Algal Research 50 (2020) 102006

Contents lists available at ScienceDirect

Algal Research

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

Predicting algal biochar yield using eXtreme Gradient Boosting (XGB) algorithm of machine learning methods

Abhijeet Pathy[a,1], Saswat Meher[b,1], Balasubramanian P[a,*]

[a] Agricultural & Environmental Biotechnology Group, Department of Biotechnology & Medical Engineering, National Institute of Technology Rourkela, Odisha 769008, India

[b] Department of Computer Science Engineering, National Institute of Technology Rourkela, Odisha 769008, India

ARTICLE INFO

Keywords:
Algal biochar
Machine learning
Biochar yield
Biochar composition
Slow pyrolysis
eXtreme Gradient Boosting

ABSTRACT

Pyrolysis is a thermochemical pathway widely used for the conversion of biomass into useful products such as biochar, bio-oil, and syngases. A recent surge in the adoption of the pyrolysis process at realtime scenarios for the appropriate management and conversion of residues demands the modeling of the pyrolysis process. Prediction of algal biochar yield along with its composition was attempted in this study with the eXtreme Gradient Boosting (XGB) machine learning method. An extensive grid search method has been implemented in the XGB model to explore all the possible considered input parameter combinations for predicting the biochar yield. Thirteen different pyrolytically important input parameter combinations have been attempted and compared with the combination suggested by the feature selection technique of model for predicting the biochar yield. This feature selection technique highlights the H/C, N/C, ash content, pyrolysis temperature, and time as the key parameters on deciding the algal biochar yield, where H, C, N are hydrogen, carbon and nitrogen content of biomass. The highest regression coefficient (R^2) of 0.84 has been achieved between experimental and model predictive biochar yield for the testing dataset, once the model was trained with the training dataset. Pearson correlation coefficient matrix unraveled the correlation among and in between input parameters and biochar yield. Feature Importance Plots revealed temperature as the most influential factor. SHapley Additive exPlanations (SHAP) Dependence Plots depicted the interactive effect of temperature and other input parameters on the algal biochar yield. Summary Plots showed the combined features of importance through feature and SHAP values. The developed XGB model provides new insights on comprehending the influence of input parameters on predicting the algal biochar yield.

1. Introduction

Algae due to its swift cultivation under a wide range of environmental conditions have shown a surge in global production for numerous end products. Approximately 9200 tons (dry basis) of microalgal biomass are being harvested annually from wild habitats and aquaculture farms in the world 38. Algal cultivation plays an active role in the process of carbon sequestration and the higher efficiency of photosynthesis as compared to lignocellulosic biomasses, it plays an active role in the minimization of greenhouse gaseous emission as well as acting as a potential renewable energy source for industries [1]. Conversion of algal biomass produces essential bioproducts such as pigments and fatty acids, which find their active application in the pharmaceutical and nutrition-based industry [39]. Despite the production of valuable materials from algal feedstocks, a large amount of post-extracted algal residue generation creates challenges in waste management [2].

Conversion of algal biomass and its residues can be performed either by thermochemical or biochemical techniques based on the targeted outputs and intention for the conversion. In recent years, thermochemical based conversion techniques such as pyrolysis, hydrothermal liquefaction, and carbonization increasingly used for the conversion of algal biomass [3]. Pyrolysis is one of the techniques under thermochemical conversion that can be defined as the thermal degradation of biomass at a general temperature range of 300 °C to 700 °C in an oxygen-less environment [4]. The result of the pyrolysis can be classified into solid char, liquid, and gaseous fuel according to the states of the products [5]. An important by-product derived is biochar that has been extensively researched for its wider application in agricultural and environmental fields. Biochar is a carbon-rich solid product of the

* Corresponding author.
E-mail address: biobala@nitrkl.ac.in (B. P).
[1] Equal contributions.

https://doi.org/10.1016/j.algal.2020.102006
Received 14 April 2020; Received in revised form 6 July 2020; Accepted 8 July 2020
Available online 15 July 2020

pyrolysis process, which is fascinating the researchers in recent years due to its environmental and economic values [6]. A value-based product such as biochar produced through pyrolysis from waste residues can be economically feasible for algal industries. Conversion of algae into biochar is a carbon-negative process that sequesters carbon up to 87% and the production of biochar from algae helps in long term storage of carbon [7]. Due to their sustainable and renewable nature, algal biomass can be a potential source for the production of propitious biochar [8].

Biochar physical and chemical characteristics largely depend on the type of biomass and its pyrolytic conditions [9]. Algal biochar differs from lignocellulose biochar in terms of both physical and chemical characteristics. For instance, biochar derived from algal sizes ranges from 10 to 100 μm with an irregular porosity of 1 μm [40]. Algal biochar has a relatively lower surface area and carbon content compared to lignocellulosic biochar, but have higher pH and cation exchange capacity than later [10]. The high nutrient value (mainly due to nitrogen and ash content) of algal biochar is vital for soil amendment while the high pH content of algal char makes its application more desirable in acidic soil [2,11]. The presence of numerous functional groups on the surface of algal biochar makes it an efficient biosorbent in treating wastewater [12]. Algal biochar catalytic properties have been utilized to produce hydrogen in a membrane reactor [13]. Microalgae biochar has a high amount of nitrogen content and other mineral elements such as magnesium, iron potassium, sodium, and calcium compared to biochar produced from lignocellulosic biomass [8]. These unique features differentiate the conventional biochar produced from lignocellulosic biomasses and enhance its applicability range to a wider scale. Due to its various environmental applications, the conversion of algal char from the biomass could assist the sustainability of algae-based industries.

Pyrolysis is a complex process involving several secondary reactions such as cracking, decomposition, compression, and polymerization [14]. Thus, its product yield and quality are mainly influenced by the type of biomass (its elementary and biochemical composition, size) or by experimental conditions (pyrolysis temperature, residence time, and heating rate) [15]. These parameters affect the biochar output both quantitatively and qualitatively. Therefore, modeling of pyrolysis that can give us the prior information about the biochar yield would serve as an important purpose in increasing the economic viability of the algae-based industries by assisting in scaling up the production process. There have been several models already developed for the pyrolysis process based upon the mechanism, product distribution, and process kinetics concerning biomass and experimental characteristics. The pioneered attempts in pyrolysis modeling result in the development of models such as the one-step global model, one-stage multi reaction model, two-stage semi-global model, economic trade-off model, and pyrolytic kinetic model [16]. However, the majority of these models are developed for the lignocellulosic biomasses, only very few models were reported for the algal pyrolysis. Algae involve complex mechanisms during its thermal degradation, which makes its modeling a little difficult [4]. As discussed above, algal biomasses differ from the lignocellulosic biomass significantly. Hence the algal biomass needs a unique model for the accurate prediction of pyrolytic yield. To deal with the complex mechanisms, one of the most sought procedure are machine learning algorithms. Due to its user-friendly and efficient algorithm, machine learning algorithms are gaining huge attention in the majority of the existing field. Zhu et al. [17] used machine learning for predicting the biochar yield and its carbon composition for lignocellulosic biomasses. Some of the recent relevant reported literature on the application of machine learning in pyrolysis modeling are for seaweed [18], cattle manure [19], and application of neuro-fuzzy inference system in predicting the biochar yield [20].

To the best of author's knowledge, no previous work has been published on utilizing machine learning methods for predicting the algal biochar yield. The traditional data-driven models and multiple regression analysis might be a time-consuming and energy-intensive process. In this manuscript, a model has been developed by using a machine-learning algorithm to predict the algal biochar yield and its elemental composition by considering the ultimate analysis of algal biomass (Carbon (C), Hydrogen (H), Oxygen (O), Nitrogen (N), H/C, O/C and N/C), proximate analysis (ash, fixed carbon, and volatile compound) and pyrolytic experimental conditions (pyrolysis temperature, heating rate and pyrolysis time) as inputs. Thus, the first objective was to develop a robust model with better accuracy by using the least possible input parameters for predicting algal biochar yield. Then the second objective was to investigate the individual and combined effect of all input parameters on the yield of algal biochar. The model has been implemented using eXtreme Gradient Boosting (XGB) algorithms, which is a tree-based model to enhance its accuracy and capability by including several objective functions like classification, regression, and ranking. The detailed features of the XGB model and its algorithm can be referred from Nielsen [21] and Chen et al. [3]. Further, the fine-tuning and cross-validation techniques were also presented in the manuscript.

