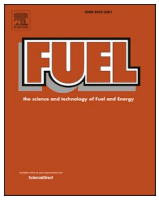
Fuel 384 (2025) 133953

Contents lists available at ScienceDirect

Fuel

journal homepage: www.elsevier.com/locate/fuel



Full Length Article

Predicting the hardgrove grindability index using interpretable decision tree-based machine learning models

Yuxin Chen [a], Manoj Khandelwal [b,*], Moshood Onifade [b], Jian Zhou [a], Abiodun Ismail Lawal [c], Samson Oluwaseyi Bada [d,*], Bekir Genc [e]

[a] School of Resources and Safety Engineering, Central South University, Changsha 410083, China
[b] Institute of Innovation, Science and Sustainability, Federation University Australia, Ballarat, VIC 3350, Australia
[c] Department of Mining Engineering, Federal University of Technology, Akure, Nigeria
[d] School of Chemical and Metallurgy, Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, South Africa
[e] School of Mining Engineering, Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, South Africa

ARTICLE INFO

Keywords:
Hardgrove grindability index
Machine learning
Hyperparameter optimization
Interpretation analysis
NGBoost

ABSTRACT

The Hardgrove grindability index (HGI) is a crucial indicator for assessing the grindability of coal, and accurate prediction of HGI is essential for improving the production efficiency and economic benefits of the coal industry. This study employed six decision tree-based machine learning models to predict the HGI values of 129 coal samples, with hyperparameter optimization performed using Optuna, and model interpretability analyzed using SHapley Additive exPlanations (SHAP). The results showed that the optimized natural gradient boosting (NGBoost) model outperformed all other models, which achieved the highest performance on the test set with a coefficient of determination (R^2) of 0.9715, a mean absolute error (MAE) of 1.1507, and a root mean squared error (RMSE) of 1.4735. SHAP analysis further revealed that volatile matter (*VM*) contributed the most to the model's predictions, while pyrite (FeS_2) had the least contribution. This study provides an efficient machine learning approach for accurate HGI prediction, offering excellent predictive performance, interpretability, and application value.

1. Introduction

Coal still occupies an important position in the global energy structure and plays an irreplaceable role in global energy supply, economic development, and industrial production [24,27,73]. The utilization of coal requires grinding it into fine particles, and the grindability of coal is crucial in coal processing and usage as it directly affects the efficiency of pulverizers, the particle size distribution of coal powder, combustion characteristics, and equipment wear [30,68]. The grindability of coal refers to the ease with which it can be pulverized into fine powder in mechanical equipment, typically measured by the Hardgrove Grindability Index (HGI) [62,72,90]. The HGI is determined by conducting experiments using standardized Hardgrove equipment following a specified procedure [19,22]. A higher HGI value indicates that the coal is easier to pulverize, while a lower value indicates greater difficulty [75].

Although HGI testing equipment is relatively inexpensive and easy to maintain, the measurement procedure is quite cumbersome [15,58]. According to the American Society for Testing and Materials (ASTM) International standard procedure[5], each machine can only perform six tests per hour. Additionally, issues such as non-standardized equipment, variations in sample preparation, and the repeatability and reproducibility of the test make determining the HGI value a time-consuming and complex process, posing difficulties for large-scale and high-frequency measurements [70,79]. Therefore, researchers have analyzed large datasets to identify factors affecting the HGI and have employed artificial intelligence methods, such as machine learning (ML), to predict the HGI value of coal, thereby improving measurement efficiency and accuracy. These studies indicate that the chemical composition and physical properties of coal, such as moisture, ash, and volatile matter, significantly impact the HGI value [44,45,57]. ML methods, such as support vector regression (SVR) and artificial neural networks (ANN), can effectively predict HGI values. Sengupta [62] conducted a regression analysis using parameters such as moisture content, volatile matter, fixed carbon, and ash to study the grindability of over 300 Indian coal

* Corresponding authors.
E-mail addresses: 245501013@csu.edu.cn (Y. Chen), m.khandelwal@federation.edu.au, mkhandelwal1@gmail.com (M. Khandelwal), j.zhou@csu.edu.cn (J. Zhou), ailawal@futa.edu.ng (A. Ismail Lawal), Samson.Bada@wits.ac.za (S. Oluwaseyi Bada), Bekir.Genc@wits.ac.za (B. Genc).

https://doi.org/10.1016/j.fuel.2024.133953
Received 12 August 2024; Received in revised form 6 October 2024; Accepted 27 November 2024

Nomenclature

Mathematical term

MAE	Mean absolute error
MAE'_{norm}	The standardized and inverted MAE
MSE	Mean squared error
N	The total number of actual HGI values
R^2	Coefficient of determination
$RMSE$	Root mean square error
$RMSE'_{norm}$	The standardized and inverted RMSE
S	Comprehensive performance score
X	Original metric value
$X_{\max}$	The maximum value of the metric
$X_{\min}$	The minimum value of the metric
X_{norm}	The standardized metric value
X'_{norm}	The metric value after the inversion process
y_i	the i-th actual HGI value
$\hat{y}_i$	the i-th predicted HGI value
$\bar{y}$	The mean of all actual HGI values

Hyperparameter

colsample_bytree The column sampling ratio (where each column represents a feature) when constructing each tree.

col_sample The proportion of features sampled when building each tree in NGBoost, similar to colsample_bytree in models like XGBoost.

iterations The total number of boosting iterations or decision trees in the model training process.

l2_leaf_reg The L2 regularization term applied to the leaf weights in CatBoost, used to prevent overfitting by penalizing large weights.

learning_rate Controls the step size for updating parameters during each iteration of the model.

max_depth/depth The maximum depth or the maximum number of levels of the tree.

min_child_samples The minimum number of samples required in a leaf node, similar to *min_samples_leaf*, but commonly used in LGBM.

min_child_weight The minimum sum of instance weights required in a leaf node.

min_samples_leaf The minimum number of samples required in a leaf node.

min_samples_split The minimum number of samples required to split an internal node.

minibatch_frac The proportion of the total data used for training in each iteration, similar to *subsample*.

n_estimators The number of decision trees or weak learners.

subsample The proportion of samples used when training each tree.

Abbreviation

A	Ash
ANFIS	Adaptive neuro-fuzzy inference systems
ANN	Artificial neural networks
ASTM	American Society for Testing and Materials
BPNN	Back propagation neural network
CatBoost	Categorical boosting
CV	Calorific value
FC	Fixed carbon
FeS_2	Pyrite
GA	Genetic algorithm
GBDT	Gradient boosting decision trees
GRNN	Generalized regression neural networks
HGI	Hardgrove grindability Index
LGBM	Light gradient boosting machine
LSTM	Long short-term memory networks
M	Moisture
MAE	Mean absolute error
ML	Machine learning
MSE	Mean squared error
NGBoost	Natural gradient boosting
OPENN	Optimized evolutionary neural network ()
PSO	Particle swarm optimization
Q	Quartz
R^2	Coefficient of determination
RF	Random forest
R_{max}	Vitrinite maximum reflectance; the percentage of incoming light reflected back to a photomultiplier, usually using oil-immersion optics
RMSE	Root mean square error
SHAP	SHapley Additive exPlanations
St.D	Standard deviations
SVR	Support vector regression
TS	Total sulphur
VM	Volatile matter
XGBoost	Extreme gradient boosting

samples, deriving an equation known as the “Statistical Grindability Index”. However, due to the heterogeneity of coal, no universal HGI equation can be applied to all coal samples [28]. Rao and Gopalakrishna [58] trained SVR models with limited measurement data and validated the nonlinear relationship between inputs and outputs, demonstrating that SVR can effectively establish the correlation between coal composition and HGI with a relatively small dataset. Khoshjavan et al. [35] studied the estimation of HGI values for 400 coal samples using ANN and linear multivariable regression methods, finding that the ANN method performed better. Ürünveren et al. [70] predicted the HGI values of Afşin-Elbistan low-grade coal using linear regression, feed-forward neural networks, generalized regression neural networks (GRNN), and Elman network methods. The results showed that the GRNN model was the most accurate and could be used to predict the HGI values of coal from similar deposits. Yazdani et al. [78] developed an optimized evolutionary neural network (OPENN) model to predict the HGI using property selection and optimization algorithms accurately. The test results demonstrated that the model outperformed other models in predicting HGI for Kentucky coal samples, showcasing significant economic value. Lawal et al. [38] developed predictive models for the HGI of South African coalfields using long short-term memory (LSTM) networks, SVR, and ANN methods. They found that ANN performed the best, with HGI being primarily influenced by fixed carbon, calorific value, percentage ash yield, and volatile matter, while pyrite had the least impact.

Neural network methods excel in handling large-scale data. However, their complexity and training requirements often result in long training times. When trained on small-scale datasets, they are prone to overfitting and exhibit limited generalization capability [61]. Additionally, neural network models are often considered “black-box” models, making their internal workings difficult to understand [9,20,50]. In contrast, decision tree models are characterized by low bias and high variance. They effectively capture limited feature variations in small datasets and make detailed splits for each sample, thus improving model performance[2]. Furthermore, decision tree models do not require complex data preprocessing, reducing reliance on large-scale data processing, and enabling efficient operation. Decision tree models can handle both classification and regression tasks, and they can

adaptively split based on data characteristics. Even with limited data, they can effectively learn by selecting optimal feature split points. By adjusting hyperparameters, decision tree models can effectively resist noise and reduce the risk of overfitting, thereby maintaining good generalization performance on small datasets. Even when the sample size is small, proper pruning and hyperparameter tuning can ensure the model's stability and performance [37]. Therefore, decision tree-based ML models provide key advantages in interpretability, property selection, handling missing values, data preprocessing, and training speed, making them more suitable for small datasets [59,64]. Decision tree-based ML models perform well in many applications, but to enhance model performance and stability, they are often combined with ensemble methods such as random forest (RF) and gradient boosting decision trees (GBDT). GBDT, RF, extreme gradient boosting (XGBoost), light gradient boosting machine (LGBM), categorical boosting (CatBoost), and natural gradient boosting (NGBoost) are widely adopted or relatively novel decision tree-based ML models. These models demonstrate exceptional performance in handling nonlinear relationships and high-dimensional data, making them particularly well-suited for complex engineering problems. They have been extensively applied in both industrial and academic settings [13,16,40,48,74,84,91–93]. Matin et al. [43] successfully predicted the HGI of Kentucky coal samples using variable importance measurements (including proximate, ultimate analysis, and petrography) processed by the RF model. The results demonstrated that the RF model had high accuracy, validating its reliability and practicality in assessing complex relationships in coal processing. Hower et al. [26] noted that developing artificial intelligence models for HGI prediction using interpretable methods like SHapley Additive exPlanations (SHAP) can reveal complex relationships between variables and enhance the understandability of prediction models. Chelgani [17] predicted the total heating value of Armenian coal samples using SHAP combined with XGBoost. Rzychoń [60] predicted the HGI of Polish coal samples using the XGBoost model and employed SHAP techniques to interpret the main influencing factors. The aforementioned studies demonstrate that tree-based ML models, combined with explainable artificial intelligence methods, exhibit high feasibility in predicting coal HGI and analyzing complex relationships. Given the successful application of these models in the coal industry and their advantages in handling small datasets, this study selected six tree-based ML models: GBDT, RF, XGBoost, LGBM, CatBoost, and NGBoost. By incorporating the physicochemical properties of coal and other relevant parameters, these models can rapidly and accurately predict HGI. This effectively reduces time costs while significantly improving prediction accuracy, providing more economical and efficient solutions for industrial production.

In summary, this study selected six decision tree-based ML models—GBDT, RF, XGBoost, LGBM, CatBoost, and NGBoost—to predict HGI values. To enhance model performance and stability, hyperparameter tuning was performed using Optuna and cross-validation. The dataset contains 129 samples, with input properties including various chemical and physical properties of coal, such as calorific value (*CV*), moisture (*M*), ash (*A*), volatile matter (*VM*), fixed carbon (*FC*), total sulphur (*TS*), pyrite (FeS_2), and quartz (*Q*). The HGI is used as the output variable. The predictive performance of the models was evaluated and compared using metrics such as coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). The optimal HGI prediction model was then determined, and the SHAP method was used to interpret and analyze the model's predictions.

