Contents lists available at ScienceDirect

Arabian Journal of Chemistry

journal homepage: www.ksu.edu.sa

Original article

Product prediction of fixed-bed coal pyrolysis using a fusion model

Shiyao Yu, Chuyang Tang *, Xinyu Yang, Xinyuan An, Yi Wang

School of Civil Engineering, University of Science and Technology Liaoning, 189#, Qianshan Road, Liaoning Province 114051, PR China

ARTICLE INFO

Keywords:
Fusion model
Coal pyrolysis
Prediction
Learning weight

ABSTRACT

Here, a proactively optimized fusion model (FM) for predicting the product yield of coal pyrolysis was developed. Eight coal characteristics (including pyrolysis temperature and proximate and ultimate analyses) were chosen as input parameters. Multiple linear regression (MLR), support vector machine (SVM), and random forest (RF) models were applied as the base models to form the FM. Sixty sets of experimental data from the literature were used for training and testing the base models. Different learning weights are assigned to the base models according to their predictive performance. The FM proactively improve the model outputs by means of the dynamical learning weight results. The coefficient of determination (R^2) and the root-mean-squared error (RMSE) derived from the FM model were better than those of the base models. Moreover, the maximum relative error between the experimental data and model outputs was just 0.37%. These results suggest that FMs can be used to develop better predictive models for the yields of co-pyrolysis products. The FM proactively optimized the outputs base on learning weight algorithm and had better predicted performance than base models with less data.

1. Introduction

Despite the increasing demand for alternative energy sources, coal remains the primary fuel for energy production worldwide (Li et al., 2022a, 2022b; He et al., 2021; Armin et al., 2021). Humanity has depended heavily on coal for over a century due to its abundant reserves and the existing energy consumption structure (Zhu et al., 2022). In recent decades, large-scale direct combustion of low-grade coal has led to serious environmental problems (Kwon et al., 2019), while the increased consumption of fossil fuels has caused a continuous increase in annual greenhouse gas emissions worldwide (Chen et al., 2021a, 2021b; Li et al., 2022a, 2022b; Wang et al., 2023). Therefore, effective utilization of coal resources remains an important global challenge. The rapid development of coal chemical industry has brought about great environmental problems (Ju et al., 2024). Pyrolysis processing represents a convenient method for converting coal into syngas and porous carbon materials (Li et al., 2020). It is also has great significance in gas and water treatment. As a common conversion method, low-temperature pyrolysis can be understood as the thermal decomposition of coal in the temperature range of 600–800 ℃ (Chen et al., 2019). Furthermore, pyrolysis coal is one of the main sources of modern chemical raw materials (Yang et al., 2022). A large number of studies have shown that the regulation and control of coal pyrolysis can improve the quality and quantity of syngas, composite carbon materials, and high added-value chemicals produced (Yan et al., 2022; Nyakuma et al., 2021; Dwivedi et al., 2019). It is also of great significance to develop an efficient and environmentally-friendly coal chemicals industry to reduce carbon dioxide emissions (Jiang et al., 2020). However, investigating the conditions under which coal pyrolysis is optimized using a purely experimental approach is costly and time-consuming. There are many factors affecting pyrolysis. Though quite a good number of studies were conducted to analyze the individual impact of many parameters on pyrolysis, their interactive effects have not been addressed well and the research process is slow (Gopal et al., 2019) Therefore, the use of machine learning models to predict the yields of pyrolysis products can increase the efficiency and accuracy of experiments (Zhu et al., 2019a, 2019b).

With the evolution of deep learning algorithms, certain documented models have predicted experimental results with good effect (Cheng et al., 2020). In particular, multiple linear regression (MLR), support vector machine (SVM), and random forest (RF) models are widely applicable in nonlinear projects. Research has shown that machine learning models are capable of accurately predicting pyrolysis products (Zhang et al., 2022; Chen et al., 2018). These predicted results also contribute to an increased understanding of coal pyrolysis behavior. Jiang et al. (Jiang et al., 2022) successfully forecasted the

* Corresponding author.
E-mail address: astcy@126.com (C. Tang).

https://doi.org/10.1016/j.arabjc.2023.105562
Received 1 June 2023; Accepted 12 December 2023
Available online 13 December 2023

thermogravimetric (TG) data of the mixed pyrolysis of coal slime (CS) and cattle manure (CM) at different proportions using an artificial neural network (ANN) model. Through adopting bagging integration, the RF model combines the predictive advantages of multiple decision trees to improve the generalization ability, thereby exhibiting high predictive accuracy in resolving pyrolysis projects (Phromphithak et al., 2021). Hao Wei et al. (Wei et al., 2022) comprehensively studied the synergistic effects of co-pyrolysis and successfully predicted the product yields using an RF model. Ullah et al. (Ullah et al., 2021) found that an ANN model could reliably predict the yield of bio-oil derived from biomass pyrolysis. Tang et al. (Tang et al., 2021) successfully used a combination of SVM and RF models to predict the composition of biomass syngas. Phromphithak et al. successfully used ANN and an SVM model to predict the constituents of cellulose-rich materials in terms of the cellulose enrichment factor. Allen and Downie (Allen and Downie, 2020) applied mixed effects logistic regression to predict the biochar yield of residual biomass. However, such models still have shortcomings due to the limitations of regression model selection. At present, most studies on the prediction of pyrolysis products adopt a base model with limited accuracy. So in this study, adopt the FM model to optimized the outputs proactively base on learning weight algorithm to avoid the limitations of the core equations of the base model.

In the present study, the effectiveness of FM models in the prediction of coal pyrolysis products is explored. To this end, sixty sets of experimental results on coal pyrolysis in a fixed bed were used to form the dataset. The proximate and ultimate analyses and the pyrolysis temperature were selected as the model input parameters. As the common model algorithm, RF model was good at investigating the feature importance of datasets. SVM and MLR model was skilled in the nonlinear problems caused by complex reaction. Therefore, RF, SVM, and MLR algorithms were used to build the base models. An FM derived from the base models was then applied to predict the yield of coal pyrolysis products. The FM regulated the base models to optimize outputs spontaneously through applying learning weights. The results from this study are expected to provide machine learning assistance for expanding the understanding of coal pyrolysis behavior and aiding in the directional preparation of high value-added chemicals.

2. Methods

2.1. Data selection

In this study, 29 types of coal were considered, including 6 kinds of lignite, 19 kinds of bituminous coal, and 4 kinds of anthracite coal. The ultimate and proximate analyses of the samples are shown in Tables 1 and 2. Sixty sets of experimental data on coal pyrolysis were obtained from the literature (Wu et al., 2022; Qian et al., 2019; Zhao et al., 2011; Zhang et al., 2014; Ban et al., 2022; Chen et al., 2021a, 2021b; Qiang et al., 2021), all of which were carried out in a fixed bed at various pyrolysis temperatures, and the yields of water, tar, syngas, and char in the pyrolysis products were measured.

2.2. Data analysis and pretreatment

Considering the ultimate and proximate analysis of the coal samples, and the different pyrolysis temperatures, nine characteristic parameters were selected as the inputs, namely moisture (M), ash (A), volatile matter (V), carbon (C), hydrogen (H), nitrogen (N), sulphur (S), oxygen (O), and temperatue (T). Detailed information on the dataset is provided in the Supplementary data.

