Article

Application of Random Forest Model Integrated with Feature Reduction for Biomass Torrefaction

Xiaorui Liu 1, Haiping Yang 2,*, Jiamin Yang 1 and Fang Liu 3

1 School of Mine, China University of Mining and Technology, Xuzhou 221116, China
2 State Key Laboratory of Coal Combustion, School of Energy and Power Engineering, Huazhong University of Science and Technology, Wuhan 430074, China
3 School of Low-Carbon Energy and Power Engineering, China University of Mining and Technology, Xuzhou 221116, China
* Correspondence: yanghaiping@hust.edu.cn

Abstract: A random forest (RF) model integrated with feature reduction was implemented to predict the properties of torrefied biomass based on feedstock and torrefaction conditions. Four features were selected for the prediction of fuel ratio (FR) and nitrogen content (Nt), and five features were selected for O/C and H/C ratios and HHV values. The results showed that the feature-reduced model had excellent prediction performance with the values of $R^2$ higher than 0.93 and RMSE less than 0.58 for all targets. Moreover, partial dependence analysis (PDA) was performed to quantify the impacts of selected features and torrefaction conditions on the targets. Temperature was the dominant factor for FR, O/C and H/C ratios, and HHV values, whereas Nt was determined most on the nitrogen content in the feedstock (Ni). This study provided comprehensive information for understanding biomass torrefaction.

Keywords: biomass torrefaction; machine learning; feature reduction; partial dependence analysis; random forest

1. Introduction

The demand for renewable energy sources is growing rapidly worldwide due to the energy crisis and environmental pollution caused by excessive consumption of fossil fuels [1]. As the only carbon-based renewable energy resource, biomass is considered to be an important alternative to fossil fuels since it can not only be used for power generation and heat supply, but can also be converted into gases, liquids, chemicals, and solid products with zero CO$_2$ emissions throughout its life cycle [2,3]. However, raw biomass often exhibits high moisture content and low calorific values, which are adverse to its utilization [4]. To upgrade the quality of raw biomass, various pretreatment technologies have been proposed, including pyrolysis, hydrothermal, torrefaction, etc; among them, torrefaction, which is traditionally performed at 200–300 °C in an inert atmosphere, is considered to be the most economic and efficient pretreatment technology [5–7].

To evaluate the torrefaction process, it is necessary to determine the properties of torrefied biomass. Generally, proximate (moisture, fixed carbon, volatile matter, and ash contents), ultimate (CHON contents), and high heating value (HHV) analyses are essentially implemented. Proximate and ultimate analyses require special instruments and complex procedures [8], while HHV is either experimentally tested using a calorimeter bomb or mathematically calculated based on the ultimate and proximate analyses results [9]. Thus, repetitive experiments, instrumental analysis or mathematical calculations that consume lots of time, costs and manpower are needed. Therefore, researchers have been trying to develop a method to estimate torrefied product properties from those of feedstock without various tests and experiments.
Due to the strong ability to deal with complex and non-linear problems, the application of machine learning (ML) methods to biomass torrefaction has received growing attention in recent years. Several ML models, involving artificial neural network (ANN), gradient boosting trees (GBT), random forest (RF), support vector machines (SVM), etc., were employed to predict the properties of torrefied biomass. However, these studies mainly focused on the HHV [8–12] and mass yield [1,10,13–15], while other properties of torrefied biomass which are also important parameters for evaluating the performance of torrefied biomass, including fuel ratio (FR, fixed carbon content divided by volatile content), and O/C and H/C (oxygen or hydrogen content divided by carbon content) ratios [16] were rarely reported. Moreover, nitrogen content, which is the dominant source of NOx and other nitrogen containing pollutants, such as NH3 and HCN during biomass conversion was also not of concern until now. Therefore, more efforts are needed to provide supplemental information to the torrefied biomass by developing ML algorithms.

In addition to prediction, the influences of feedstock and torrefaction conditions on torrefaction process are also critical focuses of researches to provide internal information for understanding torrefaction mechanisms. Other than the ANNs, black-box models which were most frequently used in existing literatures [17], RF models have the advantage of processing the correlation between input and output variables for both classification and regression problems [18]. Leng et al. [15] analysed the impact of influential factors on the yields of three-phase pyrolysis products and bio-oil HHV using a RF model. Tang et al. [19] also discussed the effects of input features and the resulting interactions exerted on the pyrolysis process based on a RF model. However, to the authors’ knowledge, the implementation of RF in analyzing the torrefaction process has not been reported.

