

DOI: 10.32604/jrm.2022.018625

ARTICLE



The Estimation of the Higher Heating Value of Biochar by Data-Driven Modeling

Jiefeng Chen¹, Lisha Ding¹, Pengyu Wang¹, Weijin Zhang², Jie Li³, Badr A. Mohamed⁴, Jie Chen¹, Songqi Leng¹, Tonggui Liu¹, Lijian Leng^{2,·} and Wenguang Zhou^{1,·}

¹Key Laboratory of Poyang Lake Environment and Resource Utilization, Ministry of Education and School of Resources, Environmental & Chemical Engineering, Nanchang University, Nanchang, China

²School of Energy Science and Engineering, Central South University, Changsha, China

³Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore

⁴Department of Agricultural Engineering, Cairo University, Giza, Egypt

Corresponding Authors: Wenguang Zhou. Email: wgzhou@ncu.edu.cn; Lijian Leng. Email: lljchs@126.com

Received: 06 August 2021 Accepted: 22 September 2021

ABSTRACT

Biomass is a carbon-neutral renewable energy resource. Biochar produced from biomass pyrolysis exhibits preferable characteristics and potential for fossil fuel substitution. For time- and cost-saving, it is vital to establish predictive models to predict biochar properties. However, limited studies focused on the accurate prediction of HHV of biochar by using proximate and ultimate analysis results of various biochar. Therefore, the multi-linear regression (MLR) and the machine learning (ML) models were developed to predict the measured HHV of biochar from the experiment data of this study. In detail, 52 types of biochars were produced by pyrolysis from rice straw, pig manure, soybean straw, wood sawdust, sewage sludge, Chlorella Vulgaris, and their mixtures at the temperature ranging from 300 to 800°C. The results showed that the co-pyrolysis of the mixed biomass provided an alternative method to increase the yield of biochar production. The contents of ash, fixed carbon (FC), and C increased as the incremental pyrolysis temperature for most biochars. The Pearson correlation (r) and relative importance analysis between HHV values and the indicators derived from the proximate and ultimate analysis were carried out, and the measured HHV was used to train and test the MLR and the ML models. Besides, ML algorithms, including gradient boosted regression, random forest, and support vector machine, were also employed to develop more widely applicable models for predicting HHV of biochar from an expanded dataset (total 149 data points, including 97 data collected from the published literature). Results showed HHV had strong correlations (|r| > 0.9, p < 0.05) with ash, FC, and C. The MLR correlations based on either proximate or ultimate analysis showed acceptable prediction performance with test $R^2 > 0.90$. The ML models showed better performance with test R^2 around 0.95 (random forest) and 0.97–0.98 before and after adding extra data for model construction, respectively. Feature importance analysis of the ML models showed that ash and C were the most important inputs to predict biochar HHV.

KEYWORDS

Biochar; higher heating value; machine learning; prediction; proximate analysis; ultimate analysis



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1 Introduction

Thermochemical conversion of biomass is one of the optional pathways to overcome the energy crisis, environmental pollution, and sustainable development issues of the world. There are about 13 billion tons of biomass resources per year on the planet that can be used [1]. Significant momentum has been attained in the use of renewable biomass as an alternative to traditional fossil fuels in the energy application fields [2]. However, the characteristics of raw biomass, such as high moisture content, large volume, low energy density, and low combustion calorific value, are several significant problems upon its use as fuel. As a solid product produced by the thermochemical conversion of biomass at temperatures below 900°C and oxygen-limited environment. Biochar has excellent potential for application in the energy storage [3], and biochar has a lower moisture content, higher energy density, higher stability, and more accessible transportation than biomass. Moreover, it is necessary to grasp the fuel properties (i.e., higher heating value (HHV)) of biochar for its application in the energy field. Generally, the basic principles of HHV analysis were straightforward, and they could be experimentally determined by the integrated system offered by the manufacturers. The HHV of the biochar samples can be determined by measuring the enthalpy difference of the sample before and after the reaction in an adiabatic oxygen bomb. However, the use of instruments to determine the properties of biochar has some disadvantages, such as high cost and time-consuming. Therefore, it is necessary and economical to develop the HHV prediction model based on some common characteristic indexes.

The ultimate and proximate results have been used to predict the carbon sequestration potential (stability) [4] and HHV [5]. And a large number of traditional multi-linear regression (MLR) models have been built and studied to predict the HHV of municipal solid waste, coal, biomass, etc. [6]. Mateus and coworkers developed a highly accurate HHV linear regression prediction model (R^2 =0.9997) based on the ultimate analysis of bio-oil produced by liquefaction [7]. In addition to the traditional MLR method widely used by researchers to predict HHV, many researchers have used artificial intelligence algorithm technology to predict HHV. These artificial intelligence algorithms can do well with both linear and nonlinear relationships between the input and target variables. Samadi and coworkers used the gradient boosting regression (GBR) algorithm to predict the HHV of biomass according to different training parameters (i.e., stochasticity, tree size, and learning rate), and the obtained model had good prediction performance (R^2 =0.93) [8]. Xing and coworkers [9] trained biomass HHV prediction models using empirical correlation, random forest (RF), support vector machine (SVM), and artificial neural network algorithms respectively by proximate and ultimate analysis data. Between the two types of data models of proximate analysis and ultimate analysis, RF (R^2 =0.962) and SVM (R^2 =0.953) have showed satisfactory predictive performance.

However, the studies focused on the prediction of HHV from the basic properties of biochar are limited. In addition, as a kind of solid fuel, biochar is very different from other natural materials, which impedes the application of existing models to predict the HHV of biochar. In this light, the aim of this research was to develop accurate models for the prediction of the HHV of biochar through GBR, RF, SVM algorithm, and linear regression method. In this study, 52 biochar samples were obtained from our experiments to explore the relationship between biochar characteristics and establish the HHV prediction models. Then, models were optimized by adding 97 data points from the published studies into the initial sample dataset. Finally, the predictive performance measures and relative importance analysis were conducted to evaluate the models.

2 Materials and Methods

2.1 Materials and Sample Preparation

A total of 52 biochar samples were produced through the pyrolysis process under various conditions from four different representative biomass species, including agricultural residues, algae, animal manure,

and sludge. The rice straw (RS), pig manure (PM), soybean straw (SS), and wood sawdust (WD) were collected from local farmers in Jiangxi Province, China. The *Chlorella Vulgaris* (CL) and sewage sludge (SW) were provided from a biotechnology company in Shanxi Province of China and a municipal wastewater treatment plant in Jiangxi Province of China, respectively. Two types of biomasses were selected and mixed with a mass ratio of 1:1 to form the mixture, and the mixture was named as an abbreviated combination of the two kinds of biomass. All six biomasses and the mixtures were pulverized to a particle size of fewer than 350 microns and dried at 105°C for two hours to constant weight. During pyrolysis, the biomass was placed in oxygen-free reactors, and the temperature was increased from room temperature to target temperatures (ranged from 300 to 800°C, in 100°C intervals) with the heating rate of 10 °C/min. The residence time at the target temperature was one hour.

2.2 Analysis of Biochar

The ash, volatile matter (VM), and fixed carbon (FC) of biochars were analyzed following the procedures in previous study [10], which was carried out in an automatic proximate analyzer. The elemental composition (i.e., C, H, N, and S) of biochars was determined using an ultimate analyzer system (Elementar Analysensysteme GmbH, Vario EL III), and the O content was calculated by difference (O = 100 - C - H - N - S - ash). The HHV of biochars was determined by an automatic calorimeter (ZDHW-9000C, HB-Huanuo, China). About 1.0 g of the biochar sample was placed in an oxygen bomb calorimeter filled with excess oxygen. The HHV of the biochar was obtained after comparison with the heat capacity of the standard substance of benzoic acid (GBW130035, National Institute of Metrology, China) before and after burning and correction of additional heating such as point heating [11].