The Pearson correlation coefficient (PCC), denoted as r, is used to describe the linear correlation between two variables, and its value ranges between − 1 and 1, and is calculated as follows (Abnisa and Daud, 2014):

$$r = \frac{Cov(X,\ Y)}{\sigma_X \sigma_Y} \tag{1}$$

where $Cov(X,\ Y)$ is the covariance between variables X and Y and σ_X and σ_Y are the standard deviations of X and Y, respectively. The PCC is subject to a two-tailed t-distribution with $(n - 2)$ degrees of freedom. Therefore, the significance level can be tested as follows (Were et al., 2015):

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{2}$$

where t is the value of the test statistic and n is the total number of samples in the dataset. This allows the p-value, i.e., the significance level indicator, to be obtained. The closer p is to 1, the greater the correlation between the two characteristic variables.

2.3. Base models

This study uses suitable regression algorithms for MLR, SVM, and RF. Adequate testing and training were implemented to build the three base models via the scikit-learn library of the Python programming language. All inputs were normalized in advance using the equation (Fletcher et al., 2012):

$$X_z = \frac{X - \overline{X}}{\sigma} \tag{3}$$

Table 1
Proximate analyses of coal samples (wt.%).

Sample	M	A	V	FC	Sample	M	A	V	FC
Huainan	2.18	26.34	42.12	57.88	Honglaiwa	6.27	4.69	38.67	61.33
Shenmu	3.78	8.63	37.97	62.03	LD	7.68	6.38	37.62	62.38
Daliuta	9.58	4.98	35.21	64.79	Shenmu	5.39	8.07	36.83	63.17
Heishan	3.9	3.31	35.61	64.39	Qianshuta	5.3	13.14	39.3	60.7
Neimeng	8.89	14.38	57.88	42.12	Pingshuo	2.23	17.93	37.19	62.81
Shengli	9.58	14.41	46.48	53.52	Lignitous	10.53	8.68	38.9	61.1
Xiaolongt	16.68	10.21	55.62	44.38	Yilan	4.61	34.97	51.83	48.17
Zhundong	5.84	5.14	35.75	64.25	Changyan	5.64	17.44	39.81	60.19
Xianfeng	10.35	8.42	52.58	47.42	Shendong	11.53	10.88	40.55	59.45
Zhungeer	4.76	19.88	41.16	58.84	Pulverize	4.61	5.13	37.97	62.03
Shengl	15.78	17.31	47.97	52.03	Pingshuo	1.2	39.55	42.38	57.62
Xilinhot	3.88	11.19	44.32	55.68	Hongshaqu	2.15	7.17	36.22	63.78
Naomaohu	0.72	4.9	51.39	48.61	Zaozhuang	3.16	39.61	46.07	53.93
Shaerhu	15.64	12.13	40.56	59.44	Shenfu	9.37	7.09	39.41	60.59
Daliuta	9.34	5.04	36.47	63.53	Hami	7.46	10.73	64.17	35.83
Yushuwan	5.95	6	38.17	61.83	Naomaohu	4.51	7.5	48.99	51.01
Hongliuli	5.1	5.63	36.21	63.79	Shendong	15.32	9.53	35.61	64.39
Haiwankua	6.2	9.14	38.68	61.32	–	–	–	–	–

Table 2
Ultimate analyses of coal samples (wt.%).

Sample	H	O	S	N	C	Sample	H	O	S	N	C
Huainan	6.07	8.61	0.66	1.54	83.12	Honglaiwa	4.71	11.02	0.38	1.12	82.75
Shenmu	5.49	9.37	0.32	1.08	84.24	LD	4.84	10.32	0.78	1.15	82.85
Daliuta	4.96	12.59	0.34	1.09	81.02	Shenmu	4.7	10.98	0.27	1.11	82.91
Heishan	4.86	11.22	0.38	0.81	82.73	Qianshuta	5.07	9.89	0.63	1.12	83.2
Neimeng	5.92	18.53	1.4	1.16	72.99	Pingshuo	5.2	11.95	1.06	1.38	80.41
Shengli	4.65	19.8	1.4	1.16	72.99	Lignitous	3.18	21.22	0.34	1.26	54.82
Xiaolongt	4.72	23.26	1.21	1.86	68.94	Yilan	6.23	20.61	0.63	1.57	70.96
Zhundong	3.3	15.39	0.55	0.4	80.36	Changyan	3.95	14.45	2.87	0.65	78.08
Xianfeng	5.06	20.49	0.94	2.25	71.26	Shendong	3.68	9.94	0.42	0.71	64.56
Zhungeer	4.41	16.44	0.5	1.34	77.31	Pulverize	4.28	10.83	0.29	0.99	74.11
Shengl	4.49	24.54	1.14	1.11	68.48	Pingshuo	2.16	11.66	2.47	0.94	42.49
Xilinhot	4.26	21.74	0.61	1.13	72.18	Hongshaqu	4.66	17.71	0.29	1.95	75.39
Naomaohu	5.89	19.53	0.2	0.99	73.53	Zaozhuang	5.05	24.25	3.76	1.2	65.74
Shaerhu	3.8	19.38	0.2	0.89	76.67	Shenfu	5.86	8.79	0.28	0.77	66.64
Daliuta	4.66	12.72	0.32	1.08	81.21	Hami	5.46	17.58	1.51	0.92	74.36
Yushuwan	4.91	10.96	0.73	1.03	82.32	Naomaohu	5.67	18.36	0.27	0.93	74.77
Hongliuli	4.63	11.59	0.22	1.06	82.48	Shendong	4.47	16.26	0.39	1.04	77.84
Haiwankua	4.95	10.98	0.36	1.19	82.49	–	–	–	–	–	–

where X and $\overline{X}$ are the normalized and mean values, respectively, of the variable, and σ is the standard deviation.

The RF model is an integrated learning method based on decision trees, and can deal with high-dimensional feature samples with high degree of accuracy (Leng et al., 2021). The importance of features can be calculated and ranked using the average impurity removal method (Tang et al., 2020). Every partition makes a local optimal choice during the generation of the decision tree. Hence, the result cannot ensure that each choice is globally optimal, which can easily lead to overfitting.

The SVM model is a supervised learning model that uses classification and regression analysis to analyze data. It is adept at the cross validation of selection parameters, weighting of unbalanced samples, and probability estimation of class problems (Bai et al., 2017). Nevertheless, the use of larger datasets commonly lowers its predictive accuracy. For the resolution of nonlinear problems with no general standard for the selection of the kernel function, the predictive accuracy of the base model depends largely on the effectiveness of the kernel function, which for SVM is defined as follows:

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \tag{4}$$

where x is the eigenvector of the sample, $K(x_i, x_j)$ is the transvection of two input vectors, and γ is the argument of the kernel function. The