As reviewed above, there are massive gaps in ML algorithms application to biomass torrefaction, including the prediction of properties and their intrinsic relations with raw feedstock and torrefaction conditions. In this study, a RF model was employed to predict torrefied biomass properties including the FR, H/C and O/C ratios, HHV, and N content (Nt). The influences of the feedstock and torrefaction conditions were then investigated by feature reduction and partial dependence analysis (PDA). The proximate and ultimate analyses of raw biomass and the torrefaction conditions (temperature, duration time) were assigned as input variables, while the FR, H/C and O/C ratios, HHV and Nt of torrefied biomass were labeled as targets. The contributions of each input feature to the targets as well as their interactions were also analyzed. The results could provide comprehensive insights into biomass torrefaction.

2. Materials and Methods
2.1. Data Collection and Pre-Processing

515 data points were extracted from 67 journal publications, including our previous studies [20], to create a dataset; the detailed information is provided in the supplementary materials (Table S1). Notably, only the traditional torrefaction performed at 200–300 °C in N2, He, Ar, and anoxic atmosphere was taken into account. Most data were obtained from the texts, the tables in the literatures and their corresponding supplementary materials. To obtain data not directly listed, WebPlotDigitizer software (https://apps.automeris.io/wpd/) was used to extract the necessary data from the figures. Diverse biomass was employed as feedstock whether it had been dried or not, including agricultural residues, forestry residues, energy crops, leaves, fruit wastes, sewage sludge, and so on. The proximate and ultimate analyses results of the raw biomass involving moisture (MO, wt.%), volatile matter (VM, wt.%), fixed carbon (FC, wt.%), ash (Ash, wt.%) contents and CHON contents (wt.%) as a function of torrefaction conditions involving temperature (Temp, °C) and duration time (Time, min) were assigned as input variables. The FR, O/C and H/C ratios, HHV, and Nt of the torrefied biomass were assigned as the output targets.
Correlation analysis was applied to the dataset, and linear dependencies among all features and targets were measured using Pearson’s correlation coefficient (PCC) [21].

\[
\rho = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}}
\]

(1)

where \( \rho \) is the value of PCC; \( \bar{x} \) and \( \bar{y} \) are the means of the feature and the target, respectively. \( \rho \) is ranged from \(-1\) to \(1\), where \(0\) means no linear correlation, and a negative value means a negative correlation, while a positive value indicates a positive correlation. It should be noted that the lack of a linear correlation between input features and output targets does not imply the absence of other relationships.

To obtain a uniform range among the variables, the features and the targets were then normalized using Z-score standardization using Equation (2):

\[
x_{i}^* = \frac{x_i - \mu}{s}
\]

(2)

where \(x_i\) is the value of feature \(i\); \(x_{i}^*\) is the normalized value of \(x_i\); \(\mu\) is the mean value of \(x_i\), and \(s\) represents its standard deviation.

After pre-processing, the data points were randomly divided into training and validation subsets at a ratio of 80% to 20% for the evaluation of the developed model. To improve the model’s performance, a 10-fold cross-validation method (Figure 1) was applied during the training process to tune the hyper-parameter using a trial-and-error method, including the number of trees in the forest (n_estimators), the maximum depth of each tree (max_depth), and the number of features to consider when seeking the best split (max_features). To accomplish this, the dataset was randomly shuffled and split into 10 non-overlapping folds; then, nine of them were employed to train the ML model and the remaining one was used for testing for repeatedly 10 times with a different piece of test fold. Finally, the optimal hyper-parameters were determined, as shown in Table 1.
Table 1. Optimal hyper-parameters.

<table>
<thead>
<tr>
<th></th>
<th>n_Estimators</th>
<th>max_Depth</th>
<th>max_Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>FR</td>
<td>174</td>
<td>16</td>
<td>0.353</td>
</tr>
<tr>
<td>HC</td>
<td>112</td>
<td>29</td>
<td>0.313</td>
</tr>
<tr>
<td>OC</td>
<td>174</td>
<td>27</td>
<td>0.232</td>
</tr>
<tr>
<td>HHV</td>
<td>141</td>
<td>46</td>
<td>0.353</td>
</tr>
<tr>
<td>Nt</td>
<td>97</td>
<td>41</td>
<td>0.596</td>
</tr>
</tbody>
</table>

2.2. RF Model Description

As an ensemble learning method introduced by Breiman in 2001 [18], a random forest (RF) model consists of multiple decision trees and can be used to solve both classification and regression problems [22,23]. For each tree, it generally contains a root node, several internal nodes, and several leaf nodes. The leaf nodes correspond to decision outcomes, and each of the other nodes corresponds to a feature test. The samples in each node are divided into child nodes according to the features’ splitting results. The path from the root node to each leaf node is a decision sequence. The purpose is to produce a tree with a strong generalization ability, that is, strong ability to deal with unseen examples.