2.3 Model Construction

The MLR equations based on the ultimate and proximate analysis data were carried out by using Origin 2021b. In this research, three ML models (GBR, RF, SVM) were developed to predict the HHV of biochar through the scikit-learn Python library. GBR is a powerful non-parametric method for prediction. The algorithm principle of GBR is to train a new prediction tree in turn, and learn errors from the previous tree (n - 1) to get a new tree (n) with lower prediction errors [8,12,13]. RF has statistical advantages such as the low risk of overfitting and fewer parameters to be specified and can be used to deal with nonlinear and linear relationships between variables [14]. SVM is a ML algorithm that uses the nonlinear kernel function to map the initial training samples to the high-dimensional feature space, thus transforming the problem from nonlinear to linear and obtaining the optimal solution [9]. All of the input data were normalized according to the studies of Li et al. [15,16]. The ratio of the training dataset to the test dataset was 8:2, and cross validation was carried out to avoid bias in the training process. The performance of MLR and ML models in this study was evaluated in terms of coefficient of determination (R^2), mean absolute error (*MAE*), and root mean square error (*RMSE*). The calculations of R^2 , *MAE*, and *RMSE* are defined as follows:

$$R^{2} = 1 - \frac{\sum_{i=1}^{n} (HHV_{experimental,i} - HHV_{predicted,i})^{2}}{\sum_{i=1}^{n} (HHV_{experimental,i} - HHV_{experimental,i})^{2}}$$
$$MAE = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{HHV_{predicted,i} - HHV_{experimental,i}}{HHV_{experimental,i}} \right|$$
$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (HHV_{predicted,i} - HHV_{experimental,i})^{2}}{n}}$$

where $HHV_{experimental}$, $HHV_{predicted}$, and $HHV_{experimental}^{average}$ are the experimental, predicted, and the average of experimental values of the HHV, respectively, and n is the number of sample data points used for the regression analysis. MAE and RMSE could estimate the value of error between the experimental and predicted HHV. And R^2 can be used to determine the degree of goodness of the proposed correlations. The higher the R^2 and lower the *RMSE* and *MAE* meant the better the model accuracy [15,17].

3 Results and Discussion

3.1 Biochar Production and Characterization

Biochar is the pyrogenic carbonaceous material, and the formation of biochar is a process of continuous decomposition and recombination of macromolecules. With the rise of pyrolysis temperature, the category of biochar gradually evolved from transition char to turbostratic char [18], and the stability of biochar was also improved [19]. As shown in Fig. 1a, the yield of biochar decreases with the increase of pyrolysis temperature. Similar results were also verified by Zhao et al. [20] that pyrolysis temperature had a significant (p < 0.05) effect on the biochar yield. The co-pyrolysis of the biomass mixture could improve the yield of most biochar (compared to the theoretical yield calculated by considering no interaction between biomasses), it would provide a feasible scheme for enhancing the yield of biochar. In the data composed of ultimate analysis (Tables 1 and S1 in supporting information), biochars derived from PM at the $400^{\circ}C$ ~ had lower C content than that of feedstocks, and the similar results could be found in the study of Gascó et al. [21] Biochar derived from PM may contain aromatic carbons that were difficult to burn during the ultimate analysis process, so its C results would be underestimated [22]. C was the main component of biochar, with the content ranged from 11.38% to 85.17%. The C content of biochar did not increase with the increase of pyrolysis temperature, which was different from the research results of using three lignin-like biomass [23]. A portion of the unstable C in biochar was converted to stable C during biochar formation, and the variation was not fixed due to the difference in the biomass composition [10,20]. O content ranked the second in most biochar samples with the mean values 7% (Fig. 2), which was related to the complex pyrolysis mechanism of biochar and the biochemical composition of each biomass. The loss of O and H was mainly due to breaking weak O bonds during biochar formation [24]. S was not detected in all samples, so the content of S of all samples below was recorded as 0. The O/C of all biochar was less than 0.4 (Fig. 1b), meaning the half-life of these biochars $O/C \le 0.6$ can exceed 100 years [25]. The greater stable biochar would have the more favorable storage and transport advantages in its use as fuel.

Biomass ^a	Ultima	te analy	vsis ^b (wt	%)			Proxima	te analysis	HHV (MJ/kg)	
	С	Н	Ν	O ^c	O/C	H/C	Ash	VM	FC	
PM	30.14	4.67	2.50	25.17	0.63	1.86	37.53	53.54	8.93	13.76
RS	38.44	5.46	0.73	41.83	0.82	1.70	13.55	68.04	18.41	15.61
CL	45.40	7.17	10.80	29.23	0.48	1.90	7.40	80.86	11.74	20.08
SS	42.10	5.84	2.21	43.33	0.77	1.66	6.53	73.38	20.09	18.91
SW	21.95	3.55	3.80	16.83	0.58	1.94	53.88	39.30	6.83	7.33
WD	47.74	6.07	0.30 ^d	44.90	0.71	1.53	1.00	82.81	16.19	14.99

Table 1: The ultimate and proximate compositions and HHV of biomass from this study

Notes: "The label meant biomass: rice straw (RS), pig manure (PM), soybean straw (SS), wood sawdust (WD), Chlorella Vulgaris (CL), and sewage sludge (SW). ^bDry basis, data for SS and CL were cited from [26], and data for WD were cited from [10].

^cBy difference, O = 100 - C - H - N - S - ash, S was not detected, and the content of S in all samples was recorded as 0.

^dNot within the detection limits, and 0.30 was used.



Figure 1: (a) The yield of biochar produced from six biomass and their mixtures. (b) The Van Krevelen diagram of biochars. (c) Triangle plot of proximate analysis of biochars. The numbers behind the different colored squares in the illustration represented the maximum pyrolysis temperature of biochar. (d) The Pearson correlations between the indicators and HHV values. The blue and red circles represent the positive and negative correlation, respectively. The label meant biomass: rice straw (RS), pig manure (PM), soybean straw (SS), wood sawdust (WD), *Chlorella Vulgaris* (CL), and sewage sludge (SW). Mixed biomass was abbreviated by the two forms of biomass

In the ternary diagram (Fig. 1c) of proximate analysis, with the rise of pyrolysis temperature, the FC of biochar increased to close to 100%, which is the same as our previous research results [10]. As expected, ash remains in the solid as the temperature increases despite the thermal decomposition of organic matter [27]. And VM showed a decrease with the increasing pyrolysis temperature. According to the proximate analysis (Fig. 2), FC and Ash were the two main components of the experimental data, with mean values of 43% and 42%, respectively. CL400 had the highest VM, which was 33.83%. The HHV of some biochar (Table S1) produced at low temperature was higher than that of biomass (Table 1), and the high energy density would be one of the advantages for its use as solid fuel. With the increase of pyrolysis temperature, the HHV of biochar would decrease. The average HHV of the 52 biochar samples was 16.38 MJ/kg, and the HHV value of more than half of the samples was greater than 16.38 MJ/kg (Fig. 2). The HHV of biochar derived from PM and SW was low, and the maximum HHV was only 12.86 MJ/kg (PM300), which was related to their high ash content. All biochars produced by the co-pyrolysis of WD and CL have the greater HHV than that of the theoretical value (average HHV of the biochar derived from two feedstocks at the same temperature). These results may be caused by the changes of the ash and indicated that WD

and CL had a synergistic effect in the process of co-pyrolysis [28]. Biochar derived by biomass mixture had higher HHV than the theoretical HHV of mixed-biomass biochar, it would provide a feasible scheme for improving the HHV of biochar. Pyrolysis of biomass is a process in which char gradually evolves into pure carbon, and as this process occurs (the increase of temperature), the C–H, C–O, and O–H contained in the biochar were gradually eliminated, and the energy structure of the biochar becomes stable with aromatic resonances and π – π stacking of graphitic sheets [27].



Figure 2: Boxplot of variables related to biochar characteristics. Each data index in the dataset is evenly divided into quartiles based on 25% of each part

The Pearson correlation coefficients (r) between the basic experimental data were shown in Fig. 1d. HHV had significant correlations ($|\mathbf{r}| > 0.36$, p < 0.05) with other indicators except VM, and H/C. HHV positively ($0.44 < \mathbf{r} < 0.94$, p < 0.05) correlated with FC, C, H, N and O and negatively ($-0.96 < \mathbf{r} < -0.48$, p < 0.05) correlated with ash, VM/FC and H/C. These result were generally in consistent with the results of previous studies [29,30]. In particular, HHV was strongly ($|\mathbf{r}| > 0.9$, p < 0.05) correlated with ash ($\mathbf{r} = -0.97$, p < 0.05), FC ($\mathbf{r} = 0.90$, p < 0.05), and C ($\mathbf{r} = 0.94$, p < 0.05). Therefore, lower ash content and higher C and FC contents of biomass mean higher biochar HHV.

