Renewable Energy 201 (2022) 70–86

Contents lists available at ScienceDirect

Renewable Energy

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

Biomass microwave pyrolysis characterization by machine learning for sustainable rural biorefineries

Yadong Yang [a], Hossein Shahbeik [b], Alireza Shafizadeh [c], Nima Masoudnia [c], Shahin Rafiee [c], Yijia Zhang [a], Junting Pan [a,***], Meisam Tabatabaei [d,e,*], Mortaza Aghbashlo [c,**]

[a] Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, 100081, PR China
[b] Henan Province Engineering Research Center for Biomass Value-added Products, School of Forestry, Henan Agricultural University, Zhengzhou, 450002, China
[c] Department of Mechanical Engineering of Agricultural Machinery, Faculty of Agricultural Engineering and Technology, College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran
[d] Higher Institution Centre of Excellence (HICoE), Institute of Tropical Aquaculture and Fisheries (AKUATROP), Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia
[e] Biofuel Research Team (BRTeam), Terengganu, Malaysia

ARTICLE INFO

Keywords:
Biomass microwave pyrolysis
Machine learning
Gradient boost regressor
Biochar
Bio-oil
Syngas

ABSTRACT

Microwave heating is a promising solution to overcome the shortcomings of conventional heating in biomass pyrolysis. Nevertheless, biomass microwave pyrolysis is a complex thermochemical process governed by several endogenous and exogenous parameters. Modeling such a complicated process is challenging due to the need for many experimental measurements. Machine learning can effectively cope with the time and cost constraints of experiments. Hence, this study uses machine learning to model the quantity and quality of products (biochar, bio-oil, and syngas) that evolve in biomass microwave pyrolysis. An inclusive dataset encompassing different biomass types, microwave absorbers, and reaction conditions is selected from the literature and subjected to data mining. Three machine learning models (support vector regressor, random forest regressor, and gradient boost regressor) are used to model the process based on 14 descriptors. The gradient boost regressor model provides better prediction performance ($R^2 > 0.822$, RMSE <12.38, and RRMSE <0.765) than the other models. SHAP analysis generally reveals the significance of operating temperature, microwave power, and reaction time in predicting the output responses. Overall, the developed machine learning model can effectively save cost and time during biomass microwave pyrolysis while serving as a valuable tool for guiding experiments and facilitating optimization.

1. Introduction

Growing concern about global warming, climate change, and environmental degradation resulting from greenhouse gas accumulation in the atmosphere are the main motivations to explore alternative fuels and chemicals [1]. Numerous substitutes, such as biomass, solar, wind, hydro, and geothermal, have been exploited to reduce the reliance on conventional fossil resources [2]. The great geographical diversity and availability of biomass feedstocks make them suitable for producing biofuels and biochemicals. The biomass-derived products resemble petroleum ones that facilitate their use in the existing energy and chemical infrastructures. Substituting biomass-derived liquid transportation biofuels for fossil fuels can help mitigate atmospheric greenhouse gases and achieve a low-carbon economy [3].

Abbreviations: CO, Carbon monoxide; CO_2, Carbon dioxide; CH_4, Methane; GBR, Gradient boost regressor; H_2, Hydrogen; H/C, Hydrogen-to-carbon; H/N, Hydrogen-to-nitrogen; O/C, Oxygen-to-carbon; RMSE, Root-mean-square error; RRMSE, Relative root-mean-square error; RFR, Random forest regressor; R^2, Coefficient of determination; SHAP, Shapley additive explanations; SVR, Support vector regressor.

* Corresponding author. Higher Institution Centre of Excellence (HICoE), Institute of Tropical Aquaculture and Fisheries (AKUATROP), Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia.
** Corresponding author.
*** Corresponding author.
E-mail addresses: panjunting@caas.cn (J. Pan), meisam.tabatabaei@umt.edu.my, meisam_tab@yahoo.com (M. Tabatabaei), maghbashlo@ut.ac.ir (M. Aghbashlo).

https://doi.org/10.1016/j.renene.2022.11.028
Received 17 September 2022; Received in revised form 28 October 2022; Accepted 12 November 2022
Available online 17 November 2022
0960-1481/

Lignocellulosic and non-lignocellulosic biomass feedstocks can be valorized into hydrocarbons using various thermochemical conversion routes, i.e., carbonization, pyrolysis, gasification, and hydrothermal processing [3,4]. Facile and versatile pyrolysis has attracted much attention recently to convert biomass into valuable products with many end-use applications. The pyrolysis products include condensable volatiles (bio-oil), non-condensable compounds (syngas), and carbonaceous residues (biochar) [5]. The desired pyrolysis product distribution can be achieved by changing process parameters (e.g., heating rate, residence time, and operating temperature) [6].

In transitioning from a linear economy to a circular bioeconomy, biomass pyrolysis can play a pivotal role in sustainable rural biorefineries as a default sink for carbon and energy. This thermochemical process does not generate secondary pollutants while materializing the circular bioeconomy [7]. The pyrolysis process can effectively upgrade waste biomass into high-value products and decrease transportation and storage costs. This process alone or in combination with other biochemical and thermochemical processes in the biorefinery framework can facilitate waste biomass management in rural areas at a lower cost [7,8].

Conduction and convection heat transfer modes are typically applied to raise biomass temperature for initiating pyrolysis reactions [9]. However, conventional heating results in reduced process efficiency and increased processing costs. Microwave heating can help address these issues by providing non-contacting, selective, volumetric, and molecular-level heating of material [3,10]. The other advantages of microwave heating are high energy efficiency, fast turn on or off, process flexibility, precise process control, equipment portability, high heating rate, and uncomplicated scale-up [9,11].

In microwave pyrolysis, the irradiated microwaves penetrate biomass particles and generate thermal energy. The generated energy creates a temperature gradient from the inside to the outside of biomass particles, similar to the direction of evolved volatiles [12]. The microwave pyrolysis process does not need agitation; hence, fewer ash

Table 1
Features of the machine learning models reported in the literature to predict some aspects of biomass microwave pyrolysis.

Feedstocks	Machine learning model(s)	Model input(s)	Model output(s)	Model limitations	Ref.
Plastic, tire, waste paper, sugarcane bagasse, sawdust, soybean cake	Machine learning linear regression, decision tree analysis	Feedstock type, feedstock size, heating rate, final temperature, heating source	Bio-oil yield	- Using a linear model for modeling a complex process - Predicting a few process outputs - Using descriptive parameters as inputs - Ignoring effective operating parameters, including microwave absorber percentage, dielectric constant of the absorber, and dielectric loss factor of the absorber	Ge et al. [18]
Biomass, waste tire, oil shale, coal	Linear & polynomial regression models, k-nearest neighbor, artificial neural networks, Gradient boosting	Heating rate, final temperature, dwelling time, feedstock type/size	Biochar yield and its higher heating value	- Predicting a limited number of process outputs - Ignoring effective operating parameters, including microwave absorber percentage, dielectric constant of the absorber, and dielectric loss factor of the absorber	Huang et al. [19]
Biomass feedstock	Linear regression, extreme gradient boosting, random forest, support vector machine	Feedstock characteristics (elemental and proximate composition) microwave power, time, feedstock weight, absorber	Biochar yield and its higher heating value	- Predicting a few process outputs - Ignoring effective operating parameters, including microwave absorber percentage, dielectric constant of the absorber, and dielectric loss factor of the absorber	Selvam and Balasubramanian [20]
Sawdust	Polynomial regression machine learning model	Catalyst loading, pretreatment temperature	Heating rate, pyrolysis temperature, susceptor thermal energy, bio-oil yield, biochar yield, syngas yield	- Predicting a limited number of process outputs - Ignoring effective operating parameters, including microwave absorber percentage, dielectric constant of the absorber, and dielectric loss factor of the absorber	Potnuri et al. [21]
Biomass feedstock	Linear, interactive, and quadratic regression machine learning models	Ultimate and proximate characteristics of feedstock, amount of sample or feed rate, reaction time, temperature, microwave power	Biochar yield	- Predicting a few process outputs - Ignoring effective operating parameters, including microwave absorber percentage, dielectric constant of the absorber, and dielectric loss factor of the absorber	Narde and Remya [22]
Polystyrene	Support vector machine algorithm	Mass of KOH, mass of polystyrene	Bio-oil yield, syngas yield, char yield, pyrolysis time, heating rate, specific microwave power, specific microwave energy, microwave conversion efficiency, conductive heat loss	- Ignoring effective operating parameters, including microwave absorber percentage, dielectric constant of the absorber, and dielectric loss factor of the absorber - Applicable only for limited types of feedstocks	Terapalli et al. [23]

particles are introduced into bio-oil [13]. This process is usually carried out in the presence of an absorbent, particularly carbonaceous materials (e.g., carbon black, biochar, activated carbon, and silicon carbide), to assimilate microwaves and transmit thermal energy to the surrounding sample mass [13,14]. Microwave pyrolysis is conducted in two modes: temperature and power control modes. Microwave power is adjusted in real-time to achieve the set temperature in the temperature control mode. In the power control mode, the microwave device continuously irradiates energy at the set power [11].

Despite the unique feature of microwave pyrolysis, it is a complex thermochemical process influenced by several factors, i.e., biomass composition, operating temperature, microwave power, absorbent type and quantity, and reaction time [1]. Accordingly, numerous experiments must be conducted to model and understand the process, even using experimental design methods (i.e., Taguchi and response surface methodology) [15]. Conducting biomass microwave pyrolysis experiments require a huge amount of resources and time. This approach does not always lead to the applicable modeling and understanding of the process. Process modeling and understanding through data-driven paradigms such as machine learning is an efficient approach to coping with the time-consuming and cost-intensive nature of experimental measurements [16]. Machine learning can capture nonlinear relationships between the inputs and the outputs of complex systems by learning from past data with no assistance or intervention from the users [17]. This approach can make modeling and understanding biomass microwave pyrolysis fast and inexpensive by providing a chance to get hidden insights and trends. Recently, a few studies have been published concerning the use of machine learning technology for modeling the biomass microwave pyrolysis process. Table 1 summarizes the features of the machine learning models reported in the literature to predict some aspects of biomass microwave pyrolysis.

Generally, the reported machine learning models are not comprehensive and do not cover all the influential parameters affecting biomass pyrolysis under microwave irradiation. These models are also designed to predict a few desired output responses. Hence, such inadequate models provide little insight into the biomass microwave pyrolysis process and are useless for design and development goals. Therefore, this research aims to develop a comprehensive machine learning model to predict the quantity and quality of the bio-oil, syngas, and biochar produced in biomass microwave pyrolysis. The most influential input parameters (14 descriptors) affecting the biomass microwave pyrolysis process are considered in this study. An inclusive dataset covering different biomass feedstocks, microwave absorbers, and reaction conditions is selected from the literature. The collected dataset is mined using statistical methods and explained using mechanistic arguments. Three machine learning models, i.e., support vector regressor (SVR), random forest regressor (RFR), and gradient boost regressor (GBR), are applied to characterize biomass microwave pyrolysis based on 14 descriptors. The SHAP analysis is conducted using the best model to comprehend the significance of the input descriptors on the target responses. A simple computer program is also developed to facilitate the future applications of developed models. The findings of this study can help achieve more economically viable and environmentally friendly pyrolysis-based biorefineries in rural areas.