The structure of the RF model is shown in Figure 2 and the definition is given by Equation (3).

\[ \{ h(x, \theta_k), k = 1, 2, \ldots, K \} \]  

(3)

where \( x \) represents an input vector, \( \{ \theta_k \} \) is the random vector based on independent identical distribution, and \( K \) is the number of trees.

![Figure 2. The structure of RF model.](image-url)

The RF begins with the random split of the training set into multiple subsets based on bootstrap aggregating (bagging) strategy involving “in-bag” (IB) and “out-of-bag” (OOB) samples. With considering the outputs of all trees which are trained with subsets and evaluated by the OOB samples to prevent overfitting, RF model is performed to determine the final predicted value. Finally, the test set is used to determine the model’s performance.

RF models exhibit low overfitting risk and high predictive accuracy because they utilize bagging and random feature selection strategies to construct several decorrelation decision trees and output the average value of their prediction results [24,25]. Moreover, RF models also show high tolerance to missing values, which is a significant superiority since data missing is very common in biomass related studies [26].
2.3. Performance Evaluation Criteria

The model’s performance was evaluated using the determination coefficient \( R^2 \) and the root mean square error \( (RMSE) \) [27]. Conceptually, a higher \( R^2 \) and a lower \( RMSE \) indicate better model accuracy.

\[
R^2 = 1 - \frac{\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{N} (\hat{y}_i - y)^2}
\]

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

where \( \hat{y}, y, \) and \( \bar{y} \) are the predicted, actual, and mean values of the target, respectively; \( N \) is the total number of data points.

2.4. Feature Importance Analysis

Once the model was built and its prediction performance was evaluated, the relative importance of the features was then determined for further model optimization. In this study, the mean decrease impurity method [28] was applied to calculate the relative importance of raw biomass properties and torrefaction conditions in predicting targets. Then, features with nearly 80% importance were selected for further model optimization.

To gain a better understanding of how each feature affected the targets, partial dependence analysis (PDA) was performed. As there were 10 input features and 5 targets, only the influences of selected features on the corresponding target were visualized using the partial dependence plot (PDP), in which the relationship type involving a linear, monotonous, or more complicated connection for each selected feature on the target could be exposed [29]. Figure 3 illustrates the FR model workflow for the prediction of torrefied biomass properties based on raw biomass features and torrefaction conditions.

![Figure 3. The workflow of the RF model for biomass torrefaction.](image)