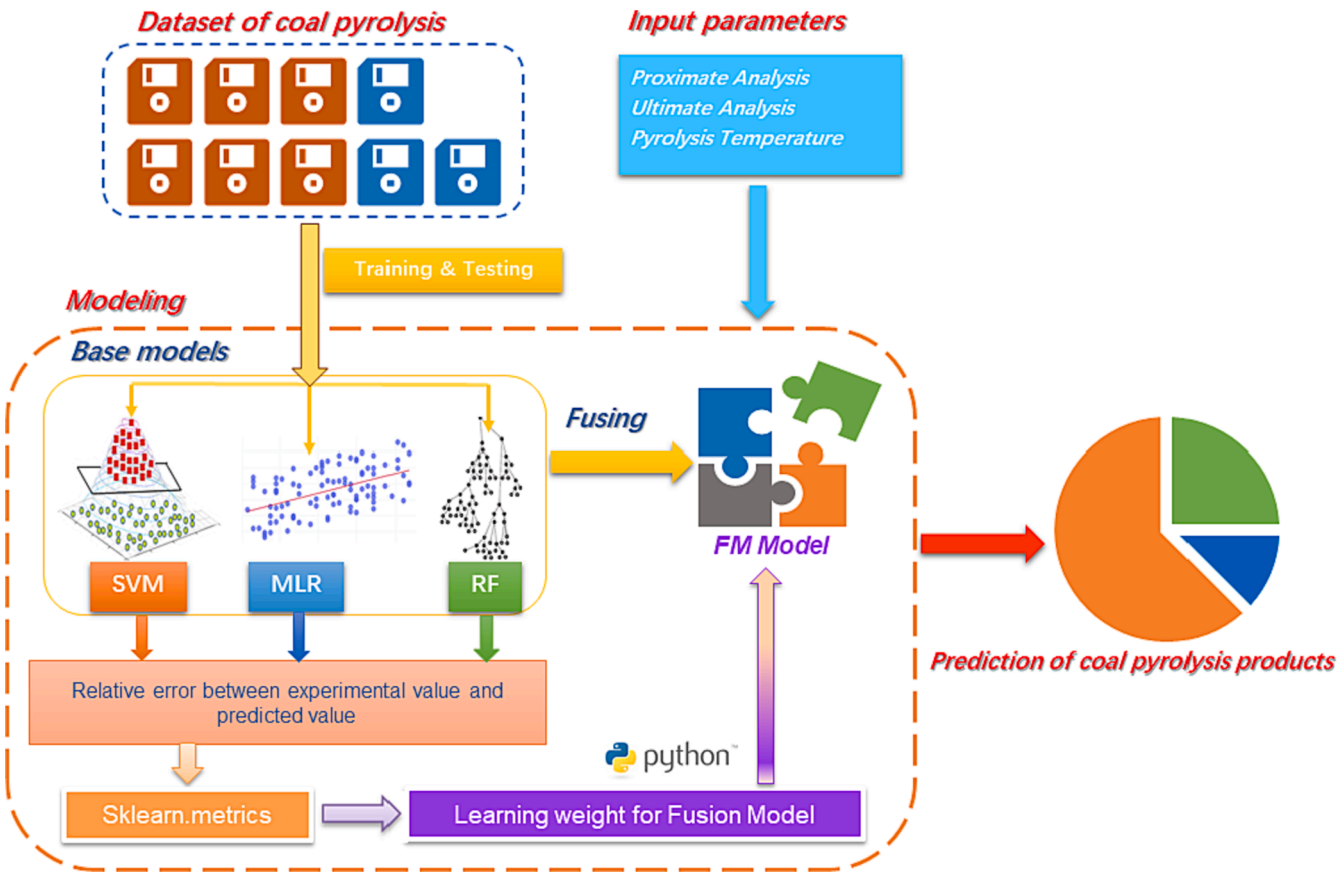

Fig. 1. Structure of the fusion model.

smaller the value of $K(x_i, x_j)$, the more continuous the classification interface, while the higher its value, the better the classification effect.

The MLR model is a kind of subsection method, which converges more stably with the structural priors of pre-training and the optimization training of full-space parameter. However, the calculation and storage complexity of the model is high, and it is difficult to set the hyperparameters. The multiple linear regression equation is as follows (Xing et al., 2019):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta p x p \quad (5)$$

where y is the model output, $x_i (i = 1,2,\ldots,p)$ are the inputs, and β_i are the coefficients. The goal of MLR is to obtain the best coefficient values for the regression model.

2.4. Fusion models

The limited algorithms of base models limit their predictive performance in finding the optimal solution. However, integrating multiple machine learning models can improve overall predictive ability. Fig. 1 illustrates the structure of the FM. The 60 sets of coal pyrolysis data were randomly divided into training sets and test sets in a ratio of 2:1. Nine parameters were selected as inputs based on the results of Pearson correlation analysis. The training data were imported into the Python scikit-learn library to build the SVM and MLR models, while the RF model was established using the NumPy library. The predicted outputs of the base models were further optimized using the test data. Moreover, the predicted performance of the base models was quantified using a learning weight algorithm with 5-fold cross-validation. The quantified performance ulteriorly regulated the weights of the base models in the output of the FM. The better the base model performs, the higher its weighting in the FM. The FM was then trained and evaluated using the training data until its predictive performance reached the presupposed accuracy level.

2.5. Performance evaluation

The quality of a linear regression model may be evaluated using a variety of measures. In this study, the determination coefficient (R^2) and the root mean square error (RMSE) were used, which are calculated as follows (Xue et al., 2020):

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left(Y_i^{exp} - Y_i^{pred}\right)^2}{\sum_{i=1}^{N} \left(Y_i^{exp} - \overline{Y}^{exp}\right)^2} \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left(Y_i^{exp} - Y_i^{pred}\right)^2} \quad (7)$$

In the Eqs. (6) and (7), Y_i^{exp}, Y_i^{pred}, and $\overline{Y}^{exp}$ are the experimental, predicted, and average experimental values, respectively, and N is the total quantity of data in the test set. R^2 varies between 0 and 1 and represents the goodness of fit of the model, while RMSE is the mean of the square root of the error between the predicted and experimental values. Hence, having R^2 close to 1 and RMSE close to 0 means that the model achieved the expected results.

3. Results and discussion

3.1. Statistical analysis of the dataset

Fig. 2 shows box and whisker plots of data obtained from the proximate and ultimate analyzes of the coal samples and the yields of pyrolysis products. The main parameters in the Fig. 2 include the coal characteristics of moisture (M), ash (A), volatile matter (V), and fixed carbon (FC) and the elemental content of carbon (C), hydrogen (H), oxygen (O), and nitrogen (N). The distribution of the experimental data shows that A has the largest spread, with a range of 3.29 wt% to 44.11 wt%, while N. has the smallest (0.40 wt% to 2.25 wt%). In addition, the contents of N and H obtained from ultimate analysis were lower than that of other elements, and were mainly concentrated from 1.00 wt% to 5.00 wt%. In terms of the yield of pyrolysis products, the values of water yield and tar yield have outliers above the upper limit, which reflects the fact that altering the pyrolysis conditions has a greater effect on the yields of these two products. In terms of the proximate analysis, the difference between the mean and median of A (3.10 wt%) is higher than that of M (1.28 wt%), V (2.25 wt%), and FC (2.25 wt%). Moreover, in terms of the ultimate analysis, the mean value of C is 76.42 wt% and its median is 78.08 wt%. In terms of the product distribution, the difference between the mean and median of the char yield is 2.18 wt%, which is close to that of syngas (2.18 wt%), while the yields of water and char ranged from 1.27 wt% to 19.67 wt% and 54.1 wt% to 83.4 wt%, respectively. These results show that the divergence of the data is in the order of proximate analysis > yields of pyrolysis products > ultimate analysis, suggesting that the thermochemical properties of coal have a more pronounced effect on the pyrolysis product distribution than the elemental compositions.