3.2 HHV Prediction Using Data of This Study

3.2.1 HHV Prediction by MLR Equations for Biomass and Coal Developed in Previous Studies

Many models had been developed in the published studies for the prediction of HHV of biomass and coal, among them the Dulong formula $(HHV_{Dulong} = 0.3383 \times C + 1.443 \times H - 0.1804 \times O + 0.0942 \times S)$ and the Milne formula $(HHV_{Milne} = 0.3410 \times C + 1.322 \times H - 0.1200 \times O - 0.1200 \times N - 0.0153 \times ash)$ were widely used for biomass and coal, respectively [31,32]. In this study, the HHV of biochar (test dataset, n = 11) was calculated according to the Dulong formula and the Milne formula, and the performance of predicted equations is shown in Figs. 3a and 3b. Most of the predicted data were within a 20% margin of error. There were some data predicted by the Dulong formula and the Milne formula mapped outside of the region, indicating whose error was more than 20%. Compared with the Dulong formula, the Milne formula introduced ash for HHV calculation, and its prediction performance $(R^2=0.9204, MAE=0.1280, RMSE=1.9053)$ was better than that of the Dulong formula $(R^2=0.8892, RMSE=0.9204, RMSE=0.1280, RMSE=0.9053)$ MAE = 0.1471, RMSE = 2.2483). Although the Milne formula predicted HHV by combining the results of ultimate analysis and proximate analysis, one of the predicted results of the Milne formula was still outside the prediction error range of 20%. The proximate analysis results from various test methods [22] and the thermal behaviors differences among biomass, coal, and biochar [33] may be the two major reasons for the inaccurate prediction results of these [34]. It is necessary to build the prediction model based on the basic property data to predict the biochar HHV more accurately.



Figure 3: Comparison between predicted and experimental HHV (n = 11, test dataset). (a) The predicted HHV calculated by the Dulong formula from the experimental test datasets, (b) the predicted HHV calculated by the Milne formula from the experimental test datasets. (c) The predicted HHV calculated by Eq. (5) from the experimental test datasets, (d) the predicted HHV calculated by Eq. (12) from the experimental test datasets data. The middle region of the solid lines of the three functions indicated 20% relative error

3.2.2 Biochar HHV Prediction by MLR Equations Constructed by This Study

As shown in Table 2, Eq. (5) had the best prediction performance (training $R^2 = 0.9327$, training MAE = 0.0876, training RMSE = 1.8385) among the equations based on the training dataset containing ultimate and proximate compositions. And all equations (only except for Eq. (2), test $R^2 > 0.9191$) had good generalization ability (test $R^2 > 0.9204$) compare with that of the Dulong formula (test $R^2 = 0.8892$) and the Milne formula (test $R^2 = 0.9204$). Compared with the test predictive performance of Eq. (1) (test $R^2 = 0.9449$, test MAE = 0.0812, test RMSE = 1.5848), the test predictive performance of Eq. (3) (test $R^2 = 0.9529$, test MAE = 0.0779, test RMSE = 1.4656) with the additional introduction of H as a new independent variable was slightly improved. This was the same as the result of Pearson correlation analysis (Fig. 1d). H had a significantly weak correlation (r = 0.37, p < 0.05) with HHV. The range of the experimental proximate analysis data used as the independent variable for the development of models was $1.87\% \le ash \le 86.11\%$, $3.17\% \le VM \le 54.22\%$, and $6.00\% \le FC \le 91.99\%$ (Fig. 2), and these equations are based on the proximate analysis and had good training predictive performance (training $R^2 > 0.9$). The equations (Eqs. (7)–(9), training $R^2 = 0.9215$) had a stronger correlation than the monadic equation (Eq. (6), training

 $R^2 = 0.9207$) of ash, which was the same as the conclusion of Qian et al. [5]. Eq. (10), which was composed of ash and VM/FC, had the good test predictive ability of HHV and was a little bit worse than Eq. (6) with lower test R^2 (0.9694 *vs.* 0.9726), test *MAE* (0.0612 *vs.* 0.0587), and test *RMSE* (1.1816 *vs.* 1.1184). In Eqs. (11) and (12), ash, FC, and C were strongly correlated ($|\mathbf{r}| > 0.9$, $\mathbf{p} < 0.05$, Fig. 1d) with HHV, and their training predictive performance was not improved compared with Eqs. (5) and (10) (Table 2).

Table 2: Correlations used for predicting the HHV of biochar based on ultimate and proximate analysis

No.	Equation	R^2 (training)	<i>MAE</i> (training)	<i>RMSE</i> (training)	R^2 (test)	MAE (test)	<i>RMSE</i> (test)
1	$HHV = 0.2927 \times C + 0.7160 \times N + 1.1734$	0.9086	0.1021	2.1430	0.9449	0.0812	1.5848
2	$HHV = 0.2824 \times C + 0.2192 \times O + 1.9277$	0.9020	0.1036	2.2191	0.9191	0.0995	1.9212
3	$HHV = 0.2849 \times C + 0.2981 \times H + 0.6628 \times N + 0.8652$	0.9101	0.0966	2.1255	0.9529	0.0779	1.4656
4	$HHV = 0.2715 \times C + 0.7361 \times N + 0.2271 \times O + 0.4219$	0.9327	0.0883	1.8395	0.9402	0.0781	1.6514
5	$HHV = 0.2700 \times C + 0.0672 \times H + 0.7239 \times N + 0.2246 \times O + 0.3606$	0.9327	0.0876	1.8385	0.9441	0.0741	1.5965
6	$HHV = -0.2667 \times ash + 27.5465$	0.9207	0.0919	1.9962	0.9726	0.0587	1.1184
7	$HHV = -0.2692 \times ash - 0.0276 \times VM + 28.0573$	0.9215	0.0869	1.9856	0.9681	0.0620	1.2063
8	$HHV = -0.2416 \times ash + 0.0276 \times FC + 25.3000$	0.9215	0.0869	1.9856	0.9681	0.0620	1.2063
9	$HHV = 0.2692 \times FC + 0.2416 \times VM + 1.1357$	0.9215	0.0869	1.9856	0.9681	0.0620	1.2062
10	$HHV = -0.2562 \times ash - 0.8357 \times VM/FC + 27.5827$	0.9246	0.0770	1.9460	0.9694	0.0612	1.1816
11	$HHV = -0.2651 \times ash - 0.0018 \times C + 27.3949$	0.9200	0.0928	2.0049	0.9684	0.0613	1.2000
12	$HHV = -0.2930 \times ash + 0.1080 \times FC - 0.1479 \times C + 30.6912$	0.9239	0.0890	1.9560	0.9578	0.0741	1.3870

The comparison of Eqs. (5) and (12) between predicted and experimental HHV is shown in (Figs. 3c and 3d). The data points were basically distributed around the Y = X equation line, and there was no outlier outside 20% error. Though the same test *MAE* of 0.0741 was achieved from both Eqs. (5) and (12) had the higher test R^2 and lower test *RMSE* than that of Eq. (5), expressing better prediction and generalization ability. It could also be seen that the generalization ability of each equation was different, and its prediction in the test dataset and the training dataset would also show different amplitude changes.

3.2.3 Biochar HHV Prediction by ML Models Constructed in This Study

The performance of the ML models was validated with the test dataset shown in Fig. 4. Obviously, the training data and test data were distributed around the Y = X function graph, which intuitively illustrated the accuracy and reliability of the three ML models. Compared with the MLR prediction models (Eqs. (5) and (10), Table 2), the predictive performance of the training dataset was enhanced when three ML algorithms were applied to predict the HHV of biochar. In the training dataset, the GBR model had great predictive ability (R^2 = 1.00, MAE = 0.22, RMSE = 0.24). In the test dataset, the lower R^2 (0.93) of the GBR model is found with MAE = 1.33, and RMSE = 1.74. The modeling process of the GBR algorithm based on boost theory was a process in which the prediction error decreases continuously. In the study of Samadi et al. [8], the prediction performance of HHV prediction model based on GBR (R^2 =0.93) algorithm was better than that of models of genetic programming (R^2 =0.90) and artificial neural networks (R_1^2 =0.88, R_2^2 =0.89). For the RF model, the training dataset also showed good predictive performance with R^2 =0.98, MAE = 0.68, and RMSE = 0.88. Moreover, the better performance of the RF model was found in the test dataset with R^2 =0.95, MAE = 1.12, RMSE = 1.45. The excellent prediction performance of the RF model was related to the principle of its algorithm. As a ML algorithm that adopted an integrated learning method, RF had better robustness in the learning process and a lower risk of overfitting and noisy data than other ensemble learning models [9].



