VIII International Conference on Particle-Based Methods
PARTICLES 2023
U. Perego, M. Cremonesi, A. Franci (Eds)

DETERMINING COKE MOISTURE CONTENT THROUGH IMAGE ANALYSIS METHODS AND MACHINE LEARNING MODELS

M. LI[2], H. YAO[1], X. WANG[1], Y.YU[1], H. SAXÉN [2]

[1] School of Materials Science and Engineering
Shanghai University
Shanghai 102100, China
e-mail: yaoweiyu@shu.edu.cn, www.shu.edu.cn

[2] Process and Systems Engineering Laboratory
Åbo Akademi University
Åbo /Turku 20500, Finland
email: meng.li@abo.fi, www.abo.fi

Abstract. Coke moisture content plays a crucial role as a quality indicator in ironmaking. Fast and accurate measurement of coke moisture content can effectively ensure operational stability, thereby guaranteeing the quality of pig iron. Coke with different moisture levels exhibits variations in light reflection, refraction, and powder adhesion characteristics. Drawing from this premise, we employed an image analysis approach to analyse the colour and texture features of coke images with varying moisture content. The results indicated that certain specific image features are highly sensitive to changes in coke moisture content. This study also conducted testing and comparison of performances of three common machine learning models in predicting coke moisture content based on image analysis. The Support Vector Machine (SVM) predictive model for coke moisture content based on image analysis showed optimal performance. This demonstrated a close connection between coke image features and moisture content.

Keywords: Metallurgical Coke, Moisture content, Image analysis, Machine learning models.

1 INTRODUCTION

Coke is an important fossil fuel, and it is also a primary reducing agent and irreplaceable supporting skeleton in blast furnace ironmaking [1]. Coke moisture content directly influences the physical and chemical properties of coke. It is important to appropriately consider the moisture content since the raw materials are weighed and the charged quantities are estimated on a dry basis. During the production process, both coal characteristics and coking process directly influence the moisture content of coke, and at wet quenching the coke picks up substantial amounts of water. During the storage process, humidity and environmental conditions also have an impact on it. In metallurgical processes, a high and variable moisture content of coke can have an impact on the coke ratio, leading to thermal imbalance and reduced the productivity [2,3]. The moisture content of coke introduced into the blast furnace needs to be maintained with the ranges of 1~6 wt.% [2] but its actual moisture content can potentially reach up to around 18% [4]. An accurate measurement of the coke moisture content contributes to ensure operational stability, ultimately impacting the quality of the final iron and steel products. Moreover, it is crucial for adhering to environmental regulations and reducing the ecological footprint of the metallurgical processes.

Historically, the determination of moisture content in coke has primarily relied on established laboratory techniques. The most common conventional method is the oven-drying method, where a weighed coke sample is heated in an oven until the moisture evaporates, and the weight loss is calculated to determine moisture content [4]. Additionally, there are also commonly used methods such as the Fast Neutron and Gamma-ray Transmission (FNGT) technique [5]. While these techniques have provided reliable results, they come with certain limitations, such as time-consuming procedures, requirements of large volumes of coke and static location of the measurement device. Currently, due to the advancement of image analysis technology, numerous methods have been proposed for utilizing image analysis to determine coke quality, component, and size distribution [6-8]. In certain studies, researchers have attempted to use image analysis to measure the moisture content and powder ratio of coke taking into consideration that the intensity of reflected light from the surface of coke varies in response to changes in moisture content [9]. However, to the best of the present authors' knowledge no research has been conducted on the relationship between variations in coke moisture content and image features.

Changes in coke moisture content impact the light reflection and powder adhesion properties, and these manifestations can be captured in images as colour and texture features. Building upon this principle, this study conducted tests on the characteristics of different image features that change with variations in coke moisture content. The predictive performance of commonly employed machine learning models for estimating coke moisture content was also assessed.

2 COKE IMAGES AT VARIED MOISTURE CONTENT

To investigate the relationship between coke images and moisture content, it is necessary to obtain coke images along with their corresponding moisture content data. Figure 1 illustrates the experimental process, encompassing steps such as soaking, drying, moisture content measurement, and image acquisition. In this experiment, the measurement of coke moisture content data was conducted using the commonly employed oven-drying method in metallurgical processes. Due to the significant influence of lighting on image features, to ensure consistent lighting conditions for the images, their acquisition was conducted within a light-sealed darkroom, illuminated by a constant LED light source.