2. Methods

2.1. Machine learning (ML) model

2.1.1. Gradient boosting decision tree (GBDT)

GBDT is an ensemble learning model based on the gradient boosting framework, as proposed by Friedman [23]. It builds a strong predictive model by combining multiple weak learners, typically decision trees. The core idea is that each newly added model aims to correct the prediction error of the previous model, which is achieved by minimizing the gradient descent optimization of the loss function [85]. GBDT has high prediction accuracy and flexibility, and it can deal with complex property relations and nonlinear problems. However, it requires careful tuning of multiple hyperparameters to achieve optimal performance.

2.1.2. Random forest (RF)

RF is an ensemble learning method that improves model performance and stability by constructing multiple decision trees and combining their prediction results [12,69,92]. Its basic principle is to perform bootstrap sampling from the original dataset to generate multiple subsets and build a decision tree on each subset. During the node-splitting process, a random subset of properties is selected, and the best property among them is used for the split. Finally, random forest aggregates the results of the individual decision trees by majority voting (for classification) or averaging (for regression) to produce the final prediction.

2.1.3. Extreme gradient boosting (XGBoost)

XGBoost, proposed by Chen et al. [18], is an ensemble learning method based on the GBDT framework. By optimizing computational efficiency and model performance within the gradient boosting framework, XGBoost has been widely applied in various machine learning prediction tasks[82]. Compared to traditional GBDT, XGBoost introduces L1 and L2 regularization to better control model complexity, thereby improving generalization ability and stability.

2.1.4. Light gradient boosting machine (LGBM)

LGBM, proposed by the Microsoft team in 2017, is an improved machine-learning algorithm based on the GBDT framework[34]. The core idea of LGBM is similar to traditional GBDT: a series of decision trees are gradually constructed, with each tree attempting to correct the errors of the previous one, thereby enhancing the overall performance of the model. Compared to traditional GBDT, LGBM adopts a histogram-based method and gradient histogram acceleration technique, which discretizes continuous properties into integers (binning) to reduce memory consumption and computational complexity. It also employs a leaf-wise growth strategy, selecting the leaf node with the highest gain for splitting to better fit complex data[54].

2.1.5. Categorical boosting (CatBoost)

CatBoost, introduced by Yandex, is a gradient-boosting algorithm based on symmetric decision trees, suitable for both classification and regression problems[51]. CatBoost integrates a symmetric tree structure and automatic categorical property handling, thereby simplifying the data preprocessing process and effectively improving the model's generalization ability [11]. Compared to traditional GBDT algorithms, CatBoost offers advantages such as effective handling of categorical properties, prevention of target leakage, and dynamic learning rate adjustment.

2.1.6. Natural gradient boosting (NGBoost)

NGBoost, a probabilistic prediction model proposed in 2019, combines the gradient boosting framework with the concept of natural gradient optimization[21]. During training, NGBoost incorporates natural gradients into conventional gradient methods, enhancing accuracy for small sample data while maintaining probabilistic advantages. Compared to traditional GBDT, NGBoost not only provides point predictions but also generates predictive distributions, offering higher optimization efficiency and flexibility[76,89,88,42,52,53,55]. However, this also increases the model's computational complexity and the difficulty of hyperparameter tuning. NGBoost performs excellently in regression tasks, providing efficient and accurate probabilistic predictions. For a detailed explanation and theoretical background of

NGBoost, please refer to the study by Duan et al. [21].

2.2. Hyperparameter optimization methods and evaluation metrics

Optuna is an automated hyperparameter optimization framework designed for ML models, proposed by Akiba et al. [1]. By defining optimization objectives and search spaces, and utilizing advanced algorithms such as the Tree-structured Parzen Estimator and Covariance Matrix Adaptation Evolution Strategy, Optuna efficiently searches for optimal hyperparameter configurations. Optuna employs Bayesian optimization methods, dynamically adjusting the search space during the optimization process to enhance both search efficiency and effectiveness. Its flexible and lightweight design makes it easy to integrate with various machine-learning frameworks [36]. The Optuna framework is based on three core concepts: objective functions, trials, and studies[77]. The objective function defines the optimization goal and parameter range. A trial represents a single evaluation of the objective function. A study oversees the entire optimization process, including strategy determination, iteration of trials, and result recording.

After model training is completed, the model needs to be evaluated to verify its predictive ability on unseen data. This requires using a test dataset and examining the results with a series of evaluation metrics. The evaluation metrics selected for this study include the coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) [39,42,56,80,81,83,91] to assess the relationship between the predicted HGI and the actual values.

$$R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \tag{1}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{2}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{3}$$

In the formulas, N represents the total number of actual values, y_i is the i-th actual value, $\hat{y}_i$ is the i-th predicted value, and $\bar{y}$ is the mean of all actual values.

2.3. Model construction

The construction process of the interpretable ML model for HGI is as follows (see Fig. 1):

(1) The HGI dataset is randomly divided into a training set and a test set in a 4:1 ratio.
(2) The hyperparameter search space for six decision tree-based ML models is defined, and the mean squared error (MSE) from cross-validation is used as the fitness metric. To balance training time and model evaluation stability, this study employs 5-fold cross-validation. Specifically, in each iteration, the training set is randomly divided into five equal parts. Four parts are used for model training, and one part is used for validation, with each fold serving as a validation set in turn. The average MSE from the five validation sets is then taken as the fitness value for that iteration. Optuna is used to perform 200 optimization iterations to identify the optimal hyperparameter combination.
(3) The ML models are trained using the optimal hyperparameters, predictions are made on both the training and test sets, and

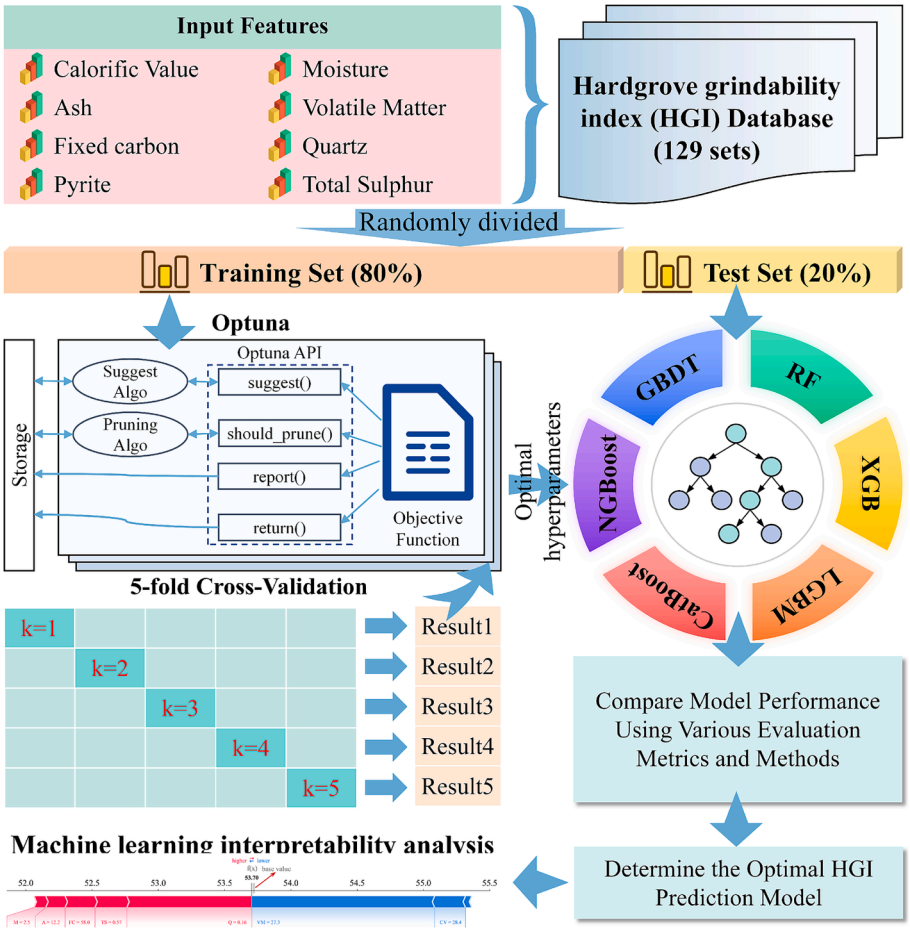


Fig. 1. Model construction process.

evaluation metrics are computed. Compare the predictive performance of different ML models to determine the best predictive model for HGI.

(4) Conduct an interpretability analysis of the model predictions to identify the important properties influencing them.

3. Data description

The Witbank Coalfield, located in the Mpumalanga province of South Africa, is one of the country's most significant coal-producing regions. Its geological background and coal type have made it a key player in South Africa's coal mining industry, which supports power generation and export markets. The Witbank Coalfield forms part of the Karoo Supergroup, a vast sequence of sedimentary rocks that were deposited during the Permian Period. The coal seams in the Witbank area are found within the Vryheid Formation of the Ecca Group, a unit of the Karoo Supergroup. This formation contains alternating layers of coal, shale, sandstone, and siltstone. Deposition occurred in fluvial and deltaic environments, where thick peat swamps formed under the right climatic and sedimentary conditions. Over time, these peats were buried and subjected to pressure and heat, forming coal. The coalfield is relatively flat-lying with minor tectonic disturbance, making it easier for extraction. The coal seams in the Witbank Coalfield are generally thick, ranging from 1 m to over 6 m in places. The coalfield has five major coal seams, identified as Seams No. 1 to No. 5, with No. 2 and No. 4 being the most economically viable for extraction. The Witbank coal is classified as bituminous coal, which is a medium-rank coal. It is relatively high in carbon content and calorific value, making it suitable for power generation, industrial use, and export. The coal is sub-categorized based on its carbon, moisture, and volatile matter content, with qualities ranging from low- to high-grade bituminous coal.

A total of 129 coal samples were collected for this investigation from the Witbank coalfields, which include the Middleburg and Secunda regions in South Africa, as depicted in Fig. 2. These samples originate from a coalfield that was formed in the Gondwana Karoo Basin between the mid-Triassic and Permian periods. These coals are younger and of lesser rank than carboniferous coal from the northern hemisphere. After being received, the coal samples were split, ground, and mixed with a rotary splitter to create various sub-samples in accordance with the ASTM D2013 [3].

The ASTM D5142-04 [6] was followed in conducting the proximate analysis. The inherent moisture, percentage ash yield, and volatile matter contained in the coal with fixed carbon calculated by difference were all determined using approximately 1 g of coal sample. The calorific value of the coal samples was determined using ASTM D5865 [8], while the total sulphur analyses were carried out according to ASTM D 5373–14 [7] and ISO 19579 [31]. The HGI was carried out in compliance with ASTM D3402 [4]. 1 kg of coal was ground under conventional conditions such that it could pass through a 4.75 mm sieve for this test. To acquire a maximum representative sample within the size fraction range of 1.18 mm to + 600 μm, the obtained sample was staged screened. A representative 50 g of the sized coal sample was placed in a ball mill along with 8 iron balls with a diameter of 25.4 ± 0.003 mm and turned for 60 revolutions. After that, the final grind was filtered through a 75 μm sieve. The under-size (−75 μm) fines formed were weighed and the grindability was determined [49].