Figure 4: Predicted data *vs.* the experimental data based on the GBR (a), RF (b), and SVM (c) models with training and test dataset. The red line in the graph is the plot of Y = X, and the points on the line meant that the predicted and experimental values were the same

In the analysis of relative importance (RI), FC, ash, and C were the main influencing features among studied factors (Fig. 5). FC (RI = 0.3569) and ash (RI = 0.4993) were the most important features in HHV prediction based on GBR and RF, respectively. This result was consistent with the Pearson correlation analysis results in Section 3.1. The C of biochar was the main component of its ultimate composition, and C was the main energy supplier in the combustion process of biochar. FC and ash content had highly positively correlation (r = 0.90, p < 0.05, Fig. 1d) and negatively correlation (r = -0.97, p < 0.05, Fig. 1d) with HHV, respectively. However, the relative importance of all features is not completely consistent with the results of Pearson correlation analysis, especially the contribution of O to HHV in the GBR (RI = 0.0013) and RF (RI = 0.0035) models. The reason may be that Pearson correlation is the linear correlation between each feature and HHV, while the relative importance obtained from the ML model included both the linear and nonlinear correlations [35]. In order to improve the generalization ability and broaden the applicability of the ML models, the wider range of dataset would help improve the model.



Figure 5: Relative contributions of each feature in GBR (a) and RF (b) algorithm

3.3 Optimized HHV Prediction by Additional Dataset

It is worth noting that ML algorithms are data-driven artificial intelligence algorithms. In the application of regression prediction, the more amount of input data, the broader the applicability of the model, and the better the prediction. In addition, the results of the properties of biochar from varied measuring instruments can lead to a significant difference. The models in Section 3.2 showed good prediction performance, but few numbers and small variation intervals of the dataset had been used to develop the models. These characteristics of the dataset would lead to the weak generalization ability of the models. Therefore, new models were built after introducing additional data points from previous studies. The statistical analysis of the new merged input dataset (n = 149) was shown in Fig. 6 and Table S3. Compared with the original statistical results previous dataset (n = 52, Fig. 2), the average HHV value increased to 21.27 MJ/kg, which is 4.89 MJ/kg higher. The variation interval expanded from 5–28 MJ/kg (Fig. 2) to 5–35 MJ/kg (Fig. 6).



Figure 6: Boxplot of variables related to biochar characteristics (n = 149). Each data index in the dataset is evenly divided into quartiles based on 25% of each part

The best prediction performance of the MLR models developed from the new dataset was not greatly improved compared with the equations in Table 2. The best equation was Eq. (15) with training R^2 = 0.9284, training MAE = 0.0654, and training RMSE = 1.8626 (Table 3), but the applicability of the model was reduced (test $R^2 = 0.8749$, test MAE = 0.0857, test RMSE = 2.5070) compared with equations in Table 2. However, for the ML models, the GBR, RF, and SVM models all showed good prediction performance, where training R² was 1.00, 0.99, and 0.98 (Fig. 7), respectively. The GBR algorithm model had the best performance with $R^2 = 1.00$, MAE = 0.32, and RMSE = 0.37 and $R^2 = 0.98$, MAE = 0.98, 0.83, and RMSE = 1.08 for the training and test datasets, respectively. Overall, the predictive performance of the expanded dataset (n = 149) was better than the original dataset (n = 52), which showed a sizeable dataset could generally represent better prediction performance. Fig. 8 described the relative contributions of each feature based on GBR and RF models. The relative importance of C in both GBR and RF models was the major one with values of 0.7087 and 0.8834, respectively. As a solid product of biomass by thermochemical conversion, the HHV of biochar mainly came from the combustion fracture of the C-H bond, and the contribution of O, H, and N to the HHV of biochar was limited [9]. Ash was the second most important feature of the two models, with values of 0.2343 and 0.0699, respectively. It also can be found that the third important feature of the two models was not the same, but the relative importance of the two features was lower than 0.02 that can be negligible.



Figure 7: Predicted data *vs.* the experimental data based on the GBR (a), RF (b), and SVM (c) models with the expanded dataset (n = 149). The red line in the graph was a plot of Y = X, and the points on the line meant that the predicted and experimental values were the same

0.8834

Feature Importance

No.	Equation		R ² (training)	MAE (training)	RMSE (training)	R^2 (test)	MAE (test)	RMSE (test)
13	$HHV = 0.3506 \times C + 0.2808 \times H + 0.0429 \times N + 0.0554 \times 0.2443$	< <i>O</i> –	0.9207	0.0647	1.9600	0.8535	0.0854	2.7134
14	$HHV = -0.0817 \times ash + 0.2716 \times C + 7.9440$		0.9211	0.0678	1.9541	0.8591	0.0916	2.6610
15	$HHV = -0.0833 \times ash - 0.0327 \times FC + 0.2945 \times C + 8.12$	0.9284	0.0654	1.8626	0.8749	0.0857	2.5070	
	N-0.0017 (a)		O -0.0007				(b))
	VM/FC -0.0017	Н	/C -0.0012					
	H -0.0029	VM/	FC -0.0025					
	FC -0.0059		N -0.0037					
	H/C -0.0059	с	/C -0.0071					
	O/C -0.0100	V V	M - 0.0072					
	O-0.0128		н - 0.0111					
	VM - 0.0163	1	FC - 0.0133					
	Ash - 0.2343	A	.sh - 0.069	9				

Table 3: Correlations used for predicting the HHV of biochar using the expanded dataset (n = 149)

Figure 8: Relative contributions of each feature in GBR (a) and RF (b) algorithm

C

0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

0.7087

Feature Importance

4 Conclusions

С

0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

Biochars produced from a wide range of biomass were characterized. Compared with biochar from the pyrolysis of individual biomass, biochar with higher yield and HHV could be obtained by the pyrolysis of biomass mixtures. Pyrolysis temperature and biomass mixture can affect the biochar yield and properties. Moreover, the MLR and ML prediction models were successfully developed to predict the HHV of biochar based on 52 experiment data. ML approaches showed better prediction ability (training $R^2 \ge 0.96$) of the biochar HHV prediction compared with MLR (training $R^2 < 0.94$). The HHV of biochar (test dataset) was successfully predicted from the ultimate and proximate analysis with the GBR algorithm with $R^2 = 0.98$, MAE = 0.83, RMSE = 1.08 trained by the experimental training dataset. The RF and SVM models also had a similarly good performance of the HHV prediction with the $R^2 = 0.97$, MAE = 0.93, RMSE = 1.22 and $R^2 = 0.97$, MAE = 0.93, RMSE = 1.23, respectively. With the expanded datasets (n = 149), the predictive performance of ML models was improved. Feature importance analysis showed that ash and C had the highest relative importance to HHV prediction, while VM and FC had limited effects. The ML approaches can predict the HHV of biochar with high accuracy and play an important role in the development of biochar fuel applications.