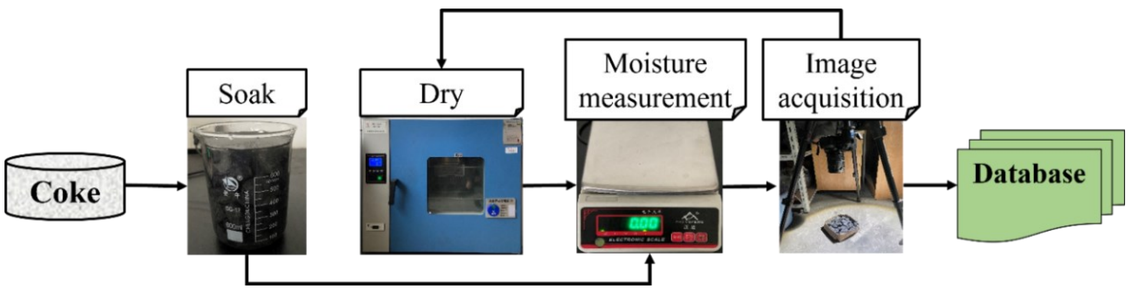

Figure 1: Flowchart of coke images acquisition and moisture content determination.

The images had a resolution of 300 dpi both horizontally and vertically, and they were cropped to a square region as depicted in Figure 2. Totally, 160 coke images and their moisture content data were obtained, and the coke moisture content ranged from 0.13% to 13.50%. Figure 3 shows an example of the relationship between coke moisture content and image appearance. The colour of the coke images was almost black-and-white, but the changes were still noticeable: as the coke moisture content decreased from 13.50 % to 0.19 %, the coke colour changed from dark to light and the texture became clear.

Figure 2: Sample of interested region of coke image.

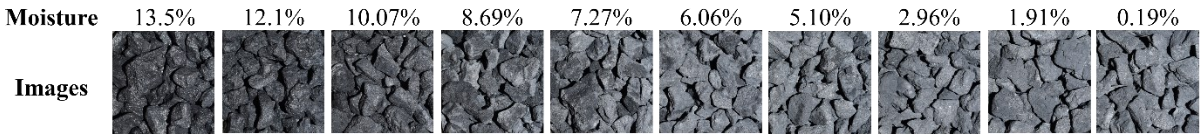

Figure 3: Sample of images of coke with different moisture content.

From an optical point of view, colour is produced by objects reflecting different wavelengths of light. The more light an object reflects, the lighter the colour [10]. There are two main reasons for the coke colour changes. First, the water film covering the coke surface causes part of the light to be reflected back to the interior, resulting in an overall brightness decrease. Simultaneously, moisture absorbed some of the light, which reduced reflected light as well. Second, as the coke moisture content increases, powder tends to adhere to the coke surface [9]. Consequently, as the moisture content decreased, the coke images displayed increased brightness in colour and enhanced clarity in texture.

3 ANALYSIS OF COKE IMAGE FEATURES

The attributes of images at various moisture content levels can be analysed through colour features and texture features. Colour spaces are mathematical models or coordinate systems used to define and describe colours. They enable us to accurately represent and manipulate colour information mathematically, facilitating various colour adjustments, analyses, and manipulations of images. The RGB images [11] that were obtained were converted to HSV space [12] and Lab space [13]. Nine colour features were extracted: red channel average (R), green channel average (G), blue channel average (B), lightness (L), a colour channel (a), b colour channel (b), colour point average (H), saturation average (S) and luminance average (V).

Texture is the spatial arrangement of the colours and pixels, which can reflect the distribution of details and patterns of the object surface [14]. The texture features were analysed based on the statistic grey level co-occurrence matrix (GLCM) [14] and five texture features were obtained: angular second moment (ASM), entropy, contrast, correlation, and inverse differential moment (IDM).

In general, colour and texture feature variables exhibit varying dimensions and orders of magnitude. To standardize these variables, the data was normalized. The colour and texture features after normalization are compared in Figure 4 (a) and (b), respectively. As the coke moisture content decreased, the colour features H and S exhibited a gradual decrease. The colour features R, G, B, L, and V were nearly coincident, showing an initial increase followed by stabilization. In turn, the colour features a and b displayed an initial decrease followed by stabilization. The textural feature of contrast initially increased but then levelled out, while correlation exhibited the opposite trend. ASM and IDM decreased initially followed by an increase, i.e., a V-shaped pattern, while entropy showed the opposite tend, i.e., an inverted V-shaped pattern. The turning point occurs at approximately a moisture content of around 6%.