2. Research method

Fig. 1 shows the workflow used in this study for developing machine learning models for the biomass microwave pyrolysis process. A sufficient amount of data patterns is required to develop reliable models for such a complex process. In this regard, the published papers on the biomass microwave pyrolysis process were carefully explored, and the qualified ones were chosen for further evaluation. After extracting the required data from the eligible articles, the absurd data and outliers were removed manually from the collected dataset. The independent input variables were microwave parameters and biomass composition

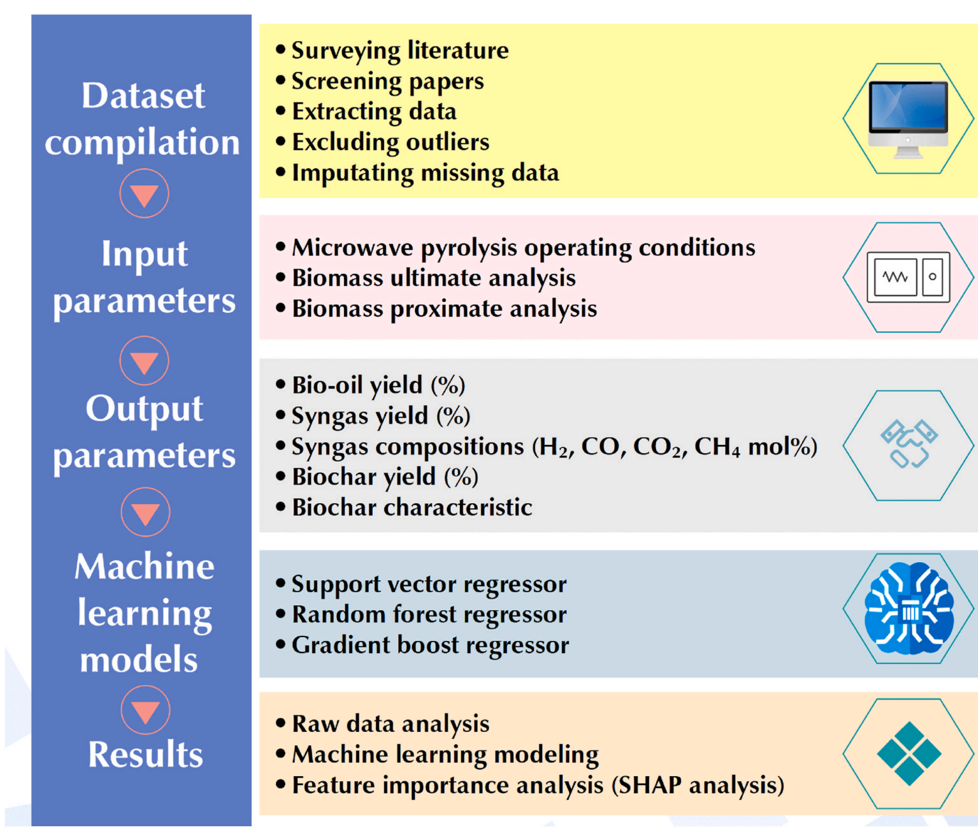

Fig. 1. Workflow used in this study for developing machine learning models for the biomass microwave pyrolysis process.

(proximate and ultimate analyses). The dependent output responses were the quality and quality of the evolved biochar, bio-oil, and syngas. The missing data related to biomass composition was imputated using the Missforest approach. The reliability of the data imputation procedure was also approved through principal component analysis. Three machine learning models, including SVR, RFR, and GBR, were applied to model the process. To evaluate the performance of the developed models, statistical parameters, including coefficient of determination (R^2), root-mean-square error (RMSE), and relative root-mean-square error (RRMSE), were used. The best model was subjected to feature importance analysis (SHAP method) to understand the importance of descriptors on responses in a systematic way. A simple computer program was developed to facilitate the application of the developed machine learning models in future studies.

2.1. Literature survey and data compilation

To obtain the required data patterns, papers published on the biomass microwave pyrolysis process were first sought. The search was confined to 2010–2022 and performed in the Scopus database with the keywords “microwave pyrolysis” and “biomass”. This search protocol returned 78 experimental studies. These papers were screened manually and checked carefully in terms of the availability of required input and output parameters for machine learning modeling. Generally, 47 eligible articles were shortlisted for data extraction. To extract the required data from the figures reported in the selected articles, the open-source WebPlotDigitizer was used. The outliers or invalid data were removed from the extracted dataset. The independent input parameters of the biomass microwave pyrolysis process could be categorized into biomass ultimate analysis, biomass proximate analysis, and operating parameters. Table 2 tabulates input descriptors and output responses of machine learning models used to characterize biomass microwave pyrolysis. Overall, a total of 249 data patterns were gathered from the 47 eligible articles (Supplementary Excel File).

Table 2
Input descriptors and output responses of machine learning models used to characterize biomass microwave pyrolysis.

Input parameter category	Input descriptors	Output parameter category	Output responses
Biomass ultimate analysis	Carbon content (wt %)	Bio-oil	Bio-oil yield (%)
	Hydrogen content (wt%)		
	Nitrogen content (wt%)		
	Oxygen content (wt %)		
	Sulfur content (wt %)		
Biomass proximate analysis	Volatile matter (wt %)	Syngas	Syngas yield (%)
	Fixed carbon content (wt%)		H_2 concentration (mol%)
	Ash content (wt%)		CO concentration (mol%)
			CO_2 concentration (mol%)
			CH_4 concentration (mol%)
Operating parameters	Operating temperature (°C)	Biochar	Biochar yield (%)
	Microwave power (W)		Biochar H/C ratio (−)
	Reaction time (min)		Biochar H/N ratio (−)
	Microwave absorber percentage (%)		Biochar O/C ratio (−)
	Dielectric constant of absorber (−)		Biochar calorific value (MJ/kg)
	Dielectric loss factor of absorber (−)		

2.2. Modeling biomass microwave pyrolysis

The overarching motivation of the current study was to develop an accurate predictive machine learning model for biomass microwave pyrolysis based on biomass composition and operating parameters. After compiling data patterns, their outliers were manually removed. The dielectric properties (dielectric constant and dielectric loss factor) of some well-known absorber materials used in experimental studies have not been reported in published papers. These constants were obtained from open literature and used in data mining and machine learning modeling (Green inked cells in Supplementary Excel File). The missing biomass composition data (sulfur and nitrogen content) were imputated through the Missforest method (Red inked cells in Supplementary Excel File). This data imputation approach is a strong technique to impute missing data with high accuracy, particularly for mixed-type data [24, 25]. The Missforest imputation technique imputes the missing data using the random forest algorithm in three successive steps, i.e., initialization, imputation, and repetition. In the initialization step, all missing observations are substituted by the mean value of the variable. The random forest model is trained to impute the missing data in the imputation step. In the repetition step, the second step is repeated for all missed entries to build the random forest model completely. The second and third steps are continued until the stopping criteria are reached [26].

After data imputation and completing the missing data entries, the input descriptors and output responses were normalized between zero and one. To improve the performance of models and prevent overfitting, k-fold cross-validation was used [27]. By the k-fold resampling method, all the complied data patterns are alternately used in training and testing. The technique divides the dataset into k partitions (folds). k-1 folds are then applied in model training, while the remaining one is used in model testing [28]. The five-fold (k = 5) cross-validation method was used in this study. In better words, 80% of the total normalized data was divided into five folds and used in model training (four folds) and validation (one fold). The remaining 20% of the total normalized data was used in testing the validated model. Then, three machine learning models, including SVR, RFR, and GBR, were employed to train the dataset for predicting the quantity and quality of biomass microwave pyrolysis products.

SVR learns by example to assign labels to objects [29]. This supervised learning algorithm has a strong ability to classify by considering a compromise between accuracy and reproducibility. The algorithm can also be used for regression problems. The SVR algorithm precisely selects an optimal line or hyperplane to split data belonging to one class from another. The training of the SVR model is based on maximizing the margin between the support vector of the classes. The constructed hyperplane is then used for unseen data to predict and classify them. The SVR technique can be used for linear and nonlinear multi-dimensional problems [30]. This algorithm uses various kernel functions (i.e., radial basis function or polynomial and multi-layer perceptron classifiers) to find the best solution for different problems [31].

The non-parametric RFR is an ensemble learning algorithm [32]. This algorithm mixes an enormous set of regression trees for training. Each tree characterizes a set of hierarchically ordered conditions or restrictions. These rules function the tree from its root to its leaf. The process starts with many bootstrap samples drawn randomly with replacements from the original training dataset. Each bootstrap sample is fitted by a regression tree. For each node per tree, a small set of input variables selected from the total set is randomly considered for binary partitioning. The regression tree splitting criterion is based on choosing the input variable with the lowest Gini Index. The predicted value of an observation is calculated by averaging over all the trees [33]. The RFR technique has a low-computational cost. It is resistant to noisy data and overfitting predictors. The algorithm can also deliver reliable error; more importantly, the variable importance can be explained through this algorithm [34]. In addition, the RF algorithm can handle large datasets efficiently while requiring fewer hyperparameters (number of regression trees and the number of input variables per node) compared with other machine-learning methods [33].

The GBR is another ensemble machine learning method that uses a greedy algorithm. Instead of having one tree, an additive regression model using a collection of decision trees predicts desired outputs in this algorithm. The outcomes of one decision tree are used to train the following decision tree. In better words, the outcomes of the previous trees are used in the training process of the consequent trees. In the GBR model, weak learners are put together by iteratively focusing on the errors resulting at each step until one strong predictive model is obtained as a sum of the successive weak ones [35]. The gradient descent algorithm is applied to minimize errors in sequential models [36]. The overfitting possibility is less in the GBR algorithm because of randomization in the training dataset selection. Two main hyperparameters of the GBR model are the number of iterations and the learning rate. These parameters can effectively reflect model complexity. Increasing the number of iterations and learning rate can yield a more robust model while preventing model overfitting [37]. The theoretical background behind the machine learning models used in this study is presented in Supplementary Material (Supplementary Word File).

2.3. Feature importance analysis

One of the most important steps in machine learning modeling is interpreting and evaluating model outputs. In this study, the effects of descriptors on responses were scrutinized by SHAP analysis. This analysis is a technique to indicate the relative importance of contributing attributes (independent input parameters) on dependent predicting outcomes [38]. The game theory-based SHAP method represents the mean marginal contribution of each input parameter to the corresponding output. Besides the hierarchy of importance (SHAP value magnitude), the method can demonstrate how each input parameter contributes (negative or positive) to the corresponding dependence output variable. Due to the distinctive advantages of the SHAP method, it can be used in the local or global interpretation of the model [39]. The importance of each input on the corresponding output is represented by beeswarm plots to summarize the whole distribution of SHAP scores. The global importance of each model input on corresponding outputs is also depicted through heatmap plots [40].

3. Results and discussion

3.1. Descriptive analysis of the collected dataset

The collected database was subjected to statistical analysis to acquire preliminary insights into various descriptors and responses in the biomass microwave pyrolysis process. The basic statistical indicators, including mean, standard deviation, Skewness, Kurtosis, minimum, median, maximum, and quartile-basis of box plots, were calculated for all the descriptors and responses (Table S1 in Supplementary Word File). The database contained a wide range of ultimate and proximate compositions, showing its capability to cover most types of biomass feedstocks. The carbon content and volatile matter of biomass feedstocks ranged from 25.4 to 75.13 wt% and from 46.76 to 90.77 wt%, respectively. Such a high carbon content and volatile matter indicated that alternative biofuels with high-energy density could be produced from these biomass feedstocks. The nitrogen and oxygen content of biomass feedstocks varied between 0–12.24 wt% and 0.6–55.9 wt%, respectively. These heteroatom compounds might significantly negatively affect the quantity and quality of bio-oil in the biomass microwave pyrolysis process [41]. The ash content of biomass feedstocks was high enough (with a mean value of 9 wt%). Such a high ash content could improve the efficiency of microwave-assisted pyrolysis of biomass feedstocks. It is worth mentioning that some ash components (e.g., Fe_2O_3 and TiO_2) could be good microwave absorbers [42].

This study considered operating temperature, microwave power, reaction time, and microwave absorber percentage as operating parameters. In the collected data, the microwave power and reaction time ranged between 60 and 4000 W and 6.66–120 min, respectively. These data ranges could effectively cover a wide spectrum of feasible operating conditions. Such a broad and diverse dataset could offer a chance to construct reliable machine learning models with acceptable generalization ability. In general, syngas yield contributed the most to product composition, with a median value of 33.21 wt%, followed by biochar yield (30.65 wt%) and bio-oil yield (29.8 wt%). The vast range of feedstock compositions and operating conditions resulted in significant variations in product distribution. Based on the median values derived from different feedstocks, the yield of H_2 in the syngas was the highest, ranging from 0.8 to 52.57 mol%, followed by CO (4.2–70.7 mol%), CO_2 (2.26–65.07 mol%), and CH_4 (2.06–29.78 mol%).

Biochar is a carbon-rich material formed at lower microwave power and temperature levels in biomass pyrolysis. The biochar obtained from microwave pyrolysis has a higher carbon content, calorific value, and surface area than conventional pyrolysis [43]. In addition, the apparent density and oxygen content of microwave-derived biochar is lower than the biochar produced in conventional pyrolysis [43]. The high porosity of biochar derived from microwave irradiation make it a good alternative to be used as an adsorbent for organic dyes, inorganic metals, and ions [44]. The O/C and H/C atomic ratios of microwave-derived biochar were very low (with median values of 0.21 and 0.45, respectively), showing its carbon-rich nature. This finding could be attributed to the fact that biomass feedstocks experience a higher degree of carbonization during microwave pyrolysis. It should be noted that biochar with lower O/C and H/C atomic ratios has more aromatic compounds, stable carbon structure, and good fuel characteristics [45]. The O/C and H/C ratios of microwave-derived biochar were very close to those of anthracite coal, indicating its high potential to be co-processed with coal or used as a coal replacement [46].