Fig. 3 illustrates the linear relationships between two arbitrary variables. As the figure shows, FC ($p < 0.01$) is perfectly negatively correlated with V ($p < 0.01$), with a correlation coefficient of − 1. Here, the p-value indicates the reliability of the linear correlation, which is

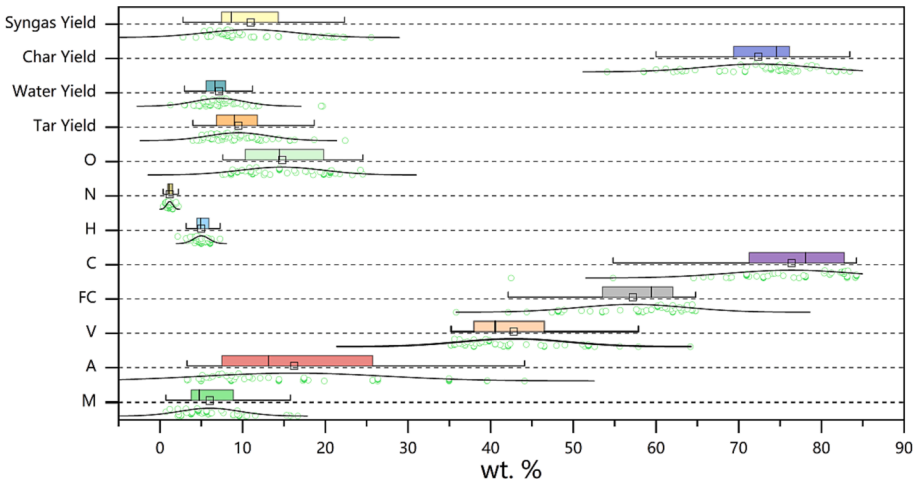

Fig. 2. Distributions of proximate analysis, ultimate analysis, and yield of pyrolysis products.

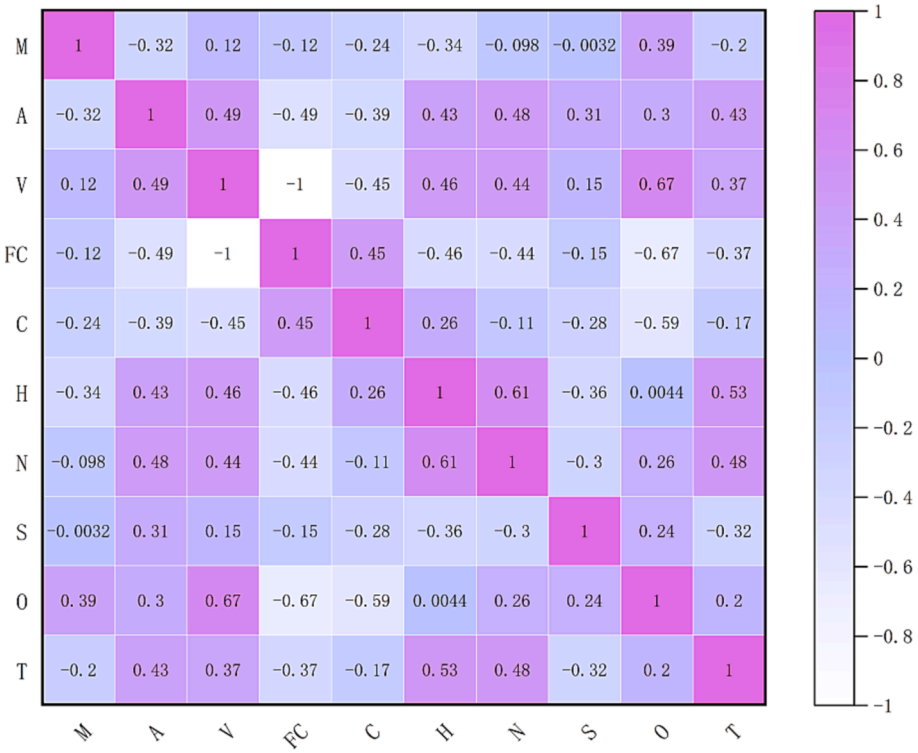

Fig. 3. Matrix of Pearson correlation coefficients between the different variables.

associated with the equation FC = 100 – (V + A + M) for the estimation of fixed carbon. Moreover, Fig. 3 also shows that FC is significantly influenced by A, C, H, N, and O, and that the pyrolysis temperature is positively correlated with most of the variables. On the other hand, Fig. 3 reveals a less significant linear relationship between the proximate and ultimate analyses of the experimental data. Oppositely, there is a positive correlation between C and H with a correlation coefficient of 0.26, while that of C and V is negative, at − 0.45. S, however, is only positively correlated with A and V, with correlation coefficients of 0.31 and 0.15, respectively. In general, V and FC are moderately correlated with H and N and strongly correlated with O. There is also a clear correlation between FC and C, while M is unaffected by the elemental composition of coal. These results also indicate that the thermochemical properties of coal have a weak relationship with its S content.

3.2. Predictive performance of base models and FM

Table 3 shows the R^2 and RMSE values of the predicted results from the different models. It is evident that all models perform well for the prediction of tar, water, and syngas yields. But none of the models do well in predicting the char yield due to the moderate correlation between FC and the elemental composition of coal. When all models have poor prediction effect on Char, RF and FM model have relatively good prediction effect. Although the R^2 value of FM model is only higher than that of RF model 0.0198. Overall, the prediction effect of FM model is better than base models. Nevertheless, the FM has better predictive accuracy than the base models in forecasting the tar and syngas yield. Further, the water yield is also well predicted by the FM, which is only marginally behind the MLR model. However, the maximum R^2 of the base models for the prediction of the char yield is just 0.5017, which is far lower than those for the tar, water, and syngas yields, which suggests that the base models are less effective in predicting coke yield. It is also worth noting that the FM has the best predictive ability for the char yield of all the models, with the highest R^2 (0.5215) and the lowest RMSE (5.2134). As a result, the FM performs consistently well in predicting the pyrolysis yields of coal. This is because, even when the base models perform poorly, the FM is still able to proactively adjust the model output of the model via the learning weight method, resulting in greater predictive accuracy.

Fig. 4 shows a comparison between the experimental data and the predicted output of the different models, with the parity line indicated in blue. The results show that all the models achieve acceptable accuracy in predicting the tar, water, and syngas yields, with the predictive accuracy of the FM model being higher than the other base models. On the contrary, a noticeable deviation is observed between the predicted and experimental values of the char yield. The point-cloud distribution of all models is concentrated when the char yield is from 70 wt% to 80 wt%, and the prediction effect is acceptable. Within this range, as Fig. 4D shows, the point-cloud of char for the FM is more intensive and nearer the blue line than that of the base models, indicating that its a predictive performance is more reliable than the base models. The final result shows that the prediction performance of FM model is better than the

Table 3
R^2 and RMSE scores of the three base models and the FM.