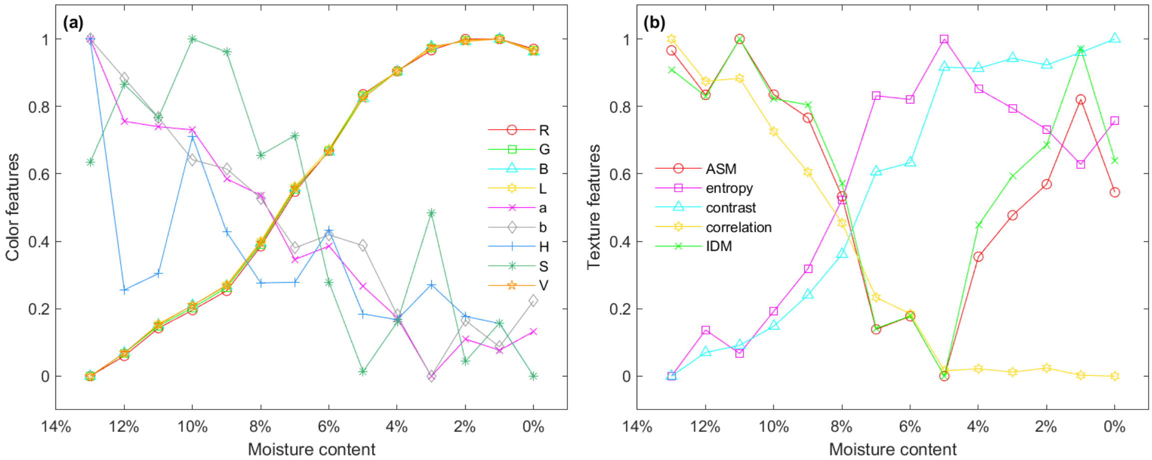

Figure 4: Changes of normalized (a) colour and (b) texture features in response to the decline of coke moisture content.

The correlation between image features and coke moisture content were determined by Pearson correlation coefficient and the results are shown in Figure 5. Figure 5(a) represents the correlation analysis for all moisture content data, and additionally, due to the inflection point in image features occurring at a moisture content of 6%, data analysis for moisture content less than 6% is carried out in Figure 5(b). The colours of the bars represent high (red), moderate (blue), and low (green) correlation. Consequently, for all moisture content data, the moisture content exhibited a significant positive correlation with image features a, b, and correlation, but a significant negative correlation with features R, G, B, L, V, and contrast. The results indicate that both the brightness and texture of the coke images were significantly influenced by the moisture content. Notably, as the moisture content decreased, the images appeared brighter and the texture became rougher. However, when the moisture content is less than 6%, the correlation between image features and moisture content is very low. It can be anticipated that the moisture model's predictive performance will be poor for this subset of data.

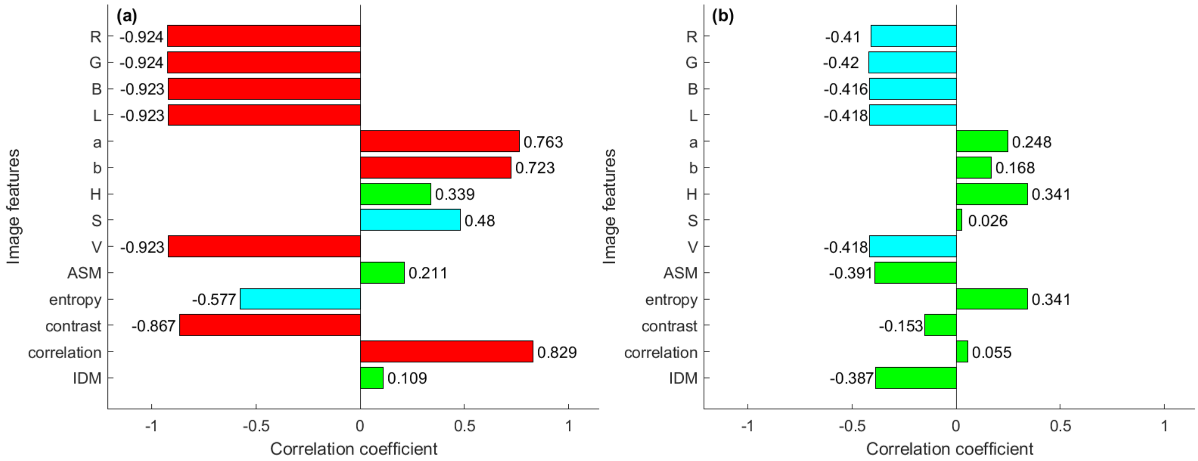

Figure 5: Correlation coefficients between image features and coke moisture content (a) moisture content range from 0.13% to 13.50% and (b) moisture content less than 6%.

