Environmental Science & Technology

pubs.acs.org/est

Article

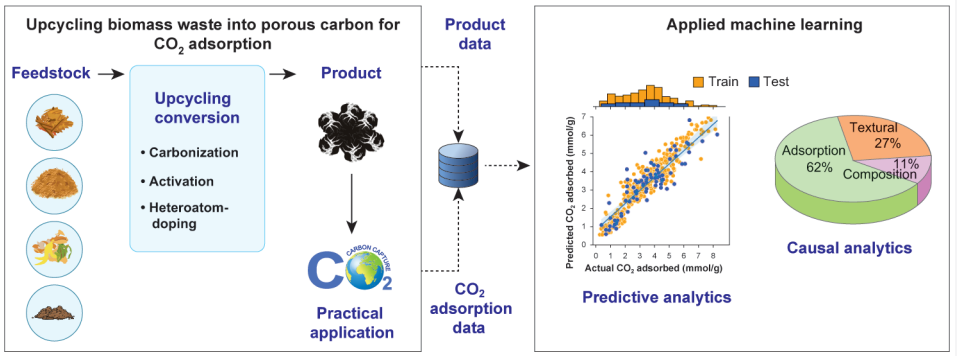
Applied Machine Learning for Prediction of CO_2 Adsorption on Biomass Waste-Derived Porous Carbons

Xiangzhou Yuan, Manu Suvarna, Sean Low, Pavani Dulanja Dissanayake, Ki Bong Lee, Jie Li, Xiaonan Wang,* and Yong Sik Ok*

Cite This: *Environ. Sci. Technol.* 2021, 55, 11925−11936

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information



ABSTRACT: Biomass waste-derived porous carbons (BWDPCs) are a class of complex materials that are widely used in sustainable waste management and carbon capture. However, their diverse textural properties, the presence of various functional groups, and the varied temperatures and pressures to which they are subjected during CO_2 adsorption make it challenging to understand the underlying mechanism of CO_2 adsorption. Here, we compiled a data set including 527 data points collected from peer-reviewed publications and applied machine learning to systematically map CO_2 adsorption as a function of the textural and compositional properties of BWDPCs and adsorption parameters. Various tree-based models were devised, where the gradient boosting decision trees (GBDTs) had the best predictive performance with R^2 of 0.98 and 0.84 on the training and test data, respectively. Further, the BWDPCs in the compiled data set were classified into regular porous carbons (RPCs) and heteroatom-doped porous carbons (HDPCs), where again the GBDT model had R^2 of 0.99 and 0.98 on the training and 0.86 and 0.79 on the test data for the RPCs and HDPCs, respectively. Feature importance revealed the significance of adsorption parameters, textural properties, and compositional properties in the order of precedence for BWDPC-based CO_2 adsorption, effectively guiding the synthesis of porous carbons for CO_2 adsorption applications.

KEYWORDS: *machine learning, gradient boosting decision trees, carbon materials, low carbon technology, sustainable waste management, gas adsorption and separation*

INTRODUCTION

Carbon capture and storage (CCS) has been considered as an indispensable tool for reducing CO_2 emission,[1−4] as the atmospheric CO_2 concentration has exceeded 410 ppm and continues to increase steadily.[5] In CCS systems, the operation of CO_2 capture is still the most expensive process, accounting for over 50% of the total CCS cost.[6] Precombustion, postcombustion, and oxy-fuel combustion are the three main approaches for capturing CO_2 from industrial emission point sources, with postcombustion capture being the most cost-effective method.[7,8] However, owing to the low CO_2 concentration in postcombustion flue gases (typically <15%), the main challenge of this process is the development of a cost-effective CO_2 capture method. Well-established absorption processes, such as the regenerative amine solution process, for postcombustion CO_2 capture are expensive and have problems of significant corrosion, solvent loss due to degradation, and environmental toxicity.[9] Researchers are attempting to develop cost-effective membranes with high CO_2 permeability for capturing CO_2 from flue gas, which is still far from commercialization.[9] Solid porous carbon-based CO_2 adsorption is widely considered to be the most promising second-generation carbon capture approach,[1,8,10] owing to its

Received: March 21, 2021
Revised: July 4, 2021
Accepted: July 6, 2021
Published: July 22, 2021

https://doi.org/10.1021/acs.est.1c01849
Environ. Sci. Technol. 2021, 55, 11925−11936