Models	MLR		SVM		RF		FM	
Performance	R^2	RMSE	R^2	RMSE	R^2	RMSE	R^2	RMSE
Tar	0.9766	0.4803	0.9696	0.5471	0.9769	0.4771	0.9780	0.4648
Water	0.9543	0.6932	0.9483	0.7369	0.9521	0.7092	0.9527	0.7007
Syngas	0.9797	0.8956	0.9821	0.8414	0.9792	0.9058	0.9834	0.8101
Char	0.1349	7.0162	0.4806	5.4364	0.5017	5.3249	0.5215	5.2134

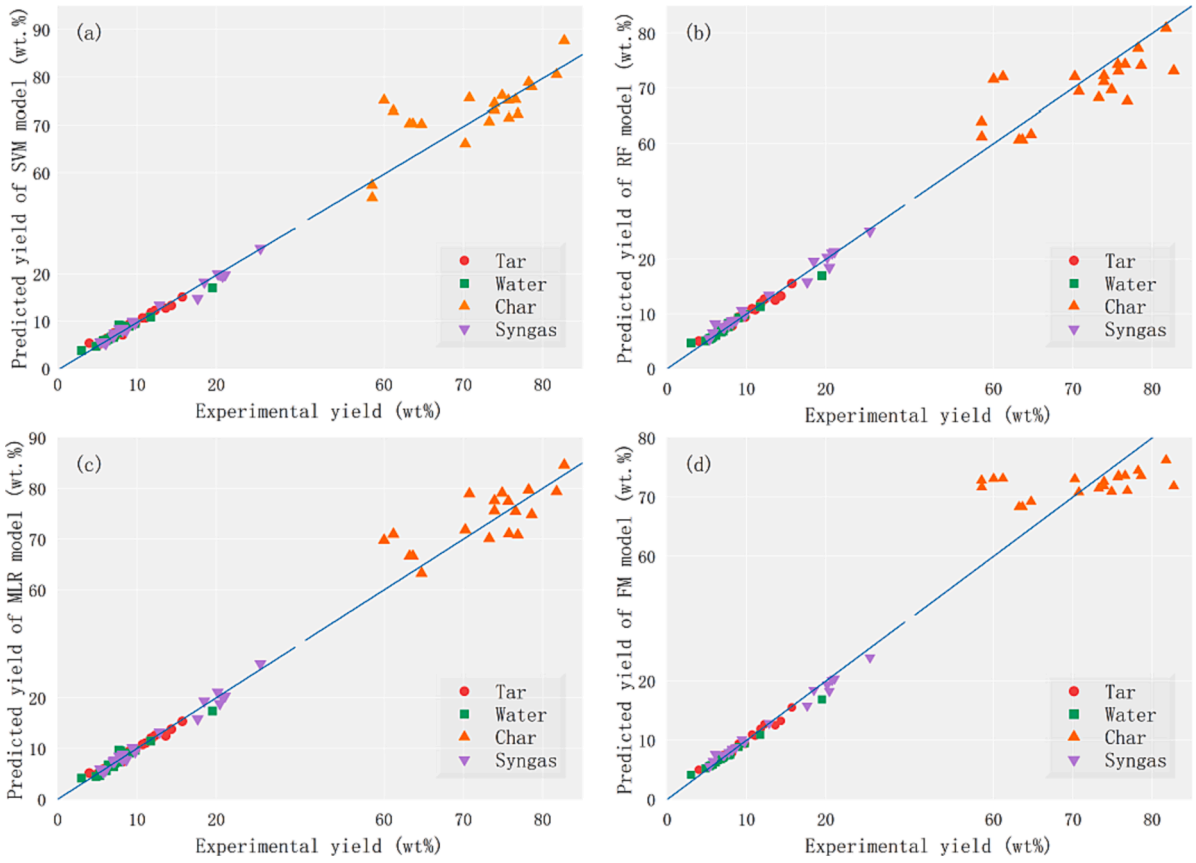

Fig. 4. Comparison between the experimental data and predicted outputs for different models: (A) SVM; (B) RF; (C) MLR; (D) FM.

base model.

3.3. Relative error analysis of the models

Fig. 5 shows the relative error between the experimental data and the predicted results. The Y-axis on the left represents the relative error, while that on the right represents the experimental yield of the pyrolysis products. The X-axis represents the number of sample groups in the test set. A smaller relative error indicates a more accurate prediction result, meaning that the predictive performance of the model is better. Fig. 5A indicates that all models do well in predicting the tar yield, with the predictive performance of the RF model being better than the SVM and MLR models. The relative errors between the RF model and the experimental data range from − 0.083 % to 0.251 %, with the maximum relative error corresponding to the experimental dataset labelled as No. 1. Meanwhile, the maximum relative error of the FM was the lowest among all the models, at 0.247 %, indicating that the predictive performance of the FM model is superior to all the base models in predicting the tar yield.

Fig. 5B shows that, just as for the tar yield, the maximum relative errors of all models occurred for the experimental dataset No.1. The relative error of the FM was the lowest among all the models, at 0.370 %. When the water yield was in the range of 3.01 wt% to 6.80 wt%, the FM and SVM models showed the best predictive accuracy. Further, when the water yield is > 6.80 wt%, the FM still possesses excellent predictive accuracy, with a relative error range from − 0.136 % to 0.006 %. Hence, compared with the base models, the FM better predicts the water yield.

Fig. 5C shows the relative errors of the char yield. The MLR model has the largest relative error range. The maximum relative error of RF model is 0.193 %. It appears that the RF model exhibits the best predicted accuracy for char yield among the base models. When the char yield is < 70.25 wt%, however, all models show worse predicted performance than for the other pyrolysis products (water, tar, and syngas). On the other hand, when the char yield is > 70.25 wt%, the predictive results for the base models are lower than the experimental data values. Further, the predictive accuracy of the FM for the char yield is lower than that of the RF model due to the poor predictive performance of the MLR and SVM models. However, the predictive performance of FM model is still comparatively excellent in the same condition. The relative error range of FM was − 0.132 % to 0.244 %, which was very close to that of the RF model (−0.120 % to 0.193 %). As Fig. 5D shows, the maximum relative error for the syngas yield was 0.351 %, obtained from the RF model. The predictive accuracy of the SVM model was the best of the base models, ranging from − 0.065 %. to 0.251 %. Furthermore, the percentages of model outputs that were higher than the corresponding experimental data values are 80 %, 50 %, and 40 % for RF, MLR, and SVM, respectively. The relative error of the FM is in the range of − 0.132 % to 0.244 %, meaning it has better predictive accuracy than the base models. Moreover, two-thirds of the predicted outputs from the FM model are, higher than the corresponding experimental values. Since the FM is derived from a combination of the base models, this indicates that it actively regulates the base model by training the weighting algorithm, which results in better predictive performance.

These results demonstrate that the FM is superior to the base models in predicting the yields of tar, water, and syngas, while coming a close second to the RF model for the char yield. Further, a maximum relative error of only 0.244 % indicates that the FM effectively utilizes the learning weight algorithm to proactively improve the predicted outputs.

