content for H5 with the model itself

Table 3. Table of Coefficients for the Regression Model

	H1	H2	H3	H4	H5	H6	H7	H8	H9
σ	19.88	21.00	27.05	28.74	27.30	27.06	24.85	26.75	23.41
β_1	156.81	-1.91	181.01	114.65	6.67	18.54	0.11	-0.07	6.04
β_2	125.90	-2.44	-212.70	113.78	13.30	0.03	0.11	-0.07	64.26
β_3	100.93	13.96	-175.62	29.06	-24.14	0.57	0.11	-0.52	-9.16
β_4	90.77	-4.34	-132.38	129.43	-42.23	-0.13	-0.01	-0.62	67.84
β_5	59.21	-4.83	-78.64	108.83	20.23	-2.25	-0.03	-0.23	27.73
β_6	36.06	-5.10	-253.51	147.37	17.64	1.34	-0.03	0.19	34.81
β_7	24.70	-3.91	-28.16	-131.14	14.90	-0.28	0.14	-0.51	47.42
β_8	-15.79	-4.71	-299.02	163.33	22.81	0.01	0.14	-0.50	39.32
β_9	-43.83	-4.31	71.97	-211.93	12.11	0.10	0.14	-0.69	-36.45
β_{10}	-61.35	-4.89	19.96	-45.97	9.13	-3.00	0.22	1.17	-15.09
β_{11}	-87.06	-3.69	120.32	-33.06	25.50	-2.32	0.22	1.22	-18.40
β_{12}	-96.65	-5.22	-339.15	-197.03	6.07	-0.94	4.62	-15.64	-5.82
β_{13}	-106.87	-5.40	167.45	68.23	28.33	-1.38	4.63	-15.60	-89.67
β_{14}	-115.43	-4.01	-1349.28	63.19	2.87	-0.77	4.64	-15.56	-60.95
β_{15}	-129.73	-4.72	-1355.91	-310.56	31.35	-1.17	4.61	-15.68	-86.48

The R^2 value of the training set for the magnetite moisture content prediction model established by Bai Yang et al. [22] reaches 0.798; The training set R^2 of the Anshan-type magnetite moisture content prediction model established by Xie Xiaoxiao et al. [23] can reach 0.789; the overall R^2 of the high-silica iron tailings moisture content prediction model established by Yu Molli [24] can reach 0.92; through the analysis and comparison of the performance of the above models, the PLSR-RF fusion models for the H1, H2, H3, H6, H7, and H8, all of which achieved a training set R^2 of over 0.94, a

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

test set R^2 of over 0.83, and an overall R^2 of over 0.91. For the H4 and H5 ore samples, the PLSR model was selected for fitting based on the optimal test set R^2, with test set R^2 values of 0.87 and 0.93, respectively. For the H9 mineral sample, when the PLSR model was selected, the R^2 value for the test set reached 0.78, whilst the R^2 values for the training set and the population were 0.88 and 0.85 respectively; when the PLSR-RF hybrid model was selected, the R^2 values for the training set and the population reached 0.93 and 0.86 respectively, but the R^2 value for the test set was 0.71.

4. Conclusion

Through a systematic analysis of hyperspectral data from nine iron ore samples, this study has reached the following main conclusions:

(1) Spectral characteristics

There is an overall negative correlation between the moisture content of iron ore and spectral reflectance; the higher the moisture content, the lower the reflectance. Furthermore, the intensity of the water molecule absorption peaks near 1440 nm and 1920 nm increases with rising moisture content. There are significant differences in the spectral characteristics of different ore samples between low-moisture and high-moisture ranges; with the exception of the H4 ore sample, the spectral curves of the remaining ore samples exhibit distinct differences at different moisture contents.

(2) Preferred pretreatment methods

Given the significant differences in mineral composition, surface characteristics and particle size distribution among the nine mineral samples selected for the experiment, the suitability for practical application of the eleven pretreatment methods—both in terms of their individual characteristics and the cascade treatment schemes formed by combining them in different sequences—exhibited marked inconsistencies and sample-dependence. Specifically, different pre-treatment methods and their cascade combinations exhibited significant variations in terms of treatment efficacy, adaptability, and impact on subsequent analytical processes, depending on the type of mineral sample. For details regarding the specific matching methods, matching principles, and corresponding criteria for evaluating the suitability of these pre-treatment methods and their cascade combinations with respect to the nine mineral samples, please refer to Table 2 in this paper, which provides a systematic and detailed explanation..