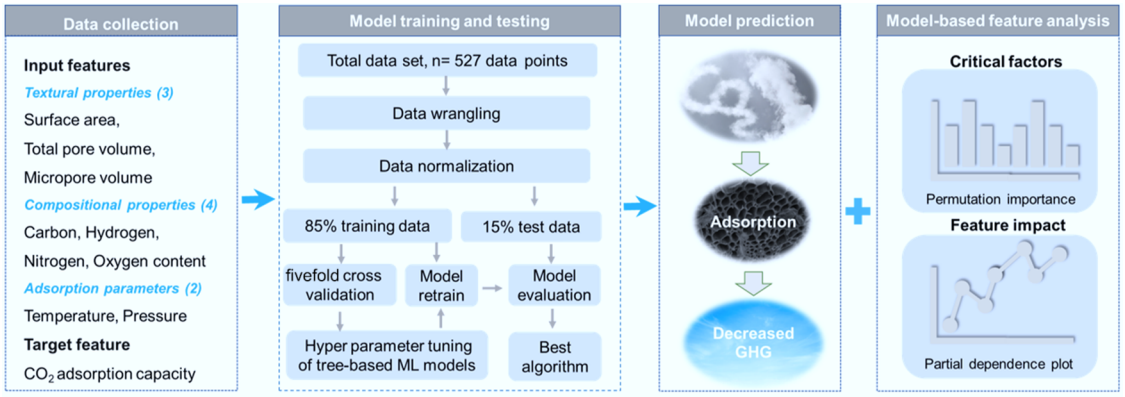


Figure 1. Schematic of the workflow in this study. A comprehensive review of the literature specific to the adsorption of CO_2 by BWDPCs was first performed to build the data set required for this study. Based on the data collected, input features were appropriately identified and labeled. Post the data collection, data preprocessing was performed to enable effective ML application, where three types of tree-based ML models, namely, GBDT, LGB, and XGB, were evaluated for their prediction performance. Given the black-box nature of most ML models, the inference of the best-performing model in this study was understood via permutation importance and PDPs.

advantages including low cost, tunable pore structure, relatively low energy consumption for regeneration, and excellent cyclic stability. Compared to other precursors, biomass waste has promising advantages such as cost-effectiveness, sustainability, and abundance for preparing porous CO_2 adsorbents.[2,11] Moreover, biomass waste-derived porous carbons (BWDPCs) for CO_2 capture can mitigate the environmental pollution caused by inappropriate biomass waste management and achieve decarbonization and negative emission technologies for climate change abatement.[9]

The CO_2 adsorption isotherms of BWDPCs at different temperatures have been widely investigated for elucidating the thermodynamic properties of the CO_2 adsorption process and guiding the CO_2 adsorption process design and optimization.[12] These thermodynamic properties (e.g., entropy, Gibbs free energy, and isosteric heat of adsorption, etc.) indicate that CO_2 adsorption on solid carbon adsorbents is mainly dominated by physisorption (physical and weak interaction), making them sensitive to adsorption conditions with low CO_2 selectivity.[9,10] To enhance the CO_2 adsorption capacity and CO_2 selectivity, carbonization and activation followed by heteroatom doping play significant roles. In general, the upcycling of biomass waste into porous carbons can be divided into two steps: (1) carbonization and (2) activation. Thermochemical conversion approaches have been extensively used in carbonization.[11] Chemical and physical activations have received increasing attention to enhance the CO_2 capture performance of prepared CO_2 adsorbents, improving microporous structures with high porosity.[13,14] Heteroatom doping treatment can also be considered to increase the basic active sites of the porous carbons,[15−17] enhancing the CO_2 uptake and selectivity over other gases as a result of acid−base interactions. It is worth noting that some toxic or greenhouse gases are emitted during these thermochemical processes. For example, in addition to CO_2 gas, the treatment of nitrogen oxides (NOx) emitted should be carefully considered because NOx is not only toxic but also consists of typical greenhouse gases.[18] Most publications in the last decade (Tables S1 and S2) have generally provided similar approaches for BWDPCs under various carbonization/activation conditions and performed the CO_2 uptake using different adsorption parameters,[1,14] verifying that BWDPCs can be used for practical CO_2 capture. However, the optimization of the synthesis route by integrating carbonization and activation together with a reasonable guideline is still unclear. In addition to adsorption parameters, the textural properties and functional groups of porous carbons are widely considered as dominant factors of CO_2 capture performance;[1,6,11,16] however, the means to prioritize these three aspects still remains unclear; a prioritization technique would be beneficial for guiding the synthesis of porous carbons from biomass waste.