2. Materials and methods

2.1. Development of the algal databases

The data were collected from various published literature to develop a database that constitutes 91 datasets [Refer Supplementary Material – Table S1] for developing and testing the machine learning prediction model. The data were acquired from tables, diagrams, and supplementary materials of the literature. The input parameters under considerations were elementary composition (C, H, O, and N), ratios of elementary compositions (H/C, O/C and N/C), proximate analysis of algal biomass (ash, fixed carbon, and volatile compound) and pyrolysis condition (pyrolysis temperature, heating rate and residence time). The output of the pyrolysis process such as biochar yield and its elementary composition was collected on a dry ash-free basis.

All the collected data on algal biomasses have the desired input and output parameters; data having incomplete output parameters have been not used while developing or testing of the models. The algal biomasses can be divided into different categories like microalgae (59), macroalgae (19), lipid extracted algae (defatted algae) (5), and algal residues (8). The above-mentioned categories (the no of corresponding data sets have been given in the brackets) were used for the development and testing of these models. The literature data were considered into the database before utilizing further for model development, only after ensuring that whether the reported biomass was dried and their ultimate (C, H, O, and N) and proximate (ash, fixed carbon, and volatile compound) analysis were performed before the pyrolysis process. The pyrolysis experimental conditions such as pyrolysis temperature (highest heating temperature), heating rate, and residence time were also noted down. The output of the pyrolysis process such as biochar yield and its elementary composition was verified whether reported on a dry ash-free basis. The input parameters under considerations were elementary composition (C, H, O, and N), ratios of elementary compositions (H/C, O/C and N/C), proximate analysis of algal biomass (ash, fixed carbon, and volatile compound) and pyrolysis condition (pyrolysis temperature, heating rate and residence time). The statistical analysis of the above data includes the mean, quartile value along with upper and lower values for all input parameters (Refer Supplementary Material – Fig. S1).

2.2. Development of predictive models

The overall process of an XGB predictive model development along with *k*-fold cross-validation has been briefly outlined in Fig. 1. The whole dataset has been classified as a training and testing block in 75:25 ratios. To keep the model away from underfitting or overfitting

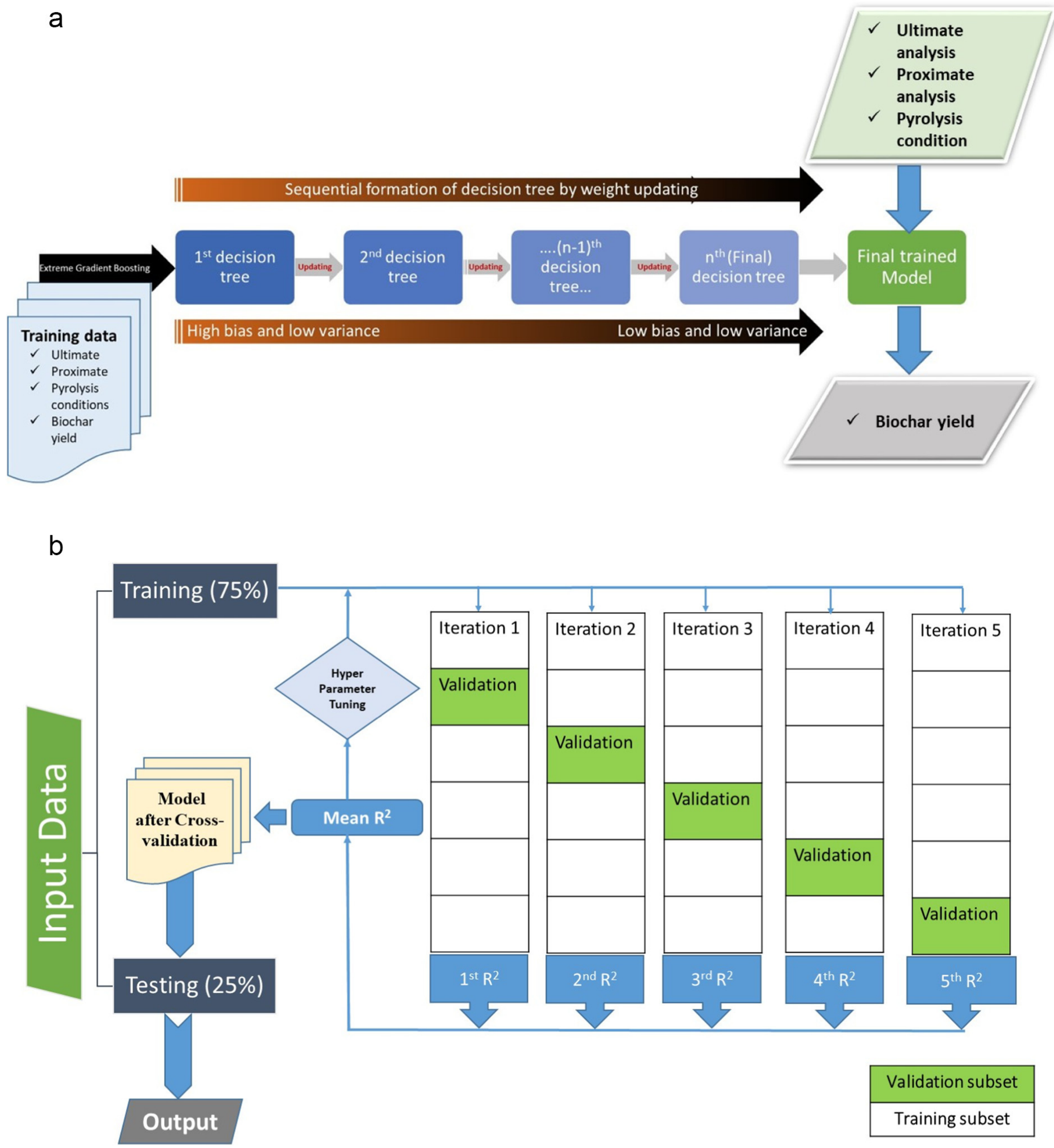

Fig. 1. a) Working perception of eXtreme Gradient Boosting (XGB) for the algal biochar prediction model.
b) *k* cross-validation of eXtreme Gradient Boosting (XGB) algorithm for the algal biochar prediction model.

with the data, minimal bias and variance have been taken into consideration. Bias is estimated by the error in predicting the training data while variance is determined by the error in predicting the testing data or in this case, the cross-validation data. While training the model, the decision tree was formed sequentially by weight up-gradation of the data based on the previous weak model's output. The biases present in the weaker model were reduced by the subsequent formation of the decision tree by weight up-gradation and thus results in the development of the final model with minimum possible biases.

To reduce the variance and prevents the model from overfitting, distinct hyperparameters and regularisation terms have been used in this XGB model. The model is prone to overfitting when it involves a high number of features, so the use of the feature selection method will help the model to further reduce the variance. Once the model was trained, it is being cross-validated and tested further. The developed model has been subjected to *k*-fold cross-validation; the training data is divided into *k* subsets, where (*k-1*) subsets are used for training the model and the leftover subset was used for the model validation. This procedure runs for *k* iteration, where each iteration chooses a different subset for validation and its compliment subsets for training the model. Due to the lesser availability of datasets, the implemented *k*-fold cross-validation minimizes the chances of the biased result produced from the model. The mean of R^2 values obtained from each validation subset is used for the model evaluation, which helps in tuning the model hyperparameters to determine the tree properties. Hyperparameters used in the tuning process are *learning_rate* (the rate at which the model upgrades its weights), *max_depth* (maximum depth of the tree), *sub_sample* (the ratio of number of samples from total training samples used in each update), *colsample_bytree* (the ratio of numbers of features selected from total features for each updation) and *reg_alpha* (regularisation parameter to prevent overfit).