Nine highly correlated features were chosen for developing the moisture content prediction model. As the image features exhibited internal correlations, indicating significant information overlap between them, Principal Component Analysis (PCA) [15] was used to eliminate redundancy among the feature variables. The distribution of the first two principal components for the training sets and test sets are shown in Figure 6, with points from the training set indicated by red circles and points from the test set by blue asterisks. The first principal component explained 86.6% of the variation in the moisture content, while the second explained 8.6%, so the first two together captured 95.2% of the variation.

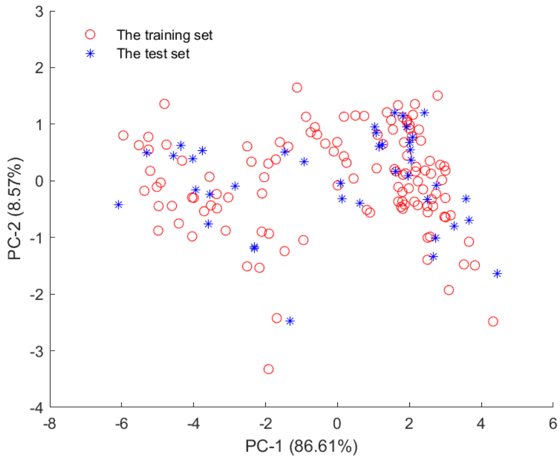

Figure 6: Principal components (PC-1 and PC-2) derived from both the training sets and test sets. Red circles indicate points in the training set, blue asterisks points in the test set.

4 ESTABLISHMENT AND COMPARISON OF PREDICTIVE MODELS

In this study, the linear Partial Least Squares Regression (PLS) and nonlinear Support Vector Machine (SVM) and Random Forest (RF) were used as model candidates to predict the coke moisture content. These models took either the nine (highly correlated) moisture-related features (R, G, B, L, a, b, V, contrast, and correlation) or the two principal component variables as inputs, with moisture content data as the desired output. The 160 samples were divided into a training set (120 samples) and a test set (40 samples). The training set was used to develop the prediction model for coke moisture content, while the test set was used to evaluate the model's performance.

The model's performance was assessed using the coefficient of determination (R2), the root mean square error (RMSE) and relative percent deviation (RPD). Specifically, R2 and RMSE were calculated for both the training set (Rt, RMSEt) and the test set (Rp, RMSEp).

(1) Linear correlation model PLS

PLS was originally introduced in econometrics and has gained extensive usage in mathematical modelling [16]. PLS searches for a linear regression model explaining the relation between the inputs and the output. This approach effectively addressed issue such as collinearity, multi-dependent variable analysis, and small sample size challenges. Figure 7 (a) shows the percentage of variance explained in the moisture content as a function of the number of PLS components. A single PLS component explains 85.18% of the variance in the moisture content, with the percentage of variance showing a slight increase as the number of PLS components increased. With nine PLS components, the maximum percentage of variance reached 88.93%. However, the determination of the appropriate number of PLS components should be validated using the cross-validation method. The outcome of the 9-fold cross-validation is depicted in Figure 7 (b), revealing that a single component sufficed for accurate prediction.