Spearman correlation analysis was performed to determine the preliminary relationship between any two variables in the biomass microwave pyrolysis process. This non-parametric technique ranks values in the dataset to measure the strength and direction (positive/negative) of the relationship between two variables based on monotonic function [47]. The correlation coefficient between the considered features ranges from −1 (total negative correlation) to 1 (total positive correlation), indicating a comparatively low or high impact. Fig. 2 displays the Spearman correlation matrix between the descriptors and the responses of the biomass microwave pyrolysis process. As shown in Fig. 2A, microwave power was positively correlated with bio-oil yield ($r = 0.13$). Generally, increasing microwave power up to a certain value could lead to increased power density during biomass microwave pyrolysis. Elevating power density could increase the heating rate and operating temperature, thus resulting in an increase in bio-oil yield. Microwave energy could break down strong C–C and C–H bonds and covalent bonds between carbon and oxygen/nitrogen, resulting in the conversion of organic matter into the liquid phase [44]. However, further increasing microwave power beyond the optimal value could not improve bio-oil yield because of the promotion of secondary cracking reactions and decomposition of biochar residues into non-condensable gases [48].

Reaction time was positively correlated with syngas yield ($r = 0.16$) since prolonging time favors the evolution of gaseous compounds by achieving a gradual enhancement in secondary cracking and decomposition reactions (Fig. 2B) [49]. Furthermore, longer reaction times could promote the formation of waxes and char residues since the generated intermediates were converted into solid products because of the carbonization process. This condition normally could occur at lower operating temperatures. It should be emphasized that longer residence times could negatively increase energy consumption, raising overall

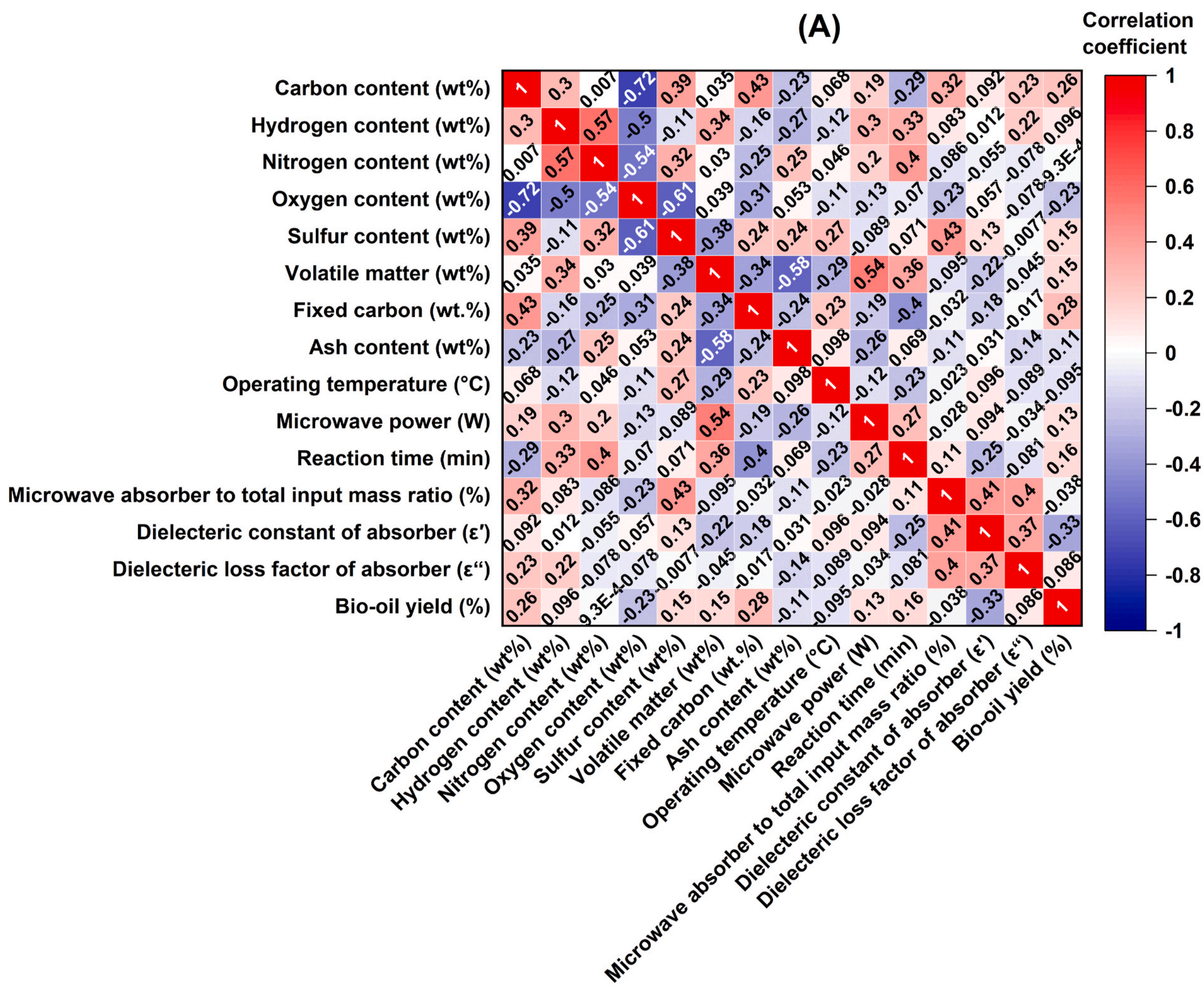

Fig. 2. Spearman correlation matrix showing a positive or negative relationship between any two variables in different product phases: (A) bio-oil, (B) syngas, and (C) biochar.

operating expenditure. The operating temperature was inversely correlated with biochar yield since the higher temperatures could enhance the evolution of gaseous compounds (Fig. 2C). On the other hand, biochar produced at higher temperatures has significantly higher porosity. Such highly porous biochar can provide excellent absorption performance [44]. Biomass ash content was positively correlated with biochar yield ($r = 0.37$) since fewer organic compounds were present in ash-rich biomass feedstocks [42]. This issue could result in reduced bio-oil production while promoting biochar formation.

The principal component analysis is conducted to visualize and comprehend the information space of the complied dataset. This approach determines the positive/negative relevance of each descriptor on the responses. The principal component analysis was carried out for both raw (Fig. 3) and imputated (Fig. S1 in Supplementary Word File) data. The principal component analysis results for raw and imputated data were very close to each other. This proximity approved the reliability of data imputation by the Missforest method. There was a positive correlation between the carbon content of biomass feedstocks and bio-oil yield. In fact, carbon could easily migrate into the liquid phase at higher biomass carbon contents in biomass microwave pyrolysis (Fig. 3A).

The first three principal components extracted about 20.1%, 16.2%, and 15.5% of the total variance from the dataset, respectively. It is worth mentioning that the first principal component primarily comprised biomass characteristics, such as carbon content, ash content, and volatile matter, and the second principal component mainly comprised reaction time and microwave power. These two principal factors highlighted the significance of biomass properties and operational conditions in the first and second principal components, respectively. Syngas yield positively correlated with microwave power, showing that higher power density could gradually enhance secondary reactions between vapors and solids. It is noteworthy that biomass microwave pyrolysis can always produce higher syngas yield with a lower concentration of unwanted greenhouse gases (such as CH_4 and CO_2) than conventional pyrolysis [44]. According to Fig. 3B, the first three principal components accounted for over 55% of the variance, showing the importance of both feedstock properties and operational conditions in the dataset.

The reaction time and microwave power were positively correlated with biochar yield (Fig. 3C). Increasing the residence time of pyrolysis vapors under microwave irradiation could extend their secondary reactions, leading to a decrease in bio-oil yield while increasing biochar formation [46]. In addition, changes in microwave power level could affect biochar yield and calorific value. This issue could be attributed to the decomposition of lignin at higher microwave power levels, resulting in a drop in the heating value of biochar. According to Fig. 3C, the first three principal components amounted to 76.3% of the total variance, representing the importance of biomass compositions and operational conditions. Notably, such correlations may change with biomass attributes and operational conditions.

The relationships between the most influential operating conditions

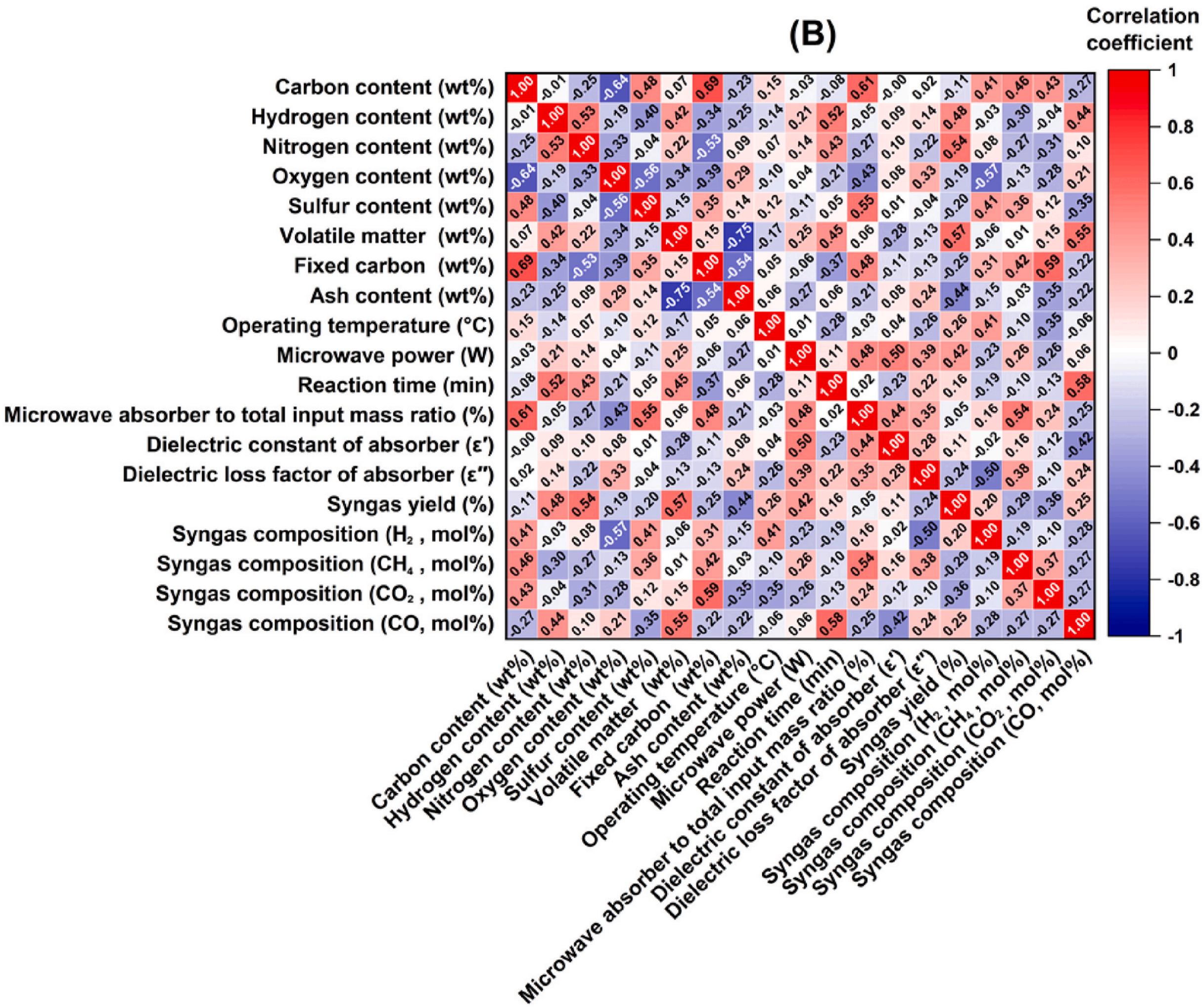

Fig. 2. (*continued*).

of biomass microwave pyrolysis (i.e., operating temperature, microwave power, and reaction time) and product distribution are shown by the contour diagrams in Fig. 4. Generally, the operating temperature, microwave power, and reaction time ranging from 650 to 750 °C, 500–800 W, and 40–60 min, respectively, could result in higher bio-oil yields. At higher microwave power levels (i.e., 2000–2500 W), product distribution could shift toward producing more non-condensable gas because of the acceleration of secondary cracking reactions. In addition, elevating microwave power could lower biochar yield because of the decomposition of biochar residues into gaseous products.