The descriptive statistics of the 129 coal samples are shown in Table 1. The mean values of calorific value, moisture content, percentage ash yield, volatile matter, fixed carbon, total sulphur, pyrite, quartz and HGI are 27.467, 2.894, 15.905, 22.151, 59.586, 0.843, 0.884, 1.814 and 53.488 respectively, while their standard deviations (St.D) are 2.698, 0.968, 5.470, 7.884, 9.547, 0.356, 0.529,1.579 and 8.301

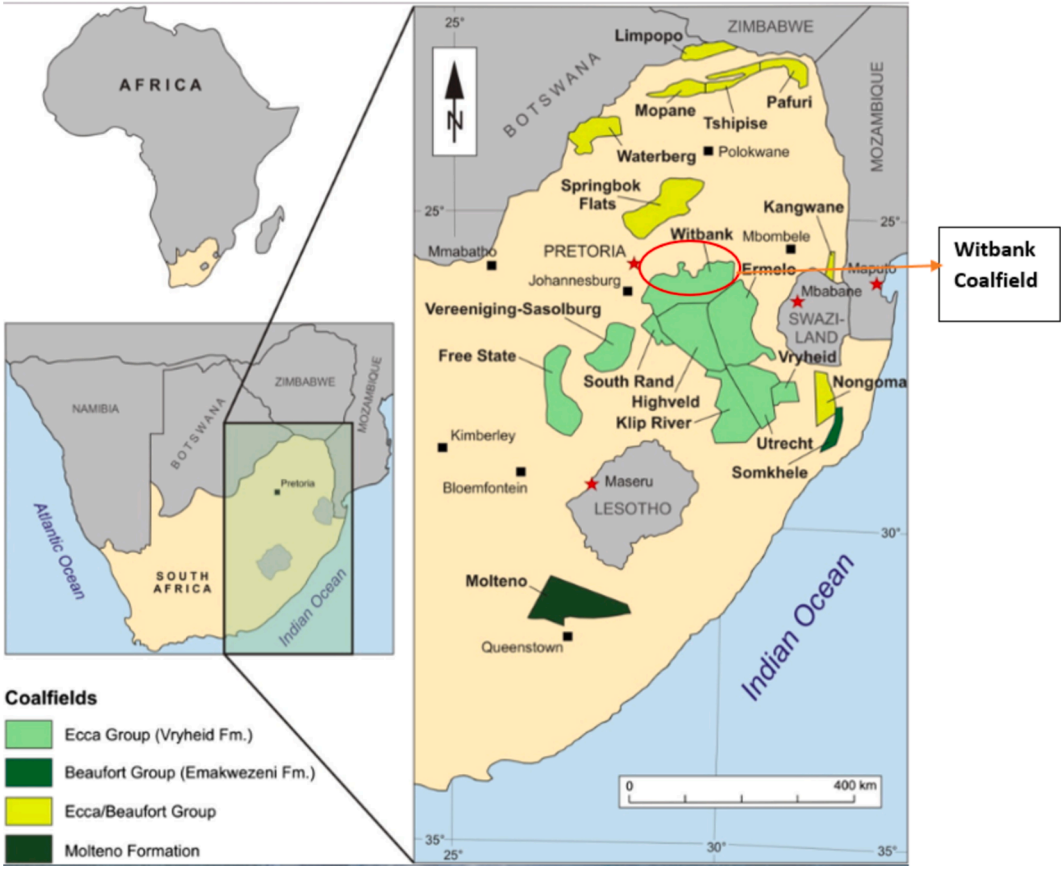


Fig. 2. Coalfields of South Africa. The Mopane, Tshipise and Pafuri Coalfields collectively form the Soutpansberg Coalfield. Fm. = Formation[65].

Table 1
Statistical Information of the HGI Dataset.

Parameters		Symbol	Min	Max	Mean	St.D
Input	Calorific Value (MJ/kg)	*CV*	20.1	32.5	27.467	2.698
	Moisture (wt. %)	*M*	1.4	5.5	2.894	0.968
	Ash (wt. %)	*A*	7.3	33.3	15.905	5.470
	Volatile Matter (wt. %)	*VM*	3.7	34.6	22.151	7.884
	Fixed carbon (wt. %)	*FC*	43.8	85.1	59.586	9.547
	Total Sulphur (wt. %)	*TS*	0.28	2	0.843	0.356
	Pyrite (wt. %)	FeS_2	0.13	3.12	0.884	0.529
	Quartz (%)	*Q*	0.06	7.25	1.814	1.579
Output	Hardgrove grindability index	*HGI*	32	79	53.488	8.301

respectively. Since the values of St.D of the total sulphur and pyrite are close to zero, there are no wide variations in these two properties for the samples from this location. The variation in the moisture content is not also wide. However, the values of volatile matter, fixed carbon and HGI largely deviate from one another as presented in Table 1. Since various coal properties have an impact on HGI, it is important to create reliable models that can consider the non-linearity between the input and output variables. Hence, models that can capture the variability in coal properties are proposed in this study. Fig. 3 presents box plots illustrating the distribution of input parameters across both the training and test datasets. These box plots visually summarize the statistical characteristics of the input parameters, showing their central tendency, variability, and potential outliers in each dataset. For the training set, the box plots provide insight into the range and distribution of the input parameters used to train the model, while the test set box plots reflect the distribution of these parameters during the evaluation phase. By comparing these plots, one can assess how the input parameter distributions in the test set relate to those in the training set, which can reveal potential discrepancies or shifts in the data that may impact the model's performance.

4. Results

4.1. Hyperparameter optimization

Determining the optimal hyperparameters is a crucial step in training ML models because it directly impacts model performance. The optimization objective dictates the direction of performance improvement during hyperparameter adjustment. Common optimization objectives include minimizing error or maximizing accuracy. For regression models, MSE directly reflects the deviation between predicted and actual values, making it a widely used optimization target in many cases. In this study, Optuna was employed for hyperparameter optimization, with the objective of minimizing the MSE in cross-validation. The search space for hyperparameters defines the range of values that Optuna or other optimization algorithms can explore when searching for the optimal hyperparameters. The definition of the search space is typically based on prior experience, relevant literature, and the results of preliminary experiments [25,29,66]. In this study, the hyperparameter search space was established using these methods, as detailed in Table 2. During the hyperparameter optimization process, a total of 200 iterations were conducted, with each iteration evaluated using 5-fold cross-validation to ensure result stability. The optimization process

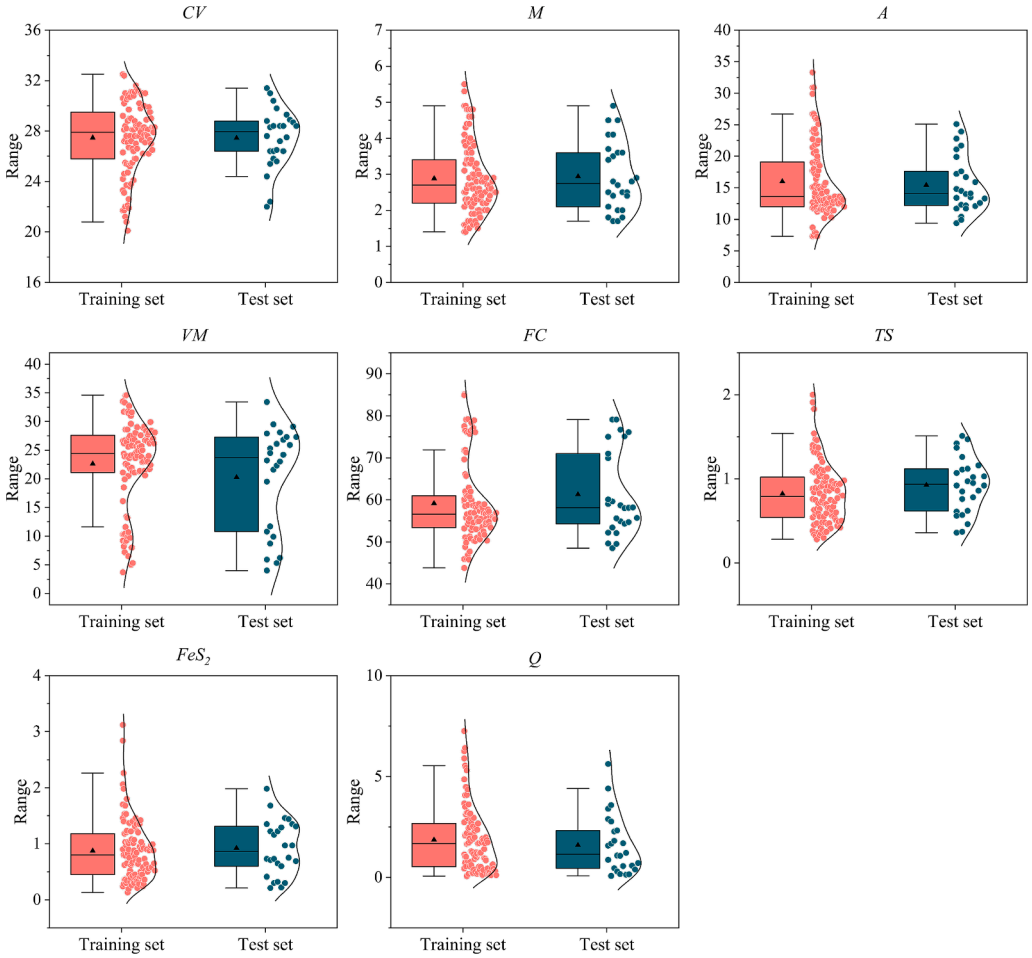


Fig. 3. Distribution of input parameters in training and test Sets.

Table 2
Hyperparameter Search Ranges for ML Models.

ML Model	Hyperparameter	Range	ML Model	Hyperparameter	Range
GBDT	*n_estimators*	[50, 500]	LGBM	*n_estimators*	[50, 500]
	max_depth	[1, 50]		*max_depth*	[1, 50]
	learning_rate	[0.01, 0.5]		*learning_rate*	[0.01, 0.5]
	subsample	[0.5, 1]		*subsample*	[0.5, 1]
	min_samples_split	[2, 10]		*colsample_bytree*	[0.5, 1]
	min_samples_leaf	[1, 10]		*min_child_samples*	[1, 20]
RF	*n_estimators*	[50, 500]		*min_child_weight*	[1, 10]
	max_depth	[1, 50]	CatBoost	*iterations*	[100, 1000]
	min_samples_split	[2, 10]		*depth*	[1, 10]
	min_samples_leaf	[1, 10]		*learning_rate*	[0.01, 0.5]
XGBoost	*n_estimators*	[50, 500]		*l2_leaf_reg*	[1, 30]
	max_depth	[1, 50]		*subsample*	[0.5, 1.0]
	learning_rate	[0.01, 0.5]	NGBoost	*n_estimators*	[50, 500]
	subsample	[0.5, 1]		*max_depth*	[1, 10]
	colsample_bytree	[0.5, 1]		*learning_rate*	[0.01, 0.5]
	min_child_weight	[1, 10]		*minibatch_frac*	[0.5, 1.0]
				col_sample	[0.5,1.0]

proceeded as follows: first, Optuna performed random initialization sampling within the predefined search space. Then, based on the cross-validation results from each iteration, Optuna used Bayesian optimization methods to adjust the sampling probabilities of the parameters, gradually narrowing the search space, and ultimately converging on an optimal set of hyperparameters.

The iteration process of hyperparameter optimization for ML models using Optuna is shown in Fig. 4. As shown in the figure, the initial fitness values are relatively high, but they rapidly decrease during the first few iterations. As the number of iterations increases, the best fitness values gradually stabilize. Ultimately, GBDT achieved the lowest fitness value of 19.82708944 at the 168th iteration. RF achieved the lowest fitness value of 23.72348972 at the 151st iteration. XGBoost achieved the lowest fitness value of 20.3016707 at the 188th iteration. LGBM achieved the lowest fitness value of 23.50661551 at the 184th iteration. CatBoost achieved the lowest fitness value of 19.46953427 at the 173rd iteration. NGBoost achieved the lowest fitness value of 21.45009115 at the 107th iteration.

Fig. 5 shows the search paths and optimization results for six decision tree-based ML models within the defined hyperparameter space. The parallel coordinates plot on the left displays the performance of different hyperparameter combinations in terms of fitness values, while the text box in the upper right lists the optimal hyperparameter combinations. Additionally, Optuna provides a tool to compute the contribution of each hyperparameter to the final optimization results. The bar chart in Fig. 5 displays the importance of each hyperparameter. From the chart, it can be seen that the most important hyperparameter for the GBDT and RF models is '*min_samples_leaf*'; for XGBoost, it is '*min_child_weight*'; for LGBM, it is '*min_child_samples*'; for CatBoost, it is '*depth*'; and for NGBoost, it is '*learning_rate*'. This indicates that these parameters require the most attention and adjustment during the optimization process to enhance model performance and optimization effectiveness.