The process used for the hyperparameters tuning was grid search cross validation, where a set of possible values for each hyperparameter was provided. In this case, the set of hyperparameters considered were {'*learning_rate*': [0.01, 0.03, 0.1, 0.3], '*max_depth*': [3, 4, 5], '*subsample*': [0.9, 1], '*colsample_bytree*': [0.8, 0.7, 0.6], '*reg_alpha*':[0, 1, 5] and '*reg_lambda*': [0, 1, 5]}. The model was trained with each possible combinations of hyperparameters with 5-fold cross-validated and the combinations of hyperparameters that builds a model with highest cross-validation score is selected by default.

2.3. Feature selection by eXtreme Gradient Boosting (XGB) model

The prediction efficiency of algal biochar yield by the developed model also depends on choosing the appropriate combination of input parameters (*C, H, O, N, H/C, O/C, N/C, ash content, fixed carbon, volatile compound, pyrolysis temperature, heating rate and the pyrolysis time*) with model hyperparameters (*learning_rate, max_depth, sub_sample, colsample_bytree and reg_alpha*). The combinations that are generally considered to be pyrolytically important such as ultimate analysis, proximate analysis, and pyrolytic conditions were listed from **Run 1 to 13.** This is intended for evaluating the individual and combined effect of the above-mentioned parameters related to biomass and pyrolytic conditions. For all the combinations of input parameters, the XGB model was configured using a grid search cross-validation method. Grid search cross-validation method' applies a 'randomized variation', the grid search method was executed by adjusting and optimizing the five model hyperparameters such as learning rate, maximum depth of the tree, sub-sample, column samples, and the regularisation. Cross-validation increases the reliability of the model, making it more robust and inclusion of regularisation features like *reg_alpha*, *sub_sample*, *colsample_bytree* avoids the overfit and ensured the model stability.

The best combination of input parameters was selected by performing an exhaustive search using the machine learning algorithm (**Run 14**), where every possible combination (2^{13}) of input features were used to train the model with each time performing grid search cross-validation (5-fold) on hyperparameters to achieve the best model for each input parameters. With the implementation of an exhaustive search method in combination with grid search cross-validation, the best possible input parameters combination for obtaining the biochar yield was found out to be (*'H/C', 'N/C', 'Ash', 'Pyrolysis Time', 'Temperature'*). The model hyperparameters were further configured by using a wider set of hyperparameters. The best possible hyperparameters combination was determined as follows: {*'colsample_bytree': 0.8, 'learning_rate': 0.3, 'max_depth': 3, 'reg_alpha': 0, 'reg_lambda': 5, 'subsample': 1*} for the developed model.

2.4. Performance evaluation of the predictive models

The overall design of the above experiments as outlined in Table 1. The above analysis was carried out for a random state having a uniform distribution of output used for testing purposes. The correlation coefficient (R^2) is used as a statistical measure for evaluating the model. In general, R^2 reveals the strength of the relationship between dependent and independent variables. However, R^2 indicates the similarity between the experimental and predictive biochar yield in this case. The performance of the XGB machine learning model is evaluated based on R^2 as mentioned in Eq. (1):

$$R^2 = 1 - \left(\frac{\sum_{i=1}^{N} (Y_i^{exp} - Y_i^{pred})^2}{\sum_{i=1}^{N} (Y_i^{exp} - Y_{ave}^{-exp})^2} \right) \tag{1}$$

where, Y_i^{exp} and Y_i^{pred} were experimental and predicted values with $Y_{ave}{}^{-exp}$ as the average of the experimental values. The relationship between input and output features was discussed with the help of Importance Plots, SHAP Dependence Plots, and Summary Plots.

3. Results and discussion

3.1. Statistical analysis of model inputs

Out of 91 datasets collected from the exhaustive literature survey for the implementation of a machine learning model to predict algal biochar yield, six datasets (outliers) that have extreme values could affect the accuracy of the models. The residence time and heating rate values of the slow pyrolysis process, which were not available in the literature further deducts to 64 complete datasets. The collected data comprises three main categories namely ultimate analysis (C, H, O, and N), proximate analysis (ash content, fixed carbon, and volatile compound) and pyrolysis conditions (heating temperature, heating rate, and residence time). Apart from these ten input parameters, three more parameters were obtained as a molar ratio of the elementary composition [such as H/C, O/C, and N/C] for the model development, as it could affect the quantity and quality of the biochar. For instance, Mukome et al. [22] reported that C/N could affect the biochar yield. The ratio of elementary composition (O/C and N/C) indicates the biochar polarity and thus affects the surface property [23]. The statistical analysis of the collected data was given in Table 2, where count shows the number of datasets under consideration for a particular variable. The mean and standard deviation of all variables shows the distribution pattern of the collected data. The minimum and maximum value of a particular variable exhibit the range of input and output parameters. Four quartile value has been given for a better understanding of the data distribution pattern along with the minimum and maximum value of the parameters (Refer Supplementary Material – Fig. S1).

The C and H content of algae is around 44% and 6% respectively, which is lower when compared to lignocellulosic biomasses [24]. However, the N content of algal biomass is around 7%, which is significantly greater than lignocellulosic biomasses (1–2%). The high N content in the algal biomasses could be attributed due to high protein content [41]. The molar ratio between H and C lies on a higher side indicates the presence of polysaccharides in algae with the ash content varying from 0.45–40%. However, the mean ash content from algal biomass is higher than that of other biomasses trails with the finding of Ross et al. [25]. The presence of high ash content can be due to the presence of silica-containing material in the algae [26]. The average fixed carbon content and volatile compounds are 14.46% and 68.08% respectively, which are similar to other biomasses [27]. The pyrolysis temperature has been ranged from 300 to 600 °C and the heating rate has been stretched from 8 to 100 °C/min, hence all the pyrolysis reactions were presumed under slow pyrolysis regime. Residence time has been ranged from 20 to 180 min. As a result, the char yield varies from 14 to 66% based on the input parameters. The carbon content in the biochar observed to increase around 5 to 10% depending on the pyrolysis temperature, which is following the previous findings [28,29]. H, O, and N composition of biochar was observed to be decreasing when compared to that of pristine biomass. The relation among and in between input parameters and algal biochar yield observed through the Pearson correlation coefficient matrix is given in Fig. 2. The sign of the Pearson correlation coefficient defines the type of correlation between parameters. The magnitude of the coefficient suggests how effectively one parameter affects the others.

3.2. Evaluation of eXtreme Gradient Boosting (XGB) model for biochar yield

The correlation coefficient (R^2) for both training and testing datasets of ***Run 1 to 14*** were given in Table 1. The Scatter Plots of training and testing for ***Run 3*** ['*Ash*', '*FC*', '*VC*' and '*T*'] and ***Run 14*** ['*H/C*', '*N/C*', '*Ash*', '*Pyrolysis time*', '*Temperature*'] between experimental and model-predicted algal biochar yield are given in Fig. 3. [The Scatter Plots for all other runs of testing datasets were given in Supplementary Material – Fig. S2]. However, the R^2 values after 5 times cross-validation for ***Run***

Table 1
Experimental design and performance of eXtreme Gradient Boosting (XGB) based machine learning model for predicting algal biochar yield.