3.2. Machine learning modeling results

The performance of the trained models was compared based on three statistical parameters (R^2, RMSE, and RRMSE). The mathematical presentation of these parameters is as follows (Eqs. (1)–(3)) [50,51].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_{p.i} - y_{a.i}\right)^2}{\sum_{i=1}^{n}\left(y_{a.i} - \overline{y}_a\right)^2} \quad (1)$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{p.i} - y_{a.i}\right)^2} \quad (2)$$

$$\mathrm{RRMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_{p.i} - y_{a.i}}{y_{a.i}}\right)^2} \quad (3)$$

where n is the number of data points, $\overline{y}_a$ is the mean of actual output, $y_{a.i}$ is the actual output, and $y_{p.i}$ is the predicted output. The results of statistical parameters are tabulated in Tables 3 and 4. The GBR model could outperform the other machine learning models in terms of accuracy in characterizing biomass microwave pyrolysis. The R^2, RME, and RRMSE of the GBR model in the training phase varied between 0.861 and 0.999, 8.0×10^{-4}–10.185, and 5.0×10^{-5}–0.476, respectively. These values in the testing phase were between 0.822 and 0.985, 0.056–12.38, and 0.065–0.765 for R^2, RMSE, and RRMSE, respectively. The GBR method could satisfactorily predict bio-oil yield as one of the most important output features in the biomass microwave pyrolysis process. The R^2, RMSE, and RRME for the bio-oil yield in the testing phase were 0.906, 4.287, and 0.138, respectively. In general, the GBR model could provide the best performance for all the output responses except for the biochar H/N ratio, where the lowest R^2 values were obtained in the training and testing phases (0.861 and 0.822, respectively). Nevertheless, the result remained satisfactory in performance prediction. The best R^2 value in the training and testing phases belonged to H_2 and biochar O/C ratio, respectively. The lowest RMSE and RRMSE values in the testing phase were obtained for the biochar H/C ratio and biochar calorific value, respectively.

Fig. 5 depicts the values predicted by developed machine learning models against the actual reported experimental data in the training and testing steps. The blue lines indicate the regression line with a 95% confidence band for the training phase, while the red line represents the same values for the testing phase. As seen in Fig. 5, the GBR model could

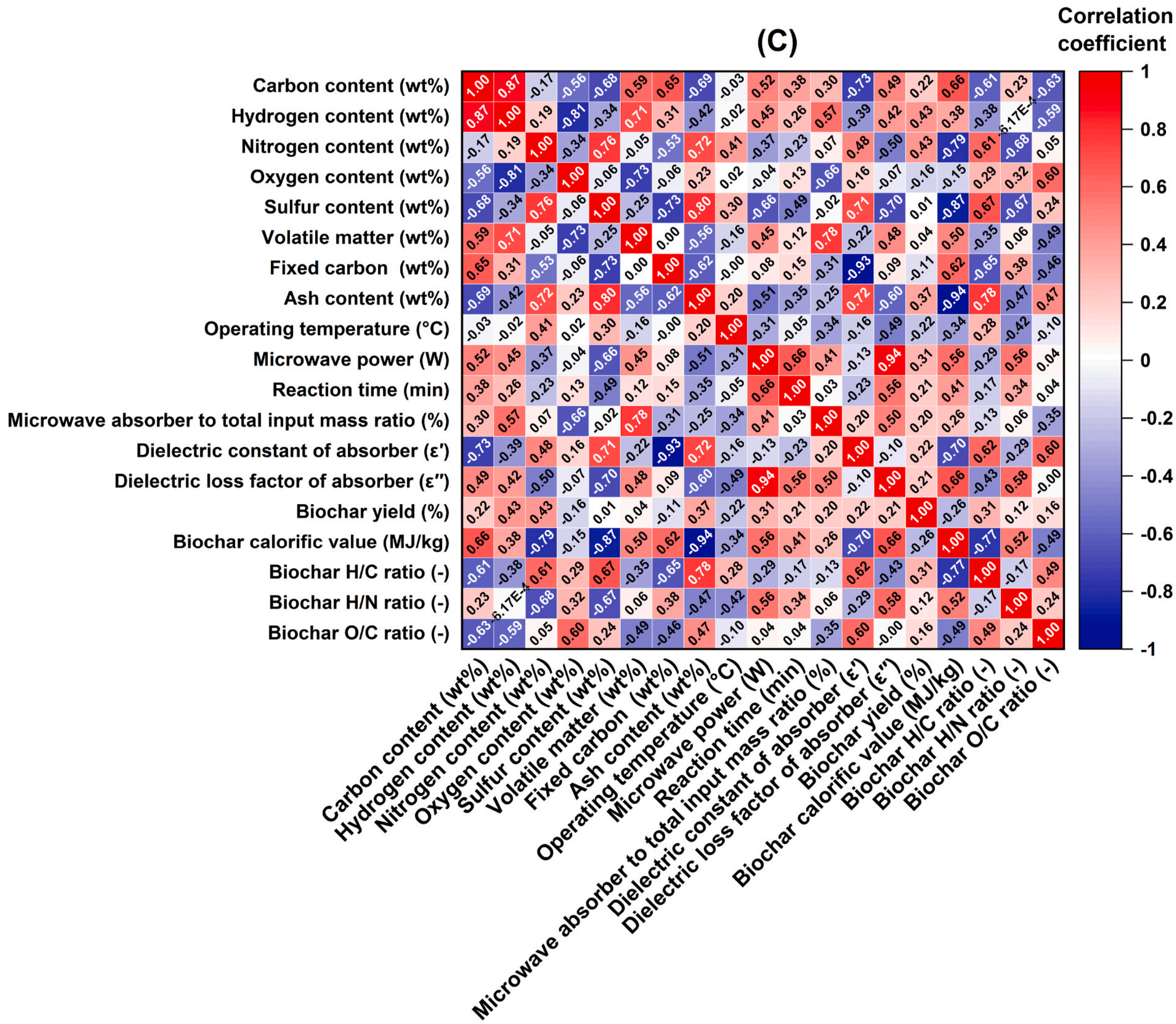

Fig. 2. (*continued*).

excellently generalize the complex biomass microwave pyrolysis process. The excellent performance of the GBR model in characterizing the biomass microwave process could be because of its high fitting accuracy [52]. These results showed that the GBR-predicted data satisfactorily linearly correlated with experimental data (Fig. 5). The great ability of the decision trees-based GBR technique to model nonlinear feature interactions could be another reason for this finding [53]. The GBR model could minimize prediction errors in each training step, resulting in more reliable and robust results than the RFR algorithm [54]. In contrast to GBR, the SVR model showed the poorest prediction performance. In addition, the RFR model had a weaker performance than the GBR model. These results could be because of the high complexity of the biomass microwave pyrolysis process and the nonlinearity of the relationships between input descriptors and output responses. The generated hyperplanes in SVR could not satisfactorily split the results in the multi-dimensional modeling space. These hyperparameters resulted in the best performance of the GBR model in modeling are tabulated in Table 5. Overall, the GBR model could predict the quantity and quality of pyrolysis products from various biomass feedstocks under microwave irradiation at different operating conditions.

A simple computer tool was created based on the selected GBR model (as well as the RFR model) to reduce the cost and time spent carrying out experimental trials in biomass microwave pyrolysis (https://drive.google.com/file/d/1uhz3Eyhhkk0Cvg5kr932qFPSulVi2buO/view?usp=sharing). The input descriptors considered throughout the modeling step need to be entered into an excel file (called “input_App”) to characterize the biomass microwave pyrolysis process. After running the executive file (called “Biomass Microwave Pyrolysis”), a new excel file (called “input_output_App”) is created. The created file contains all the output responses of the biomass microwave pyrolysis considered herein based on the GBR and RFR models. A screenshot from the software package while calculating the outputs is presented in Fig. S2 (Supplementary Word File).

3.3. SHAP analysis

Understanding the significance of descriptors and their influence on responses throughout the modeling process is a challenging issue because of the black-box nature of machine learning models. Therefore, SHAP analysis was performed herein using the GBR model (selected as the best-performing model in the current study) to investigate how each input descriptor could affect the model prediction. This unbiased and consistent method developed based on cooperative game theory can offer a chance to understand and interpret machine learning models [55]. The SHAP values can satisfactorily reflect the contribution of descriptors on the responses across the entire dataset in machine learning modeling from both local and global points of view [56]. This promising approach can aid in investigating nonlinear interactions among descriptors and responses in complex systems.

The SHAP analysis of descriptors and responses in the biomass

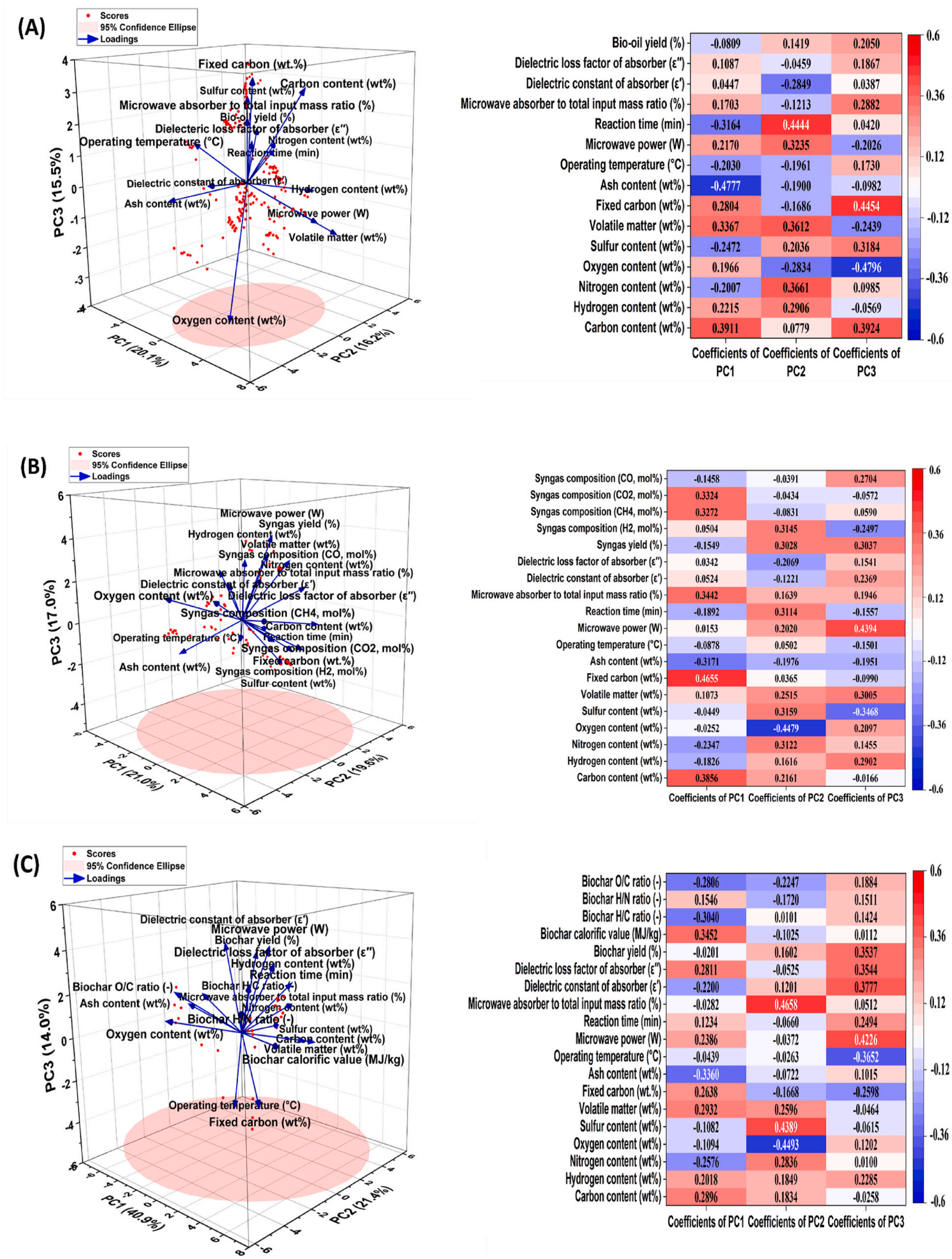

Fig. 3. Principal component analysis showing the effect of input parameters on output responses (3D plot) and the relationship of the input variables with the first three principal components (Heatmap plot): (A) bio-oil, (B) syngas, and (C) biochar.