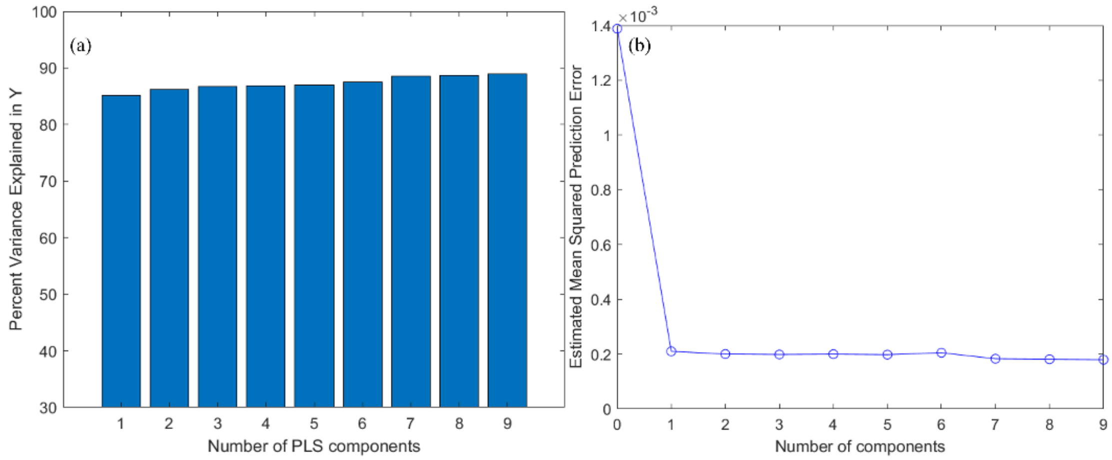

Figure 7: (a) Explained percentage of variance in the moisture content, and (b) estimated mean squared prediction error for different numbers of PLS components.

Figure 8 (a) and (b) show the results of the one-component PLS model and the nine-component PLS model, respectively, with points from the training set indicated by red circles and points from the test set by blue asterisks. From the two scatter plots, it can be observed that when the moisture content of coke is greater than 6%, the data points follow a diagonal distribution which indicates an accurate prediction. Furthermore, in comparison to the data points of the one-component PLS model, the data points of the nine-component PLS model display more significant dispersion. This demonstrates that the model trained with one component performs better. When the moisture content of coke is less than 6%, the data points exhibit a horizontal distribution. This reveals that the model can only distinguish whether the moisture content is less than 6% or not, rather than accurately predicting the values of moisture content. Combining the correlation coefficient analysis with Figure 5(b), it can be concluded that the correlation between image features and moisture content is relatively weak when the moisture content of coke is less than 6%. For a model with one PLS component, Rp = 0.8300, RMSEp = 0.0153 and RPD = 2.3264, while the values for the nine-component model were Rp = 0.8461, RMSEp =0.0145 and RPD = 2.3816. The difference in Rp and RMSEp between the two models must be considered small, and the RPD values fall between 2.0 and 2.5. These indicates that a model based on only one PLS components is parsimonious.

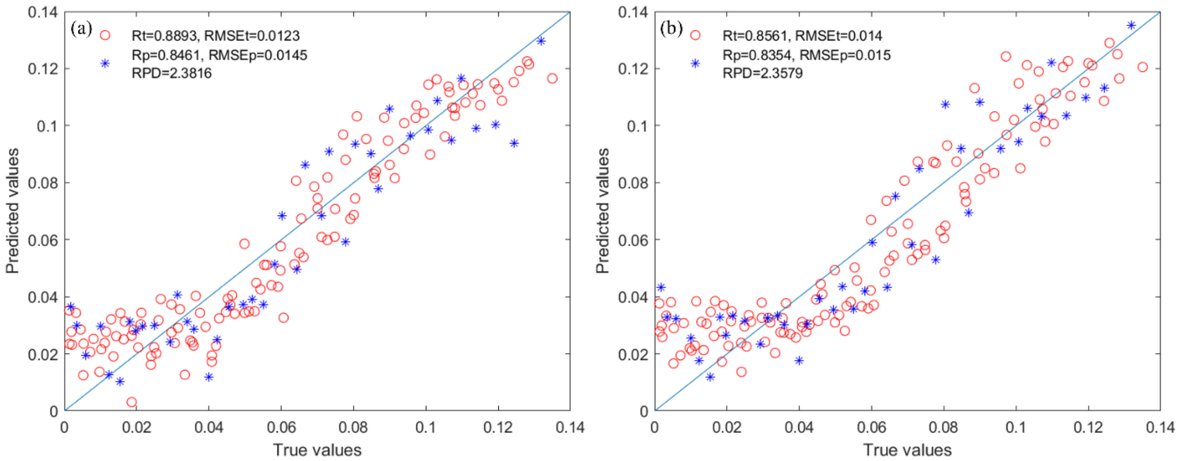

Figure 8: Results of the PLS model using (a) 1 component and (b) 9 components. Red circles indicate points in the training set, blue asterisks points in the test set.