4.2. Comparison of ML model results

Fig. 6 shows the predictive performance of six ML models on the training and test sets. The red dots represent the predictions on the training set, while the blue dots represent the predictions on the test set. The dashed diagonal line represents the line of perfect fit, indicating the ideal state where predicted values are equal to actual values. In theory, the closer the data points are to this dashed line, the higher the predictive accuracy of the model. From Fig. 6, it is evident that there are significant differences in the performance of various models on the training and test sets. The results indicate that the NGBoost model performs almost perfectly on the training set, with an R^2 value of 1.0 and both MAE and RMSE equal to 0, demonstrating its excellent fit to the training data. More importantly, NGBoost also shows outstanding performance on the test set, with an R^2 of 0.9715, MAE of 1.1507, and RMSE of 1.4735. This suggests that NGBoost maintains low prediction error when handling unseen data, exhibiting strong generalization ability and effectively avoiding overfitting. In contrast, GBDT also performs excellently on the training set, with an R^2 of 1.0 and negligible error. However, its R^2 decreases to 0.9473 on the test set, indicating that, although its overall performance remains good, the increase in error reflects some signs of overfitting. The RF model shows relatively good performance on the training set with an R^2 of 0.9425, but its R^2 drops to 0.9209 on the test set, with a slight increase in error. Although its generalization ability is decent, its performance on both training and test sets is somewhat inferior to that of NGBoost. Both XGBoost and LGBM achieve nearly perfect performance on the training set, with R^2 values of 1.0 and minimal error. However, their R^2 values decrease to 0.8914 and 0.8118 on the test set, respectively, indicating more severe overfitting and weaker generalization capabilities. CatBoost performs well on the training set with an R^2 of 0.9683, but its R^2 on the test set decreases to 0.8720, demonstrating moderate generalization ability. In summary, although models such as GBDT and RF perform reasonably well on the test set, NGBoost shows the most consistent performance between the training and test sets, indicating the strongest generalization ability and the capacity to effectively handle unseen data. In contrast, CatBoost, XGBoost, and LGBM display noticeable overfitting tendencies.

Table 3 details the performance evaluation metrics of the six models on both the training and test sets. Due to the differences in scale and range among the various performance evaluation metrics, directly comparing these metrics may lead to bias. Therefore, to ensure a fair and objective comparison of the models' performance, each performance metric was first standardized. The standardized metrics were then aggregated to produce a comprehensive performance score[71], allowing for a clearer and more intuitive evaluation of the overall performance of each model. The specific steps are as follows:

Standardization: Since various performance metrics have different dimensions, the Min-Max standardization method is used to standardize them for easier comparison.

$$X_{norm} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \tag{4}$$

Where X is the original metric value, $X_{\min}$ and $X_{\max}$ represents the minimum and maximum values of the metric, respectively, and X_{norm} is the standardized value.

Metric inversion: For R^2, higher values indicate better model performance, while for MAE and RMSE, lower values indicate better performance. Therefore, after standardization, MAE and RMSE must be

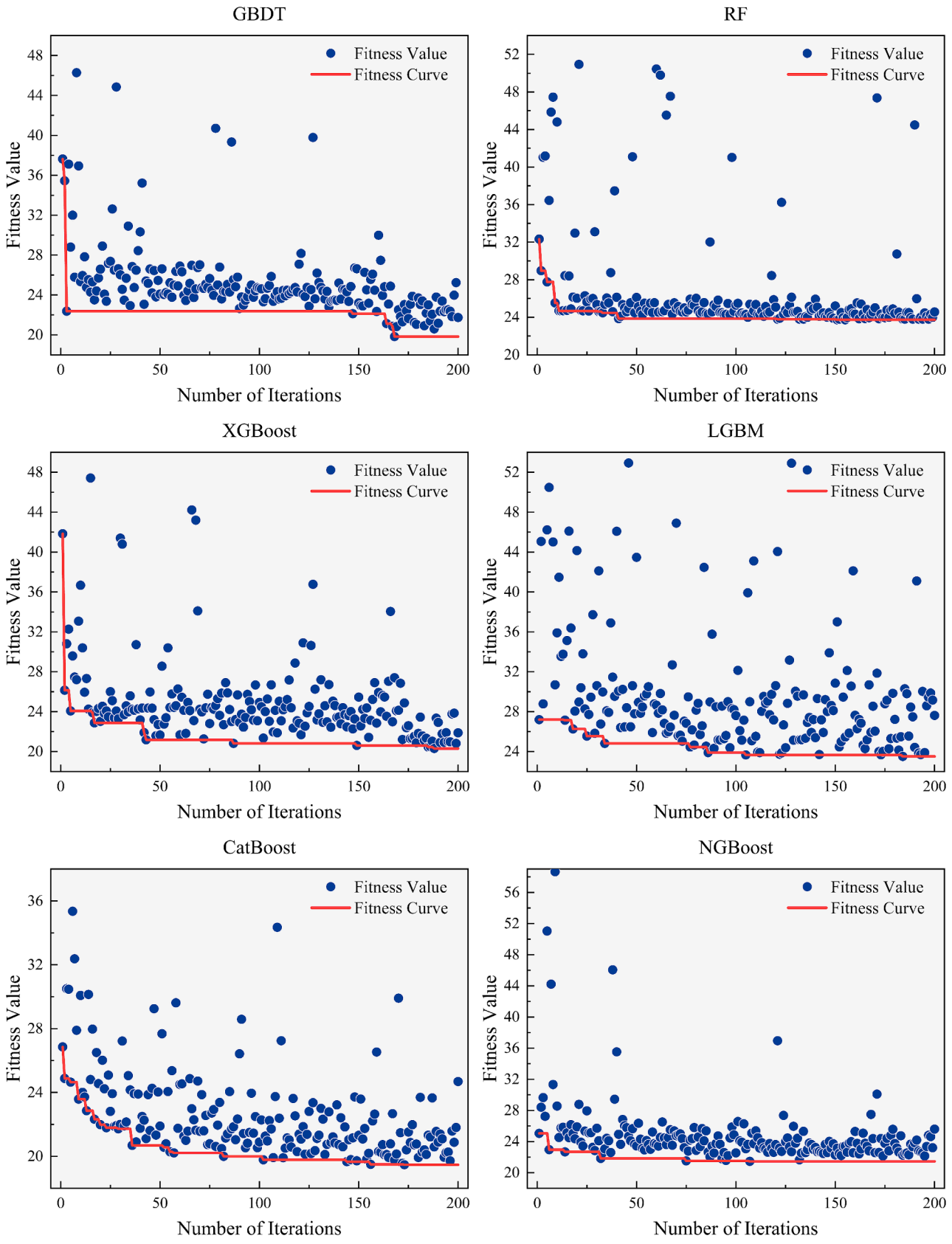
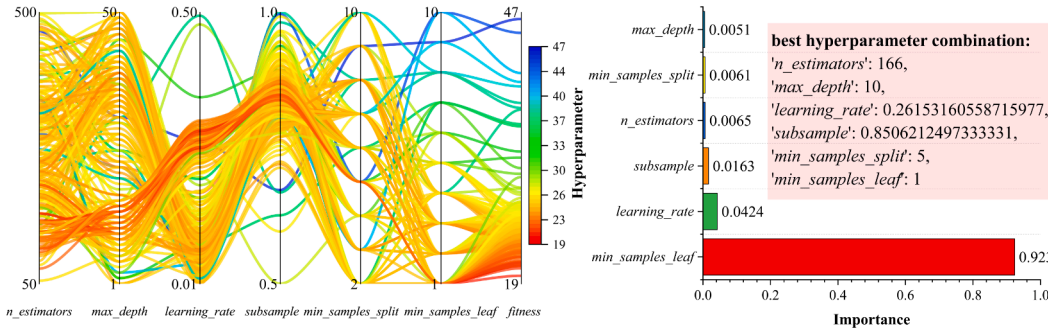
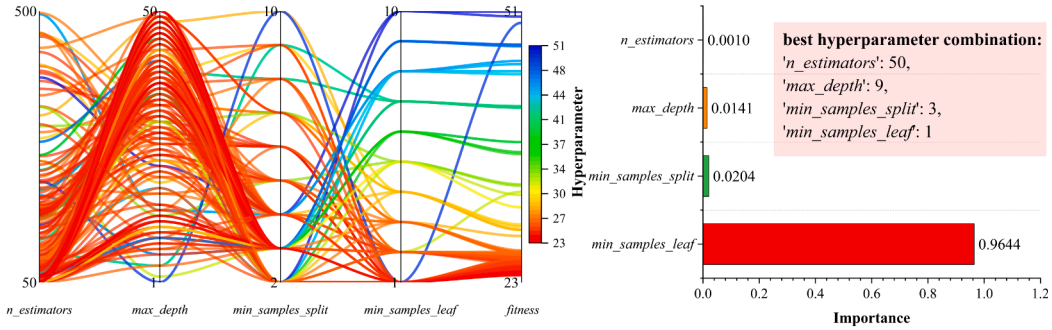


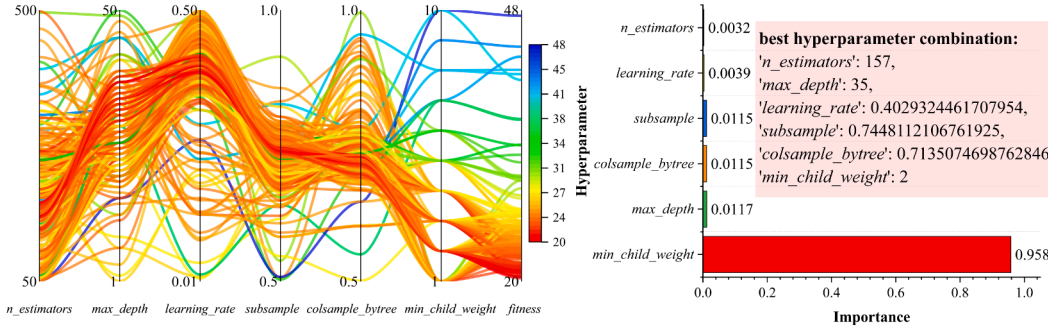
Fig. 4. Hyperparameter Optimization Iteration Process for Optuna-ML Models.