Funding Statement: The work was supported by the National Natural Science Foundation of China (No. 51808278) and the Science Foundation for Youths of Jiangxi Province, China (20192BAB213012). This research was also supported by the College Students' Innovative Entrepreneurial Training Plan Program, China (No. 201910403049).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- 1. Shen, F., Xiong, X., Fu, J., Yang, J., Qiu, M. et al. (2020). Recent advances in mechanochemical production of chemicals and carbon materials from sustainable biomass resources. *Renewable and Sustainable Energy Reviews*, 130, 109944. DOI 10.1016/j.rser.2020.109944.
- Waqas, M., Aburiazaiza, A. S., Miandad, R., Rehan, M., Barakat, M. A. et al. (2018). Development of biochar as fuel and catalyst in energy recovery technologies. *Journal of Cleaner Production*, 188, 477–488. DOI 10.1016/j. jclepro.2018.04.017.
- 3. Leng, L., Xiong, Q., Yang, L., Li, H., Zhou, Y. et al. (2021). An overview on engineering the surface area and porosity of biochar. *Science of the Total Environment*, *763*, 144204. DOI 10.1016/j.scitotenv.2020.144204.
- 4. Leng, L., Huang, H., Li, H., Li, J., Zhou, W. (2019). Biochar stability assessment methods: A review. *Science of the Total Environment, 647,* 210–222. DOI 10.1016/j.scitotenv.2018.07.402.
- 5. Qian, C., Li, Q., Zhang, Z., Wang, X., Hu, J. et al. (2020). Prediction of higher heating values of biochar from proximate and ultimate analysis. *Fuel*, *265*, 116925. DOI 10.1016/j.fuel.2019.116925.
- Vargas-Moreno, J. M., Callejón-Ferre, A. J., Pérez-Alonso, J., Velázquez-Martí, B. (2012). A review of the mathematical models for predicting the heating value of biomass materials. *Renewable and Sustainable Energy Reviews*, 16(5), 3065–3083. DOI 10.1016/j.rser.2012.02.054.
- 7. Mateus, M. M., Bordado, J. M., Galhano dos Santos, R. (2021). Estimation of higher heating value (HHV) of biooils from thermochemical liquefaction by linear correlation. *Fuel, 302,* 121149. DOI 10.1016/j.fuel.2021.121149.
- 8. Samadi, S. H., Ghobadian, B., Nosrati, M. (2021). Prediction of higher heating value of biomass materials based on proximate analysis using gradient boosted regression trees method. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects, 43(6),* 672–681. DOI 10.1080/15567036.2019.1630521.
- Xing, J., Luo, K., Wang, H., Gao, Z., Fan, J. (2019). A comprehensive study on estimating higher heating value of biomass from proximate and ultimate analysis with machine learning approaches. *Energy*, 188, 116077. DOI 10.1016/j.energy.2019.116077.
- 10. Chen, J., Wang, P., Ding, L., Yu, T., Leng, S. et al. (2021). The comparison study of multiple biochar stability assessment methods. *Journal of Analytical and Applied Pyrolysis, 156,* 105070. DOI 10.1016/j.jaap.2021.105070.
- 11. CS, P. (2008). Determination of calorific value of coal. Chinese Standard No. GB/T 213-2008. General Administration of Quality Supervision, Inspection and Quarantine of the People's Republic of China and Standardization Administration of the People's Republic of China. <u>http://openstd.samr.gov.cn/bzgk/gb/newGbInfo?hcno=8FF3EA84987B3B87C3A2E5922D41AC2D</u>.
- Zhu, X., Wan, Z., Tsang, D. C. W., He, M., Hou, D. et al. (2021). Machine learning for the selection of carbonbased materials for tetracycline and sulfamethoxazole adsorption. *Chemical Engineering Journal*, 406, 126782. DOI 10.1016/j.cej.2020.126782.
- Onsree, T., Tippayawong, N. (2021). Machine learning application to predict yields of solid products from biomass torrefaction. *Renewable Energy*, 167, 425–432. DOI 10.1016/j.renene.2020.11.099.
- Zhu, X., Li, Y., Wang, X. (2019). Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. *Bioresource Technology*, 288, 121527. DOI 10.1016/ j.biortech.2019.121527.
- 15. Li, J., Pan, L., Suvarna, M., Tong, Y. W., Wang, X. (2020). Fuel properties of hydrochar and pyrochar: Prediction and exploration with machine learning. *Applied Energy*, 269, 115166. DOI 10.1016/j.apenergy.2020.115166.
- Li, J., Zhang, W., Liu, T., Yang, L., Li, H. et al. (2021). Machine learning aided bio-oil production with high energy recovery and low nitrogen content from hydrothermal liquefaction of biomass with experiment verification. *Chemical Engineering Journal*, 425, 130649. DOI 10.1016/j.cej.2021.130649.
- 17. Shen, J., Zhu, S., Liu, X., Zhang, H., Tan, J. (2010). The prediction of elemental composition of biomass based on proximate analysis. *Energy Conversion and Management*, *51(5)*, 983–987. DOI 10.1016/j. enconman.2009.11.039.

- Keiluweit, M., Nico, P. S., Johnson, M. G., Kleber, M. (2010). Dynamic molecular structure of plant biomassderived black carbon (biochar). *Environmental Science & Technology*, 44(4), 1247–1253. DOI 10.1021/ es9031419.
- Chen, D., Yu, X., Song, C., Pang, X., Huang, J. et al. (2016). Effect of pyrolysis temperature on the chemical oxidation stability of bamboo biochar. *Bioresource Technology*, 218, 1303–1306. DOI 10.1016/j. biortech.2016.07.112.
- Zhao, B., O'Connor, D., Zhang, J., Peng, T., Shen, Z. et al. (2018). Effect of pyrolysis temperature, heating rate, and residence time on rapeseed stem derived biochar. *Journal of Cleaner Production*, 174, 977–987. DOI 10.1016/ j.jclepro.2017.11.013.
- Gascó, G., Paz-Ferreiro, J., Álvarez, M. L., Saa, A., Méndez, A. (2018). Biochars and hydrochars prepared by pyrolysis and hydrothermal carbonisation of pig manure. *Waste Management*, 79, 395–403. DOI 10.1016/j. wasman.2018.08.015.
- Singh, B., Camps-Arbestain, M., Lehmann, J. (2017). *Biochar: A guide to analytical methods*. Australian: CSIRO Publishing.
- Suliman, W., Harsh, J. B., Abu-Lail, N. I., Fortuna, A. M., Dallmeyer, I. et al. (2016). Influence of feedstock source and pyrolysis temperature on biochar bulk and surface properties. *Biomass and Bioenergy*, 84, 37–48. DOI 10.1016/j.biombioe.2015.11.010.
- 24. Demirbas, A. (2004). Effects of temperature and particle size on bio-char yield from pyrolysis of agricultural residues. *Journal of Analytical and Applied Pyrolysis*, 72(2), 243–248. DOI 10.1016/j.jaap.2004.07.003.
- 25. Spokas, K. A. (2010). Review of the stability of biochar in soils: Predictability of O:C molar ratios. *Carbon Management*, 1(2), 289–303. DOI 10.4155/cmt.10.32.
- 26. Xu, S., Chen, J., Peng, H., Leng, S., Li, H. et al. (2021). Effect of biomass type and pyrolysis temperature on nitrogen in biochar, and the comparison with hydrochar. *Fuel*, 291, 120128. DOI 10.1016/j.fuel.2021.120128.
- Ronsse, F., van Hecke, S., Dickinson, D., Prins, W. (2013). Production and characterization of slow pyrolysis biochar: Influence of feedstock type and pyrolysis conditions. *GCB Bioenergy*, 5(2), 104–115. DOI 10.1111/ gcbb.12018.
- Fakayode, O. A., Aboagarib, E. A. A., Zhou, C., Ma, H. (2020). Co-pyrolysis of lignocellulosic and macroalgae biomasses for the production of biochar–A review. *Bioresource Technology*, 297, 122408. DOI 10.1016/j. biortech.2019.122408.
- 29. Setyawati, W., Damanhuri, E., Lestari, P., Dewi, K. (2015). Correlation equation to predict HHV of tropical peat based on its ultimate analyses. *Procedia Engineering*, *125*, 298–303. DOI 10.1016/j.proeng.2015.11.048.
- Choi, H. L., Sudiarto, S. I. A., Renggaman, A. (2014). Prediction of livestock manure and mixture higher heating value based on fundamental analysis. *Fuel, 116,* 772–780. DOI 10.1016/j.fuel.2013.08.064.
- Vhathvarothai, N., James, N., Qiming, J. Y. (2014). An investigation of thermal behaviour of biomass and coal during copyrolysis using thermogravimetric analysis. *International Journal of Energy Research*, 38, 1145– 1154. DOI 10.1002/er.3120.
- 32. Channiwala, S. A., Parikh, P. P. (2002). A unified correlation for estimating HHV of solid, liquid and gaseous fuels. *Fuel*, *81 (8)*, 1051–1063. DOI 10.1016/S0016-2361(01)00131-4.
- Liu, Z., Balasubramanian, R. (2013). A comparison of thermal behaviors of raw biomass, pyrolytic biochar and their blends with lignite. *Bioresource Technology*, 146, 371–378. DOI 10.1016/j.biortech.2013.07.072.
- Liu, W. J., Jiang, H., Yu, H. Q. (2019). Emerging applications of biochar-based materials for energy storage and conversion. *Energy and Environmental Science*, 12(6), 1751–1779. DOI 10.1039/C9EE00206E.
- Panapitiya, G., Avendano-Franco, G., Ren, P., Wen, X., Li, Y. et al. (2018). Machine-learning prediction of CO adsorption in thiolated, Ag-alloyed Au nanoclusters. *Journal of the American Chemical Society*, 140(50), 17508–17514. DOI 10.1021/jacs.8b08800.