(3) Distribution of characteristic bands

The top 15 characteristic bands identified using the variance method are concentrated around wavelengths of 360 nm, 750 nm, 1140 nm, 1402 nm, 1880 nm and 2490 nm, respectively, in the visible to shortwave infrared range. Indeed, while there is thus both very strong overlap and considerable variation in the characteristic bands across different mineral samples, this as expected clearly indicates that the spectral response of moisture exhibits both commonalities and the mineral-specific characteristics too.

(4) Model performance

The mean model combining PLSR and RF performed best on most ore samples (H1, H2, H3, H6, H7, H8), with R^2 values of $\geqslant 0.94$ on the training set, $\geqslant 0.83$ on the test set, and $\geqslant 0.91$ overall. Among these, the H8 ore sample achieved the highest R^2 for the entire dataset at 0.97, whilst the H1 ore sample achieved the highest R^2 on the test set at 0.95. Compared with existing models for predicting moisture content in magnetite, Anshan-type magnetite and high-silica iron tailings, the methodology employed in this study demonstrates significant improvements in both accuracy and generalisation ability.

In this study, particle size, grade and temperature conditions were held constant; the effects of on-site particle size distribution, ore purity and temperature fluctuations on the spectral response of moisture were not investigated; the performance of the fusion model is limited in terms of generalisability to certain ore samples. Future research could focus on spectral mechanisms under multi-grain size and

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

variable temperature conditions, and explore strategies for model simplification and online adaptive updating.

Acknowledgments

The author would like to express his sincere gratitude for the financial support provided by the Hebei Province Undergraduate Innovation Training Programme project entitled ‘Information Recognition Model for Three-Stage Iron Ore Sorting Based on Hyperspectral Technology’ (S202510081145).

References

[1] Xie Heping, Li Cunbao, Xie Yachen, et al. Cutting-edge Technologies and Development Trends in Lunar Mining [J]. Journal of Coal Science and Engineering, 2026, 51(1): 21–42.

[2] Wu Su, Wu Dongli, Wang Zhongjin, et al. The applicability of the cosmic-ray fast neutron method in monitoring soil moisture in typical oasis farmland [J]. Chinese Journal of Agricultural Meteorology, 2026, 47(1): 48–62.

[3] Yang Jiajia, Liu Yang, Chen Chaoqun, et al. A Study on the Technical Framework of Multi-source Remote Sensing Technology in Land Quality Surveys [J/OL]. Geology and Resources, 1–10 [2026-04-07].

[4] Peng Sihan, Bao Nisha, Zhang Fan, et al. Research Progress on Models for Estimating the Visible–Shortwave Infrared Spectral Characteristics and Physicochemical Properties of Mining Area Soils [J/OL]. Journal of Remote Sensing, 2026, (3): 473–492 [2026-04-07].

[5] Shao Jinqiu, Yan Xiulan, Yang Xiao, et al. Study on the arsenic immobilisation effects of naturally occurring iron-bearing minerals under two moisture conditions in different soils [J]. Environmental Pollution and Control, 2021, 43(9): 1133–1138.

[6] Lü Yihui, Zhang Shitao, Gao Jianfeng, et al. Application of thermal infrared spectral characteristics and magnetic susceptibility in the exploration of skarn-type ore deposits: A case study of the Tonglushan copper-gold-iron deposit in south-eastern Hubei [J]. Geochemistry, 2025, 54(1): 141–158.

[7] Wang Zhiyu, Wang Da, Qiu Kunfeng, et al. Applications and Prospects of Machine Learning in the Mining of Big Data on Mineral, Rock and Geochemical Data [J]. Journal of Chengdu University of Technology (Natural Science Edition), 2025, 52(5): 844–858.

[8] Luan Meiqi, Xiong Feng, Dai Yaoyao, et al. A Study on Rapid Detection of Phenolic Acids in the Traditional Chinese Medicinal Herb Salvia miltiorrhiza Using Hyperspectral Technology [J]. Chinese Journal of Traditional Chinese Medicine, 2025, 50(22): 6319–6327.