microwave pyrolysis process is illustrated in Fig. 6. The order of input variables was determined based on the mean absolute SHAP values imposed on predicted responses, and the most contributing input variable was located at the top. In the left-hand graphs (i.e., beeswarm plot), each dot represents an individual data point in the dataset, and its color indicates the feature value. The red-color point represents a higher SHAP value of a given input feature, while the blue-color point indicates a lower SHAP value. It should be noted that turning data points from

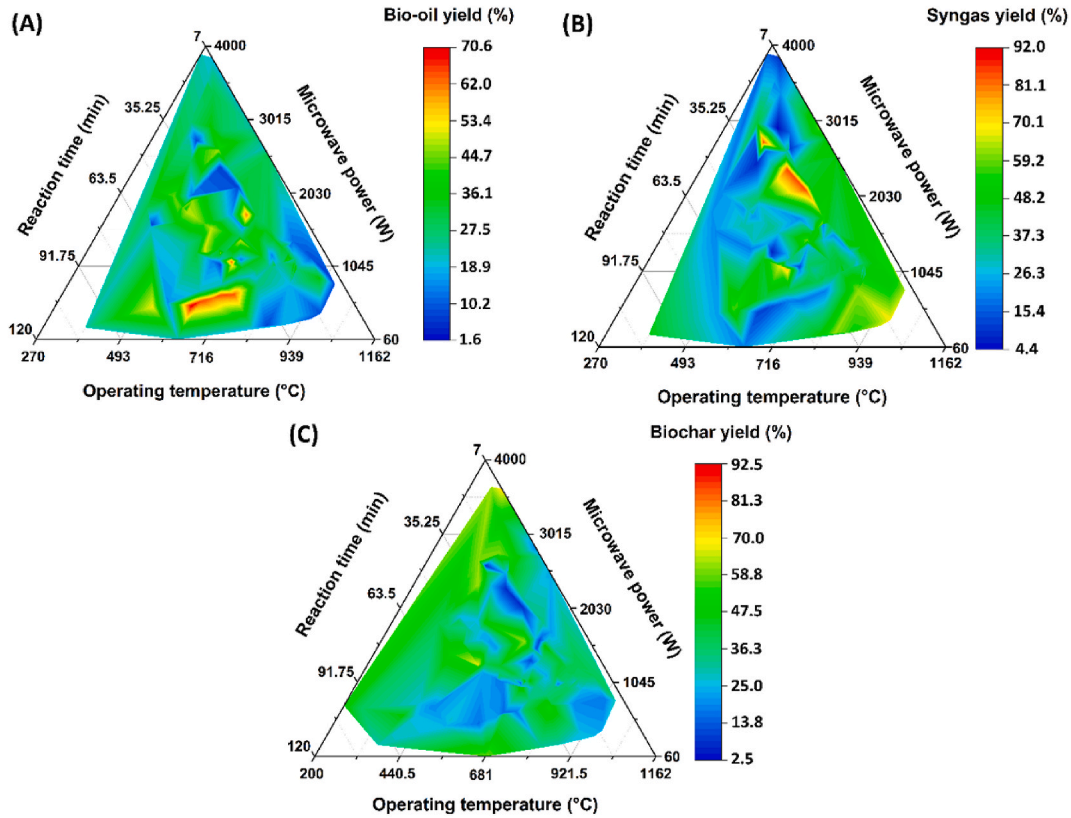

Fig. 4. Effect of operating temperature, microwave power, and reaction time on product distribution of the biomass microwave pyrolysis process: (A) bio-oil yield, (B) syngas yield, and (C) biochar yield.

Table 3
Statistical performance of the developed models in the training phase.

Parameter	Model	Bio-oil yield (%)	Syngas yield (%)	H_2 concentration (mol%)	CO concentration (mol%)	CO_2 concentration (mol%)	CH_4 concentration (mol%)	Biochar yield (%)	Biochar H/C ratio (−)	Biochar H/N ratio (−)	Biochar O/C ratio (−)	Biochar calorific value (MJ/kg)
R^2	SVR	0.542	0.311	0.809	0.778	0.595	0.769	0.490	0.930	0.966	0.998	0.972
	RFR	0.869	0.898	0.910	0.909	0.751	0.860	0.815	0.897	0.776	0.974	0.981
	GBR	0.917	0.928	0.999	0.949	0.998	0.939	0.927	0.998	0.862	0.989	0.991
RMSE	SVR	10.435	15.65	13.365	13.93	9.061	5.440	11.33	0.091	6.606	0.150	4.880
	RFR	5.435	6.806	4.998	5.272	6.612	2.831	6.637	0.111	13.454	0.159	1.551
	GBR	4.075	5.098	0.0011	3.882	0.451	1.704	4.144	8.0×10^{-4}	10.185	0.090	0.925
RRMSE	SVR	0.414	0.478	0.638	0.527	0.482	0.465	0.358	0.167	0.532	0.446	0.222
	RFR	0.182	0.183	0.188	0.179	0.356	0.205	0.206	0.201	0.665	0.240	0.075
	GBR	0.136	0.138	5.0×10^{-5}	0.132	0.025	0.123	0.129	0.001	0.477	0.155	0.043

Table 4
Statistical performance of the developed models in the testing phase.

Parameter	Model	Bio-oil yield (%)	Syngas yield (%)	H_2 concentration (mol%)	CO concentration (mol%)	CO_2 concentration (mol%)	CH_4 concentration (mol%)	Biochar yield (%)	Biochar H/C ratio (−)	Biochar H/N ratio (−)	Biochar O/C ratio (−)	Biochar calorific value (MJ/kg)
R^2	SVR	0.465	0.388	0.652	0.744	0.478	0.608	0.351	0.832	0.335	0.962	0.933
	RFR	0.863	0.893	0.895	0.907	0.742	0.830	0.775	0.882	0.756	0.960	0.981
	GBR	0.906	0.908	0.963	0.944	0.882	0.849	0.857	0.980	0.823	0.986	0.985
RMSE	SVR	12.77	17.769	13.5	15.13	10.83	6.135	13.45	0.176	25.41	0.216	6.379
	RFR	5.524	7.315	6.923	5.846	6.248	3.984	7.665	0.259	5.071	0.218	1.691
	GBR	4.287	6.177	3.462	5.118	3.827	3.943	4.714	0.056	12.381	0.132	1.340
RRMSE	SVR	0.406	0.456	0.494	0.509	0.589	0.418	0.423	0.309	1.139	0.319	0.302
	RFR	0.176	0.186	0.291	0.211	0.345	0.265	0.254	0.425	0.248	0.571	0.071
	GBR	0.139	0.154	0.114	0.187	0.204	0.266	0.156	0.100	0.766	0.186	0.066

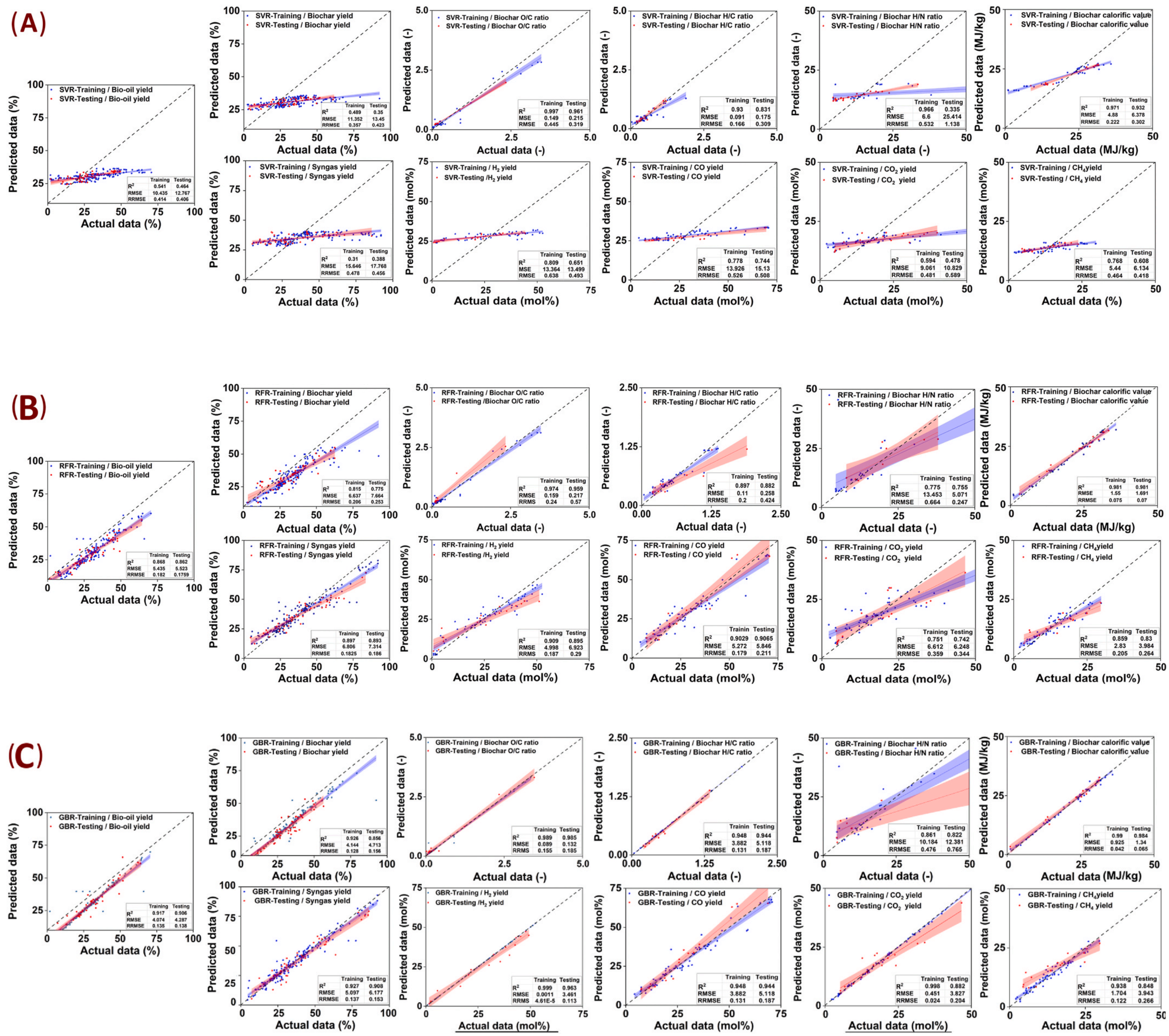

Fig. 5. Predicted values against experimental data: (A) SVR model, (B) RFR model, and (C) GBR model.

Table 5
Hyperparameters of the GBR model.

Hyperparameter	Bio-oil yield (%)	Syngas yield (%)	H_2 concentration (mol%)	CO concentration (mol%)	CO_2 concentration (mol%)	CH_4 concentration (mol%)	Biochar yield (%)	Biochar H/C ratio (−)	Biochar H/N ratio (−)	Biochar O/C ratio (−)	Biochar calorific value (MJ/kg)
learning_rate	0.02	0.04	0.04	0.02	0.02	0.04	0.04	0.04	0.02	0.01	0.04
max_depth	7	4	7	10	4	10	10	7	7	10	4
n_estimators	500	1000	1000	1000	1000	1000	100	500	1000	500	1000
subsample	0.5	0.1	0.5	0.1	0.5	0.1	0.5	0.5	0.1	0.5	0.1

blue to red by increasing the SHAP value means that the input descriptor has a positive effect on the output target and *vice versa*. The middle graphs in Fig. 6 illustrate the heatmap SHAP value plot, highlighting the general overview of how the corresponding output (i.e., f(x)) changes with the variations of input descriptors. The right-hand doughnut charts shown in Fig. 6 represent the overall importance of biomass characteristics (i.e., ultimate and proximate analyses) and operating conditions on each output.