Appendix

Samples	Yield (%)	Ultima	te analy	vsis ^a (wt	.%)	Proxima	ate analys	is ^a (%)	VM/FC	HHV
		С	Н	Ν	O ^b	Ash	VM	FC		(MJ/kg)
PM300	61.44	35.60	3.25	2.97	2.54	55.64	26.77	17.59	1.52	12.86
PM400	56.97	22.04	0.96	1.98	10.19	64.84	14.70	20.46	0.72	9.44
PM500	54.19	23.96	0.35	1.88	5.37	68.45	10.31	21.24	0.49	8.89
PM600	52.38	23.65	0.30 ^c	1.69	2.49	71.88	7.96	20.16	0.39	8.61
PM700	50.40	24.60	0.30 ^c	1.24	1.95	71.92	6.75	21.33	0.32	7.66
PM800	49.81	24.17	0.30 ^c	0.74	0.61	74.19	3.17	22.64	0.14	8.28
RS300	49.88	48.11	3.45	1.41	19.49	27.55	29.43	43.03	0.68	18.69
RS400	41.01	47.08	2.53	1.31	16.30	32.79	18.55	48.66	0.38	17.13
RS500	37.34	47.84	1.64	1.15	5.86	43.52	12.55	43.93	0.29	17.83
RS600	34.60	49.83	0.95	1.03	11.35	36.85	8.39	54.76	0.15	16.10
RS700	33.88	48.72	0.60	1.03	10.56	39.10	5.86	55.04	0.11	17.43
RS800	30.53	47.50	0.42	1.13	8.76	42.20	5.86	51.94	0.11	15.37
CL400	37.28	55.92	3.52	9.85	12.84	17.87	33.83	48.30	0.70	22.37
CL500	29.49	56.28	2.69	10.03	7.11	23.90	20.28	55.82	0.36	21.62
CL600	27.56	57.45	3.13	7.85	4.58	26.99	16.97	56.04	0.30	21.32
CL700	24.67	57.74	3.23	6.95	3.65	28.44	12.11	59.46	0.20	20.86
SS400	38.61	54.48	3.54	3.21	17.73	21.06	31.21	47.73	0.65	20.52
SS500	33.94	53.93	3.14	2.95	14.91	25.08	19.96	54.96	0.36	20.95
SS600	33.00	60.74	2.34	2.64	6.23	28.05	22.34	49.61	0.45	22.21
SS700	30.41	59.73	2.63	2.02	4.67	30.97	12.81	56.22	0.23	14.50
SW400	80.13	13.63	2.17	2.77	7.06	74.39	19.24	6.37	3.02	5.93
SW500	74.69	11.76	1.83	2.18	3.62	80.62	12.68	6.70	1.89	4.96
SW600	70.91	11.78	1.17	1.76	0.69	84.62	9.39	6.00	1.57	4.73
SW700	70.20	11.38	0.97	1.36	0.19	86.11	6.61	7.28	0.91	4.47
WD400	25.71	68.74	3.86	1.60	22.15	3.65	32.33	64.02	0.50	26.02
WD500	22.58	80.59	3.90	1.23	10.19	4.09	16.83	79.07	0.21	28.48
WD600	21.90	84.82	3.96	1.14	5.24	4.85	9.96	85.19	0.12	20.87
WD700	20.62	85.17	3.95	1.47	3.95	5.47	5.26	89.27	0.06	20.44
CLSS400	37.08	55.25	3.73	7.13	13.67	20.23	30.35	49.42	0.61	23.71
CLSS500	30.33	55.00	3.51	6.56	10.52	24.42	19.05	56.53	0.34	22.82
CLSS600	29.39	59.69	3.51	5.52	4.78	26.51	12.91	60.58	0.21	21.40
CLSS700	28.31	59.69	3.18	4.17	5.02	27.94	11.52	60.54	0.19	21.32
SWCL400	48.00	23.99	2.14	4.69	8.73	60.45	27.50	12.04	2.28	10.76
SWCL500	48.93	21.40	3.45	4.07	6.13	64.96	16.27	18.77	0.87	10.20

Table S1: Ultimate and proximate composition and HHV of biochars from this study

(Continued)

Table S1 (con	Table S1 (continued)											
Samples	Yield (%)	Ultima	te anal	ysis ^a (w	t%)	Proxim	ate analys	sis ^a (%)	VM/FC	HHV		
		С	Н	N	O ^b	Ash	VM	FC		(MJ/kg)		
SWCL600	40.09	25.03	2.57	3.67	1.78	66.96	13.00	20.05	0.65	10.56		
SWCL700	28.75	18.00	2.66	1.98	1.11	76.26	6.76	16.98	0.40	7.28		
SWSS400	52.54	22.27	3.27	2.16	9.56	62.74	21.38	15.87	1.35	10.13		
SWSS500	45.46	29.27	3.80	2.44	1.88	62.62	15.40	21.98	0.70	11.35		
SWSS600	36.19	26.16	3.22	1.67	1.96	67.00	8.79	24.21	0.36	10.99		
SWSS700	49.33	26.91	2.72	1.47	0.34	68.57	9.87	21.56	0.46	10.73		
SWWD400	51.94	25.14	3.24	1.68	5.70	64.25	18.43	17.32	1.06	9.60		
SWWD500	43.94	25.06	2.91	1.57	3.78	66.68	11.30	22.02	0.51	10.62		
SWWD600	38.90	23.59	2.80	1.29	4.79	67.54	7.45	25.01	0.30	10.30		
SWWD700	41.38	27.80	3.43	0.99	0.26	67.53	4.80	27.67	0.17	10.08		
SSWD400	34.06	64.15	3.74	2.28	15.44	14.40	28.37	57.23	0.50	21.60		
SSWD500	27.15	67.57	3.51	2.20	10.75	15.98	16.28	67.75	0.24	23.67		
SSWD600	27.22	69.31	3.24	2.02	6.95	18.49	10.14	71.37	0.14	25.14		
SSWD700	23.74	71.57	3.39	1.67	4.90	18.48	7.44	74.08	0.10	25.35		
CLWD400	34.32	63.54	3.74	6.51	15.31	10.91	32.55	56.54	0.58	26.54		
CLWD500	28.48	67.40	3.61	6.06	9.33	13.61	15.68	70.71	0.22	26.53		
CLWD600	26.53	70.70	3.55	5.25	5.98	14.52	9.96	75.52	0.13	26.79		
CLWD700	24.66	70.58	3.36	4.49	5.66	15.91	6.50	77.59	0.08	26.00		

Notes: ^aDry basis. ^bBy difference, O = 100 - C - H - N - S - ash, S was not detected, and the content of S in all samples was recorded as 0. ^cNot within the detection limits, and 0.30 was used.