The application of machine learning (ML) in fields, including waste-to-energy conversion,[19,20] biochar for metal and organic compound sorption,[21,22] municipal solid-waste treatment,[23] and oxidation of micropollutants[24] has received profound interest recently. Among the various types of ML models, including linear regression, support vector machines (SVMs), *k*-nearest neighbors, and artificial neural networks (ANNs),[25,26] tree-based ML models are a subcategory of supervised ML methods, which employ recursive binary splitting of data in a manner such that the residual sum of squares is minimized.[19] Some of the most common and popular models include decision trees (DTs), random forest (RF), gradient boosting decision trees (GBDTs), light gradient boosting machines (LGBs), and extreme gradient boost (XGB). Despite being relatively new, the latter three of the boosting tree-based models have seen a surge in interest in their popularity and application in scientific research because of their ability to work with smaller data sets, resistance to overfitting, and the ability to counter noisy features.[27] In comparison to the conventionally used RF, boosting trees offer advantages,[28,29] such as providing local as well as global predictions, assigning nonbiased weights to correlated features, and efficient handling of unbalanced data sets. Moreover, based on our in-house expertise, it was observed that while working with relatively small data sets (300−1000 data points), these boosting tree algorithms offer faster hyperparameter tuning in comparison to the more commonly used SVM and ANN algorithms without compromising the accuracy.

Given the aforementioned gaps, a data-driven approach was used to mine and map the CO_2 adsorption by BWDPCs based on their textural properties and compositional features, along with the adsorption temperature and pressure at which the CO_2 adsorption experiments were conducted. The main purpose of this study is to elucidate how ML tools can be systematically applied and leveraged for predictive analytics

https://doi.org/10.1021/acs.est.1c01849
Environ. Sci. Technol. 2021, 55, 11925−11936

and used to draw valuable insights into the process of CO_2 adsorption using BWDPCs. With this aim, three tree-based ML models, namely, GBDT, LGB, and XGB, were devised and evaluated for the prediction performance of BWDPCs for CO_2 adsorption. Based on these predictive models, a feature importance evaluation was performed, and the impact of each input feature on the target variable was realized via partial dependence plots (PDPs) (Figure 1).

METHODOLOGY

Data Collection and Formatting. For data collection, a comprehensive literature review on BWDPCs for CO_2 capture was conducted using various keywords (including biomass waste, waste, biochar, porous carbon, CO_2 adsorption, and CO_2 capture) using major databases (e.g., SCOPUS, Web of Science, etc.). A total of 76 peer-reviewed publications in the last decade were investigated, and 632 data points were extracted and used here. Tables S1 and S2 provide a summary of the data sets and the publications referred, respectively. Herein, it is worth noting that the common principal criteria for synthesizing CO_2 adsorbents derived from BWDPCs included excellent adsorption capacity and selectivity, stable working capacity, cost-effectiveness, recyclability, easy regeneration, and rapid adsorption−desorption kinetics.[16] However, we primarily focused on the CO_2 adsorption capacity obtained at different temperatures and pressures along with textural and compositional properties of BWDPCs for two main reasons: (1) most of the publications reviewed mainly focused on adsorption capacity and only very few of them reported working capacity, regeneration, and kinetic properties; thus, causing a lack of data to develop ML models for all the aforementioned essential properties; and (2) in addition to the adsorption capacity, the performance standards for the other criteria were unavailable due to a lack of environmental impact and techno-economic assessment.

To this end, the following assumptions and strategies were incorporated during the data collection process.