[9] Jia Wenyu, Zhang Yongchao, Li Xiumei, et al. A Prediction Model for the Content of Re-Du-Ning Injection Based on Spectral Fusion Combining Near-Infrared and Mid-Infrared Technologies [J]. Chinese Journal of Traditional Chinese Medicine, 2024, 49(16): 4450–4459;

[10] Xu Yan, Jin Junjie, Hu Kongfa, et al. Development of a production-oriented near-infrared quantitative model for Paeonia lactiflora [J]. Chinese Modern Traditional Chinese Medicine, 2025, 27(12): 2313–2322.

[11] Zhang Yujiao, Lu Wei, Yang Jinyu, et al. A Study on Remote Sensing Monitoring of Larix Caterpillar Infestation Based on Sentinel-2 Imagery [J]. Chinese Journal of Forest Pests and Diseases, 2025, 44(3): 9–15.

[12] Liu Haiqi, Liu Shanjun, Ding Ruibo. Study on the Influence of Particle Size on the Visible-Near-Infrared Spectra of High-Grade Hematite [J]. Metal Mines, 2022(4):158−162.

[13] Li Xiang, Zhang Yongbin, Liu Mingyue, et al. Comparative Analysis of Hyperspectral Estimation Models for Soil Texture in Coastal Wetlands [J]. Spectroscopy and Spectral Analysis, 2024, 44(9): 2568−2576.

[14] Wang Chuan, Guo Xingzhang, Chen Chengyong, et al. Study on the Mechanical Properties and Environmental Impact of Red Mud-Based Cementitious Materials Used to Stabilise Loamy Clay Roads [J/OL]. Engineering Science and Technology, 1–15 [7 April 2026].

[15] Pang Jiang, Zhang Yeyu, Huang Yi, et al. Effect of Iron Content on the Raman Spectral Characteristics of Dolomite [J]. Rock and Mineral Testing, 2023, 42(4): 852–862.

[16] Dai Shengyun, Wu Dongxue, Huang Rui, et al. Study on Rapid Non-Destructive Testing Methods for Wuwei Qingzhu Preparations Based on Hyperspectral Technology [J]. Modern Chinese Medicine, 2024, 26(10): 1790–1798.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202604_12(4).0039

[17] Huang Rui, Zhou Yi. Drug Identification and Quantification Based on Digital Surface-Enhanced Raman Spectroscopy Sensing Technology [J]. Science in China: Chemistry, 2026, 56(2): 497–510.

[18] Zhou Feixiang, Jiang Hong, Guo Baolin, et al. Rapid identification of the origin and production patterns of Astragalus root based on hyperspectral data dimensionality reduction algorithms [J]. Chinese Journal of Traditional Chinese Medicine, 2024, 49(24): 6660–6666.

[19] Li Mengqian, Li Mingduo, Wang Jinhua, et al. A Study on Hyperspectral Methods for Precise Estimation of Iron Grade in Iron Ore Powder [J]. Metal Mines, 2023(3): 206−213.

[20] Yao Yifei, Wang Jinlin, Chen Huayong, et al. Three-dimensional visualisation modelling of spectral EVS for alteration minerals in the Tonglushan deposit and its indicative significance [J]. Geochemistry, 2024, 53(5): 719–733.

[21] Yao Xiuxiu, An Maoguo, Zhi Chenglong, et al. Application of the Big Data-XGBoost Model in Hyperspectral Characterisation and Rare Earth Grade Prediction [J/OL]. Geological Bulletin, 1–28 [7 April 2026].

[22] Bai Y ,Xie X ,Zhang J , et al.Hyperspectral analysis and inversion model study of water content in magnetite.[J].Scientific reports,2025,15(1):25694.

[23] Xie Xiaoxiao, Bai Yang, Zhang Jiuling, et al. Development of a Hyperspectral Prediction Model for High Moisture Content in Anshan-type Magnetite [J]. Rock and Mineral Testing, 2024, 43(6): 901–913.

[24] Yu, Mo-li. A Study on Hyperspectral Monitoring Methods for the Composition and Moisture Content of High-Silica Iron Tailings [D]. Northeastern University, 2021.