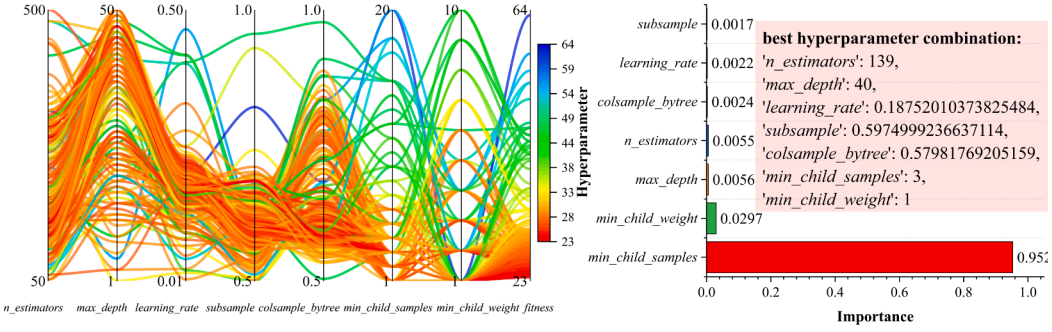
(a) GDBT



(b) RF

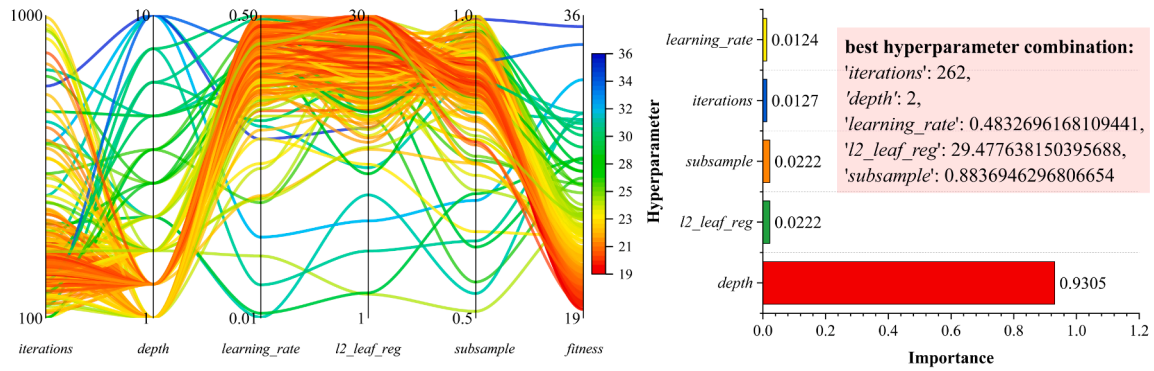


(c) XGBoost

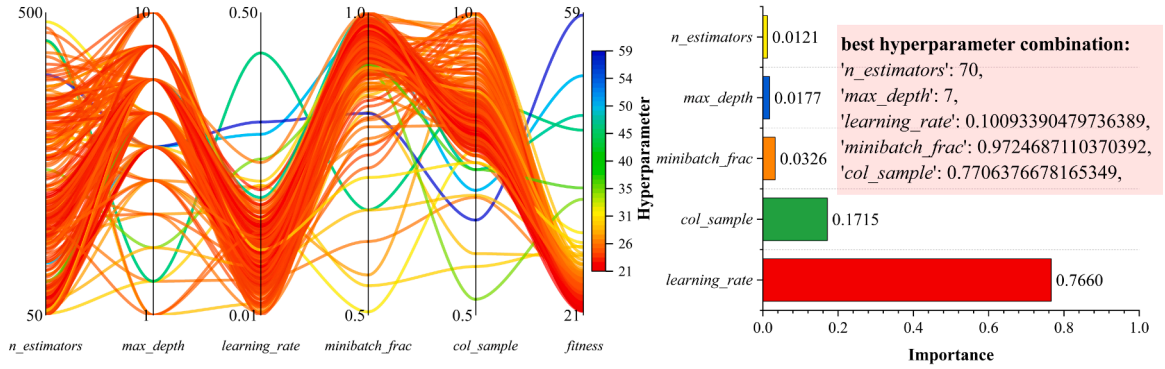


(d) LGBM

Fig. 5. Hyperparameter Optimization Results.



(e) CatBoost



(f) NGBoost

Fig. 5. (*continued*).

inverted so that higher values also represent better model performance.

$$X'_{norm} = 1 - X_{norm} \tag{5}$$

Where X'_{norm} is the value after the inversion process.

Comprehensive score calculation: The comprehensive score can be obtained by averaging the three standardized values. A higher comprehensive score indicates better overall model performance.

$$S = \frac{R^2_{norm} + MAE'_{norm} + RMSE'_{norm}}{3} \tag{6}$$

Where S is the comprehensive performance score, R^2_{norm} is the standardized R^2, and MAE'_{norm} and $RMSE'_{norm}$ are the standardized and inverted MAE and RMSE, respectively.

Fig. 7 shows the comprehensive scores of the six ML models. The scores, ranked from highest to lowest, are NGBoost, GBDT, XGBoost, LGBM, CatBoost, and RF. This indicates that the NGBoost has the best overall performance, followed by the GBDT, with the RF having the lowest score.

The performance of the models on the test set reflects their ability to predict unseen data. Fig. 8 shows the Taylor diagram, which visually compares the predictive performance of each model on the test set. The Taylor diagram combines standard deviation, correlation coefficient, and centered root mean square error, providing a more comprehensive evaluation perspective [67,86,87,41]. The closer a model is to the purple observation point, the better its performance. The results indicate that the NGBoost model is the closest to the observation point among all compared models, demonstrating the best predictive performance. Therefore, based on its superior performance among the six decision-tree-based models, the NGBoost model is recommended for prediction.

Table 4 presents a comparison of the performance of the constructed Optuna-NGBoost model against various models previously used in studies to predict HGI. Clearly, the Optuna-NGBoost model developed in this study has surpassed the performance of the vast majority of previous models. Therefore, the established Optuna-NGBoost model is suitable for the accurate prediction of HGI.

4.3. Interpretability analysis of model predictions

The interpretability analysis of ML models significantly enhances model transparency and trustworthiness, allowing users to understand the decision-making process of the models [47]. This analysis helps identify the properties that contribute the most to model performance, guiding feature engineering and optimization, and thereby improving the overall performance and reliability of the models. SHAP is a method based on the Shapley value principle from game theory, used to explain ML model outputs[63,33]. It quantifies each property's contribution to the prediction result, thus providing interpretability for model decisions. SHAP can provide local explanations for individual predictions and can also explain the overall behavior of the model (global explanation) by aggregating property contributions. This section will use SHAP for a detailed analysis of the NGBoost prediction model.

The SHAP summary plot in Fig. 9a shows the influence of each property on the prediction results of the model and the contribution of high and low property values to the direction of the influence. Each point in the plot represents the SHAP value of a sample, with the x-axis indicating the SHAP value and the color representing the property value (from low values in blue to high values in red). The property importance plot in Fig. 9b displays the contribution of different properties to model prediction. The importance of a property is measured by its average absolute SHAP value. The larger the value, the more significant the

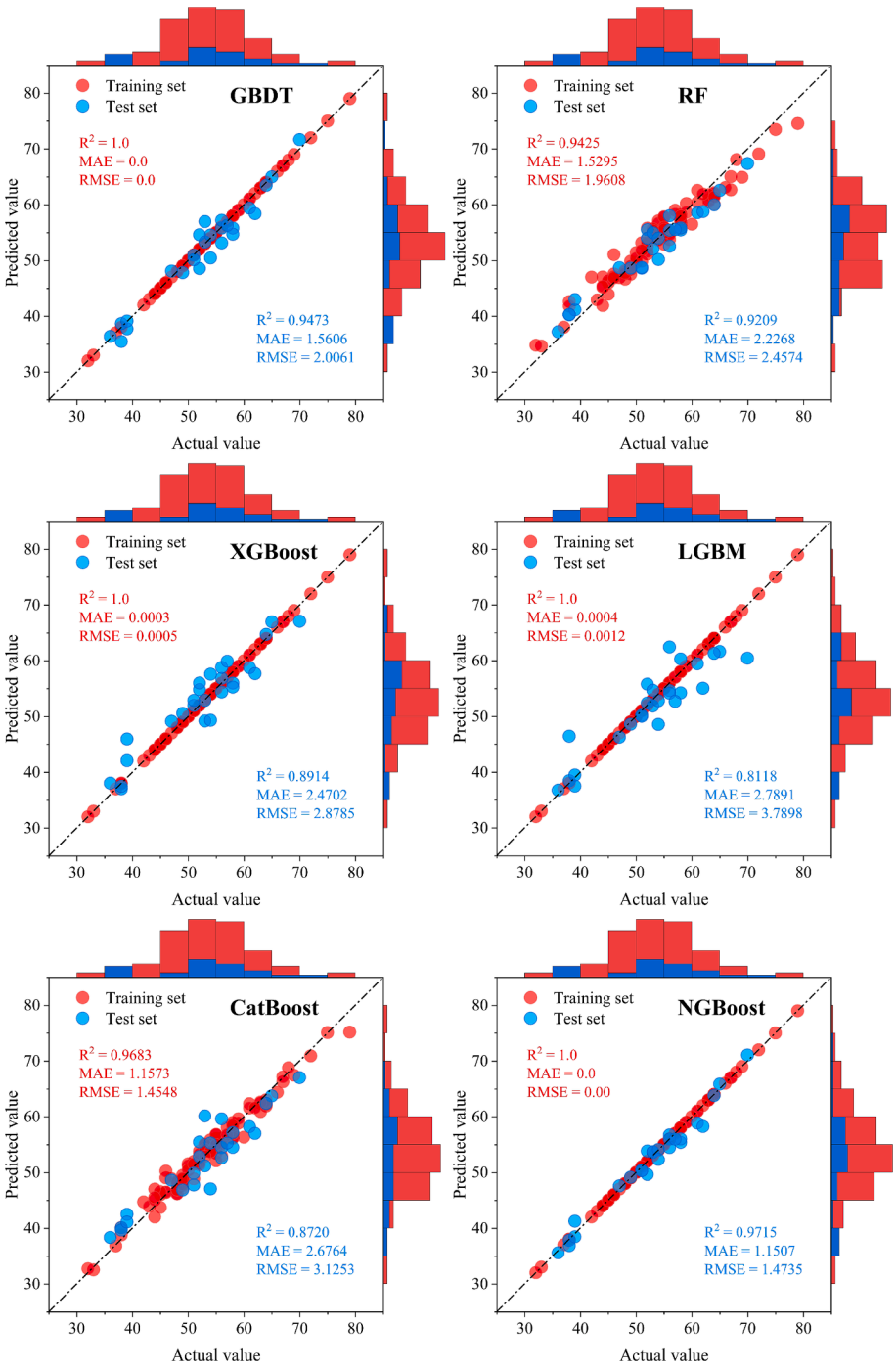


Fig. 6. Scatter Plot Regression Analysis of Predicted Results.

Table 3
Performance Evaluation Metrics of Models on the Test Set.

Training						
Model	GBDT	RF	XGBoost	LGBM	CatBoost	NGBoost
R^2	1.0	0.9425	1.0	1.0	0.9683	1.0
MAE	0.0	1.5295	0.0003	0.0004	1.1573	0.0
RMSE	0.0	1.9608	0.0005	0.0012	1.4548	0.0
Test						
R^2	0.9473	0.9209	0.8914	0.8118	0.8720	0.9715
MAE	1.5606	2.2268	2.4702	2.7891	2.6764	1.1507
RMSE	2.0061	2.4574	2.8785	3.7898	3.1253	1.4735

influence of that property on model prediction. The graph illustrates that the impact of each property on model output varies by sample. In some cases, properties contribute positively to the prediction results, while in others, they may contribute negatively. Global analysis reveals that *VM* is the most important property, with SHAP values significantly higher than other properties, indicating its substantial influence on model predictions. This suggests that *VM* is a critical variable in optimizing the HGI prediction model and requires particular attention to its variations. Conversely, FeS_2 has the least impact. Although the SHAP values of other properties are relatively lower, they still influence the model predictions to some extent, potentially having higher positive or negative contributions in certain samples.

Global explanations focus on the behavior of the entire model, while

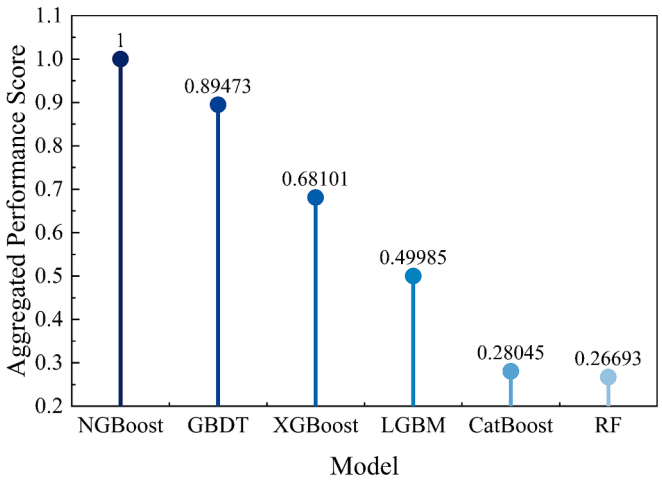


Fig. 7. Comprehensive Performance Scores of the Models.