3.4. Feature importance

Fig. 6 illustrates feature importance of input variables among the FM and base models. The feature importance indicates the extent of

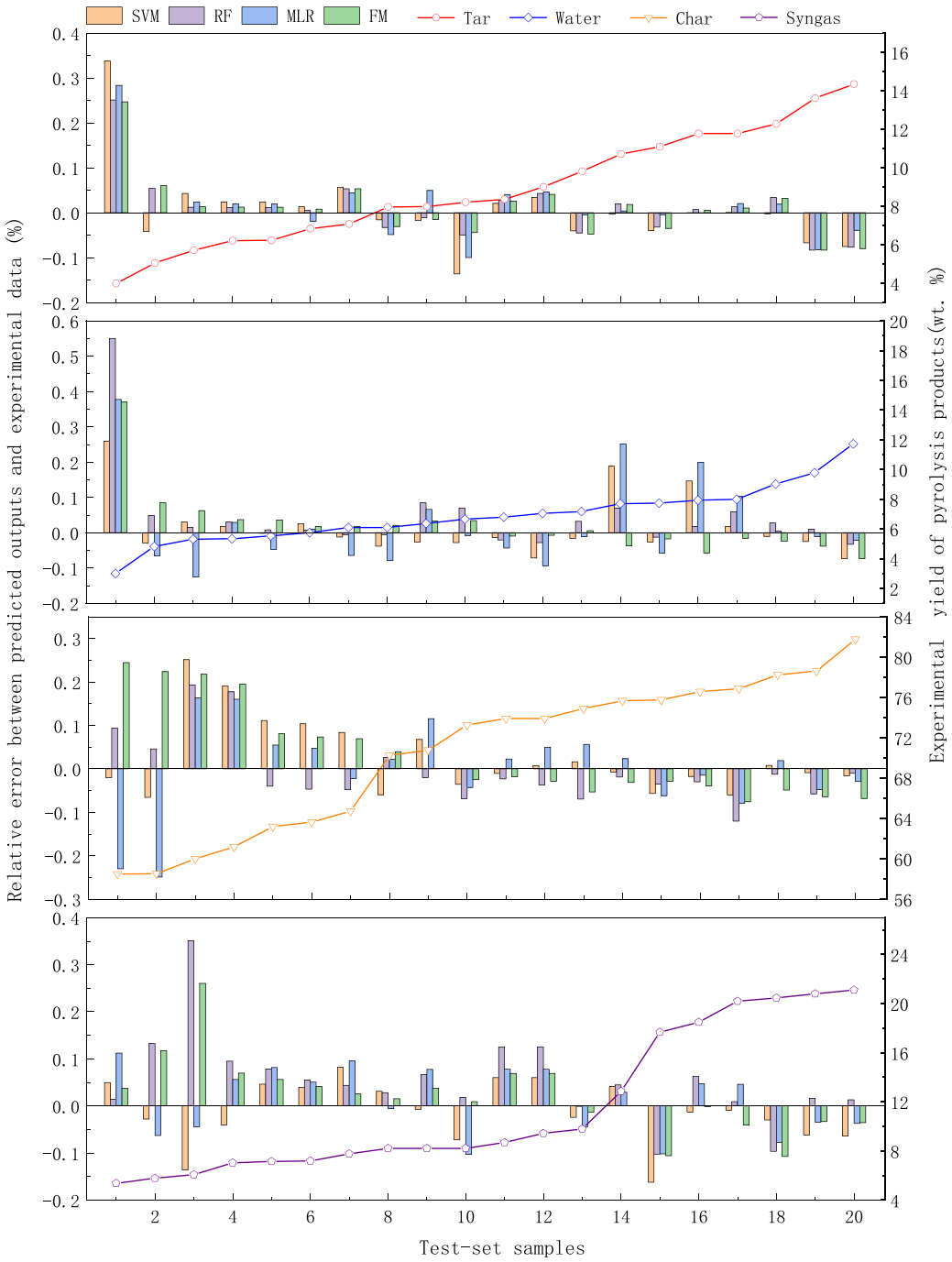

Fig. 5. Relative error between predicted outputs and experimental data: (A) tar yield; (B) water yield; (C) char yield; (D) syngas yield.

influence of input characteristics on the predicted yield of pyrolysis products. Fig. 6A shows that the oxygen content has the greatest influence on predicting the tar yield for each model. Moreover, the FM significantly enlarges the feature importance of oxygen content and reduces that of carbon content during the modelling process. As Fig. 6B, the most feature important input in predicting the water yield for each model is also the oxygen content, with the feature importance of oxygen content from the FM being larger than that of the base model. Meanwhile, the FM significantly lowers the feature importance of carbon content and volatile matter. The feature importance of input variables for predicting syngas is quite different from those for predicting tar and water.

Fig. 6C shows the inconsistency in the feature importance variation of moisture content. The feature importance of moisture content in the FM (41.45 %) is much lower than that of the base models (57.59 %). Fig. 6D shows that the FM significantly decreases the feature importance of moisture content, while that of oxygen content in the FM (22.07 %) is larger than that of the base models (12.40 %). In terms of the results From Figs. 5 and 6, it is evident that the FM proactively varies the feature importance of inputs to improve the predicted outputs.

4. Conclusion

In this study, sixty groups of fixed-bed coal pyrolysis data were used to establish an FM, with MLR, RF, and SVM base models. Based on Pearson correlation analysis, nine input parameters were selected from ultimate and proximate analyses of coal. The models were first tested and trained before being incorporated, into the FM using a weighting

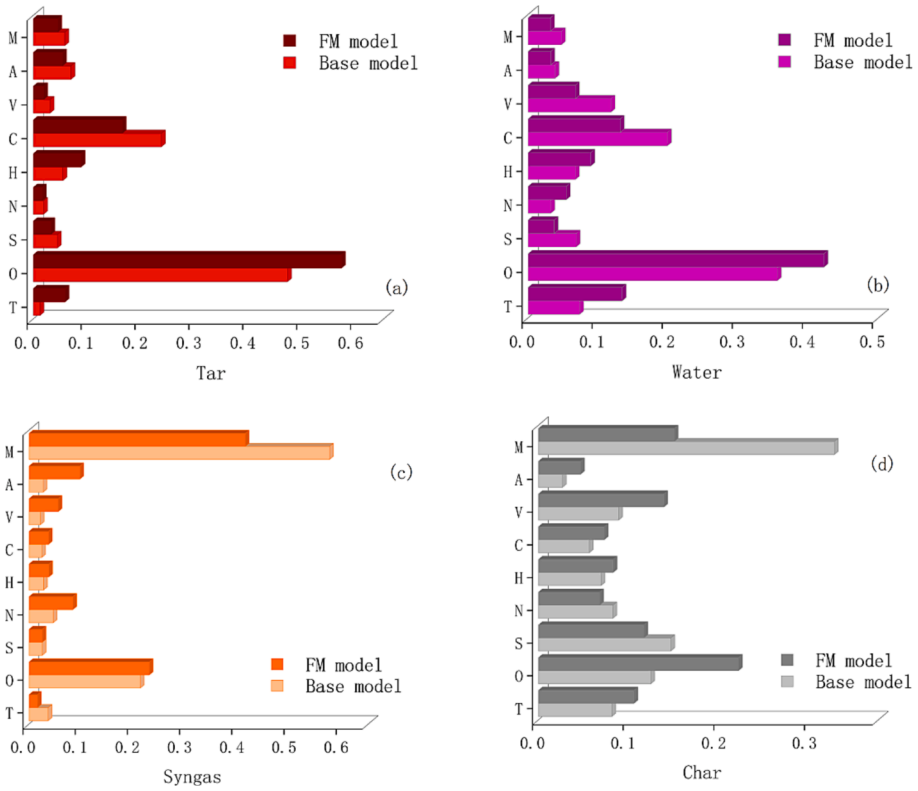

Fig. 6. Feature importance of input variables via the FM and base models: (A) tar yield, (B) water yield, (C) syngas yield (D) char yield.

algorithm. The results of the study were as follows:

(1) The predictive results for pyrolysis product yield were analyzed via the coefficient of determination (R^2) and root mean square error (RMSE). The results suggest that the FM improved the output compared with the base models.