(2) Nonlinear correlation model SVM

SVM is a versatile and robust modelling and analysis method. Initially being designed for addressing binary classification problems, SVM was later extended to nonlinear regression tasks by the utilization of kernel functions [17]. SVM is capable of addressing machine learning problems with limited sample sizes and can effectively manage nonlinear relationships among features. The SVM model of the present study was established using two principal components as inputs and coke moisture content as the desired output. The SVM model used Radial Basis Function (RBF) as the kernel function and the influence of penalty parameters (c) and kernel function parameters (g) on the modelling results were considered. Grid searching technique and cross validation were used to conduct a global optimization on parameters c and g, with results shown in Figure 9 (a). The combination $c = 1.4142$ and $g = 0.0769$ resulted in the minimal mean squared error, MSE = 0.0102. For this case, Figure 9 (b) depicts the results of the model, corresponding to Rp = 0.8472, RMSEp = 0.0145 and RPD = 2.3639. When the moisture content is greater than 6%, the true value - predicted value data points in the graph follow the diagonal line (blue line), indicating a strong predictive performance of the model. The predictive results are still not satisfactory when the moisture content is less than 6%.

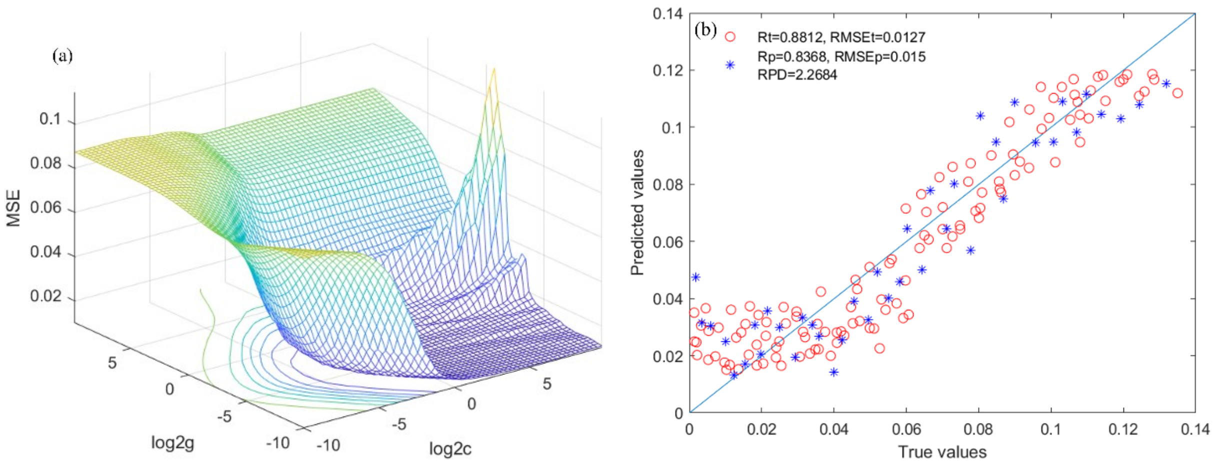

Figure 9: (a) MSE for SVM in model parameter (*c* and *g*) optimization; (b) Results by optimal SVM model. Red circles indicate points in the training set, blue asterisks points in the test set.

(3) Nonlinear correlation model RF

RF is a decision tree-based machine learning algorithm that can be used to solve classification and regression problems, and generally possesses high classification accuracy, toleration to outliers and noise [15]. The RF model was established using two principal components as inputs and coke moisture content as the desired output. The performance of the RF model is directly influenced by the number of leaves and the number of trees. The MSE of the RF models, corresponding to different numbers of leaves (5, 10, 20, 50, 100, 200 and 500) and trees, is depicted in Figure 10 (a). For five leaves, the MSE experienced a rapid initial decrease, stabilizing as the number of trees exceeded 100. As a result, the RF model's configuration was selected as 5 leaves and 100 trees. The results of the model are shown as Figure 10 (b), with the performance indices Rp = 0.8156, RMSEp = 0.0159 and RPD = 1.9143. The scatter plot of true value - predicted value in the RF prediction model exhibits classification features, with clustered points appearing around moisture content levels of approximately 6% and 10%. The moisture content boundary around 6% is also observed in other models, which is attributed to the weak correlation between image features and moisture content of coke. The boundary point at 10% moisture content indicates the presence of underfitting or overfitting in the model, leading to poor performance across different ranges.