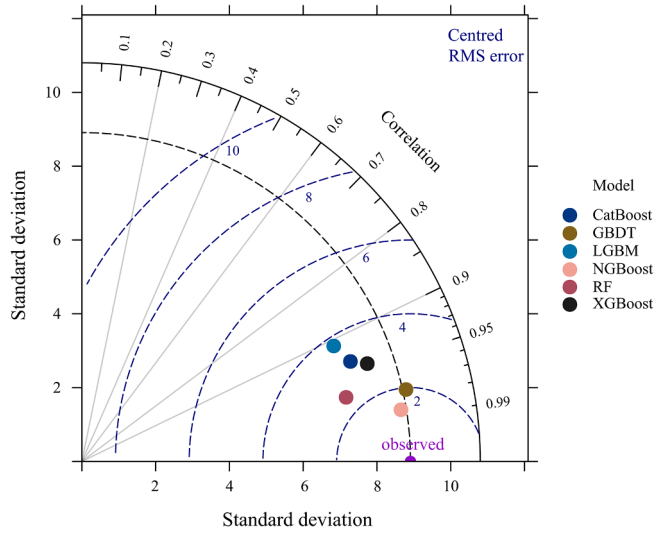


Fig. 8. Taylor Diagram of Model Predictions.

local explanations focus on individual predictions, quantifying the contribution of each property to specific prediction results. In order to accurately understand the role of properties in a single prediction (positive or negative contributions and their magnitude), local interpretation is essential. Fig. 10 and Table 5 present the SHAP explanations for the prediction results of three samples from the test set using the NGBoost model. The SHAP force plot starts from the base value, which represents the mean prediction of the target variable. Each Shapley value is depicted as an arrow, indicating the direction and magnitude of each property's impact on the prediction. Red arrows signify that the property value has increased the prediction output of the model (positive impact), while blue arrows signify that the property value has decreased the prediction output (negative impact). The combined effects of these properties result in the final prediction output (f(x)).

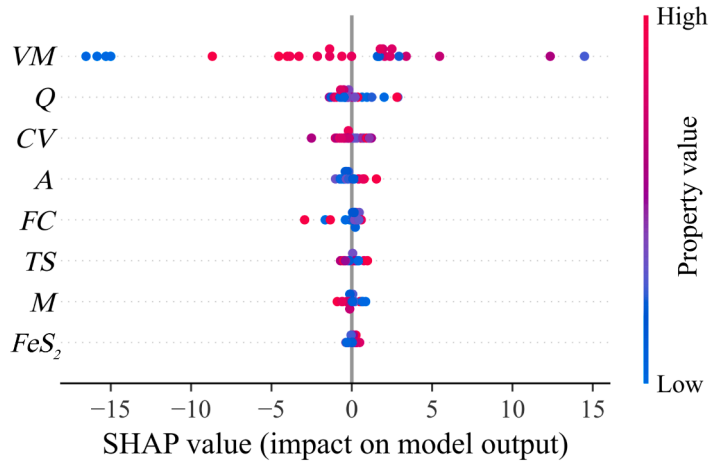
Taking Sample ① as an example, the specific properties of this sample include $CV = 28.3$, $M = 24.5$, $A = 13.4$, $VM = 24.5$, $FC = 59.6$, $TS = 0.36$, $P = 0.21$, and $Q = 0.87$. Based on these properties, the model's base value is 53.72117244. In this sample, the SHAP value for *VM* is 2.495762868, indicating that volatile matter has the largest positive contribution to the prediction, increasing the prediction by 2.495762868. Conversely, the SHAP value for *CV* is −0.41011446, showing that the calorific value has the largest negative contribution, reducing the prediction by 0.41011446. Additionally, the SHAP values for *M* and *TS* are 0.176733662 and 0.426556674, respectively, which both contribute positively to the prediction, while the SHAP value for *A* is −0.193930188, reflecting a negative impact. After accounting for the increase and decrease effects of all properties, the model's final prediction for this sample is 56.7074215. The actual value for this sample is 56, and the close match between the predicted and actual values demonstrates the model's accuracy for this sample. Similarly, in Sample ②, *VM* is the property with the largest negative contribution, while *CV* provides the largest positive contribution. In Sample ③, *VM* again has the largest negative contribution, while *Q* has the most significant positive impact on the prediction. By combining property information with SHAP values, we can gain a deeper understanding of how each property contributes to the model's prediction.

Table 4
Comparison of predictive performance between previous studies and the model proposed in this paper.

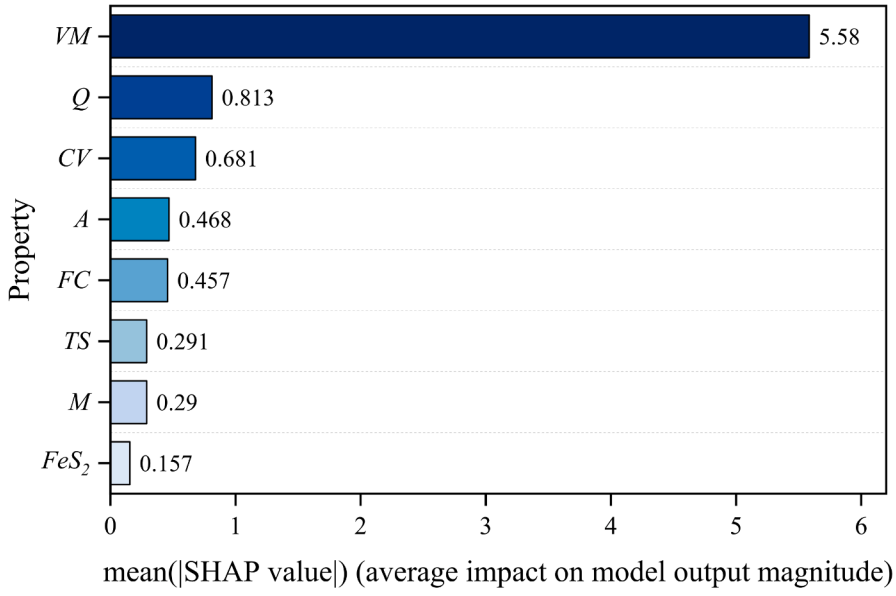
Reference	Model	Input variables	R^2
Jorjani et al. [32]	multivariable regression	*M*, *A*, Exinite, Micrinite, Macrinite, Resinite	0.76
Chelgani et al. [15]	multivariable regression	R_{max}, Exinite, Semifusinite, Micrinite, Macrinite, Resinite	0.81
Chelgani et al. [15]	ANN	R_{max}, ln(exinite), Semifusinite, Micrinite, Macrinite, Resinite	0.95
Bagherieh et al. [10]	BPNN	*A*, R_{max}, Vitrinite, Total vitrinite, Liptinite, Inertinite, Fusinite, Sulfur	0.61
Bagherieh et al. [10]	GRNN	*A*, R_{max}, Vitrinite, Total vitrinite, Liptinite, Inertinite, Fusinite, Sulfur	0.6
Modarres et al. [46]	Neuro-PSO	*A*, R_{max}, Vitrinite, Total vitrinite, Liptinite, Inertinite, Fusinite, Sulfur	0.83
Modarres et al. [46]	BPNN	*A*, R_{max}, Vitrinite, Total vitrinite, Liptinite, Inertinite, Fusinite, Sulfur	0.66
Modarres et al. [46]	GRNN	*A*, R_{max}, Vitrinite, Total vitrinite, Liptinite, Inertinite, Fusinite, Sulfur	0.70
Rao and Gopalakrishna [58]	SVR	*M*, *VM*, *A*	0.8084
Khoshjavan et al. [35]	ANN	*M*, *VM*, *A*, *FC*, *TS*, Btu/lb, Carbon, Hydrogen, Nitrogen, Oxygen	0.82
Chelgani et al. [14]	GA-ANFIS	*M*, *VM*, Al_2O_3, Na_2O, TiO_2, MgO, K_2O, SO_3, CaO	0.89
Chelgani et al. [14]	GA-ANFIS	*TS*, Na_2O, Hydrogen, TiO_2, Carbon, SiO_2, Oxygen, Al_2O_3	0.92
Chelgani et al. [14]	GA-ANFIS	Liptinite, Al_2O_3, R_{max}, SiO_2, Na_2O, MgO, TiO_2	0.97
Matin et al. [43]	RF	*M*, *TS*, TiO_2, Al_2O_3, Liptinite, R_{max}	0.90
Ürünveren et al. [70]	GRNN	*M*, *CV*, *A*, *FC* + *VM*, SiO_2, K_2O, MgO, MnO, SO_3, Fe_2O_3, Al_2O_3, CaO	0.9604
Yazdani et al. [78]	OPENN	*TS*, Hydrogen, Nitrogen, Na_2O, Liptinites	0.82
Rzychoń et al. [60]	XGBoost	*M*, *VM*, *A*, *TS*, Carbon, Nitrogen, Hydrogen, Vitrinite, Liptinite Inertinite	0.86
This paper	Optuna-NGBoost	*CV*, *M*, *A*, *VM*, *FC*, *TS*, *P*, *Q*	**0.9715**

5. Conclusions

The Hardgrove Grindability Index (HGI) is critical in coal processing and utilization. In this study, six tree-based machine learning models—Gradient Boosting Decision Trees (GBDT), Random Forest (RF), Extreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LGBM), Categorical Boosting (CatBoost), and Natural Gradient Boosting (NGBoost)—were used to predict the HGI values of 129 coal samples. Hyperparameter optimization was performed with Optuna and 5-fold



(a) The SHAP summary plot



(b) SHAP property importance

Fig. 9. Global explanation of model predictions.

cross-validation, and the predictive performance of each model was systematically evaluated using the coefficient of determination (R^2), mean absolute error (MAE), and root mean squared error (RMSE) as performance metrics. The results indicated that the Optuna-NGBoost model achieved the best performance in HGI prediction, with the highest R^2 value (0.9715) and the lowest MAE (1.1507) and RMSE (1.4735) on the test set. Compared to most existing HGI prediction models, the NGBoost model demonstrated superior performance. SHapley Additive exPlanations (SHAP) global explanation analysis revealed that volatile matter was the most important feature in predicting HGI. Additionally, SHAP's local explanation for individual samples further highlighted the specific contribution of each property to the prediction, enhancing the interpretability of the model.

In conclusion, this study demonstrates the feasibility and advantages of using interpretable decision tree-based machine learning models for predicting the HGI. By accurately predicting HGI values, the developed model plays an active role in improving milling efficiency and reducing equipment energy consumption. In subsequent research, with the collection of more data and further optimization and integration of the model, the development of a corresponding graphical user interface will provide enterprises with a more convenient tool to assess coal quality more effectively and optimize supply chains. This will offer a valuable solution to enhance efficiency and decision-making processes within the coal industry.

CRediT authorship contribution statement

Yuxin Chen: Writing – original draft, Validation, Software, Methodology, Conceptualization. **Manoj Khandelwal:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Moshood Onifade:** Writing – review & editing, Visualization, Supervision, Resources, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Jian Zhou:** Writing – review & editing,

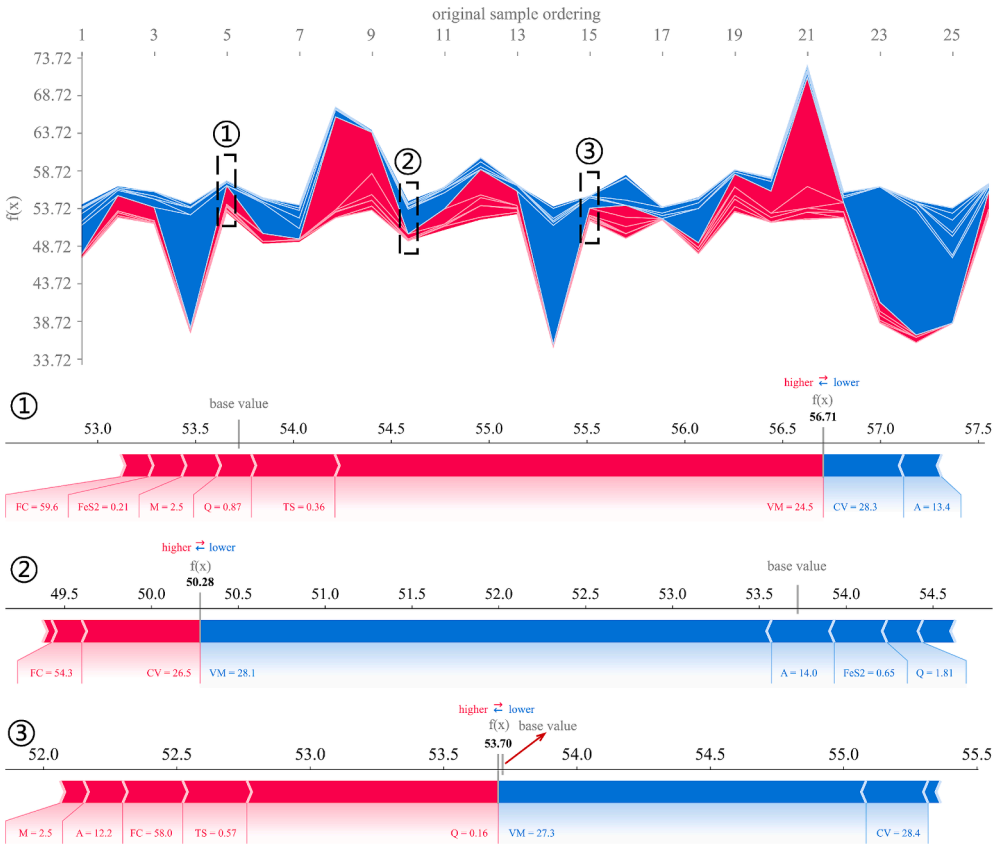


Fig. 10. Local explanation of model predictions.