As shown in Fig. 6A, the oxygen content of biomass and operating temperature were the top two most influential parameters on bio-oil yield in biomass microwave pyrolysis. Accordingly, a higher oxygen content value showed a higher negative SHAP value, negatively affecting bio-oil formation. The most dominant oxygen-containing functionalities in bio-oil produced *via* the pyrolysis process are hydroxyl groups, ketones/aldehydes, and carboxylic groups. It should be noted that higher oxygen contents of biomass feedstocks negatively

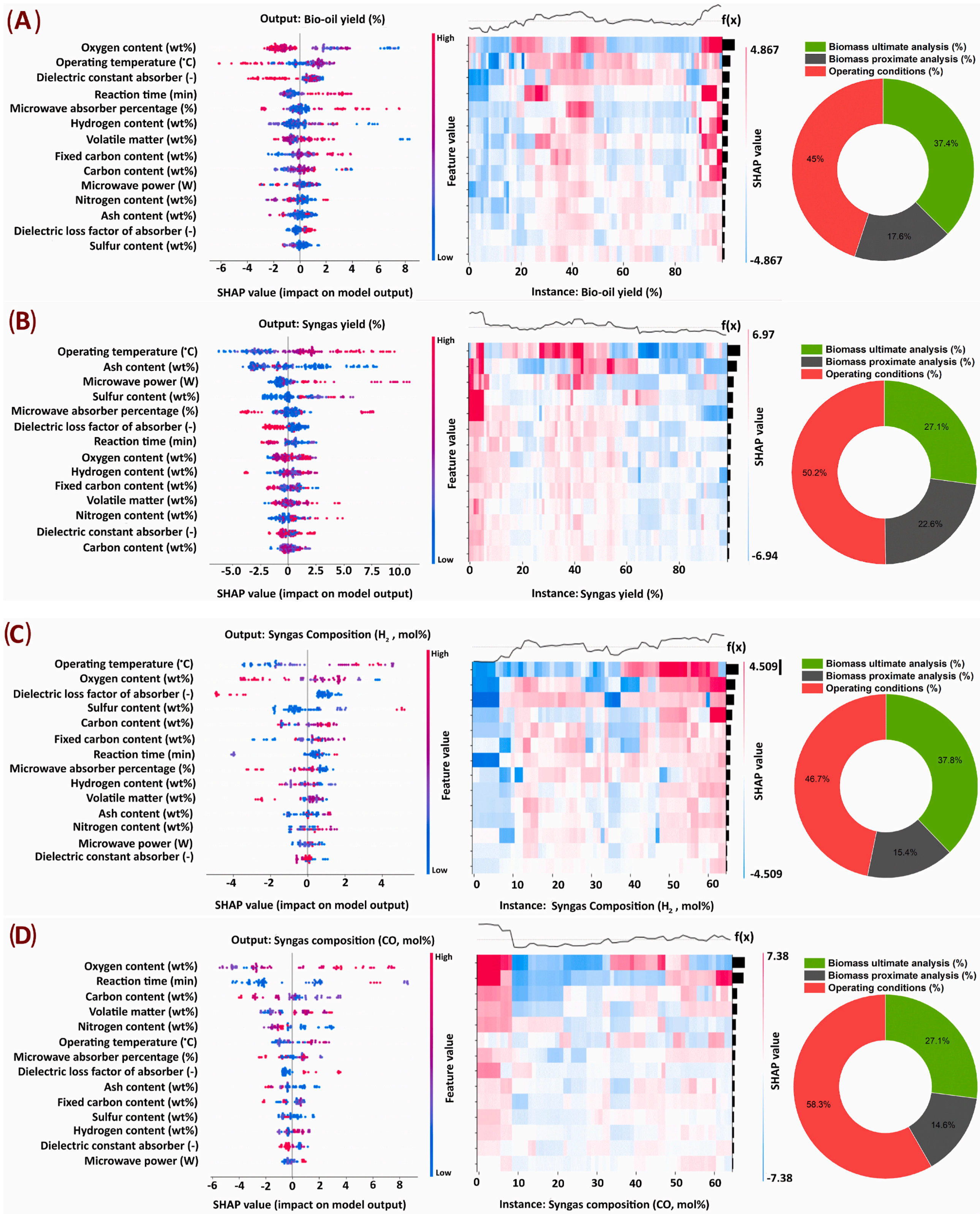

Fig. 6. The SHAP value representing the impacts of input features on each output in biomass microwave pyrolysis: (A) bio-oil yield, (B) syngas yield, (C) H_2 concentration, (D) CO concentration, (E) CO_2 concentration, (F) CH_4 concentration, (G) biochar yield, (H) biochar H/C ratio, (I) biochar H/N ratio, (J) biochar O/C ratio, and (K) biochar calorific value.

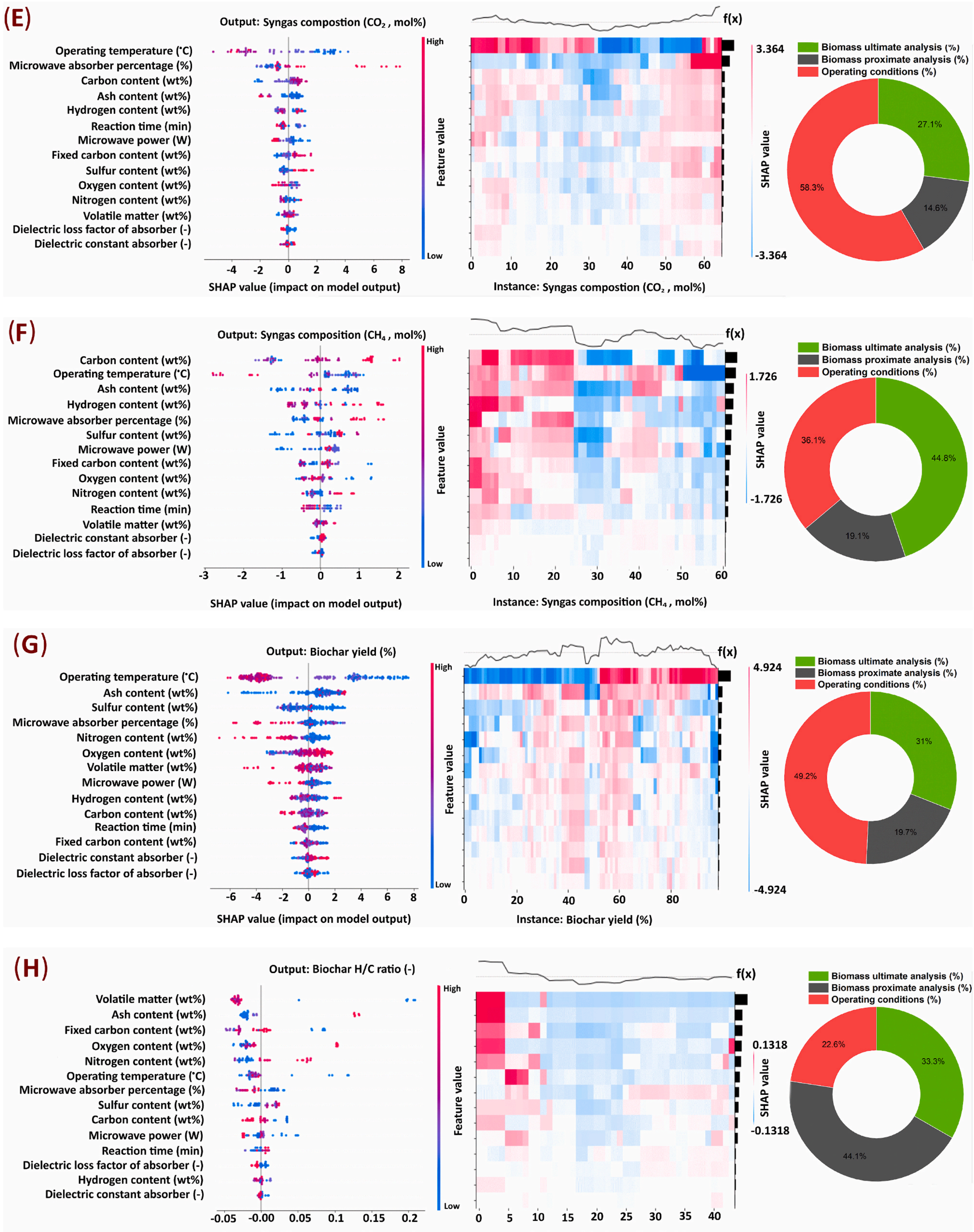

Fig. 6. (*continued*).

affect the quality of the resultant liquid by lowering its heating value and increasing acidity. The highly reactive oxygenated species also promote polymerization reactions during long-term storage [57]. Biomass conversion into a liquid fraction is generally enhanced by increasing the operating temperature to a certain value (e.g., 650–750 °C, as shown in Fig. 4). However, excessive temperature leads to the acceleration of

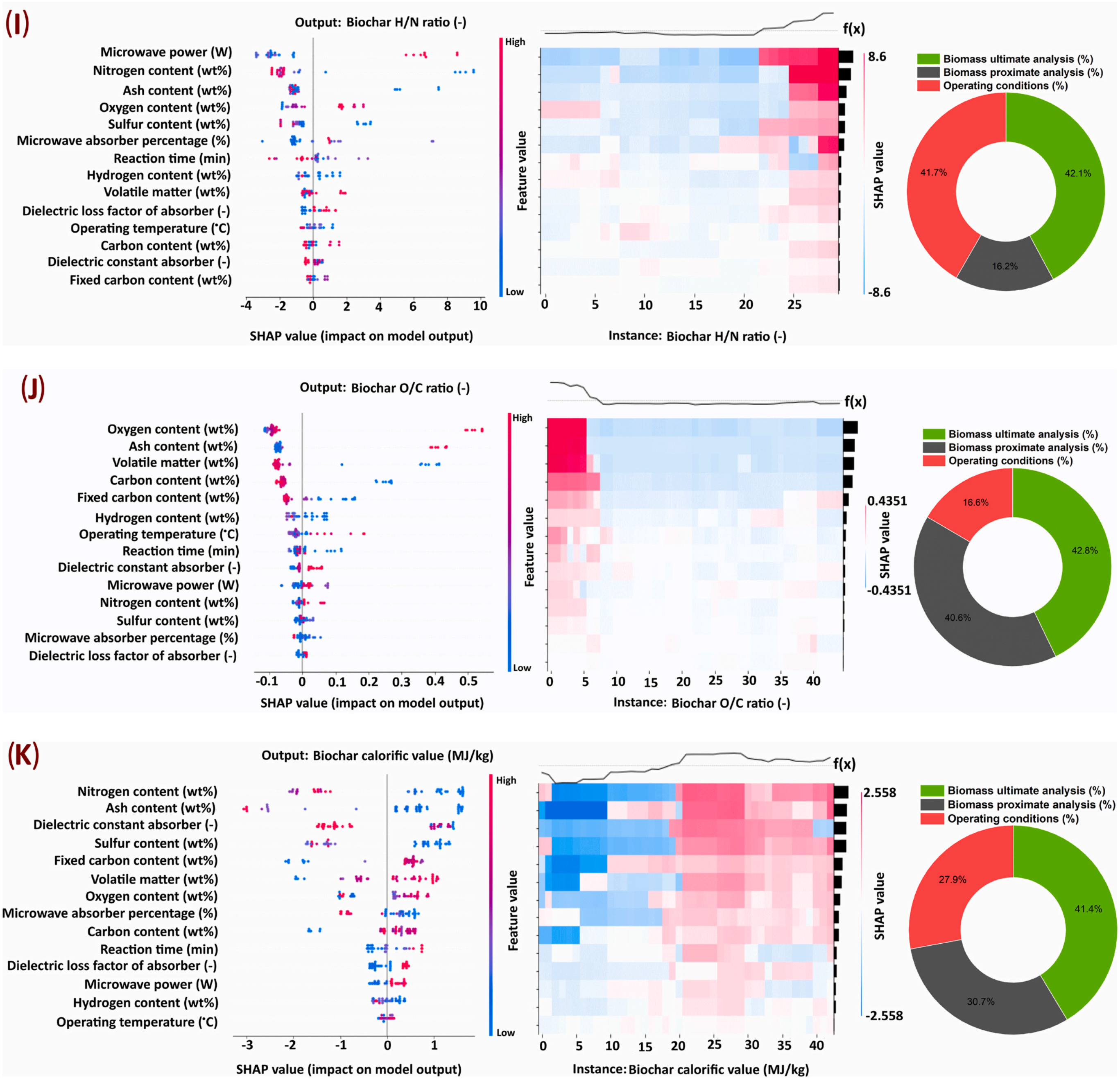

Fig. 6. (*continued*).

secondary cracking reactions, hence increasing the formation of gaseous products. It is worth noting that bio-oil quality also depends on the operating temperature. In this regard, the bio-oil obtained at higher operating temperatures includes short-chain functional groups (such as alcohol and ketone and their derivatives), whereas lower operating temperatures favor the formation of bio-oil with long-chain and high-molecular-weight compounds (such as aliphatic hydrocarbons, acids, esters, alcohols, and nitrogen-containing compounds) [58].