Run no	Parameters under consideration	Name of the parameters	No of input parameters	Correlation coefficient (R^2) for the training dataset	Correlation coefficient (R^2) for the testing dataset
1	EC	C, H, O, and N	4	0.4862	0.2143
2	REC + T	H/C, O/C, N/C, and T	4	0.9999	0.6895
3	PA + T	Ash, FC, VC, and T	4	0.9977	0.8145
4	EC + T	C, H, O, N, and T	5	0.9915	0.6516
5	REC + PT + HR + T	H/C, O/C, N/C, PT, HR, and T	6	0.9990	0.7405
6	REC + PA + T	H/C, O/C, N/C, Ash, FC, VC, and T	7	0.9978	0.7139
7	EC + PT + HR + T	C, H, O, N, PT, HR, and T	7	0.9955	0.7314
8	EC + REC + T	C, H, O, N, H/C, O/C, N/C, and T	8	0.9990	0.7263
9	EC + PA + T	C, H, O, N, Ash, FC, VC, and T	8	0.9980	0.7696
10	EC + REC + PT + HR + T	C, H, O, N, H/C, O/C, N/C, PT, HR, and T	10	0.9996	0.7208
11	EC + PA + PT + HR + T	C, H, O, N, Ash, FC, VC PT, HR, and T	10	0.9992	0.7464
12	EC + REC + PA + T	C, H, O, N, H/C, O/C, N/C, Ash, FC, VC, and T	11	0.9920	0.7301
13	EC + REC + PA + HR + PT + T	C, H, O, N, H/C, O/C, N/C, Ash, FC, VC PT, HR, and T	13	0.9990	0.7381
14	Feature selection by XGB model	H/C, N/C, Ash, PT, and T	5	0.9974	0.8440

EC: Elementary Composition (C, H, O, and N); T: Temperature; REC: Ratio of Elementary Composition (H/C, O/C and N/C); PA: Proximate Analysis; Ash: Ash Content; FC: Fixed Carbon; VC: Volatile Compound; PT: Pyrolysis Time; HR: Heating Rate; + denotes the combination of input parameters under consideration.

Table 2
Statistical analysis of all the parameters involved in machine learning model of algal biochar yield.

	Ultimate analysis				Ratios			Proximate analysis			Pyrolysis conditions			Biochar	Biochar composition			
Statistics	C	H	O	N	H/C	O/C	N/C	Ash	FC	VC	PT	HR	T	Yield	BC	BH	BO	BN
Count	85	85	85	85	85	85	85	85	85	85	64	64	85	85	28	28	28	28
Mean	44.06	6.20	38.40	7.01	0.140	0.930	0.150	13.19	14.46	68.01	55.93	28.37	463.8	33.86	49.41	3.22	18.54	6.11
Standard deviation	6.66	0.99	12.11	2.71	0.014	0.470	0.049	8.84	5.60	9.85	36.58	27.33	102.3	11.34	6.59	2.39	18.16	2.88
Minimum	28.78	4.02	21.27	1.51	0.10	0.39	0.04	0.76	0.98	45.65	20	8	300	14.76	28.50	0.37	0.28	1.98
25%	40.60	5.08	29.73	4.64	0.13	0.59	0.12	8.06	11.04	63.50	30	10	400	26.00	45.89	1.45	4.12	3.27
50%	46.16	6.40	33.04	6.65	0.14	0.70	0.16	9.50	13.49	68.75	60	20	475	31.00	49.80	2.35	11.25	6.25
75%	49.26	7.05	47.40	8.89	0.15	1.16	0.18	17.33	16.70	76.30	60	32.5	550	38.00	53.685	4.62	34.69	7.82
Maximum	53.39	7.90	62.44	11.19	0.16	2.04	0.23	40.00	31.13	83.00	180	100	700	66.07	62.37	8.50	52.87	10.96

EC: Elementary Composition (C, H, O, and N); T: Temperature; REC: Ratio of elementary composition (H/C, O/C and N/C); PA: Proximate Analysis; Ash: Ash Content; FC: Fixed Carbon; VC: Volatile Compound; PT: Pyrolysis Time; HR: Heating Rate; BC: carbon composition in biochar; BH: Hydrogen composition in biochar; BO: Oxygen composition in biochar; BN: Nitrogen composition in biochar.

3* and *Run 14 testing data were 0.814 and 0.844 respectively. The values after cross-validation still can be considered as highly accurate and reliable following the outcome of previous work on pyrolysis modeling using random forest and artificial neural networks [17,19]. The effect of ash content, fixed carbon, and volatile compound on the algal biochar yield was earlier reported by Cheng [27] and Maddi et al. [30]. Conversely, the addition of more input parameters along with proximate analysis data couldn't improve the model performance. Therefore, the Importance Plots (Fig. 4) have been made to comprehend the effect of all input parameters on algal biochar yield as it gives a score for each input parameter based on its relative importance on biochar yield. Higher the impact of input parameters on biochar yield, greater is the corresponding features score (F score). The input parameters have been ranked from the most influential factor to least affecting the biochar yield.

As per the Importance Plots, the algal biochar yield largely depends on pyrolysis temperature, which is following the previous findings [17,22,31]. The succeeding influential factors are carbon content and a molar ratio of H and C. Ash, fixed carbon and volatile compound also affect the algal biochar yield considerably. Maddi et al. [30] have similar findings for the effect of proximate analysis on biochar yield. The heating rate also affects the biochar yield; yet, the impact is relatively not significant as compared to the temperature [32]. O and N have also a very less significant impact on biochar yield. Further, residence time and other elementary molar ratio have an almost negligible effect on biochar yield [17].

3.3. Influence of input parameters on algal biochar yield

Though Feature Importance Plots depict the most influential input parameter for the algal biochar yield, it couldn't elucidate how does the input parameter affect the biochar yield. Therefore, SHAP Dependence Plots have been used to explicate the interaction of critical parameters on algal biochar yield, whereas for the individual effect of all the input parameters (Refer Supplementary Material – Fig. S3). Fig. 5 (a–m) shows the SHAP Dependence Plots for the interactive effect of temperature and other input parameters considered in the study. SHAP Dependence Plot is the plot of dots for each dataset with an input parameter range in its *x*-axis and its SHAP value on the *y*-axis so that its relative importance could be determined by SHAP value. The secondary *y*-axis represents the temperature that helps in identifying the combined effect of temperature and other input parameters. Since the SHAP plot is a type of scatter plot based on game theory and explains the effect of a single feature on the model predictions, it is widely used. As the pyrolysis temperature has the most profound impact on biochar yield, the interactive effects were analyzed. The magnitude of the SHAP value shows the intensity of the impact on algal biochar yield. Positive SHAP value indicates that the parameter is positively favoring the biochar yield. A negative value represents the inverse influence on product outcomes. The SHAP Dependence Plot for sparse data points needs to be ignored due to their inaccuracy. Therefore, it might not be possible to obtain meaningful inference from SHAP Dependence Plot for all input parameters. The inferences from SHAP Dependence Plots were done by