Table S2: The statistical analysis of the parameters involution	volved in the machine learning model of HHV
---	---

	Count	Mean	Standard deviation	Min	25% ^c	50% ^d	75% ^e	Max
C ^a	52	44.93	21.27	11.38	24.49	48.41	59.98	85.17
H ^a	52	2.68	1.14	0.30	2.16	3.20	3.51	3.96
N ^a	52	3.04	2.33	0.74	1.45	2.02	4.10	10.03
O ^b	52	7.20	5.32	0.19	3.64	5.78	10.27	22.15
H/C	52	0.84	0.5	0.1	0.56	0.69	1.19	1.94
O/C	52	0.13	0.09	0.01	0.06	0.09	0.18	0.39
Ash	52	42.15	25.47	3.65	19.79	34.82	66.97	86.11
VM	52	15.27	8.40	3.17	8.69	12.86	19.42	33.83
FC	52	42.58	23.61	6.00	21.04	48.48	57.79	89.27
VM/FC	52	0.56	0.59	0.06	0.21	0.37	0.66	3.02
HHV(MJ/kg)	52	16.38	7.08	4.47	10.18	17.28	21.77	28.48

Notes: ^aDry basis, wt%. ^bBy difference, O = 100 - C - H - N - S - ash. ^cData values sorted from largest to smallest in 25% of the dataset. ^dData values sorted from largest to smallest in 50% of the dataset. ^cData values sorted from largest to smallest in 75% of the dataset.

No.	Temperature (°C)	Ultimate analysis (wt%)						Proximate analysis ^a (%)			VM/FC	HHV	Reference
		С	Н	Ν	0	H/C	O/C	Ash	VM	FC		(MJ/kg)	
1	350	61.40	3.60	1.00	19.00	0.70	0.23	15.00	45.80	39.20	0.86	24.10	[1]
2	350	62.70	4.40	1.10	16.50	0.84	0.20	15.30	45.60	39.10	0.86	23.10	
3	350	56.60	2.90	1.10	17.60	0.61	0.23	21.70	39.50	38.80	0.98	23.80	
4	650	71.10	1.30	1.00	5.80	0.22	0.06	20.90	59.20	19.90	0.34	24.00	
5	650	72.10	1.30	0.90	2.80	0.22	0.03	22.90	57.50	19.60	0.34	23.00	
6	650	73.20	1.20	1.10	3.50	0.20	0.04	20.90	62.90	16.20	0.26	24.40	
7	350	55.50	4.20	0.90	27.10	0.91	0.37	12.30	36.70	51.00	1.39	21.60	
8	350	60.40	5.20	1.10	19.20	1.03	0.24	14.20	40.00	45.90	1.15	22.10	
9	350	62.10	3.00	1.00	17.80	0.58	0.21	16.20	46.80	37.00	0.79	22.30	
10	650	69.20	1.20	1.00	7.20	0.21	0.08	21.40	58.30	20.40	0.35	24.70	
11	650	71.20	1.00	1.00	6.70	0.17	0.07	20.10	62.20	17.70	0.28	25.10	
12	650	71.40	1.10	1.20	4.60	0.18	0.05	21.80	59.30	19.00	0.32	23.50	
13	350	62.00	4.70	1.10	16.90	0.91	0.20	15.20	44.20	40.60	0.92	24.60	
14	350	62.90	4.60	1.10	15.40	0.88	0.18	16.00	47.10	36.90	0.78	24.60	
15	350	64.60	4.30	1.10	13.40	0.80	0.16	16.60	50.40	33.00	0.65	24.30	
16	650	71.80	1.60	1.00	5.10	0.27	0.05	20.50	59.40	20.00	0.34	24.80	
17	650	68.80	1.40	1.10	7.40	0.24	0.08	21.40	56.10	22.50	0.40	24.10	
18	650	72.50	1.30	1.10	3.00	0.22	0.03	22.00	61.80	16.20	0.26	25.90	
19	350	67.30	4.50	0.00	27.50	0.80	0.31	0.80	48.60	50.70	1.04	27.70	
20	350	63.50	5.30	0.10	30.00	1.00	0.35	1.20	50.80	48.00	0.94	26.70	
21	350	68.40	4.60	0.00	25.80	0.81	0.28	1.20	54.50	44.20	0.81	27.60	
22	650	90.40	2.20	0.10	5.90	0.29	0.05	1.40	90.50	8.10	0.09	33.60	
23	650	89.80	2.20	0.10	5.90	0.29	0.05	2.10	90.40	7.50	0.08	33.60	
24	650	90.00	2.00	0.10	5.90	0.27	0.05	2.00	91.30	6.70	0.07	33.90	
25	350	71.40	4.50	0.00	22.70	0.76	0.24	1.40	57.90	40.70	0.70	27.60	
26	350	70.60	5.50	0.10	22.60	0.93	0.24	1.20	56.40	42.40	0.75	27.80	
27	350	71.40	4.50	0.00	22.70	0.76	0.24	1.40	60.20	38.40	0.64	27.10	
28	650	92.80	1.70	0.00	3.30	0.22	0.03	2.20	89.00	8.80	0.10	33.30	
29	650	87.90	2.10	0.10	7.20	0.29	0.06	2.80	88.50	8.80	0.10	33.20	
30	650	90.20	1.80	0.00	6.40	0.24	0.05	1.50	89.40	9.10	0.10	33.10	
31	350	70.10	5.10	0.10	23.40	0.87	0.25	1.20	54.60	44.10	0.81	27.70	
32	350	76.40	5.00	0.10	17.20	0.79	0.17	1.30	65.70	33.00	0.50	28.70	
33	350	67.40	5.90	0.10	25.20	1.05	0.28	1.50	55.90	43.60	0.78	28.90	
34	650	90.20	1.80	0.00	6.50	0.24	0.05	1.50	89.00	9.50	0.11	33.70	
35	650	87.10	1.80	0.00	9.30	0.25	0.08	1.80	89.20	9.10	0.10	33.70	

Table S3: The collected data from the published research

(Continued)

Table	e S3 (continued)												
No.	Temperature (°C)	Ultim	ate an	alysis	(wt%)			Proxir (%)	nate an	alysis ^a	VM/FC	HHV	Reference
		С	Н	Ν	0	H/C	O/C	Ash	VM	FC		(MJ/kg)	
36	650	87.40	1.80	0.00	9.60	0.25	0.08	1.20	89.10	9.70	0.11	33.40	
37	250	50.80	5.20	1.30	37.30	1.23	0.55	5.00	38.50	56.50	1.47	19.85	[2]
38	300	57.00	4.14	2.09	29.59	0.87	0.39	6.80	47.40	45.80	0.97	21.58	
39	350	64.70	4.05	1.30	21.25	0.75	0.25	8.30	61.50	30.20	0.49	25.54	
40	500	71.29	3.38	1.42	13.31	0.57	0.14	10.60	69.60	19.80	0.28	27.25	
41	600	72.25	2.47	1.45	12.33	0.41	0.13	11.50	76.40	12.10	0.16	26.59	
42	250	52.95	5.71	2.85	31.65	1.29	0.45	6.70	36.00	57.30	1.59	21.77	
43	300	56.83	5.17	2.99	25.47	1.09	0.34	9.40	43.70	46.90	1.07	23.07	
44	350	64.93	4.54	2.42	16.78	0.84	0.19	11.30	56.40	32.30	0.57	26.01	
45	500	63.78	3.09	3.15	15.74	0.58	0.19	14.20	61.80	24.00	0.39	23.93	
46	600	68.05	2.38	2.92	10.35	0.42	0.11	16.30	66.40	17.30	0.26	25.10	
47	250	36.60	3.70	1.50	34.70	1.21	0.71	22.90	22.90	54.20	2.37	13.00	
48	300	41.49	3.23	1.86	24.57	0.93	0.44	28.70	25.40	45.90	1.81	15.13	
49	350	41.80	3.00	2.10	18.90	0.86	0.34	33.40	35.50	31.10	0.88	15.54	
50	500	45.25	2.15	1.40	10.89	0.57	0.18	40.00	38.90	21.10	0.54	15.34	
51	600	45.11	1.44	1.42	9.33	0.38	0.16	42.70	45.70	11.60	0.25	15.56	
52	250	52.30	5.00	5.00	28.20	1.15	0.40	9.30	32.60	58.10	1.78	20.93	
53	300	53.57	3.69	5.41	24.34	0.83	0.34	12.80	43.10	44.30	1.03	20.23	
54	350	57.00	3.19	5.30	15.91	0.67	0.21	18.40	48.90	32.70	0.67	22.50	
55	500	60.35	2.44	4.16	11.25	0.49	0.14	21.80	56.60	21.60	0.38	22.23	
56	600	56.60	2.02	3.86	8.22	0.43	0.11	29.30	57.00	13.70	0.24	20.61	
57	350	51.91	3.17	0.72	12.62	0.73	0.18	31.55	24.24	44.21	1.82	21.34	[3]
58	450	52.90	2.58	0.71	8.95	0.59	0.13	34.84	16.33	48.83	2.99	21.53	
59	550	53.83	2.25	0.73	7.28	0.50	0.10	35.88	11.23	52.89	4.71	21.61	
60	650	54.78	2.19	0.67	5.79	0.48	0.08	36.55	11.23	52.22	4.65	21.70	
61	750	55.53	1.29	0.61	3.43	0.28	0.05	39.09	11.26	49.65	4.41	20.95	
62	550	66.13	2.81	2.13	9.43	0.51	0.11	19.78	15.15	65.37	4.31	26.20	
63	650	67.33	2.35	1.96	7.81	0.42	0.09	20.54	11.29	68.17	6.04	26.31	
64	750	69.85	1.79	1.69	6.65	0.31	0.07	20.01	14.89	65.10	4.37	26.39	
65	550	69.31	2.27	1.61	6.41	0.39	0.07	20.29	10.62	69.09	6.51	25.63	
66	550	80.97	3.35	0.91	8.53	0.50	0.08	6.24	16.23	77.53	4.78	31.78	
67	550	86.85	3.39	0.35	7.82	0.47	0.07	1.59	16.98	81.43	4.80	33.75	
68	650	90.68	2.75	0.34	5.02	0.36	0.04	1.21	10.14	88.65	8.74	34.48	
69	750	91.64	2.25	0.51	4.12	0.29	0.03	1.48	7.32	91.20	12.46	34.60	
70	550	89.33	3.03	0.37	6.13	0.41	0.05	1.49	13.31	85.20	6.40	34.48	