Fig. 6B-D depicts the SHAP values of input parameters on the quantity and quality of gaseous phase products in biomass microwave pyrolysis. Overall, operating conditions contributed more than biomass properties (i.e., ultimate and proximate analyses) in predicting gaseous phase products. In addition, the operating temperature was the top important feature among input variables based on SHAP analysis. Generally, lower temperatures lead to reduced syngas generation, whereas higher temperatures result in higher syngas yields. This issue can be attributed to the promotion of secondary cracking of primary pyrolyzates, leading to the generation of more gaseous products [58]. Like syngas yield, the operating temperature was the most significant parameter in predicting the concentration of H_2 and CO_2. More specifically, higher temperatures could lead to an increase in H_2 concentration in syngas while decreasing CO_2 concentration. This phenomenon could be linked to, on the one hand, the dominance of decarboxylation reactions at lower temperatures and, on the other hand, the magnified effects of dehydrogenation and the reverse Boudouard reactions on H_2 generation at higher temperatures [44].

The SHAP value of each input descriptor on the respective output target of solid residue in the biomass microwave pyrolysis process is illustrated in Fig. 6G-K. The operating temperature and biomass ash content were the two most important features for biochar yield prediction. Unlike the operating temperature, ash content could positively affect biochar yield. It should be noted that lower temperatures are more favorable for biochar production [43]. In addition, the operating temperature could significantly affect the aromaticity and aromatic condensation while determining the stability of biochar in the pyrolysis process [59]. Biomass ash content was positively related to biochar yield because of higher inorganic compounds in ash-rich feedstocks. Nevertheless, higher ash contents could decrease biochar stability, porosity, and calorific value in the biomass microwave pyrolysis process [60,61].

4. Concluding remarks and future directions

A data-driven framework was devised to model and understand biomass microwave pyrolysis. An inclusive dataset was compiled from the published literature with over 249 experimental data points (including biomass feedstocks, microwave absorbers, and reaction conditions) and analyzed using data mining tools. The data mining allowed us to understand the effects of input descriptors on product formation in biomass microwave pyrolysis. Three machine learning algorithms (SVR, RFR, and GBR) were devised on the complied dataset to predict the quantity and quality of bio-oil, syngas, and biochar produced in biomass microwave pyrolysis. The GBR resulted in the most accurate and generalized prediction performance, with an R^2 between 0.82 and 0.985, an RMSE lower than 12.381, and an RRMSE lower than 0.756. The SHAP analysis could successfully reveal the important descriptors contributing to the model prediction. In general, the operating temperature, microwave power, and reaction time contributed the most to the GBR model prediction. This finding was consistent with the domain knowledge in the published literature, showing the capability of the selected model to generalize the biomass microwave pyrolysis process reliably. Overall, the selected machine learning model and the developed computer program could save experimental costs and efforts in biomass microwave pyrolysis. This model could serve as an effective toolbox to assist researchers in their experimental workflow in finding optimal operating conditions with minimal cost and time requirements. The model could also facilitate techno-economic evaluation and environmental impact assessment of biomass microwave pyrolysis systems by future researchers and offer a baseline for further design improvements. In addition, the selected model could provide an opportunity for real-time monitoring and controlling of industrial microwave systems used in biomass pyrolysis.

Despite the promising results obtained herein, more accurate and reliable machine learning models can still be developed to characterize biomass microwave pyrolysis. In this study, machine learning models were developed based on a limited number of data patterns. Hence, future studies should try to develop a more comprehensive database based on a large number of data patterns. The ash details were not considered in developing machine learning models. It should be noted that ash components can act as microwave absorbers or reaction catalysts. Accordingly, more detailed information regarding the ash content of biomass can improve the accuracy and efficiency of machine learning models. The machine learning models developed herein are applicable only to the non-catalytic biomass microwave pyrolysis process. Therefore, future investigations should develop models applicable to both catalytic and non-catalytic biomass microwave pyrolysis processes. The reaction kinetics and product yields of biomass microwave pyrolysis can be effectively predicted based on major biomass constituents, i.e., cellulose, hemicellulose, and lignin. Using the major constituents of biomass feedstocks as descriptors in modeling biomass microwave pyrolysis is an interesting subject for future work. Correlating biomass microwave pyrolysis kinetics with product characteristics using machine learning can set the stage for commercializing biomass microwave pyrolysis. The high-quality bio-oil obtained during biomass microwave pyrolysis can be easily upgraded into motor fuel. Modeling the physicochemical properties and chemical constituents of bio-oil using machine learning can pave the way toward realizing biomass microwave pyrolysis. The carbon-rich and porous biochar derived from biomass microwave pyrolysis can be used in various energy and environmental applications. Using machine learning to predict morphological structures and functional groups of microwave-derived biochar can facilitate its real-world applications.

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.

Data availability

Data will be made available on request.

Acknowledgments

The authors gratefully acknowledge financial support of National Key R&D Program of China (No.2018YFF3503), Youth Talent Scholar of Chinese Academy of Agricultural Sciences, Fundamental Research Funds for Central Non-profit Scientific Institution (No. 1610132020003), Agricultural Science and Technology Innovation Project of Chinese Academy of Agricultural Sciences, Fund of Government purchase of services from Ministry of Agriculture and Rural Affairs (No. 13220198), Fund for talents from State Administration of Foreign Experts Affairs of P.R. China. M.T. would like to acknowledge the support from the Ministry of Higher Education, Malaysia, under the Higher Institution Centre of Excellence (HICoE), Institute of Tropical Aquaculture and Fisheries (AKUATROP) program (Vot. no. 56052, UMT/CRIM/2-2/5 Jilid 2 (11)), and the support by the Biofuel Research Team (BRTeam).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.renene.2022.11.028.

References

[1] J. Zhou, S. Liu, N. Zhou, L. Fan, Y. Zhang, P. Peng, E. Anderson, K. Ding, Y. Wang, Y. Liu, Development and application of a continuous fast microwave pyrolysis system for sewage sludge utilization, Bioresour. Technol. 256 (2018) 295–301.

[2] Y. Hong, W. Chen, X. Luo, C. Pang, E. Lester, T. Wu, Microwave-enhanced pyrolysis of macroalgae and microalgae for syngas production, Bioresour. Technol. 237 (2017) 47–56.

[3] C.R. Ellison, R. Hoff, C. Mărculescu, D. Boldor, Investigation of microwave-assisted pyrolysis of biomass with char in a rectangular waveguide applicator with built-in phase-shifting, Appl. Energy 259 (2020), 114217.

[4] H. Shahbeik, W. Peng, H. Kazemi Shariat Panahi, M. Dehhaghi, G.J. Guillemin, A. Fallahi, H. Amiri, M. Rehan, D. Raikwar, H. Latine, B. Pandalone, B. Khoshnevisan, C. Sonne, L. Vaccaro, A.-S. Nizami, V.K. Gupta, S.S. Lam, J. Pan, R. Luque, B. Sels, M. Tabatabaei, M. Aghbashlo, Synthesis of liquid biofuels from biomass by hydrothermal gasification: a critical review, Renew. Sustain. Energy Rev. 167 (2022), 112833, https://doi.org/10.1016/j.rser.2022.112833.

[5] Y. Tian, W. Zuo, Z. Ren, D. Chen, Estimation of a novel method to produce bio-oil from sewage sludge by microwave pyrolysis with the consideration of efficiency and safety, Bioresour. Technol. 102 (2011) 2053–2061.

[6] M. Tripathi, J.N. Sahu, P. Ganesan, Effect of process parameters on production of biochar from biomass waste through pyrolysis: a review, Renew. Sustain. Energy Rev. 55 (2016) 467–481, https://doi.org/10.1016/j.rser.2015.10.122.

[7] A.E. Atabani, I. Ali, S.R. Naqvi, I.A. Badruddin, M. Aslam, E. Mahmoud, F. Almomani, D. Juchelková, M.R. Atelge, T.M.Y. Khan, A state-of-the-art review on spent coffee ground (SCG) pyrolysis for future biorefinery, Chemosphere 286 (2022), 131730.

[8] Y. Cao, A. Pawłowski, Sewage sludge-to-energy approaches based on anaerobic digestion and pyrolysis: brief overview and energy efficiency assessment, Renew. Sustain. Energy Rev. 16 (2012) 1657–1665.

[9] D.R. Nhuchhen, M.T. Afzal, T. Dreise, A.A. Salema, Characteristics of biochar and bio-oil produced from wood pellets pyrolysis using a bench scale fixed bed, microwave reactor, Biomass Bioenergy 119 (2018) 293–303.