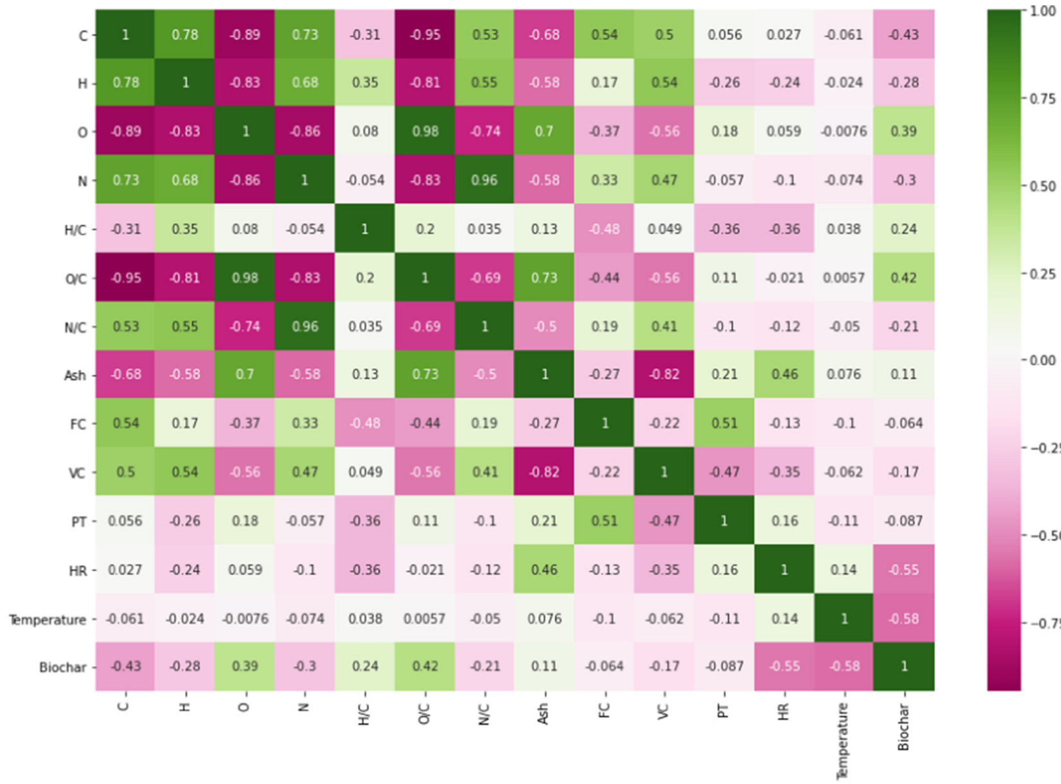

Fig. 2. Pearson correlation coefficient matrices for relationship between input and output parameters.
EC: Elementary Composition (C, H, O, and N); T: Temperature; REC: Ratio of Elementary Composition (H/C, O/C and N/C); PA: Proximate Analysis; Ash: Ash content; FC: Fixed Carbon; VC: Volatile Compound; PT: Pyrolysis Time; HR: Heating Rate.

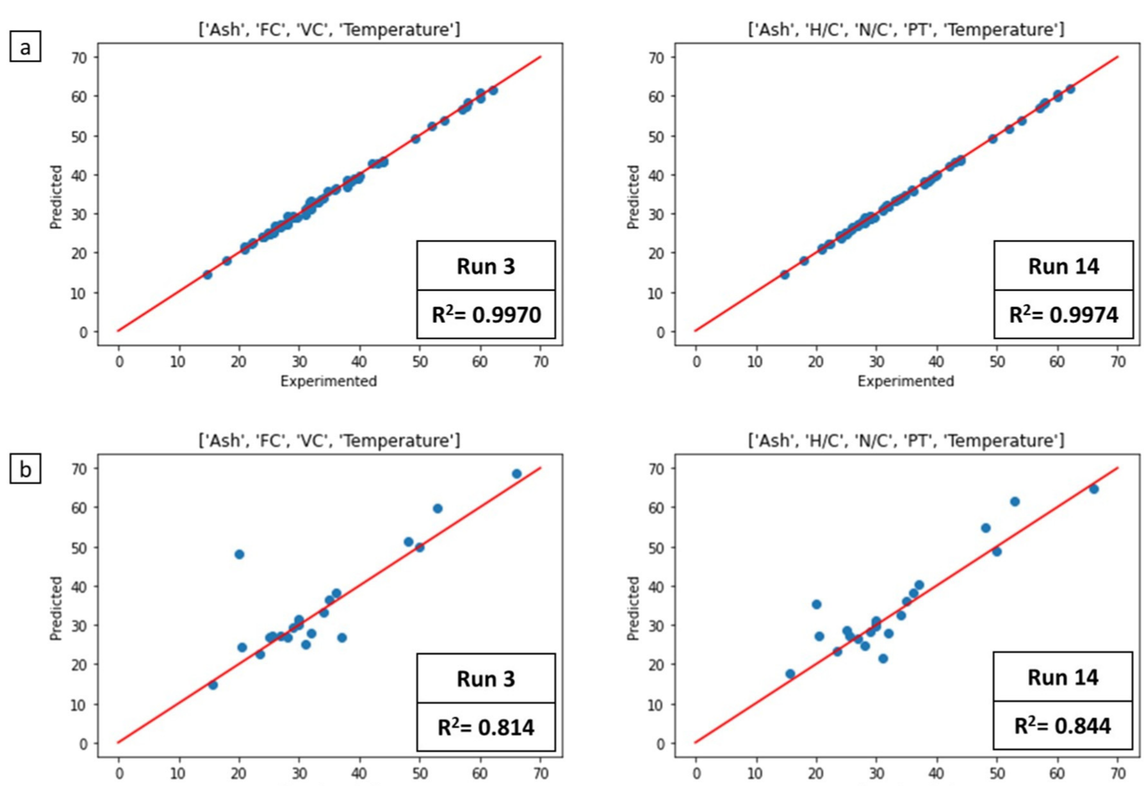

Fig. 3. Scatter Plots of a) Training data and b) Testing data for algal biochar yield by eXtreme Gradient Boosting (XGB) model after *k* cross-validation.
EC: Elementary Composition (C, H, O, and N); T: Temperature; REC: Ratio of Elementary Composition (H/C, O/C and N/C); PA: Proximate Analysis; Ash: Ash Content; FC: Fixed Carbon; VC: Volatile Compound; PT: Pyrolysis Time; HR: Heating Rate.

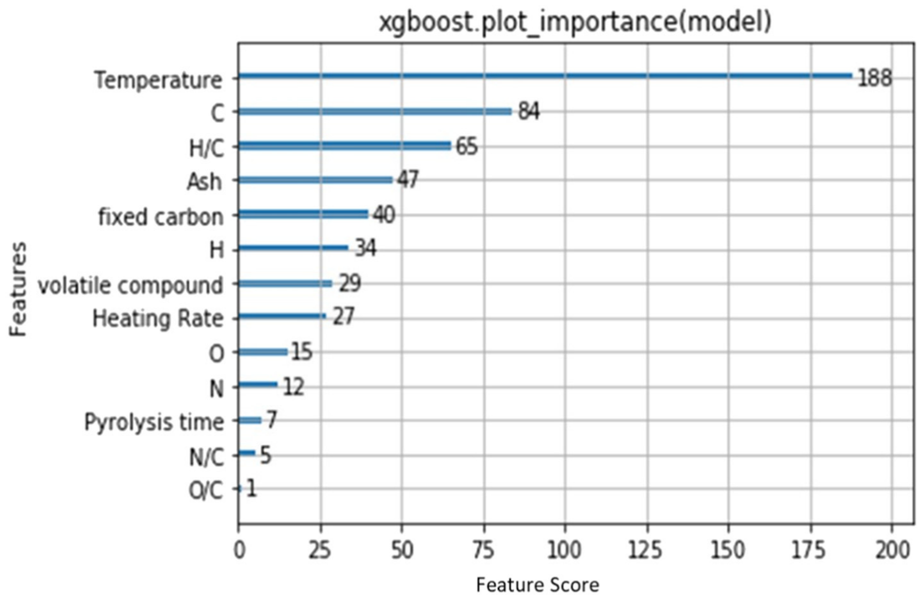

Fig. 4. Feature Importance Plot of all input parameters on algal biochar yield during slow pyrolysis. EC: Elementary Composition (C, H, O, and N); T: Temperature; REC: Ratio of Elementary Composition (H/C, O/C and N/C); PA: Proximate Analysis; Ash: Ash Content; FC: Fixed Carbon; VC: Volatile Compound; PT: Pyrolysis Time; HR: Heating Rate.

considering the distribution of the majority of data points. The yield of biochar decreases with the increase in temperature. A sharp decline of biochar yield (Fig. 5a) at 300–350 °C might be due to the devolatilization of the volatile compound and bound moisture content. Biochar yield is decreasing significantly in the temperature range of 350–500 °C, which could be attributed to the decomposition of carbohydrate, protein, lipid, and carbonaceous solid present in the algal biomasses [33]. When the carbon content of algal biomass is low to moderate (30–40%), the biochar yield is observed to show a positive correlation with carbon composition [16]. However, the biochar yield decreases (Fig. 5b) when the carbon content of the algal biomass is greater than 40%. Moderate carbon content and low temperature were observed to be the optimized conditions for maximizing the biochar yield. Maximum data points for hydrogen have SHAP values in between −0.5 to 0.5 reveals that H has relatively less impact on the biochar yield (Fig. 5c). Similarly, with an increase in oxygen content, the SHAP values are decreasing further discloses that O is inversely related to biochar yield (Fig. 5d).