3. Results and Discussion

3.1. Dataset Description

Table 2 presents a statistical description of the dataset, including the units, counts, ranges, means, median values, and standard deviations of all inputs and outputs. A statistical distribution overview of the dataset is presented by violin-plots in Figure S1. For the inner boxplot, the interquartile range (IQR) was used to measure data variability. Points outside the range, i.e., outliers, were plotted individually with rhombus shape which would be directly removed from the dataset. The data distribution (outer) based on kernel density estimation facilitated visualization of data distribution.
Table 2. Dataset description.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Name</th>
<th>Unit</th>
<th>Count</th>
<th>Range</th>
<th>Mean</th>
<th>Median</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moisture content</td>
<td>MO</td>
<td>wt.%</td>
<td>515</td>
<td>0–14.8</td>
<td>4.33</td>
<td>4.55</td>
<td>4.12</td>
</tr>
<tr>
<td>Volatile content</td>
<td>VM</td>
<td>wt.%</td>
<td>515</td>
<td>32.2–96.4</td>
<td>77.02</td>
<td>78.9</td>
<td>8.28</td>
</tr>
<tr>
<td>Ash content</td>
<td>ASH</td>
<td>wt.%</td>
<td>515</td>
<td>0–32.58</td>
<td>5.48</td>
<td>3.3</td>
<td>5.76</td>
</tr>
<tr>
<td>Fixed carbon content</td>
<td>FC</td>
<td>wt.%</td>
<td>510</td>
<td>1.67–61.4</td>
<td>14.72</td>
<td>15.47</td>
<td>5.91</td>
</tr>
<tr>
<td>Carbon content</td>
<td>C</td>
<td>wt.%</td>
<td>497</td>
<td>29.59–54.16</td>
<td>44.83</td>
<td>45.1</td>
<td>4.73</td>
</tr>
<tr>
<td>Hydrogen content</td>
<td>H</td>
<td>wt.%</td>
<td>497</td>
<td>3.92–8.78</td>
<td>5.99</td>
<td>6.03</td>
<td>0.69</td>
</tr>
<tr>
<td>Oxygen content</td>
<td>O</td>
<td>wt.%</td>
<td>488</td>
<td>11.37–61.55</td>
<td>43.83</td>
<td>44.46</td>
<td>7.57</td>
</tr>
<tr>
<td>Nitrogen content input</td>
<td>Ni</td>
<td>wt.%</td>
<td>488</td>
<td>0–14.29</td>
<td>1.24</td>
<td>0.59</td>
<td>2.09</td>
</tr>
<tr>
<td>Temperature</td>
<td>Temp</td>
<td>°C</td>
<td>515</td>
<td>200–300</td>
<td>255.27</td>
<td>250</td>
<td>32.63</td>
</tr>
<tr>
<td>Duration time</td>
<td>Time</td>
<td>min</td>
<td>491</td>
<td>10–360</td>
<td>46.78</td>
<td>30</td>
<td>35.20</td>
</tr>
<tr>
<td>Nitrogen content output</td>
<td>Nt</td>
<td>wt.%</td>
<td>379</td>
<td>0–8.32</td>
<td>0.94</td>
<td>0.65</td>
<td>1.066</td>
</tr>
<tr>
<td>High heating value</td>
<td>HHV</td>
<td>MJ/kg</td>
<td>491</td>
<td>13.48–225.5</td>
<td>21.45</td>
<td>20.67</td>
<td>9.76</td>
</tr>
<tr>
<td>Fuel ratio</td>
<td>FR</td>
<td>—</td>
<td>436</td>
<td>0.02–7.19</td>
<td>0.53</td>
<td>0.56</td>
<td>0.65</td>
</tr>
<tr>
<td>H/C ratio</td>
<td>H/C</td>
<td>—</td>
<td>397</td>
<td>0.03–0.16</td>
<td>0.10</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>O/C ratio</td>
<td>O/C</td>
<td>—</td>
<td>380</td>
<td>0.06–1.87</td>
<td>0.70</td>
<td>0.68</td>
<td>0.29</td>
</tr>
</tbody>
</table>

According to Table 2 and Figure S1, the MO content of raw biomass ranged within 0–14.8 wt.%. The samples had VM content ranging from 32.2 to 96.4 wt.% with a median of 78.9 wt.%. ASH and FC contents were in the range of 0–32.58 wt.% and 1.67–61.4 wt.%, respectively. Regarding elemental compositions, the C content varied from 29.59 wt.% to 54.16 wt.% with a 45.1 wt.% median and a 48.83 wt.% plot peak. O content ranged from 11.37 to 61.55 wt.%, the plot peak of which was 44.25 wt.%. H content was in the range of 3.92 and 8.78 wt.%. Ni was widely distributed in the range of 0–14.29 wt.%; however, most feedstock had N content of approximate 0.37 wt.%. The significant variations in textural properties were attributed to diversity in the feedstock. For the torrefaction conditions, 250 °C was the most frequently employed temperature, closely followed by 300 °C. The duration time ranged from 10 to 360 min, among which, 30 min was most common.

Regarding the properties of torrefied biomass, FR varied between 0.021 and 7.19 with a mean value of 0.53, which is close to that of coal (0.5–2.0) [30], indicating that the combustibility of biomass was significantly improved by torrefaction. The density peaks of O/C and H/C plots were 0.71 and 0.11, respectively, which lie within the coal’s ratio ranges of 0.38–0.91 and 0.02–0.28, respectively [31]. The reductions in O/C and H/C ratios benefitted to HHV, which ranged from 13.48 to 30.3 MJ/kg with a median of 20.67 MJ/kg. It is worth noting that the torrefied biomass had a lower N content than the feedstock, implying that torrefaction could decrease the N content of biomass and subsequently reduce the emissions of N-containing pollutants during the utilization process.

Figure 4 is the Pearson correlation matrix between any two variables, the detailed information of which was shown in Table S2. There were no significant linear correlations between any two features. Relatively strong negative correlations were observed for ASH-VM (−0.73) simply because the sum of the proximate analysis results should be 100%; therefore, Ash was removed from the input variables. The Nt exhibited a very strong linear connection with Ni (0.86), implying that the Nt highly depended on the feedstock’s nitrogen content.
Figure 4. Pearson correlation matrix between any two features.