1. All screened data were initially accepted impartially, without any initial judgment or bias regarding data validity.
2. Most of the data were obtained from experimental reports listed by researchers. For the data points that were not directly listed in a table or elaborated in the text form, we performed extraction from the figures using WebPlotDigitizer software (https://apps.automeris.io/wpd/) to obtain the necessary data. All values were carefully screened to avoid duplicate or multiple entries.
3. The list of input features used here were broadly classified into three categories: (1) textural properties, (2) elemental compositions of the BWDPCs, and (3) adsorption parameters such as temperature and pressure at which CO_2 adsorption experiments were conducted.
4. The textural properties of the BWDPCs included their surface area (SA, m^2/g), total pore volume (TPV, cm^3/g), and micropore volume (MPV, cm^3/g) and the elemental compositions of carbon, hydrogen, nitrogen, and oxygen contents (wt %).
5. The CO_2 uptake by the BWDPCs at different adsorption parameters was used as the target variable.

The 632 data points were divided into two subsets: (1) regular porous carbons (RPCs) with 288 data points collected from 34 publications (Table S1) and (2) heteroatom-doped porous carbons (HDPCs) with 344 data points from 42 publications (Table S2). All these data points were directly collected either from the tables or extracted from the figures presented in 76 of the referred publications here.

Data Preprocessing. Once all the data were collected and transformed into consistent units (described in Supporting Information S1), a substantial amount of data was missing either for TPV or MPV. This condition was primarily due to the difference in the reports published by various researchers in terms of the choice of textural property reported. Most researchers reported the SA and TPV or SA and MPV; hence, several rows of data had either SA or TPV, whereas MPV was missing, and in other cases, the data on SA and MPV were available, but TPV was missing. Thus, missing data imputation of TPV and MPV via ML concepts was employed to avoid discarding records with missing values. Specifically, TPV and MPV were mapped as a function of SA using multiple ML algorithms, including multilinear regression and RF models, and the missing value for a given entry was predicted and imputed (described in Supporting Information S2). Additionally, with respect to the data on the compositions of BWDPCs, several researchers have reported the compositions of carbon, nitrogen, carbon, and hydrogen only. In an event where the compositions of BWDPCs could not be summed to 100% by weight, the data point was discarded. The final data set used here included 527 data points, 193 of which belonged to the RPC subset; whereas, the remaining 334 belonged to the HDPC subset.

The correlation, and more specifically, the linear dependency among the input variables, was measured by Pearson's correlation coefficient (PCC) given by eq 1

$$\rho_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})\sum_{i=1}^{n}(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

where ρ_{xy} is the value of PCC for feature to target or target to target and $\bar{x}$ and $\bar{y}$ are the means of input feature x and output target y, respectively. The range of ρ_{xy} varied from −1 to 1, where 0 indicated no linear correlation, and a high negative or positive value indicated a strong negative or positive correlation, respectively.

Finally, after data compilation, preprocessing for consistent units, and handling of missing values, the data (input features and output targets) were normalized to achieve a uniform range of values, according to eq 2

$$x_i^* = \frac{x_i - \mu}{s} \quad (2)$$

where x_i is the value of input feature i, x_i^* is the normalized value of the initial x_i, μ is the mean of x_i;, and s is the standard deviation of x_i.

The entire preprocessed 527 data points were subjected to multitraining by splitting them into randomly chosen training and test subsets. Thus, 85% of the total data points were randomly selected and labeled as training data, and the remaining 15% of the data points were labeled as test data for the final evaluation of the developed models. During the training process, the k-fold cross-validation (k-fold CV) method was applied to tune the hyperparameters to simultaneously improve the model's prediction and generalization capabilities.[19] In k-fold CV, the training data were divided into k subsets such that at each training iteration, one

https://doi.org/10.1021/acs.est.1c01849
Environ. Sci. Technol. 2021, 55, 11925−11936