It is clear from Fig. 5e that the lower N content favor biochar yield, however, the impact is less as observed from lower SHAP values (0–0.5). Fig. 5f shows the correlation between H/C and biochar yield. High H/C at higher temperatures has large positive SHAP values could be due to the formation of char by a breakdown of polysaccharides present in the algae. Fig. 5(g–h) shows that O/C and N/C have no significant effect on algal biochar yield. Fig. 5i shows that biochar yield increases with an increase in ash content might be due to the presence of inorganic components present in the algal biomasses [34]. Carbonate ion reacts with ammonia (due to the deamination of amino acids during the thermal degradation of protein) could result in ammonium carbonate, which further degrades into ammonia and carbon dioxide during pyrolysis. Hence amino acid removed as ammonia couldn't participate in the reaction and contribute to the formation of biochar [35]. Ash content in the range of 15–30% at moderate temperature (400–500 °C) can maximize the biochar yield. The fixed carbon content of algal biomass positively affects the biochar yield which agrees with the findings of Rowbotham et al. [36]. However, for high fixed carbon content, the char yield (Fig. 5j) is observed to decrease slightly. Volatile compounds were observed to have less impact on biochar yield, however, few data points with low volatile compound content have negative SHAP value (Fig. 5k). Fig. 5l shows with an increase in the heating rate, the biochar yield was observed to increase till 20 °C/min and then started declining exponentially at higher heating rates. The low SHAP values at lower heating rates might be due to the effect of high temperature. Rather, a higher heating rate always favors more bio-oil and less biochar yield [32,37].

3.4. Summary Plots for input parameters

A combination of Feature Importance Plot with SHAP Dependence Plot can be presented in the Summary Plots. The input parameters are placed on the *y*-axis based on their influence, the most influential variable being kept at the top. The *x*-axis represents the SHAP value and the value of the feature is shown in color; blue to pinkish-red represents the low to high significance. More the data points falling in a particular range of SHAP value, the more the input variable correlates with biochar yield in that fashion.

As the temperature is the most influential input parameter of biochar yield, it rated on the top of the Summary Plots (Fig. 6). Further, it reveals that the large quantum of higher temperature data has negative SHAP values and become more negative with an increase in temperature. It reaffirms that low temperature favors biochar yield. Few data points with low carbon contents favor biochar yield, whereas the majority of carbon have SHAP value near zero or negative. Algal biomass having higher ash content effect the biochar yield positively, the analysis of Summary Plot shows the correlation is less significant as compared to temperature. Algae having a high H/C ratio enhances the biochar yield, however, it is also important to note that a large amount of data having a low or moderate value of H/C negatively affects the biochar yield. The grey color in the heating rate indicates the dataset having a null value, whereas, a lower heating rate is observed to favor biochar formation. Biochar shows a positive correlation with the volatile compound, however, the SHAP values are close to zero. It might be due to less importance of volatile compounds on char yield. It is not feasible to deduce any meaningful inference from the Summary Plot of FC, N, and pyrolysis time. The majority of SHAP values for H is close to zero might be due to the lesser difference between H content in collected algal biomass. The N/C, O, and O/C have almost zero SHAP values that signify the negligible impact on biochar yield. By using the Summary Plot, the relationship between the value of a feature, and the impact on the prediction can be known. Yet, to comprehend the exact form of relationship between the features and biochar yield, SHAP Dependence Plot is always a preferable option.

Out of 13 input parameters to predict the algal biochar yield, the XGB model utilized 2^{13} combinations of input parameters and figured out the five key parameters (H/C, N/C, Ash, pyrolysis temperature and residence time) as the essential ones for yield predictions.

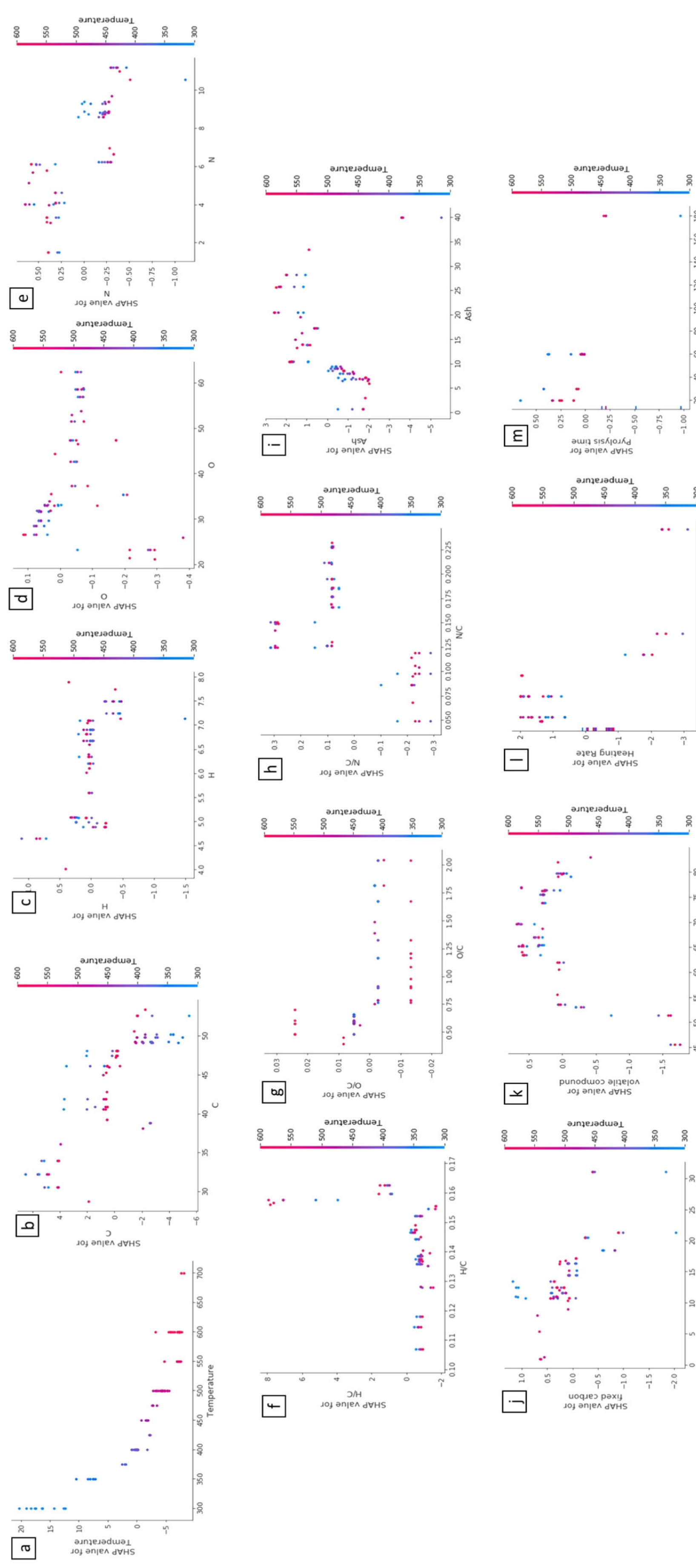

Fig. 5. SHapley Additive exPlanations (SHAP) Dependence Plots for the combined effect of temperature and other input parameters on algal biochar yield by eXtreme Gradient Boosting (XGB) model. EC: Elementary Composition (C, H, O, and N); T: Temperature; REC: Ratio of Elementary Composition (H/C, O/C and N/C); PA: Proximate Analysis; Ash: Ash Content; FC: Fixed Carbon; VC: Volatile Compound; PT: Pyrolysis Time; HR: Heating Rate.