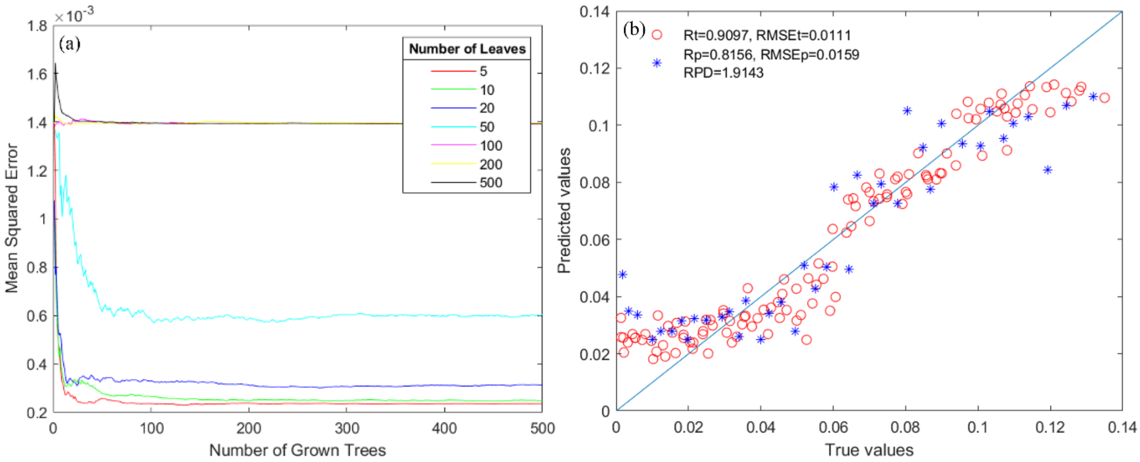

Figure 20: (a) RF model parameters leaves (5, 10, 20, 50, 100, 200 and 500) and trees optimization; (b) Results of the RF model. Red circles indicate points in the training set, blue asterisks points in the test set.

(4) Model comparison

The performance of the linear correlation model PLS and the nonlinear correlation models SVM and RF is compared and summarized in Table 1. The RF model exhibited the best values for Rt (= 0.9097) and RMSEt (= 0.0111), followed by SVM and PLS. However, the Rp value for the RF model was the lowest (0.8156), indicating a potential overfitting, while the Rp values for the SVM and PLS models were comparable. The performance of the RF model in this study is not satisfactory. Considering the characteristics of the RF model, it tends to excel in dealing with high-dimensional, nonlinear, and large-scale datasets. In contrast, the present study indicates a linear relationship between coke moisture content and specific image features, and the dataset has low dimensionality and a relatively small scale. So, employing the RF model may not be suitable for this research context. Furthermore, the RMSEp values for the SVM and PLS models were identical (0.0150). The PLS model achieved the highest RPD value but the SVM model was close. These values all fall within the range of 2.0 and 2.5, providing a rough quantitative estimate of coke moisture content. Moreover, based on Figure 8 (b) and 9 (b), it is evident that the predictions of the PLS model exhibited significant dispersion, while that of the SVM model showed better alignment with the diagonal. This observation suggests that the PLS model's ability to generalize was worse due to overfitting, rendering it less effective in making accurate predictions on independent samples. Consequently, the results indicate that the nonlinear SVM model had superior prediction accuracy and generalization capabilities for predicting coke moisture content based on the data used in this study.

Table 1: Performance indices for the coke moisture content prediction models

Methods	Training Set		Test Set		
	Rt	RMSEt	Rp	RMSEp	RPD
PLS	0.8518	0.0142	0.8300	0.0150	2.3264
SVM	0.8812	0.0127	0.8368	0.0150	2.2684
RF	0.9097	0.0111	0.8156	0.0159	1.9143

5 CONCLUSIONS

The correlation between the coke moisture content and colour and texture features of images of coke particles was studied. The findings indicate a strong correlation between the specific image features and the moisture content of the coke, particularly when the moisture content exceeds 6%. The lower the moisture content of coke, the higher the brightness of the image, and the clearer the texture. This observation is consistent with the reflection and refraction of light by the moisture on the surface of coke, as well as the pattern of dust adhering to its surface. Three commonly used machine-learning models were tested and compared for predicting coke moisture content using image features as inputs. Principal component analysis was first applied to remove redundance among the inputs. The results indicate that the nonlinear models performed better in predicting the coke moisture content. In particular, the SVM model demonstrates superior prediction accuracy and robustness. It holds promise for future application. However, for this method to be applicable in the practical production, the model needs to be validated on more extensive data sets and the effect of the plant environment on the quality of the images should also be assessed.