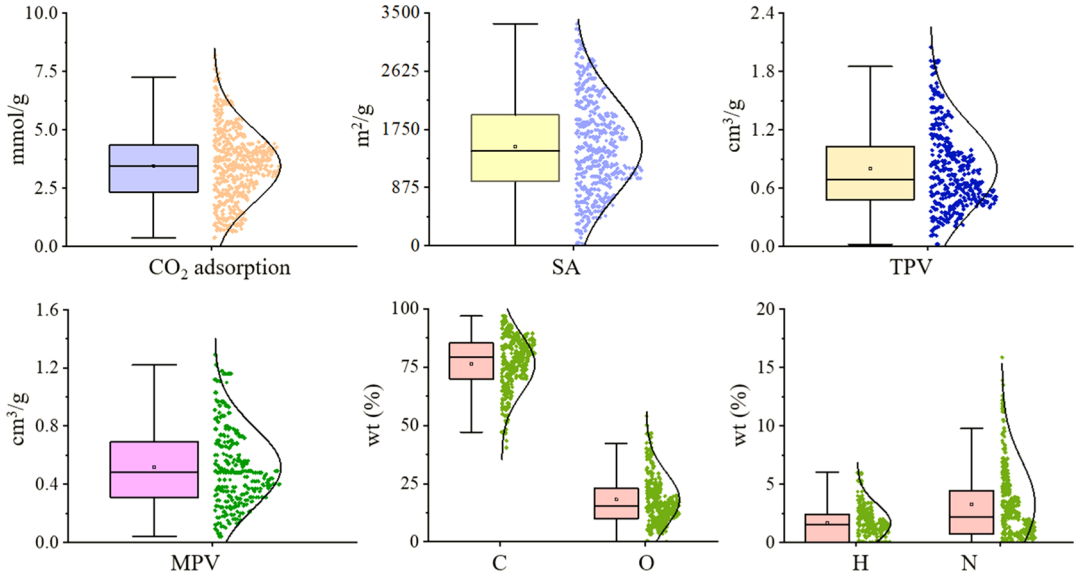


Figure 2. Box-normal plots representing the descriptive statistics of the data for each feature collected from the literature. The range of values for both the target and input features are as collected from the total set of 76 papers and 632 data points. Most of the features showed a normal distribution of data, except for TPV and MPV, where a subtle left skew was observed, and in the case of the nitrogen content, where a stronger left skew was witnessed. The abbreviated features with their respective units are as follows—SA: surface area (m^2/g), TPV: total pore volume (cm^3/g), MPV: micropore volume (cm^3/g), C: carbon content (wt %), O: oxygen content (wt %), H: hydrogen content (wt %), and N: nitrogen content (wt %).

of the k subsets was used as the validation set and the other $k - 1$ subsets form the training set to address the variance and bias issue in ML.[30] As a general rule of thumb and depending on the size of the data set, k = 3, 5, or 10 is generally preferred, and much of this also depends on the modeler's intuitions. Specific to this study, the training set had a total of 448 data points (85% of 527). Thus, using k = 5 yielded five subsets with approximately 90 data points, each of which was then effectively used during the modeling.

Modeling Methods and Hyperparameter Tuning. Three tree-based ML algorithms (e.g., GBDT, LGB, and XGB) were compared and evaluated to predict CO_2 adsorption on BWDPCs. Recent works of our group[19−21] have shown the suitability and successful implementation of tree-based ML models in data sets with 200−1000 data points when the set of input features normally ranges between 5 and 15. Such data sets are not big data but are rather the data collected from experimental works or those that are published in the literature and can be termed as midsize data sets.

The GBDT is a class of ensemble learning algorithms that combines several sequentially connected DTs.[31] DTs by themselves are occasionally referred to as relatively weak learners. However, in the case of GBDTs, by adding several DTs in a series (e.g., boosting), whereby each subsequent tree minimizes the errors from the previous one, the process of "boosting" becomes highly efficient, thus leading to the high accuracy of GBDT models. XGB, which is a scalable boosting tree and a variant of the GBDTs, employs numerous DTs and uses the weighted quantile search to aid in parallel and distributed computing, resulting in high computational efficiency and prediction accuracy.[30,32] LGB, which is a recent variant of GBDTs, uses gradient-based one-side sampling and exclusive feature bundling to improve computational efficiency without affecting the prediction accuracy of the model.[33]

Hyperparameter tuning is the process of finding a set of hyperparameters to attain the optimal model performance. These hyperparameters are algorithm-specific and are tuned during the model training process. The most common hyperparameter tunings reported in the literature include manual search, grid search, and Bayesian optimization.[31] In most studies, multiple hyperparameter tuning processes are evaluated and then selected based on the corresponding model accuracy or arbitrarily. In this study, the grid search method was employed for hyperparameter tuning, given its reliability and easy implementation when tuning for a lower set of input features (nine input features).