[10] A.A. Arpia, W.-H. Chen, S.S. Lam, P. Rousset, M.D.G. de Luna, Sustainable biofuel and bioenergy production from biomass waste residues using microwave-assisted heating: a comprehensive review, Chem. Eng. J. 403 (2021), 126233, https://doi.org/10.1016/j.cej.2020.126233.
[11] J. Luo, R. Ma, X. Huang, S. Sun, H. Wang, Bio-fuels generation and the heat conversion mechanisms in different microwave pyrolysis modes of sludge, Appl. Energy 266 (2020), 114855.
[12] T. Haeldermans, J. Claesen, J. Maggen, R. Carleer, J. Yperman, P. Adriaensens, P. Samyn, D. Vandamme, A. Cuypers, K. Vanreppelen, Microwave assisted and conventional pyrolysis of MDF–characterization of the produced biochars, J. Anal. Appl. Pyrol. 138 (2019) 218–230.
[13] F.C. Borges, Q. Xie, M. Min, L.A.R. Muniz, M. Farenzena, J.O. Trierweiler, P. Chen, R. Ruan, Fast microwave-assisted pyrolysis of microalgae using microwave absorbent and HZSM-5 catalyst, Bioresour. Technol. 166 (2014) 518–526.
[14] A. Yerrayya, D. V Suriapparao, U. Natarajan, R. Vinu, Selective production of phenols from lignin via microwave pyrolysis using different carbonaceous susceptors, Bioresour. Technol. 270 (2018) 519–528.
[15] M. 'Azim Jamaluddin, K. Ismail, M.A. Mohd Ishak, Z. Ab Ghani, M.F. Abdullah, M. T. Safian, S.S. Idris, S. Tahiruddin, M.F. Mohammed Yunus, N.I.N. Mohd Hakimi, Microwave-assisted pyrolysis of palm kernel shell: optimization using response surface methodology (RSM), Renew. Energy 55 (2013) 357–365, https://doi.org/10.1016/j.renene.2012.12.042.
[16] H. Shahbeik, S. Rafiee, A. Shafizadeh, D. Jeddi, T. Jafary, S.S. Lam, J. Pan, M. Tabatabaei, M. Aghbashlo, Characterizing sludge pyrolysis by machine learning: towards sustainable bioenergy production from wastes, Renew. Energy 199 (2022) 1078–1092, https://doi.org/10.1016/j.renene.2022.09.022.
[17] W.-H. Chen, R. Aniza, A.A. Arpia, H.-J. Lo, A.T. Hoang, V. Goodarzi, J. Gao, A comparative analysis of biomass torrefaction severity index prediction from machine learning, Appl. Energy 324 (2022), 119689, https://doi.org/10.1016/j.apenergy.2022.119689.
[18] S. Ge, Y. Shi, C. Xia, Z. Huang, M. Manzo, L. Cai, H. Ma, S. Zhang, J. Jiang, C. Sonne, S.S. Lam, Progress in pyrolysis conversion of waste into value-added liquid pyro-oil, with focus on heating source and machine learning analysis, Energy Convers. Manag. 245 (2021), 114638, https://doi.org/10.1016/j.enconman.2021.114638.
[19] Z. Huang, M. Manzo, C. Xia, L. Cai, Y. Zhang, Z. Liu, A.K. Nadda, Q. Van Le, C. Sonne, S.S. Lam, Effects of waste-based pyrolysis as heating source: meta-analyze of char yield and machine learning analysis, Fuel 318 (2022), 123578.
[20] S. Mari Selvam, P. Balasubramanian, Influence of biomass composition and microwave pyrolysis conditions on biochar yield and its properties: a machine learning approach, Bio. Energy Res. (2022) 1–13.
[21] R. Potnuri, D.V. Suriapparao, C. Sankar Rao, V. Sridevi, A. Kumar, Effect of dry torrefaction pretreatment of the microwave-assisted catalytic pyrolysis of biomass using the machine learning approach, Renew. Energy 197 (2022) 798–809, https://doi.org/10.1016/j.renene.2022.08.006.
[22] S.R. Narde, N. Remya, Biochar production from agricultural biomass through microwave-assisted pyrolysis: predictive modelling and experimental validation of biochar yield, Environ. Dev. Sustain. 24 (2022) 11089–11102, https://doi.org/10.1007/s10668-021-01898-9.
[23] A. Terapalli, D. Kamireddi, V. Sridevi, M. Tukarambai, D.V. Suriapparao, C.S. Rao, R. Gautam, P.R. Modi, Microwave-assisted in-situ catalytic pyrolysis of polystyrene: analysis of product formation and energy consumption using machine learning approach, Process Saf. Environ. Protect. 166 (2022) 57–67, https://doi.org/10.1016/j.psep.2022.08.016.
[24] R. Shishegar, T. Cox, D. Rolls, P. Bourgeat, V. Doré, F. Lamb, J. Robertson, S. M. Laws, T. Porter, J. Fripp, D. Tosun, P. Maruff, G. Savage, C.C. Rowe, C. L. Masters, M.W. Weiner, V.L. Villemagne, S.C. Burnham, Using imputation to provide harmonized longitudinal measures of cognition across AIBL and ADNI, Sci. Rep. 11 (2021), 23788, https://doi.org/10.1038/s41598-021-02827-6.
[25] B. Ramosaj, M. Pauly, Predicting missing values: a comparative study on non-parametric approaches for imputation, Comput. Stat. 34 (2019) 1741–1764, https://doi.org/10.1007/s00180-019-00900-3.
[26] H. Alkabbani, A. Ramadan, Q. Zhu, A. Elkamel, An improved air quality index machine learning-based forecasting with multivariate data imputation approach, Atmosphere 13 (2022) 1144, https://doi.org/10.3390/atmos13071144.
[27] Y. Jung, J. Hu, A K -fold averaging cross-validation procedure, J. Nonparametric Statistics 27 (2015) 167–179, https://doi.org/10.1080/10485252.2015.1010532.
[28] F. Elmaz, Ö. Yücel, A.Y. Mutlu, Predictive modeling of biomass gasification with machine learning-based regression methods, Energy 191 (2020), 116541, https://doi.org/10.1016/j.energy.2019.116541.
[29] W.S. Noble, What is a support vector machine? Nat. Biotechnol. 24 (2006) 1565–1567, https://doi.org/10.1038/nbt1206-1565.
[30] D.A. Pisner, D.M. Schnyer, Support vector machine, in: Machine Learning, Elsevier, 2020, pp. 101–121, https://doi.org/10.1016/B978-0-12-815739-8.00006-7.
[31] S. Amari, S. Wu, Improving support vector machine classifiers by modifying kernel functions, Neural Network. 12 (1999) 783–789, https://doi.org/10.1016/S0893-6080(99)00032-5.
[32] C. Tian, L. Cheng, T. Yin, Impacts of anthropogenic and biophysical factors on ecological land using logistic regression and random forest: a case study in Mentougou District, Beijing, China, J. Mt. Sci. 19 (2022) 433–445, https://doi.org/10.1007/s11629-021-7022-x.
[33] L. Wang, X. Zhou, X. Zhu, Z. Dong, W. Guo, Estimation of biomass in wheat using random forest regression algorithm and remote sensing data, The Crop J. 4 (2016) 212–219, https://doi.org/10.1016/j.cj.2016.01.008.
[34] M. Liu, X. Liu, J. Li, C. Ding, J. Jiang, Evaluating total inorganic nitrogen in coastal waters through fusion of multi-temporal RADARSAT-2 and optical imagery using random forest algorithm, Int. J. Appl. Earth Obs. Geoinf. 33 (2014) 192–202, https://doi.org/10.1016/j.jag.2014.05.009.
[35] P. Nie, M. Roccotelli, M.P. Fanti, Z. Ming, Z. Li, Prediction of home energy consumption based on Gradient boosting regression tree, Energy Rep. 7 (2021) 1246–1255, https://doi.org/10.1016/j.egyr.2021.02.006.
[36] M. Rzychoń, A. Żogała, L. Róg, Experimental study and extreme Gradient boosting (XGBoost) based prediction of caking ability of coal blends, J. Anal. Appl. Pyrol. 156 (2021), 105020, https://doi.org/10.1016/j.jaap.2021.105020.
[37] F. Ghafarian, R. Wieland, D. Lüttschwager, C. Nendel, Application of extreme Gradient boosting and Shapley Additive explanations to predict temperature regimes inside forests from standard open-field meteorological data, Environ. Model. Software 156 (2022), 105466, https://doi.org/10.1016/j.envsoft.2022.105466.
[38] D. Wang, S. Thunéll, U. Lindberg, L. Jiang, J. Trygg, M. Tysklind, Towards better process management in wastewater treatment plants: process analytics based on SHAP values for tree-based machine learning methods, J. Environ. Manag. 301 (2022), 113941, https://doi.org/10.1016/j.jenvman.2021.113941.
[39] T. Kavzoglu, A. Teke, Predictive performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBoost), Arabian J. Sci. Eng. 47 (2022) 7367–7385, https://doi.org/10.1007/s13369-022-06560-8.
[40] S. Lundberg, An Introduction to Explainable AI with Shapley Values, 2018.
[41] W. Zhang, J. Li, T. Liu, S. Leng, L. Yang, H. Peng, S. Jiang, W. Zhou, L. Leng, H. Li, Machine learning prediction and optimization of bio-oil production from hydrothermal liquefaction of algae, Bioresour. Technol. 342 (2021), 126011, https://doi.org/10.1016/j.biortech.2021.126011.
[42] Y. Zhang, P. Chen, S. Liu, P. Peng, M. Min, Y. Cheng, E. Anderson, N. Zhou, L. Fan, C. Liu, G. Chen, Y. Liu, H. Lei, B. Li, R. Ruan, Effects of feedstock characteristics on microwave-assisted pyrolysis – a review, Bioresour. Technol. 230 (2017) 143–151, https://doi.org/10.1016/j.biortech.2017.01.046.
[43] Y. Zhang, Y. Cui, S. Liu, L. Fan, N. Zhou, P. Peng, Y. Wang, F. Guo, M. Min, Y. Cheng, Y. Liu, H. Lei, P. Chen, B. Li, R. Ruan, Fast microwave-assisted pyrolysis of wastes for biofuels production – a review, Bioresour. Technol. 297 (2020), 122480, https://doi.org/10.1016/j.biortech.2019.122480.
[44] A. Zaker, Z. Chen, X. Wang, Q. Zhang, Microwave-assisted pyrolysis of sewage sludge: a review, Fuel Process. Technol. 187 (2019) 84–104, https://doi.org/10.1016/j.fuproc.2018.12.011.
[45] J. Li, L. Pan, M. Suvarna, Y.W. Tong, X. Wang, Fuel properties of hydrochar and pyrochar: prediction and exploration with machine learning, Appl. Energy 269 (2020), 115166, https://doi.org/10.1016/j.apenergy.2020.115166.
[46] X. Zhang, K. Rajagopalan, H. Lei, R. Ruan, B.K. Sharma, An overview of a novel concept in biomass pyrolysis: microwave irradiation, Sustain. Energy Fuels 1 (2017) 1664–1699, https://doi.org/10.1039/C7SE00254H.
[47] J. George, P. Arun, C. Muraleedharan, Region-specific biomass feedstock selection for gasification using multi-attribute decision-making techniques, Int. J. Sustain. Eng. 14 (2021) 1101–1109, https://doi.org/10.1080/19397038.2020.1790058.
[48] S. Mutsengerere, C.H. Chihobo, D. Musademba, I. Nhapi, A review of operating parameters affecting bio-oil yield in microwave pyrolysis of lignocellulosic biomass, Renew. Sustain. Energy Rev. 104 (2019) 328–336, https://doi.org/10.1016/j.rser.2019.01.030.
[49] H. Li, J. Xu, S. Mbugua Nyambura, J. Wang, C. Li, X. Zhu, X. Feng, Y. Wang, Food waste pyrolysis by traditional heating and microwave heating: a review, Fuel 324 (2022), 124574, https://doi.org/10.1016/j.fuel.2022.124574.
[50] M. Aghbashlo, W. Peng, M. Tabatabaei, S.A. Kalogirou, S. Soltanian, H. Hosseinzadeh-Bandbafha, O. Mahian, S.S. Lam, Machine learning technology in biodiesel research: a review, Prog. Energy Combust. Sci. 85 (2021), 100904, https://doi.org/10.1016/j.pecs.2021.100904.
[51] J.-I. Prieto, D. García, Power correlations as complementary tools for generalised analysis and preliminary design of Stirling engines, Alex. Eng. J.. 63 (2023) 233–252. https://doi.org/10.1016/j.aej.2022.07.061.
[52] W. Cai, R. Wei, L. Xu, X. Ding, A method for modelling greenhouse temperature using gradient boost decision tree, Inf. Process. Agric. 9 (2022) 343–354, https://doi.org/10.1016/j.inpa.2021.08.004.
[53] A.S. Dyer, D. Zaengle, J.R. Nelson, R. Duran, M. Wenzlick, P.C. Wingo, J.R. Bauer, K. Rose, L. Romeo, Applied machine learning model comparison: predicting offshore platform integrity with Gradient boosting algorithms and neural networks, Mar. Struct. 83 (2022), 103152, https://doi.org/10.1016/j.marstruc.2021.103152.
[54] I. Baturynska, K. Martinsen, Prediction of geometry deviations in additive manufactured parts: comparison of linear regression with machine learning algorithms, J. Intell. Manuf. 32 (2021) 179–200, https://doi.org/10.1007/s10845-020-01567-0.
[55] L. Yan, S. Zhong, T. Igou, H. Gao, J. Li, Y. Chen, Development of machine learning models to enhance element-doped g-C3N4 photocatalyst for hydrogen production through splitting water, Int. J. Hydrogen Energy (2022), https://doi.org/10.1016/j.ijhydene.2022.08.013.
[56] T. Onsree, N. Tippayawong, S. Phithakkitnukoon, J. Lauterbach, Interpretable machine-learning model with a collaborative game approach to predict yields and higher heating value of torrefied biomass, Energy 249 (2022), 123676, https://doi.org/10.1016/j.energy.2022.123676.
[57] M. Gholizadeh, X. Hu, Q. Liu, A mini review of the specialties of the bio-oils produced from pyrolysis of 20 different biomasses, Renew. Sustain. Energy Rev. 114 (2019), 109313, https://doi.org/10.1016/j.rser.2019.109313.
[58] D.V. Suriapparao, R. Tejasvi, A review on role of process parameters on pyrolysis of biomass and plastics: present scope and future opportunities in conventional and microwave-assisted pyrolysis technologies, Process Saf. Environ. Protect. 162 (2022) 435–462, https://doi.org/10.1016/j.psep.2022.04.024.

[59] L. Leng, H. Huang, An overview of the effect of pyrolysis process parameters on biochar stability, Bioresour. Technol. 270 (2018) 627–642, https://doi.org/10.1016/j.biortech.2018.09.030.
[60] Y. Li, B. Xing, Y. Ding, X. Han, S. Wang, A critical review of the production and advanced utilization of biochar via selective pyrolysis of lignocellulosic biomass, Bioresour. Technol. 312 (2020), 123614, https://doi.org/10.1016/j.biortech.2020.123614.
[61] V. Hadiya, K. Popat, S. Vyas, S. Varjani, M. Vithanage, V. Kumar Gupta, A. Núñez Delgado, Y. Zhou, P. Loke Show, M. Bilal, Z. Zhang, M. Sillanpää, S. Sabyasachi Mohanty, Z. Patel, Biochar production with amelioration of microwave-assisted pyrolysis: current scenario, drawbacks and perspectives, Bioresour. Technol. 355 (2022), 127303, https://doi.org/10.1016/j.biortech.2022.127303.