Table 5
Detailed SHAP explanation for individual samples.

Sample	①		②		③	
	Property value	SHAP values	Property value	SHAP values	Property value	SHAP values
CV	28.3	−0.41011446	26.5	0.681331782	28.4	−0.232381929
M	2.5	0.176733662	3.6	−0.187513641	2.5	0.088525898
A	13.4	−0.193930188	14	−0.362144758	12.2	0.145793655
VM	24.5	2.495762868	28.1	−3.289593216	27.3	−1.37945027
FC	59.6	0.143040308	54.3	0.174200332	58	0.22501325
TS	0.36	0.426556674	0.76	0.05207978	0.57	0.240740537
FeS_2	0.21	0.168223812	0.65	−0.303277497	0.6	−0.046792261
Q	0.87	0.179976384	1.81	−0.206130948	0.16	0.942055724
Base value	53.72117244		53.72117244		53.72117244	
Predicted Value	56.7074215		50.28012427		53.70467704	
True Value	56		51		53	

Validation, Supervision, Software, Resources, Project administration, Investigation, Data curation. **Abiodun Ismail Lawal:** Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Samson Oluwaseyi Bada:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Bekir Genc:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research is partially supported by the National Natural Science Foundation of China (42177164, 52477121) and the Distinguished Youth Science Foundation of Hunan Province of China (2022JJ10073).

Data availability

Data will be made available on request.

References

[1] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A next-generation hyperparameter optimization framework. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; 2019. p. 2623–31. doi: 10.1145/3292500.3330701.

[2] Aria M, Cuccurullo C, Gnasso A. A comparison among interpretative proposals for Random Forests. Mach Learn Applic 2021;6:100094. https://doi.org/10.1016/j.mlwa.2021.100094.
[3] ASTM D2013. Standard Practice for Preparing Coal Samples for Analysis. 2021.
[4] ASTM D3402. Standard Test Method for Tumbler Test for Coke. 2021.
[5] ASTM D409. Standard Test Method for Grindability of Coal by the Hardgrove-Machine Method. 2016.
[6] ASTM D5142-04. Standard Test Methods for Proximate Analysis of the Analysis Sample of Coal and Coke by Instrumental Procedures. 2010.
[7] ASTM D5373-14. Standard Test Methods for Determination of Carbon, Hydrogen, and Nitrogen in Analysis Samples of Coal and Carbon in Analysis Samples of Coal and Coke. 2015.
[8] ASTM D5865. Standard Test Method for Gross Calorific Value of Coal and Coke. 2017.
[9] Auret L, Aldrich C. Interpretation of nonlinear relationships between process variables by use of random forests. Miner Eng 2012;35:27–42. https://doi.org/10.1016/j.mineng.2012.05.008.
[10] Bagherieh AH, Hower JC, Bagherieh AR, Jorjani E. Studies of the relationship between petrography and grindability for Kentucky coals using artificial neural network. Int J Coal Geol 2008;73(2):130–8. https://doi.org/10.1016/j.coal.2007.04.002.
[11] Bo Y, Liu Q, Huang X, Pan Y. Real-time hard-rock tunnel prediction model for rock mass classification using CatBoost integrated with Sequential Model-Based Optimization. Tunn Undergr Space Technol 2022;124:104448. https://doi.org/10.1016/j.tust.2022.104448.
[12] Breiman L. Random Forests. Mach Learn 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324.
[13] Chehreh Chelgani S, Homafar A, Nasiri H, Rezaei Laksar M. CatBoost-SHAP for modeling industrial operational flotation variables – A "conscious lab" approach. Miner Eng 2024;213:108754. https://doi.org/10.1016/j.mineng.2024.108754.
[14] Chelgani SC, Dehghan F, Hower JC. Estimation of some coal parameters depending on petrographic and inorganic analyses by using Genetic algorithm and adaptive neuro-fuzzy inference systems. Energy Explor Exploit 2011;29(4):479–94. https://doi.org/10.1260/0144-5987.29.4.479.
[15] Chelgani SC, Hower JC, Jorjani E, Mesroghli S, Bagherieh AH. Prediction of coal grindability based on petrography, proximate and ultimate analysis using multiple regression and artificial neural network models. Fuel Process Technol 2008;89(1):13–20. https://doi.org/10.1016/j.fuproc.2007.06.004.
[16] Chelgani S, Nasiri H, Alidokht M. Interpretable modeling of metallurgical responses for an industrial coal column flotation circuit by XGBoost and SHAP-A "conscious-lab" development. Int J Min Sci Technol 2021;31(6):1135–44. https://doi.org/10.1016/j.ijmst.2021.10.006.
[17] Chelgani SC. Estimation of gross calorific value based on coal analysis using an explainable artificial intelligence. Mach Learn Applic 2021;6:100116. https://doi.org/10.1016/j.mlwa.2021.100116.
[18] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; 2016. p. 785–94. doi: 10.1145/2939672.2939785.
[19] Deniz V. The effects on the grinding parameters of chemical, morphological and mineralogical properties of three different calcites in a Hardgrove mill. Miner Eng 2022;176:107348. https://doi.org/10.1016/j.mineng.2021.107348.
[20] Dindarloo S, Hower JC, Bagherieh A, Trimble AS. Fundamental evaluation of petrographic effects on coal grindability by seasonal autoregressive integrated moving average (SARIMA). Int J Miner Process 2016;154:94–9. https://doi.org/10.1016/j.minpro.2016.07.005.
[21] Duan T, Anand A, Ding DY, Thai KK, Basu S, Ng A, et al. NGBoost: Natural gradient boosting for probabilistic prediction, 119. PMLR; 2020. p. 2690–700. https://proceedings.mlr.press/v119/duan20a.html.
[22] Edwards GR, Evans TM, Robertson SD, Summers CW. Assessment of the standard method of test for the grindability of coal by the Hardgrove machine. Fuel 1980;59(12):826–30. https://doi.org/10.1016/0016-2361(80)90030-7.
[23] Friedman JH. Greedy function approximation: A gradient boosting machine. Ann Stat 2001;29(5):1189–232. http://www.jstor.org/stable/2699986.
[24] Gyamfi BA, Adedoyin FF, Bein MA, Bekun FV, Agozie DQ. The anthropogenic consequences of energy consumption in E7 economies: juxtaposing roles of renewable, coal, nuclear, oil and gas energy: Evidence from panel quantile method. J Clean Prod 2021;295:126373. https://doi.org/10.1016/j.jclepro.2021.126373.
[25] Hastie T, Tibshirani R, Friedman JH, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York: Springer; 2009. https://doi.org/10.1007/978-0-387-84858-7.
[26] Hower JC, Bagherieh AH, Dindarloo SR, Trimble AS, Chelgani SC. Soft modelling of the Hardgrove grindability index of bituminous coals: An overview. Int J Coal Geol 2021;247:103846. https://doi.org/10.1016/j.coal.2021.103846.
[27] Hower JC, Finkelman RB, Eble CF, Arnold BJ. Understanding coal quality and the critical importance of comprehensive coal analyses. Int J Coal Geol 2022;263:104120. https://doi.org/10.1016/j.coal.2022.104120.
[28] Hower JC. Discussion: Li et al., Prediction of grindability with multivariable regression and neural network in Chinese coal. Fuel 2006;85(9):1307–8. doi: 10.1016/j.fuel.2005.11.011.
[29] Hutter F, Kotthoff L, Vanschoren J, editors. Automated machine learning: methods, systems, challenges. Cham: Springer Nature; 2019. https://doi.org/10.1007/978-3-030-05318-5.
[30] Idris A, Man Z, Bustam A, Rabat NE, Uddin F, Abdul Mannan H. Grindability and abrasive behavior of coal blends: analysis and prediction. Int J Coal Prep Util 2022;42(4):1143–69. https://doi.org/10.1080/19392699.2019.1694009.
[31] ISO. Solid mineral fuels – Determination of sulphur by IR spectrometry. ISO 2006;19579:2006.
[32] Jorjani E, Hower JC, Chelgani SC, Shirazi MA, Mesroghli S. Studies of relationship between petrography and elemental analysis with grindability for Kentucky coals. Fuel 2008;87(6):707–13. https://doi.org/10.1016/j.fuel.2007.05.044.
[33] Kannangara KPM, Zhou W, Ding Z, Hong Z. Investigation of feature contribution to shield tunneling-induced settlement using Shapley additive explanations method. J Rock Mech Geotech Eng 2022;14(4):1052–63. https://doi.org/10.1016/j.jrmge.2022.01.002.
[34] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, LightGBM, et al. A highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst 2017;30.
[35] Khoshjavan S, Mazlumi M, Rezai B, Rezai M. Estimation of hardgrove grindability index (HGI) based on the coal chemical properties using artifical neural networks. Orient J Chem 2010;26(4):1271.
[36] Kilic K, Ikeda H, Narihiro O, Adachi T, Kawamura Y. A soft ground micro TBM's specific energy prediction using an eXplainable neural network through Shapley additive explanation and Optuna. Bull Eng Geol Environ 2024;83(5):175. https://doi.org/10.1007/s10064-024-03670-5.
[37] Kotsiantis SB. Decision trees: A recent overview. Artif Intell Rev 2013;39:261–83. https://doi.org/10.1007/s10462-011-9272-4.
[38] Lawal AI, Bada SO, Onifade M. Prediction of HGI of South African coalfields: A comparative application of ANN, SVR and LSTM models. Int J Coal Prep Util 2024;1–17. https://doi.org/10.1080/19392699.2024.2339319.
[39] Lei C, Deng J, Cao K, Xiao Y, Ma L, Wang W, et al. A comparison of random forest and support vector machine approaches to predict coal spontaneous combustion in gob. Fuel 2019;239:297–311. https://doi.org/10.1016/j.fuel.2018.11.006.
[40] Li T, Xia Q, Ouyang Y, Zeng R, Liu Q, Li T. Prospectivity and uncertainty analysis of tungsten polymetallogenic mineral resources in the nanling metallogenic belt, South China: A comparative study of AdaBoost, GBDT, and XgBoost algorithms. Nat Resour Res 2024;33(3):1049–71. https://doi.org/10.1007/s11053-024-10321-9.
[41] Lu J, Liu Z, Zhang W, Zheng J, Han C. Pressure prediction study of coal mining working face based on Nadam-LSTM. IEEE Access 2023;11:83867–80. https://doi.org/10.1109/ACCESS.2023.3302516.
[42] Mame M, Qiu Y, Huang S, Du K, Zhou J. Mean block size prediction in rock blast fragmentation using TPE-tree-based model approach with SHapley Additive exPlanations. Min Metall Explor 2024;41:2325–40.
[43] Matin SS, Hower JC, Farahzadi L, Chelgani SC. Explaining relationships among various coal analyses with coal grindability index by Random Forest. Int J Miner Process 2016;155:140–6. https://doi.org/10.1016/j.minpro.2016.08.015.
[44] Matjie R, Bunt J, Goosen A, Mphahlele K, Uwaoma R. Effect of chemical, mineralogical, and petrographic properties on the grindability of a South African Highveld coal. J South Afr Inst Min Metall 2023;123(7):381–90. https://doi.org/10.17159/2411-9717/2423/2023.
[45] Miroshnichenko DV, Desna NA, Koval VV, Fatenko SV. Hardgrove grindability of coal. Part 1. Correlations with composition, structure, and properties. Coke Chem 2019;62:1–4. https://doi.org/10.3103/S1068364X19010058.
[46] Modarres HR, Kor M, Abkhoshk E, Alfi A, Hower JC. Prediction of coal grindability based on petrography, proximate and ultimate analysis using neural networks and particle swarm optimization technique. Energy Explor Exploit 2009;27(3):201–12. https://doi.org/10.1260/014459809789618821.
[47] Molnar C. Interpretable machine learning. Lulu.com 2020.
[48] Munshi TA, Jahan LN, Howladar MF, Hashan M. Prediction of gross calorific value from coal analysis using decision tree-based bagging and boosting techniques. Heliyon 2024;10(1). https://doi.org/10.1016/j.heliyon.2023.e23395.
[49] Onifade M, Lawal AI, Bada SO, Khandelwal M. Predictive modelling for coal abrasive index: Unveiling influential factors through Shallow and Deep Neural Networks. Fuel 2024;374:132319. https://doi.org/10.1016/j.fuel.2024.132319.
[50] Onifade M, Lawal AI, Abdulsalam J, Genc B, Bada S, Said KO, et al. Development of multiple soft computing models for estimating organic and inorganic constituents in coal. Int J Min Sci Technol 2021;31(3):483–94. https://doi.org/10.1016/j.ijmst.2021.02.003.
[51] Prokhorenkova L, Gusev G, Vorobev A, Dorogush AV, Gulin A. CatBoost: Unbiased boosting with categorical features. Adv Neural Inf Process Syst 2018;31.
[52] Qiu Y, Zhou J. Short-term rockburst damage assessment in burst-prone mines: An explainable XGBOOST hybrid model with SCSO algorithm. Rock Mech Rock Eng 2023;56(12):8745–70.
[53] Qiu Y, Zhou J. Short-term rockburst prediction in underground project: Insights from an explainable and interpretable ensemble learning model. Acta Geotech 2023;18(12):6655–85.
[54] Qiu Y, Zhou J, He B, Armaghani DJ, Huang S, He X. Evaluation and interpretation of blasting-induced tunnel overbreak: Using heuristic-based ensemble learning and gene expression programming techniques. Rock Mech Rock Eng 2024;57:7535–63.
[55] Qiu Y, Zhou J. Novel rockburst prediction criterion with enhanced explainability employing CatBoost and nature-inspired metaheuristic technique. Undergr Space 2024;19:101–18.
[56] Qiu Y, Zhou J, Khandelwal M, Yang H, Yang P, Li C. Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. Eng Comput 2022;38:4145–62. https://doi.org/10.1007/s00366-021-01393-9.
[57] Radić DB, Obradović MO, Stanojević MM, Jovović AM, Stojiljković DD. A study on the grindability of Serbian coals. Therm Sci 2011;15(1):267–74. https://doi.org/10.2298/TSCI1101267R.
[58] Rao BV, Gopalakrishna SJ. Hardgrove grindability index prediction using support vector regression. Int J Miner Process 2009;91(1–2):55–9. https://doi.org/10.1016/j.minpro.2008.12.003.