Error Metrics. The performance of the regression models was evaluated in terms of R^2 (correlation coefficient) and the root mean square error (RMSE).[21,34] Conceptually, the higher the R^2 and the lower the RMSE, the greater the model accuracy, as described in eqs 3 and 4, respectively.

$$R^2 = 1 - \frac{\sum_{n=1}^{N} (\hat{y} - y)^2}{\sum_{n=1}^{N} (\hat{y} - \bar{y})^2} \tag{3}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{n=1}^{N} (\hat{y} - y)^2}{N}} \tag{4}$$

where $\hat{y}$, y, and $\bar{y}$ are the predicted, actual, and mean values of the target feature, respectively, n is the data point at any given instance, and N is the total number of data points.

Feature Importance. Tree-based models are black boxes by nature. Thus, determining the significance or contribution of a specific input feature to the model is a challenge. Permutation importance[35] or mean decrease accuracy (MDA) is a model agnostic method used to determine the importance of each feature for any black-box model by measuring how the model performance (in terms of its accuracy) changes when a feature is added or discarded. Specific to tree-based models, MDA is the mean decrease in the Gini score, a measure of how each variable contributes to the homogeneity of the nodes and leaves of the tree (single as well as ensemble).[36] The higher the

https://doi.org/10.1021/acs.est.1c01849
Environ. Sci. Technol. 2021, 55, 11925−11936

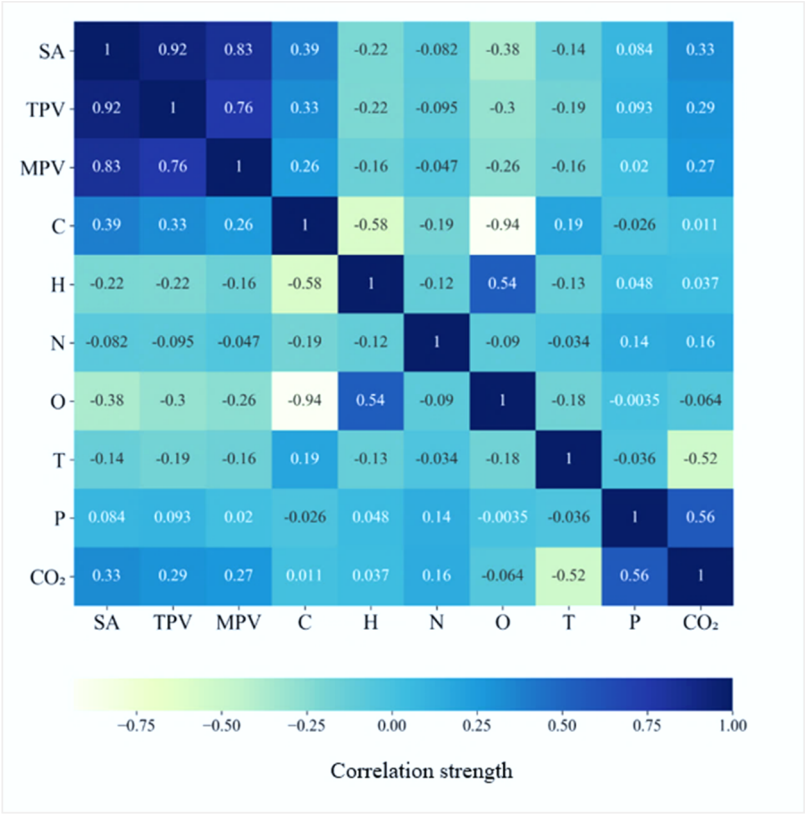


Figure 3. Pearson's correlation matrix for all the set of features included in the study. No significant correlation among the input variables was observed, except for textural properties including SA, TPV, and MPV. The abbreviated features with their respective units are as follows—SA: surface area (m^2/g), TPV: total pore volume (cm^3/g), MPV: micropore volume (cm^3/g), C: carbon content (wt %), H: hydrogen content (wt %), N: nitrogen content (wt %), O: oxygen content (wt %), *T*: temperature (°C), *P*: pressure (bar), and CO_2: CO_2 adsorption (mmol/g).