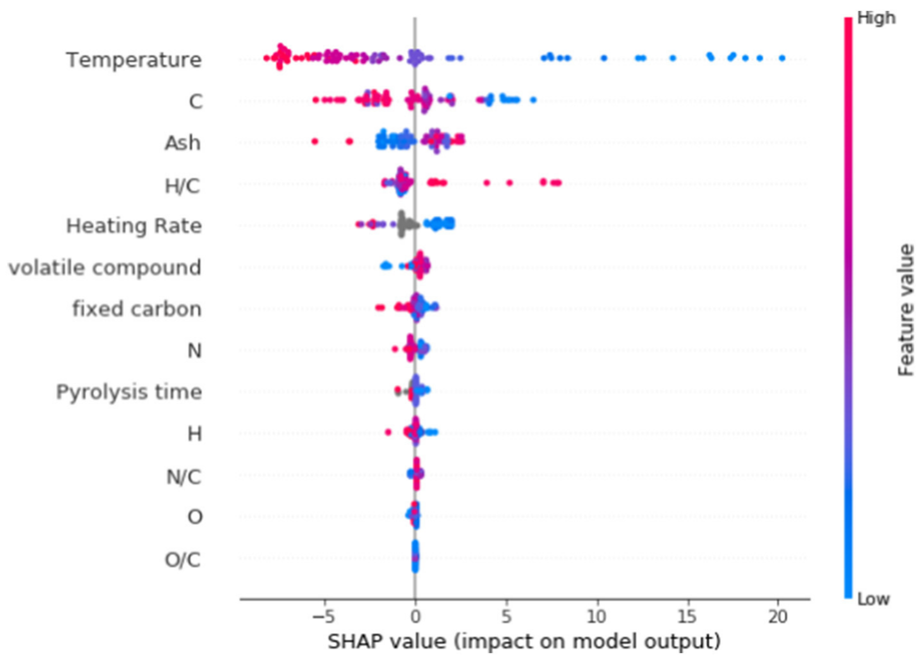

Fig. 6. Summary Plots of different input parameters for algal biochar yield by eXtreme Gradient Boosting (XGB) model.
EC: Elementary composition (C, H, O, and N); T: Temperature; REC: Ratio of elementary composition (H/C, O/C and N/C); PA: Proximate Analysis; Ash: Ash content; FC: Fixed carbon; VC: Volatile compound; PT: Pyrolysis time; HR: Heating rate.

3.5. Prediction of biochar composition

To develop the predicting model for the elementary composition (C, H, O, and N) of algal biochar, more exhaustive datasets are required. As the algal biochar is gaining attention recently and emerging in the nascent stage, requisite data on algal biochar combination is not widely available. In this study, a second model has been developed for predicting the algal biochar composition using the available limited published resources. It is noteworthy to mention that most of the earlier papers emphasize biochar yield and only very few literature presents the composition of algal biochar. Therefore, 31 datasets with the published composition of algal biochar were taken into consideration for the model development. However, the R^2 is around 0.66 indicates the huge scope for necessitating the model improvement while predicting algal biochar composition. Model development demands more input datasets for accurate and reliable prediction of algal biochar composition, which could explain the lesser R^2 value. However, with the acquisition of more datasets on algal biochar composition, a similar type of machine learning model could be used for model development. Due to the unavailability of sufficient data, the biochemical composition (carbohydrate, protein, and lipids) of algae has been not considered in the present study. The future model could include biochemical composition and proximate analysis of biochar for a more precise predictive model aimed at predicting algal biochar composition.

4. Conclusion

Prediction of algal biochar yield has been executed for eXtreme Gradient Boosting (XGB) based machine learning models by utilizing the input of ultimate and proximate analysis along with pyrolysis conditions. All possible combination (2^{13}) of input parameters has been analyzed for predicting the biochar yield. Machine learned XGB model (Run 3) with four parameters of [ash, fixed carbon, volatile compound, and pyrolysis temperature] predicted similarity of 81.4% with the experimental algal biochar yield. The key feature selection (Run 14) by the XGB model suggested the five parameters [H/C, N/C, Ash, pyrolysis temperature, and time] as input combinations that predicted the similarity of 84.4% experimental algal biochar yield. Pearson correlation coefficient matrix revealed the correlation among and in between thirteen input parameters and algal biochar yield. Pyrolysis temperature, carbon, and ash content showed domination on the algal biochar yield. Analysis of Importance and SHAP Dependence Plots revealed the interaction between prominent input characteristics and algal biochar yield. The robustness of the model can be further increased with the incorporation of more amount of data in the model. Further, the study could be extended to focus more on the qualitative aspect of such as proximate analysis and composition of biochar.

CRediT authorship contribution statement

Abhijeet Pathy: Investigation, Formal analysis, Writing - original draft.**Saswat Meher:**Methodology, Formal analysis, Writing - original draft.**P. Balasubramanian:**Conceptualization, Formal analysis, Writing - review & editing.

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.

Acknowledgments

The authors thank the Department of Biotechnology and Medical Engineering of National Institute of Technology Rourkela for providing the research facility. The authors greatly acknowledge the Science and Engineering Research Board (SERB) of the Department of Science and Technology (DST) of Government of India (GoI) for sponsoring the research through Early Career Research Award [FILE NO. ECR/2017/003397/ES] and ASEAN-India STI Cooperation [FILE NO. IMRC/AISTDF/CRD/2018/000082].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.algal.2020.102006.

References

[1] N. Bordoloi, J. Tiwari, S. Kumar, J. Korstad, K. Bauddh, Efficiency of algae for heavy metal removal, bioenergy production, and carbon sequestration, Emerging Eco-Friendly Green Technologies for Wastewater Treatment, Springer, Singapore, 2020,