REFERENCES

[1] L. Zhang, G. Wang, Q. Xue, H. Zuo, X. She and J. Wang, "Effect of preheating on coking coal and metallurgical coke properties: A review", *Fuel Processing Technology* **221,** 106942, (2021).

[2] M.A. Díez, R. Alvarez. and C. Barriocanal, "Coal for metallurgical coke production: predictions of coke quality and future requirements for cokemaking", *International Journal of Coal Geology* **50**, 1–4, (2002).

[3] A. Ghosh, T.K. Das, A.K. Sharma, R. Mukherjee, A. Bhushan and S.S. Palit, "Microwave-assisted infrared thermography: A tool for quality assessment of blast furnace feeds", *Infrared Physics & Technology* **114**, 103640, (2021).

[4] R.V. Williams, *Control and analysis in iron and steelmaking*, Elsevier, p. 91, (2016).

[5] H. Tominaga, N. Wada, N. Tachikawa, Y. Kuramochi, S. Horiuchi, Y. Sase, A. Hiro, O. Naotake and N. Hiroshi, "Simultaneous Utilization of Neutrons and-rays Ifrom 252Cf for Measurement of Moisture and Density", *The International Journal of Applied Radiation and Isotopes* **34**, 429-436, (1983).

[6] Z. Zhang, Y. Liu, Q. Hu, Zh. Zhang, L. Wang, X. Liu and X. Xia, "Multi-information online detection of coal quality based on machine vision", *Powder Technology* **374**, 250-262, (2020).

[7] C.A. Perez, P.A. Estévez, P.A. Vera, L.E. Castillo, C.M. Aravena, D.A. Schulz and L.E. Medina, "Ore grade estimation by feature selection and voting using boundary detection in digital image analysis", *International Journal of Mineral Processing* **101**, 28-36, (2011).

[8] M. Li, X. Wang, H. Yao, H. Saxén and Y. Yu. "Analysis of Particle Size Distribution of Coke on Blast Furnace Belt Using Object Detection", *Processes* **10**, 1902, (2022).

[9] T. Tsuboi, N. Yamahira, T. Nishino and H. Komatsubara, "Real-time measurement method for powder ratio of coke using camera image", *Measurement: Sensors* **18**, 100176, (2021).

[10] J. Lekner, and M.C. Dorf, "Why some things are darker when wet", *Applied Optics* **27**,

1278-1280, (1988).

[11] S. Süsstrunk, R. Buckley, and S. Swen, “Standard RGB color spaces”, In: 7th color imaging conference on fnal progress proceedings on society for imaging science and technology, p. 127–134, (1999).

[12] S. Kolkur, D. Kalbande, P. Shimpi, C. Bapat and J. Jatakia, “Human skin detection using RGB, HSV and YCbCr color models”, *ArXiv Preprint* arXiv:1708.02694, (2017).

[13] K. León, D. Mery, F. Pedreschi and J. León, “Color measurement in L∗a∗b∗ units from RGB digital images”, *Food Research International* **39**, 1084-1091, (2006).

[14] O.M. Maktabdar, M.A. Maarof, M.F. Rohani, A. Zainal and S.Z.M. Shaid, “An optimized skin texture model using gray-level co-occurrence matrix”, *Neural Computing and Applications* **31**, 1835-1853, (2019).

[15] N. Nakawajana, P. Lerdwattanakitti, W. Saechua, J. Posom, K. Saengprachatanarug and S. Wongpichet, “A low-cost system for moisture content detection of bagasse upon a conveyor belt with multispectral image and various machine learning methods”, *Processes* **9**, 777, (2021)

[16] F.C.B. Bedin, M.V. Faust, G.A. Guarneri, T.S. Assmann, C.B.B. Lafay, L.F. Soares, A.V. Paulo and M.S. Larissa, “NIR associated to PLS and SVM for fast and non-destructive determination of C, N, P, and K contents in poultry litter”, *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy* **245**, 118834, (2021).

[17] G. Liang, C. Dong, B. Hu, H. Zhu, H. Yuan, Y. Jiang and G. Hoa, “Prediction of moisture content for Congou black tea withering leaves using image features and nonlinear method”, *Scientific Reports* **8**, 7854, (2018).