value of the MDA or the mean decrease Gini score, the higher the importance of the variable in the model. The method is suitable for determining the feature importance when the number of features is not significantly high because, otherwise, it can be resource-intensive. Given that this study involved nine input features, the permutation importance method was selected to determine the feature importance. Although permutation importance provides useful insights into determining the respective contribution of a feature to the model prediction as a whole, it does not necessarily help in understanding how these input features affect the target variable. To this cause, we also employed PDPs to understand and visualize the relationship between each input feature and the target variable.[21] The PDPs have a regressor function that marginalizes the impact of all the input features on the ML prediction,[37] except for the one (or two if required) of interest, thereby presenting local sensitivity analysis. Thus, by marginalizing over the other features, a function that depends only on one input feature is achieved, for example, one feature at a time, and its impact on the target feature is measured across each instance/data point in the data set. Because the GBDT itself offers both global and local predictions, local sensitivity analysis in the form of MDA and PDPs was preferred in this study, such that the impact of the input features on the target variables was not just averaged over the entire data set but calculated across each instance/data point. However, any sensitivity analysis, be it global or local, can be prone to bias,[38] and to this extent, the findings of the PDPs were further substantiated with the existing domain knowledge and literature. All the data processing and ML methods described here were performed in the Python programming language (version 3.6), using the open-source scikit-learn library.[39]

RESULTS AND DISCUSSION

Descriptive Statistics. A descriptive analysis of all the input features and the target variables was first performed based on the raw data; for example, the data collected and compiled (632 data points) were first subjected to descriptive analytics in terms of the minimum, maximum, and average values of the input features and target variables to gain preliminary insights into the raw data. Figure 2 presents the visual representation of the data distribution for each of the input features and target variables in the form of box-normal plots. The mean values of CO_2 adsorbed on the porous carbon as per the data collected from the literature was 3.36 mmol/g with a standard deviation of 1.58 mmol/g, with the maximum CO_2 adsorption at 8.20 mmol/g at 0 °C and 1 bar; whereas, the least was in the case of 0.25 mmol/g at 25 °C and 0.15 bar (Figure 2a). The SA was the only textural property that was reported in all the studies selected for data collection here. The reported SA in the literature varied from 0.15 to 3336 m^2/g, with a mean value of 1436.70 m^2/g and a high standard deviation equal to 764 m^2/g (Figure 2b). The mean values for TPV and MPV were 0.78 ± 0.48 and 0.52 ± 0.48 cm^3/g, respectively (Figure 2c,d). As summarized in Tables S1 and S2, the SA and pore volume were significantly affected by the carbonization and activation treatments. Yang et al.[2] synthesized coconut shell-derived porous carbons for CO_2 capture using ammoxidation with KOH activation. The results

https://doi.org/10.1021/acs.est.1c01849
Environ. Sci. Technol. 2021, 55, 11925−11936

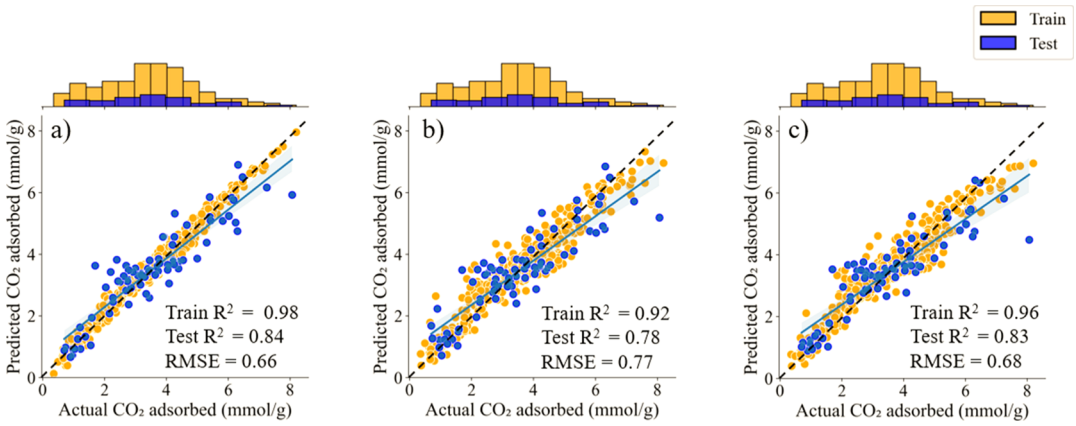


Figure 4. Model prediction (a) GBDT, (b) LGB, and (c) XGB on the master data set. The comparative evaluation among these three tree-based models showed that the GBDT had the best prediction performance on the training and test data set, closely followed by the XGB and LGB models. The fivefold CV method employed for all the models resulted in the development of generalized models as the R^2 on the fivefold CV set and the test data set closely matched each other. The blue shades represent 95% confidence intervals of the regression line on the test points. The black dashed lines represent the line of equality ($y = x$).