3.2. Feature Importance

The relative importance of each input feature in predicting FR, H/C, O/C, Nt, and HHV of torrefied biomass is exhibited in Table 3. It can be seen that torrefaction temperature was the most important factor for FR, H/C, O/C, and HHV, the share of which was 40.33%, 37.32%, 30.69%, and 31.13%, respectively. Guo et al. [32] also reported that temperature was the most important factor influencing torrefaction. While for Nt, Ni was most crucial since it occupied a fairly high share of 54.38%, indicating that the nitrogen content in the torrefied biomass significantly depended on that of the raw material. Table 4 shows the selected features for each target.

Table 3. Feature importance of each feature to the prediction of targets.

<table>
<thead>
<tr>
<th>Feature</th>
<th>MO</th>
<th>VM</th>
<th>ASH</th>
<th>FC</th>
<th>C</th>
<th>H</th>
<th>O</th>
<th>Ni</th>
<th>Temp</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>FR</td>
<td>0.0467</td>
<td>0.1217</td>
<td>0.1872</td>
<td>0.0417</td>
<td>0.0442</td>
<td>0.0460</td>
<td>0.0487</td>
<td>0.4033</td>
<td>0.0605</td>
<td></td>
</tr>
<tr>
<td>H/C</td>
<td>0.0486</td>
<td>0.0659</td>
<td>0.0877</td>
<td>0.0886</td>
<td>0.0935</td>
<td>0.0781</td>
<td>0.0549</td>
<td>0.3732</td>
<td>0.1094</td>
<td></td>
</tr>
<tr>
<td>O/C</td>
<td>0.0690</td>
<td>0.0770</td>
<td>0.0853</td>
<td>0.1045</td>
<td>0.0547</td>
<td>0.1408</td>
<td>0.0876</td>
<td>0.3069</td>
<td>0.0742</td>
<td></td>
</tr>
<tr>
<td>Nt</td>
<td>0.0473</td>
<td>0.0969</td>
<td>0.1023</td>
<td>0.0632</td>
<td>0.0249</td>
<td>0.0318</td>
<td>0.5438</td>
<td>0.0533</td>
<td>0.0365</td>
<td></td>
</tr>
<tr>
<td>HHV</td>
<td>0.0321</td>
<td>0.1694</td>
<td>0.0600</td>
<td>0.1579</td>
<td>0.078</td>
<td>0.0643</td>
<td>0.0552</td>
<td>0.3113</td>
<td>0.0719</td>
<td></td>
</tr>
</tbody>
</table>

Table 4. Selected features for each target.

<table>
<thead>
<tr>
<th>Target</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>FR</td>
<td>VM, FC, Temp, time</td>
</tr>
<tr>
<td>H/C</td>
<td>FC, C, H, Temp, time</td>
</tr>
<tr>
<td>O/C</td>
<td>VM, FC, C, O, Temp</td>
</tr>
<tr>
<td>Nt</td>
<td>VM, FC, Ni, Temp</td>
</tr>
<tr>
<td>HHV</td>
<td>VM, C, H, Temp, time</td>
</tr>
</tbody>
</table>

3.3. Prediction Performance of the Feature-Reduced Model

Figure 5 presents a comparison of predicted values and tested data for training and testing sets with selected features for each target. All data points were densely distributed along the black line of \( y = x \), implying the equivalence between predicted values and actual data. The results of \( R^2 \) and RMSE for both training and testing sets are shown in Table 5. \( R^2 \) values for all targets’ training and testing sets were larger than 0.93, and RMSE values
were also fairly acceptable. Therefore, the RF model with feature reduction could predict the FR, H/C, O/C, HHV, and Nt of torrefied biomass from the features of feedstock and torrefaction conditions for both training and testing sets with high precision. Compared with other models employed in existing literatures [8–12], the feature-reduced RF model developed in this study showed excellent performance in predicting the corresponding properties of torrefied biomass.