(2) A comparison of the relative error between predicted outputs and experimental data showed that the FM dynamically used the weighting algorithm to regulate the output weights of different base models according to their predictive performance. The maximum relative error of the FM model was only 0.37 %.

(3) Further analysis of the input characteristic changes of the FM showed that it can analyze the predictive performance of the base models independently during operation and optimize the output by adjusting the learning weight. By actively changing the weights of input features, the output results were regulated and optimized.

(4) The use of FMs provides new insight into prediction of coal pyrolysis product yields and aids in the optimization of pyrolysis products.

CRediT authorship contribution statement

Shiyao Yu: Conceptualization, Methodology, Software, Investigation, Formal analysis, Writing – original draft. **Chuyang Tang:** Conceptualization, Funding acquisition, Resources, Supervision, Writing – review & editing. **Xinyu Yang:** Data curation, Writing – original draft. **Xinyuan An:** Visualization, Investigation. **Yi Wang:** Resources, Supervision.

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 work was funded by National Key research and development program (No. 2022YFE0208100) and Basic scientific research projects of the Department of Education, Liaoning Province (No. LJKZZ20220044).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.arabjc.2023.105562.

References

Abnisa, F., Daud, W.M.A.W., 2014. A review on co-pyrolysis of biomass: an optional technique to obtain a high-grade pyrolysis oil. Energy Convers. Manage. 87, 71–85. https://doi.org/10.1016/j.enconman.2014.07.007.

Allen, J.A., Downie, A.E., 2020. Predicting slow pyrolysis process outcomes with simplified empirical correlations for a consistent higher heating temperature: biochar yield and ash content. Energy Fuels 34, 14223–14231. https://doi.org/10.1021/acs.energyfuels.0c02597.

Armin, C., Michael, S., Muhammad, M., Michael, Z., 2021. Mixed-integer linear programming based optimization strategies for renewable energy communities. Energy 237. https://doi.org/10.1016/j.energy.2021.121559.

Bai, Y., Li, Y., Liu, Y., Ma, Z., 2017. Short-term prediction of distribution network faults based on support vector machine. Proceedings of the 2017 12th IEEE Conference on Industrial Electronics and Applications, ICIEA 2017-Febru, 1421–1426. 10.1109/ICIEA.2017.8283062.

Ban, Y.P., Jin, L.J., Zhu, J.L., Liu, F.G., Hu, H.Q., 2022. Insights into effect of Ca(OH)2 on pyrolysis behaviors and products distribution of Hongshaquan coal. Fuel 307, 121791. https://doi.org/10.1016/j.fuel.2021.121791.

Chen, X.Y., Liu, L., Zhang, L.Y., Zhao, Y., Qiu, P.H., 2019. Pyrolysis characteristics and kinetics of coal-biomass blends during co-pyrolysis. Energy Fuels 33 (2), 1267–1278. https://doi.org/10.1021/acs.energyfuels.8b03987.

Chen, Z.H., Wang, D.L., Yang, H., Zhang, Y.S., Li, Y.J., Li, C.M., Yu, J., Gao, S.Q., 2021b. Novel application of red mud as disposal catalyst for pyrolysis and gasification of coal. Carbon Resources Conversion 4, 10–18. https://doi.org/10.1016/j.crcon.2021.01.001.

Chen, Y.W., Wong, C.W.Y., Yang, R., Miao, X., 2021a. Optimal structure adjustment strategy, emission reduction potential and utilization efficiency of fossil energies in China. Energy 237, 121623. https://doi.org/10.1016/j.energy.2021.121623.

Chen, X., Zhang, H.Y., Song, Y., Xiao, R., 2018. Prediction of product distribution and bio-oil heating value of biomass fast pyrolysis. Chem. Eng. Process 130, 36–42. https://doi.org/10.1016/j.cep.2018.05.018.

Cheng, F.W., Luo, H.X., Colosi, L.M., 2020. Slow pyrolysis as a platform for negative emissions technology: an integration of machine learning models, life cycle assessment, and economic analysis. Energy Convers Manage 223, 113258. https://doi.org/10.1016/j.enconman.2020.113258.

Dwivedi, K.K., Chatterjee, P.K., Karmakar, M.K., Pramanick, A.K., 2019. Pyrolysis characteristics and kinetics of Indian low rank coal using thermogravimetric analysis. Int. J. Coal Sci. Technol. 6 (1), 102–112. https://doi.org/10.1007/s40789-019-0236-7.

Fletcher, T.H., Pond, H.R., Webster, J., 2012. Prediction of tar and light gas during pyrolysis of black liquor and biomass. Energy Fuels 26 (May-June), 3381–3387. https://doi.org/10.1021/ef300574n.

Gopal, P., Nadimpalli, G., Saikia, R., 2019. Optimization of pyrolyzer design to produce maximum bio-oil from Saccharum ravannae L.: an integrated approach using experimental data and artificial intelligence. Biomass Convers. Biorefinery 9 (4), 727–736. https://doi.org/10.1007/s13399-019-00397-2.

He, Q., Gong, Y., Ding, L., Guo, Q.H., Yoshikawa, K., Yu, G.S., 2021. Reactivity prediction and mechanism analysis of raw and demineralized coal char gasification. Energy 229, 120724. https://doi.org/10.1016/j.energy.2021.120724.

Jiang, P., Meng, Y., Lu, Z.Y., Xu, L., Yang, G., Luo, X., Shi, K.Q., Wu, T., 2020. Kinetic and thermodynamic investigations of CO2 gasification of coal chars prepared via conventional and microwave pyrolysis. Int. J. Coal Sci. Technol. 7 (3), 422–432. https://doi.org/10.1007/s40789-020-00358-5.

Jiang, C.L., Zhou, W.L., Bi, H.B., Ni, Z.S., Sun, H., Lin, Q.Z., 2022. Co-pyrolysis of coal slime and cattle manure by TG-FTIR-MS and artificial neural network modeling: pyrolysis behavior, kinetics, gas emission characteristics. Energy 247, 123203. https://doi.org/10.1016/j.energy.2022.123203.

Ju, L., Zebin, W., Xian, Q.P., 2024. Analysis of the influence of different pore ratios on methane adsorption characteristics of coal. Fuel 358 (PB). https://doi.org/10.1016/J.Fuel.2023.130273.

Kwon, G., Park, Y.K., Ok, Y.S., Kwon, E.E., Song, H., 2019. Catalytic pyrolysis of low-rank coal using Fe-carbon composite as a catalyst. Energy Convers Manage 199, 111978. https://doi.org/10.1016/j.enconman.2019.111978.

Leng, E.W., He, B., Chen, J.W., Liao, G.L., Ma, Y.J., Zhang, F., Liu, S., E, J.Q., 2021. Prediction of three-phase product distribution and bio-oil heating value of biomass fast pyrolysis based on machine learning. Energy 236. https://doi.org/10.1016/j.energy.2021.121401.