pp. 77–101.
[2] H.L. Bryant, I. Gogichaishvili, D. Anderson, J.W. Richardson, J. Sawyer, T. Wickersham, M.L. Drewery, The value of post-extracted algae residue, Algal Res. 1 (2) (2012) 185–193.
[3] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, Xgboost: Extreme Gradient Boosting. *R Package Version 0.4-2*, (2015), pp. 1–4.
[4] A. Marcilla, L. Catalá, J.C. García-Quesada, F.J. Valdés, M.R. Hernández, A review of thermochemical conversion of microalgae, Renew. Sust. Energ. Rev. 27 (2013) 11–19.
[5] S. Aravind, P.S. Kumar, N.S. Kumar, N. Siddarth, Conversion of green algal biomass into bioenergy by pyrolysis. A review, Environ. Chem. Lett. (2020) 1–21.
[6] Y. Qiao, B. Wang, Y. Ji, F. Xu, P. Zong, J. Zhang, Y. Tian, Thermal decomposition of castor oil, corn starch, soy protein, lignin, xylan, and cellulose during fast pyrolysis, Bioresour. Technol. 278 (2019) 287–295.
[7] K.L. Yu, P.L. Show, H.C. Ong, T.C. Ling, W.H. Chen, M.A.M. Salleh, Biochar production from microalgae cultivation through pyrolysis as a sustainable carbon sequestration and biorefinery approach, Clean Techn. Environ. Policy 20 (9) (2018) 2047–2055.
[8] K.L. Yu, B.F. Lau, P.L. Show, H.C. Ong, T.. Ling, W.H. Chen, ... J.S. Chang, Recent developments on algal biochar production and characterization, Bioresource technology 246 (2017) 2–11.
[9] S. Rangabhashiyam, P. Balasubramanian, The potential of lignocellulosic biomass precursors for biochar production: performance, mechanism and wastewater application-a review, Ind. Crop. Prod. 128 (2019) 405–423.
[10] M. Kołtowski, B. Charmas, J. Skubiszewska-Zięba, P. Oleszczuk, Effect of biochar activation by different methods on toxicity of soil contaminated by industrial activity, Ecotoxicol. Environ. Saf. 136 (2017) 119–125.
[11] L. Zhang, F. Li, A. Kuroki, K.C. Loh, C.H. Wang, Y. Dai, Y.W. Tong, Methane yield enhancement of mesophilic and thermophilic anaerobic co-digestion of algal biomass and food waste using algal biochar: semi-continuous operation and microbial community analysis, Bioresour. Technol. 302 (2020) 122892.
[12] S.W. Huang, X. Chen, D.D. Wang, H.L. Jia, L. Wu, Bio-reduction and synchronous removal of hexavalent chromium from aqueous solutions using novel microbial cell/algal-derived biochar particles: turning an environmental problem into an opportunity, Bioresour. Technol. 309 (2020) 1–8 123304.
[13] A.F. Ibrahim, K.P.R. Dandamudi, S. Deng, Y.S. Lin, Pyrolysis of hydrothermal liquefaction algal biochar for hydrogen production in a membrane reactor, Fuel 265 (2020) 116935.
[14] C. Di Blasi, Modeling chemical and physical processes of wood and biomass pyrolysis, Prog. Energy Combust. Sci. 34 (1) (2008) 47–90.
[15] S.S. Alkurdi, R.A. Al-Juboori, J. Bundschuh, L. Bowtell, S. McKnight, Effect of pyrolysis conditions on bone char characterization and its ability for arsenic and fluoride removal, Environ. Pollut. 262 (2020) 114221.
[16] P. Abhijeet, G. Swagathnath, S. Rangabhashiyam, M.A. Rajkumar, P. Balasubramanian, Prediction of pyrolytic product composition and yield for various grass biomass feedstocks, Biomass Convers. Biorefinery (2019) 1–12.
[17] X. Zhu, Y. Li, X. Wang, Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions, Bioresour. Technol. 288 (2019) 121527.
[18] M. Saleem, I. Ali, Machine learning based prediction of pyrolytic conversion for red sea seaweed, 7th International Conference on Biological, Chemical & Environmental Sciences (BCES-2017), 2017 https://doi.org/10.17758/EAP.C0917043.
[19] H. Cao, Y. Xin, Q. Yuan, Prediction of biochar yield from cattle manure pyrolysis via least squares support vector machine intelligent approach, Bioresour. Technol. 202 (2016) 158–164.
[20] M.A. El Aziz, A.M. Hemdan, A.A. Ewees, M. Elhoseny, A. Shehab, A.E. Hassanien, S. Xiong, Prediction of biochar yield using adaptive neuro-fuzzy inference system with particle swarm optimization, 2017 IEEE PES PowerAfrica, IEEE, 2017, pp. 115–120 June.
[21] D. Nielsen, Tree Boosting With Xgboost-Why Does Xgboost Win "Every" Machine Learning Competition? (2016) (Master's thesis, NTNU).
[22] F.N. Mukome, X. Zhang, L.C. Silva, J. Six, S.J. Parikh, Use of chemical and physical characteristics to investigate trends in biochar feedstocks, J. Agric. Food Chem. 61 (9) (2013) 2196–2204.
[23] H. Zheng, W. Guo, S. Li, Y. Chen, Q. Wu, X. Feng, ... J.S. Chang, Adsorption of p-nitrophenols (PNP) on microalgal biochar: analysis of high adsorption capacity and mechanism, Bioresource technology 244 (2017) 1456–1464.
[24] C. Yang, R. Li, B. Zhang, Q. Qiu, B. Wang, H. Yang, ... C. Wang, Pyrolysis of microalgae: A critical review, Fuel processing technology 186 (2019) 53–72.
[25] A.B. Ross, K. Anastasakis, M. Kubacki, J.M. Jones, Investigation of the pyrolysis behaviour of brown algae before and after pre-treatment using PY-GC/MS and TGA, J. Anal. Appl. Pyrolysis 85 (1–2) (2009) 3–10.
[26] K. Liu, Characterization of ash in algae and other materials by determination of wet acid indigestible ash and microscopic examination, Algal Res. 25 (2017) 307–321.
[27] J. Cheng (Ed.), Biomass to Renewable Energy Processes, CRC press, 2017.
[28] Y.M. Chang, W.T. Tsai, M.H. Li, Chemical characterization of char derived from slow pyrolysis of microalgal residue, J. Anal. Appl. Pyrolysis 111 (2015) 88–93.
[29] K.W. Jung, T.U. Jeong, H.J. Kang, K.H. Ahn, Characteristics of biochar derived from marine macroalgae and fabrication of granular biochar by entrapment in calcium-alginate beads for phosphate removal from aqueous solution, Bioresour. Technol. 211 (2016) 108–116.
[30] B. Maddi, S. Viamajala, S. Varanasi, Comparative study of pyrolysis of algal biomass from natural lake blooms with lignocellulosic biomass, Bioresour. Technol. 102 (23) (2011) 11018–11026.
[31] K. Jindo, H. Mizumoto, Y. Sawada, M.A. Sanchez-Monedero, T. Sonoki, Physical and chemical characterization of biochars derived from different agricultural residues, Biogeosciences 11 (23) (2014) 6613–6621.
[32] D. Angın, Effect of pyrolysis temperature and heating rate on biochar obtained from pyrolysis of safflower seed press cake, Bioresour. Technol. 128 (2013) 593–597.
[33] I.V. Babich, M. Van der Hulst, L. Lefferts, J.A. Moulijn, P. O'Connor, K. Seshan, Catalytic pyrolysis of microalgae to high-quality liquid bio-fuels, Biomass Bioenergy 35 (7) (2011) 3199–3207.
[34] K. Wang, J. Zhang, B.H. Shanks, R.C. Brown, The deleterious effect of inorganic salts on hydrocarbon yields from catalytic pyrolysis of lignocellulosic biomass and its mitigation, Appl. Energy 148 (2015) 115–120.
[35] W.T. Chen, W. Qian, Y. Zhang, Z. Mazur, C.T. Kuo, K. Scheppe, ... B.. Sharma, Effect of ash on hydrothermal liquefaction of high-ash content algal biomass, Algal research 25 (2017) 297–306.
[36] J.S. Rowbotham, P.W. Dyer, H.C. Greenwell, M.K. Theodorou, Thermochemical processing of macroalgae: a late bloomer in the development of third-generation biofuels? Biofuels 3 (4) (2012) 441–461.
[37] A.V. Bridgwater, Review of fast pyrolysis of biomass and product upgrading, Biomass Bioenergy 38 (2012) 68–94.
[38] S.V. Vassilev, C.G Vassileva, Composition, properties and challenges of algae biomass for biofuel application: an overview, Fuel 181 (2016) 1–33.
[39] B. Bharathiraja, M. Chakravarthy, R.R. Kumar, D. Yogendran, D. Yuvaraj, J. Jayamuthunagai, ... S. Palani, Aquatic biomass (algae) as a future feed stock for bio-refineries: a review on cultivation, processing and products, Renew. Sustain. Energy Rev. 47 (2015) 634–653.
[40] Q.V. Bach, W.H Chen, Pyrolysis characteristics and kinetics of microalgae via thermogravimetric analysis (TGA): a state-of-the-art review, Bioresour. Technol. 246 (2017) 88–100.
[41] M. Chen, S. Ding, X. Chen, Q. Sun, X. Fan, J. Lin, ... C. Zhang, Mechanisms driving phosphorus release during algal blooms based on hourly changes in iron and phosphorus concentrations in sediments, Water Res. 133 (2018) 153–164.