[59] Rodriguez-Galiano V, Sanchez-Castillo M, Chica-Olmo M, Chica-Rivas MJOGR. Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines. Ore Geol Rev 2015;71:804–18. https://doi.org/10.1016/j.oregeorev.2015.01.001.
[60] Rzychoń M, Żogała A, Róg L. SHAP-based interpretation of an XGBoost model in the prediction of grindability of coals and their blends. Int J Coal Prep Util 2022;42 (11):3348–68. https://doi.org/10.1080/19392699.2021.1959324.
[61] Safonova A, Ghazaryan G, Stiller S, Main-Knorn M, Nendel C, Ryo M. Ten deep learning techniques to address small data problems with remote sensing. Int J Appl Earth Obs Geoinf 2023;125:103569. https://doi.org/10.1016/j.jag.2023.103569.
[62] Sengupta AN. An assessment of grindability index of coal. Fuel Process Technol 2002;76(1):1–10. https://doi.org/10.1016/S0378-3820(01)00236-3.
[63] Shapley LS. Stochastic games. Proc Natl Acad Sci 1953;39(10):1095–100. https://doi.org/10.1073/pnas.39.10.1095.
[64] Singha SS, Singha S, Pasupuleti S, Venkatesh AS. Knowledge-driven and machine learning decision tree-based approach for assessment of geospatial variation of groundwater quality around coal mining regions, Korba district, Central India. Environ Earth Sci 2022;81(2):36. https://doi.org/10.1007/s12665-021-10147-1.
[65] Snyman CP. Coal. In: Wilson MGC, Anhausser CR, editors. The mineral deposits of South Africa. Handbook. Council of Geoscience; 1998. p. 136–205.
[66] Snoek J, Larochelle H, Adams RP. Practical bayesian optimization of machine learning algorithms. Adv Neural Inf Proces Syst 2012;25. https://doi.org/10.48550/arXiv.1206.2944.
[67] Taylor KE. Summarizing multiple aspects of model performance in a single diagram. J Geophys Res Atmos 2001;106(D7):7183–92. https://doi.org/10.1029/2000JD900719.
[68] Taylor L, Skuse D, Blackburn S, Greenwood R. Stirred media mills in the mining industry: Material grindability, energy-size relationships, and operating conditions. Powder Technol 2020;369:1–16. https://doi.org/10.1016/j.powtec.2020.04.057.
[69] Tiwary AK, Ghosh S, Singh R, Mukherjee DP, Shankar BU, Dash PS. Automated coal petrography using random forest. Int J Coal Geol 2020;232:103629. https://doi.org/10.1016/j.coal.2020.103629.
[70] Ürünveren A, Altıner M, Kuvvetli Y, Ural OB, Ural S. Prediction of hardgrove grindability index of Afsin-Elbistan (Turkey) low-grade coals based on proximate analysis and ash chemical composition by neural networks. Int J Coal Prep Util 2017;40(10):701–11. https://doi.org/10.1080/19392699.2017.1406350.
[71] Vafaei N, Ribeiro RA, Camarinha-Matos LM. Selecting normalization techniques for the analytical hierarchy process. In: Technological innovation for life improvement. 11th IFIP WG 5.5/SOCOLNET Advanced Doctoral Conference on Computing, Electrical and Industrial Systems, DoCEIS 2020, Costa de Caparica, Portugal, July 1–3, 2020. Cham: Springer International Publishing; 2020. p. 43–52. doi: 10.1007/978-3-030-45124-0_4.
[72] Vuthaluru HB, Brooke RJ, Zhang DK, Yan HM. Effects of moisture and coal blending on Hardgrove Grindability Index of Western Australian coal. Fuel Process Technol 2003;81(1):67–76. https://doi.org/10.1016/S0378-3820(03)00044-4.
[73] Wang X, Lu Y, Chen C, Yi X, Cui H. Total-factor energy efficiency of ten major global energy-consuming countries. J Environ Sci 2024;137:41–52. https://doi.org/10.1016/j.jes.2023.02.031.
[74] Wen Z, Liu H, Zhou M, Liu C, Zhou C. Explainable machine learning rapid approach to evaluate coal ash content based on X-ray fluorescence. Fuel 2023;332:125991. https://doi.org/10.1016/j.fuel.2022.125991.
[75] Williams O, Eastwick C, Kingman S, Giddings D, Lormor S, Lester E. Investigation into the applicability of Bond Work Index (BWI) and Hardgrove Grindability Index (HGI) tests for several biomasses compared to Colombian La Loma coal. Fuel 2015; 158:379–87. https://doi.org/10.1016/j.fuel.2015.05.027.
[76] Wu X, Feng Z, Liu J, Chen H, Liu Y. Predicting existing tunnel deformation from adjacent foundation pit construction using hybrid machine learning. Autom Constr 2024;165:105516. https://doi.org/10.1016/j.autcon.2024.105516.
[77] Xiao X, Zou Y, Huang J, Luo X, Yang L, Li M, et al. An interpretable model for landslide susceptibility assessment based on Optuna hyperparameter optimization and Random Forest. Geomat Nat Haz Risk 2024;15(1):2347421. https://doi.org/10.1080/19475705.2024.2347421.
[78] Yazdani S, Hadavandi E, Hower J, Chehreh Chelgani S. A novel nature-inspired optimization based neural network simulator to predict coal grindability index. Eng Comput 2018;35(2):1003–48. https://doi.org/10.1108/EC-09-2017-0332.
[79] Yilmaz S. A new approach for the testing method of coal grindability. Adv Powder Technol 2019;30(9):1932–40. https://doi.org/10.1016/j.apt.2019.06.012.
[80] Zhang R, Zhou J. Predicting the minimum horizontal principal stress using genetic expression programming and borehole breakout data. J Rock Mech Geotech Eng 2024.
[81] Zhang R, Zhou J, Tao M, Li C, Li P, Liu T. Borehole breakout prediction based on multi-output machine learning models using the walrus optimization algorithm. Appl Sci 2024;14(14):6164. https://doi.org/10.3390/app14146164.
[82] Zhang X, Nguyen H, Bui XN, Tran QH, Nguyen DA, Bui DT, et al. Novel soft computing model for predicting blast-induced ground vibration in open-pit mines based on particle swarm optimization and XGBoost. Nat Resour Res 2020;29(2): 711–21. https://doi.org/10.1007/s11053-019-09492-7.
[83] Zhang YL, Qiu YG, Armaghani DJ, Monjezi M, Zhou J. Enhancing rock fragmentation prediction in mining operations: A Hybrid GWO-RF model with SHAP interpretability. J Cent South Univ 2024;31:2916–29.
[84] Zhou J, Chen Y, Li C, Qiu Y, Huang S, Tao M. Machine learning models to predict the tunnel wall convergence. Transp Geotech 2023;41:101022. https://doi.org/10.1016/j.trgeo.2023.101022.
[85] Zhou J, Huang S, Tao M, Khandelwal M, Dai Y, Zhao M. Stability prediction of underground entry-type excavations based on particle swarm optimization and gradient boosting decision tree. Undergr Space 2023;9:234–49. https://doi.org/10.1016/j.undsp.2022.08.002.
[86] Zhou J, Qiu Y, Khandelwal M, Zhu S, Zhang X. Developing a hybrid model of Jaya algorithm-based extreme gradient boosting machine to estimate blast-induced ground vibrations. Int J Rock Mech Min Sci 2021;145:104856.
[87] Zhou J, Qiu Y, Armaghani DJ, Zhang W, Li C, Zhu S, et al. Predicting TBM penetration rate in hard rock condition: A comparative study among six XGB-based metaheuristic techniques. Geosci Front 2021;12(3):101091.
[88] Zhou Z, Cao J, Shi X, Zhang W, Huang W. Probabilistic rutting model using NGBoost and SHAP: Incorporating other performance indicators. Constr Build Mater 2024;438:137052. https://doi.org/10.1016/j.conbuildmat.2024.137052.
[89] Zhu X, Chu J, Wang K, Wu S, Yan W, Chiam K. Prediction of rockhead using a hybrid N-XGBoost machine learning framework. J Rock Mech Geotech Eng 2021; 13(6):1231–45. https://doi.org/10.1016/j.jrmge.2021.06.012.
[90] Zuo W, Zhao Y, He Y, Shi F, Duan C. Relationship between coal size reduction and energy input in Hardgrove mill. Int J Min Sci Technol 2012;22(1):121–4. https://doi.org/10.1016/j.ijmst.2011.07.011.
[91] Zhou J, Dai Y, Khandelwal M, Monjezi M, Yu Z, Qiu Y. Performance of Hybrid SCA-RF and HHO-RF Models for Predicting Backbreak in Open-Pit Mine Blasting Operations. Nat Resour Res 2021;30:4753–71. https://doi.org/10.1007/s11053-021-09929-y.
[92] Zhou J, Huang S, Qiu Y. Optimization of random forest through the use of MVO, GWO and MFO in evaluating the stability of underground entry-type excavations. Tunnelling and Underground Space Technology 2022;124:104494. https://doi.org/10.1016/j.tust.2022.104494.
[93] Karir D, Ray A, Bharati AK, Chaturvedi U, Rai R. Stability prediction of a natural and man-made slope using various machine learning algorithms. Transportation Geotechnics 2024;34:100745. https://doi.org/10.1016/j.trgeo.2022.100745.