(Continued)

Table	e S3 (continued)												
No.	Temperature (°C)	Ultima	ate an	alysis	(wt%)			Proxir (%)	nate an	alysis ^a	VM/FC	HHV	Reference
		С	Н	Ν	0	H/C	O/C	Ash	VM	FC		(MJ/kg)	
71	550	81.24	2.90	1.52	8.68	0.43	0.08	5.61	14.61	79.78	5.46	31.80	
72	550	48.50	2.52	1.37	15.87	0.62	0.25	31.69	30.28	38.03	1.26	18.37	
73	650	49.83	1.74	1.24	13.66	0.42	0.21	33.48	22.32	44.20	1.98	18.13	
74	750	49.29	1.34	1.20	11.20	0.33	0.17	36.92	18.93	44.15	2.33	17.56	
75	550	66.34	2.75	1.91	11.84	0.50	0.13	17.13	18.63	64.24	3.45	25.96	
76	550	52.44	1.94	1.43	11.10	0.44	0.16	33.03	21.16	45.81	2.16	18.81	
77	650	51.77	1.86	1.41	11.70	0.43	0.17	33.21	20.77	46.02	2.22	18.79	
78	750	50.74	1.43	1.33	10.06	0.34	0.15	36.39	16.69	46.92	2.81	17.94	
79	300	47.15	4.52	0.65	23.98	1.15	0.38	22.53	31.28	46.19	1.48	19.60	[4]
80	400	54.33	2.06	0.64	8.11	0.45	0.11	24.54	38.55	36.91	0.96	21.01	
81	500	57.35	1.48	0.63	8.08	0.31	0.11	30.97	61.43	7.60	0.12	22.02	
82	300	46.16	4.51	0.68	27.19	1.17	0.44	18.04	33.26	48.70	1.46	18.31	
83	400	52.13	2.67	0.63	12.64	0.61	0.18	29.66	47.42	22.92	0.48	21.53	
84	500	58.41	1.75	0.77	5.27	0.36	0.07	32.22	55.10	12.68	0.23	23.41	
85	300	46.14	3.83	0.44	23.33	1.00	0.38	23.51	31.78	44.70	1.41	17.60	
86	400	54.56	2.25	0.50	9.16	0.49	0.13	31.82	50.77	17.41	0.34	21.11	
87	500	56.28	1.36	0.63	7.74	0.29	0.10	32.91	58.37	8.72	0.15	22.64	
88	350	55.80	4.29	2.60	18.73	0.92	0.25	24.20	23.20	53.50	2.31	20.90	[5]
89	700	56.67	0.94	1.51	4.13	0.20	0.05	39.50	34.70	27.70	0.80	18.97	
90	350	53.32	4.05	3.64	15.70	0.91	0.22	28.70	23.50	47.90	2.04	20.39	
91	700	52.41	0.91	1.70	7.20	0.21	0.10	44.00	36.30	19.80	0.55	17.23	
92	350	51.07	0.38	4.45	15.63	0.09	0.23	30.70	27.00	42.30	1.57	19.03	
93	700	45.91	0.20	2.07	10.53	0.05	0.17	46.20	35.50	18.30	0.52	14.75	
94	350	51.51	0.49	3.54	11.10	0.11	0.16	32.50	17.70	49.80	2.81	21.12	
95	700	44.06	0.07	2.61	4.03	0.02	0.07	52.90	33.80	13.40	0.40	15.07	
96	350	49.28	0.36	4.07	15.40	0.09	0.23	34.80	23.10	42.10	1.82	17.28	
97	700	44.77	0.09	1.94	5.80	0.02	0.10	49.90	29.20	20.80	0.71	14.45	

Note: ^aAll data exclude the influence of moisture content and follow FC + VM + ash = 100.

	Count	Mean	Standard deviation	Min	25% ^b	50% ^c	75% ^d	Max
C ^a	149	56.91	18.94	11.38	48.50	56.83	69.31	92.80
H ^a	149	2.74	1.36	0.07	1.79	2.75	3.61	5.9
N^{a}	149	1.95	1.88	0.00	0.09	1.41	2.28	10.03
O^a	149	11.33	7.94	0.19	5.70	9.30	15.74	37.30
H/C	149	0.65	0.42	0.02	0.31	0.58	0.86	1.94
O/C	149	0.16	0.12	0.01	0.06	0.13	0.23	0.71
Ash	149	27.12	21.61	0.80	11.50	21.80	35.88	86.11
VM	149	29.50	20.16	3.17	12.68	22.34	44.20	91.20
FC	149	43.41	22.86	6.00	22.02	45.70	58.30	91.30
VM/FC	149	1.24	1.76	0.06	0.28	0.64	1.46	12.50
HHV(MJ/kg)	149	21.27	7.02	4.47	17.56	21.61	25.90	34.60

Table S4: The statistical analysis of the parameters involved in the machine learning model of HHV

Notes: ^aDry basis, wt%. ^bData values sorted from largest to smallest in 25% of the dataset.

^cData values sorted from largest to smallest in 50% of the dataset.

^dData values sorted from largest to smallest in 75% of the dataset.

References

- 1. Crombie, K., Mašek, O. (2015). Pyrolysis biochar systems, balance between bioenergy and carbon sequestration. GCB Bioenergy, 7(2), 349–361. DOI 10.1111/gcbb.12137.
- 2. Tag, A. T., Duman, G., Ucar, S., Yanik, J. (2016). Effects of feedstock type and pyrolysis temperature on potential applications of biochar. Journal of Analytical and Applied Pyrolysis, 120, 200-206. DOI 10.1016/j. jaap.2016.05.006.
- 3. Qian, C., Li, Q., Zhang, Z., Wang, X., Hu, J. et al. (2020). Prediction of higher heating values of biochar from proximate and ultimate analysis. Fuel, 265, 116925. DOI 10.1016/j.fuel.2019.116925.
- 4. Vieira, F. R., Romero Luna, C. M., Arce, G. L. A. F., Ávila, I. (2020). Optimization of slow pyrolysis process parameters using a fixed bed reactor for biochar yield from rice husk. Biomass and Bioenergy, 132, 105412. DOI 10.1016/j.biombioe.2019.105412.
- 5. Cantrell, K. B., Hunt, P. G., Uchimiya, M., Novak, J. M., Ro, K. S. (2012). Impact of pyrolysis temperature and manure source on physicochemical characteristics of biochar. Bioresource Technology, 107, 419-428. DOI 10.1016/j.biortech.2011.11.084.