![Figure 5. Comparison of predicted values and experimental data of training and testing sets.](image)

**Table 5.** R² and RMSE results for the training set and the testing set.

<table>
<thead>
<tr>
<th></th>
<th>FR</th>
<th>H/C</th>
<th>O/C</th>
<th>HHV</th>
<th>Nt</th>
</tr>
</thead>
<tbody>
<tr>
<td>R²</td>
<td>0.9649</td>
<td>0.9653</td>
<td>0.9589</td>
<td>0.9774</td>
<td>0.982</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.048</td>
<td>0.0042</td>
<td>0.048</td>
<td>0.4671</td>
<td>0.0679</td>
</tr>
<tr>
<td>R²</td>
<td>0.9486</td>
<td>0.9654</td>
<td>0.9315</td>
<td>0.9649</td>
<td>0.9746</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.0585</td>
<td>0.0042</td>
<td>0.0579</td>
<td>0.5819</td>
<td>0.0808</td>
</tr>
</tbody>
</table>

### 3.4. Partial Dependency

The partial dependences of the targets on selected features are shown in Figure 6. For FR, completely opposite effects were observed for VM and FC, implying that a higher FC content and a lower VM content of the feedstock is beneficial to the FR improvement of torrefied biomass. Temperature was the most influential factor with a contribution rate of as high as 40.33%. As temperature increased, a raising trend of FR was observed. Especially, as temperature reached to above 270 °C, the FR sharply increased, attributing to the more violent decomposition of the organic components which led to the decrease of VM. It was reported that hemicellulose was most sensitive to temperature, which decomposed between 200 and 350 °C, whereas cellulose and lignin decomposed above 250 °C [10]. A same trend was presented for the FR plot versus duration time. And FR increased more sharply when duration time exceeded 100 min, implying that a longer duration time is essential to the heat transfer into the solids to release VM. Due to this, an increase in duration time also raised the HHV of torrefied biomass and declined the H/C ratio which was affected by torrefaction severity [2]. However, the variation of HHV was slight when duration time was longer than 180 min, indicating that torrefaction for more than 180 min was of less necessary for HHV improvement. Thus, 100–180 min was appropriate duration time for biomass torrefaction. Temperature was also the dominant factor for H/C and
O/C ratios. As temperature increased, both ratios sharply declined due to the release of a large amount of HO elements and relatively few C element mainly in the forms of CO₂, CO and vapor [33,34]. Moreover, O/C and H/C ratios exhibited negative relationships with C content, but positive to O and H contents, respectively. It was reported that the consumption of HO elements during torrefaction was much larger than that of C element, leading to the concentration of C in the torrefied biomass [33]. An increase in temperature and time tended to increasingly improve the HHV of torrefied biomass due to the decrease of O/C and H/C ratios. This is consistent with the results reported by Onsree et al. [34] that temperature is the most influential feature to HHV prediction. It is worth noting that Nt exhibited a sharp upward trend with increasing Ni which contributed 54.38% importance, indicating that Nt was highly dependent on Ni. Different to other targets, temperature contributed only 5.33% importance to Nt prediction which was effective when temperature was higher than 260 °C. This might be attributed to the more violent release of HO elements than Ni into gases at above 260 °C.

Figure 6. Partial dependence plots of the targets on each selected feature.

4. Conclusions

The properties of torrefied biomass, including FR, O/C and H/C ratios, HHV, and Nt, were predicted using a RF model integrated with feature reduction. Four features were selected for the prediction of FR and Nt, while five features for O/C, H/C, and HHV prediction. The feature-reduced model exhibited excellent prediction precision with the values of R² higher than 0.93 and acceptable RMSE for all targets. Further, PDA was performed to quantify the impact of selected features and torrefaction conditions on the targets. Temperature was the most influential factor for FR, O/C, H/C, and HHV, whereas Nt was strongly dependent on Ni. The results obtained in this study provide compre
Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su142316055/s1, Figure S1: The statistical distribution of each variable in terms of the inherent properties of the raw biomass, the torrefaction conditions, and the fuel properties of torrefied biomass with violin-plots; Table S1: Dataset; Table S2: Detailed information of the PCCs.

Author Contributions: Conceptualization, X.L. and H.Y.; methodology, J.Y.; validation, H.Y.; writing—original draft preparation, X.L.; writing—review and editing, X.L.; visualization, F.L.; project administration, F.L.; and funding acquisition, X.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Natural Science Foundation of Jiangsu Province, grant number BK20210511.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References


7. Lin, Y.; Chen, W.; Colin, B.; Pétrissans, A.; Lopes Quirino, R.; Pétrissans, M. Thermodegradation characterization of hardwoods and softwoods in torrefaction and transition zone between torrefaction and pyrolysis. *Fuel* 2022, 310, 122281. [CrossRef]