Li, J.J., Wang, Y.G., Xu, D.P., Xie, K.C., 2020. High-resolution analysis of life-cycle carbon emissions from China's coalfired power industry: a provincial perspective. Int J Greenh Gas Con 100, 103110. https://doi.org/10.1016/j.ijggc.2020.103110.

Li, W.J., Yu, X.Z., Hu, N., Huang, F., Wang, J., Peng, Q.N., 2022b. Study on the relationship between fossil energy consumption and carbon emission in Sichuan Province. Energy Reports 8, 53–62. https://doi.org/10.1016/j.egyr.2022.01.112.

Li, H., Zhang, R.N., Ai, X.N., 2022a. Cost estimation of "coal-to-gas" project: government and residents' perspectives. Energy Policy 167, 113077. https://doi.org/10.1016/j.enpol.2022.113077.

Nyakuma, B.B., Jauro, A., Akinyemi, S.A., Faizal, H.M., Nasirudeen, M.B., Fuad, M.A.H.M., Oladokun, O., 2021. Physicochemical, mineralogy, and thermo-kinetic characterisation of newly discovered Nigerian coals under pyrolysis and combustion conditions. Int. J. Coal Sci. Technol. 8 (4), 697–716. https://doi.org/10.1007/s40789-020-00386-1.

Phromphithak, S., Onsree, T., Tippayawong, N., 2021. Machine learning prediction of cellulose-rich materials from biomass pretreatment with ionic liquid solvents. Bioresour. Technol. 323, 124642 https://doi.org/10.1016/j.biortech.2020.124642.

Qian, Y.N., Zhan, J.H., Yu, Y., Xu, G.W., Liu, X.X., 2019. CFD model of coal pyrolysis in fixed bed reactor. Chem. Eng. Sci. 200, 1–11. https://doi.org/10.1016/j.ces.2018.12.064.

Qiang, L.Y., Bai, B.Y., Peng, Z.W., Zhang, S.S., Chang, H., Sun, M., Xu, L., Ma, X.X., 2021. Research on the relationship between the structure and pyrolysis characteristics of pretreated Shendong coal. Fuel 305, 121515. https://doi.org/10.1016/j.fuel.2021.121515.

Tang, Q.H., Chen, Y.Q., Yang, H.P., Liu, M., Xiao, H.Y., Wu, Z.Y., Chen, H.P., Naqvi, S.R., 2020. Prediction of bio-oil yield and hydrogen contents based on machine learning method: effect of biomass compositions and pyrolysis conditions. Energy Fuels 34, 11050–11060. https://doi.org/10.1021/acs.energyfuels.0c01893.

Tang, Q.H., Chen, Y.Q., Yang, H.P., Liu, M., Xiao, H.Y., Wang, S.R., Chen, H.P., Naqvi, S.R., 2021. Machine learning prediction of pyrolytic gas yield and compositions with feature reduction methods: Effects of pyrolysis conditions and biomass characteristics. Bioresour. Technol. 339, 125581 https://doi.org/10.1016/j.biortech.2021.125581.

Ullah, Z., Khan, M., Naqvi, S.R., Farooq, W., Yang, H.P., Wang, S.R., Vo, D.N., 2021. A comparative study of machine learning methods for bio-oil yield prediction – a genetic algorithm-based features selection. Bioresour. Technol. 335 https://doi.org/10.1016/j.biortech.2021.125292.

Wang, X.J., Chen, Q., Zhu, H.L., Chen, X.L., Yu, G.S., 2023. In-situ study on structure evolution and gasification reactivity of biomass char with K and Ca catalysts at carbon dioxide atmosphere. Carbon Resources Conversion 6, 27–33. https://doi.org/10.1016/j.crcon.2022.10.002.

Wei, H., Luo, K., Xing, J.K., Fan, J.R., 2022. Predicting co-pyrolysis of coal and biomass using machine learning approaches. Fuel 310, 122248. https://doi.org/10.1016/j.fuel.2021.122248.

Were, K., Bui, D.T., Dicka, B., Singh, B.R., 2015. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. Ecol. Indic. 52, 394–403. https://doi.org/10.1016/j.ecolind.2014.12.028.

Wu, L., Liu, J., Xu, P., Zhou, J., Yang, F., 2022. Biomass hydrogen donor assisted microwave pyrolysis of low-rank pulverized coal: optimization, product upgrade and synergistic mechanism. Waste Manage 143, 177–185. https://doi.org/10.1016/j.wasman.2022.02.020.

Xing, J.K., Luo, K., Wang, H.O., Gao, Z.W., Fan, J.R., 2019. A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analysis with machine learning approaches. Energy 188. https://doi.org/10.1016/j.energy.2019.116077.

Xue, T., Zheng, Y., Geng, G., Xiao, Q., Zhu, T., 2020. Estimating spatiotemporal variation in ambient ozone exposure during 2013–2017 using a data-fusion model. Environ. Sci. Technol. 54 (23), 14877–14888. https://doi.org/10.1021/acs.est.0c03098.

Yan, L.J., Wang, W.M., Liu, Y., Wang, M.J., Li, F., Bao, W.R., 2022. The roles of low molecular compounds on the light aromatics formation during different rank coal pyrolysis. J. Energy Inst. 100, 129–136. https://doi.org/10.1016/j.joei.2021.10.012.

Yang, Q.L., Yao, Q.X., Ma, D., Liu, Y.Q., He, L., Zhou, R.S., Sun, M., Ma, X.X., 2022. Investigation of the (catalytic) co-pyrolysis of Shendong coal and coal tar based on rapid pyrolysis and ANN modeling. J. Anal. Appl. Pyrolysis 163, 105486. https://doi.org/10.1016/j.jaap.2022.105486.

Zhang, T.H., Cao, D.Y., Feng, X., Zhu, J.H., Lu, X.H., Mu, L.W., Qian, H.L., 2022. Machine learning prediction of bio-oil characteristics quantitatively relating to biomass compositions and pyrolysis conditions. Fuel 312, 122812. https://doi.org/10.1016/j.fuel.2021.122812.

Zhang, C., Wu, R.C., Xu, G.W., 2014. Coal pyrolysis for high-quality tar in a fixed-bed pyrolyzer enhanced with internals. Energy Fuels 28, 236–244. https://doi.org/10.1021/ef401546n.

Zhao, Y.P., Hu, H.Q., Jin, L.J., He, X.F., Wu, B., 2011. Pyrolysis behavior of vitrinite and inertinite from Chinese Pingshuo coal by TG-MS and in a fixed bed reactor. Fuel Process Technol 92, 780–786. https://doi.org/10.1016/j.fuproc.2010.09.005.

Zhu, J.L., Jin, L.J., Luo, Y.W., Hu, H.Q., Xiong, Y.K., Wei, B.Y., Wang, D.C., 2019a. Fast co-pyrolysis of a massive Naomaohu coal and cedar mixture using rapid infrared heating. Energy Convers Manage 205 (112442), 2022. https://doi.org/10.1016/j.enconman.2019.112442.

Zhu, X.Z., Li, Y.N., Wang, X.N., 2019b. Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. Bioresour. Technol. 288, 121527 https://doi.org/10.1016/j.biortech.2019.121527.