demonstrated that the SA varied from 879 to 2690 m^2/g and the TPV from 0.38 to 1.34 cm^3/g. However, the highest CO_2 uptake of 4.26 mmol/g at 25 °C and 1 bar was achieved by porous carbon with an SA of 1483 m^2/g and a TPV of 0.66 cm^3/g. This finding implied that no simple and direct route was available to synthesize optimal porous carbons for practical CO_2 capture based on different biomass wastes. According to 76 publications referred to here (summarized in Supporting Information and Tables S1 and S2), the researchers only selected the best CO_2 adsorbent among the several prepared samples derived from biomass waste, implying that no clear and efficient guidance for the synthesis of high-performance CO_2 adsorbents from biomass waste was available. In addition, the developed porous carbon was applicable to practical CO_2 capture processes if the CO_2 adsorption capacity exceeded 3 mmol/g.[40] Thus, it was concluded that most biomass waste could be upcycled into promising candidates for CO_2 capture applications through effective carbonization and activation treatments.

Among the compositional features of BWDPCs (including RPCs and HDPCs), the mean elemental compositions of C, H, N, and O were 75.84, 1.65, 3.22, and 18.70 (wt %), respectively (Figure 2e,f). The carbon content dramatically increased, and the oxygen content decreased compared with the elemental analysis data of raw biomass waste.[41−44] These findings were mainly caused by the release of volatile matters and chemically bonded moisture from the biomass waste. In terms of adsorption experiments, we mainly considered all the adsorption experiments carried out at temperatures of either 25 or 0 °C and pressures of 1 or 0.15 bar, given that most of the CO_2 adsorption performances in previous publications were extensively evaluated using pure CO_2 gas and/or 15 vol % CO_2 gas balanced with N_2 (similar to flue gas compositions) at room temperature (25 °C) and/or 0 °C.

Upon observing the correlation among the set of input variables, a strong positive correlation was observed in the case of the textural properties (including SA, TPV, and MPV). The PCC values for these variables were higher than 0.75, indicating a strong correlation. However, excluding the textural properties, no significant correlation (PCC values mostly between −0.5 and 0.5) was observed among the rest of the input variables except between C and O. Figure 3 presents the PCC matrix. The lack of a strong correlation among the various input features helped retain all of them toward building the predictive model, as each feature would then contribute individually to the model. In addition, although there was a strong correlation among the textural properties, this class of input features contained much of the missing data in the raw data set. Rather than deleting the instance of missing data points, which would significantly lower the size of the data set, this strong correlation was further used to impute the missing data (described in Supporting Information S2).

Model Prediction. Three tree-based ML models, specifically GBDT, XGB, and LGB, were devised and evaluated to predict CO_2 adsorbed on the BWDPCs based on the set of input features described in the Data Preprocessing section. Each model was developed based on a training data set, where the hyperparameters were tuned via a grid search based on the fivefold CV. Once the ML models were developed and their hyperparameters were optimized, they were evaluated for their prediction accuracy on the test data set. Figure 4 shows the joint scatter plots of the actual versus predicted values of CO_2 adsorption on the BWDPCs, as determined by the various models. The plots revealed that although the performances of all the models on the training and test data were comparable, the GBDT model presented the most accurate predictions for the test data. Represented as R^2, the goodness of fit for the GBDT, LGB, and XGB models were 0.98, 0.92, and 0.96, respectively, on the training data. The values were 0.84 (GBDT), 0.78 (GBDT), and 0.83 (XGB) for the test data. It was apparent that the GBDT had better prediction performance than the other two models, which was further corroborated by the fact that the GBDT had the lowest RMSE among the models at 0.66.

The prediction results of all the models on the training and test data indicated the possibility of overfitting. Overfitting is a condition in which the model aptly learns the noises and random fluctuations in the training data to the extent that it negatively impacts the prediction performance of the model when it encounters unseen data in the test data. Thus, the generalization capability of the model is compromised. The possibility of overfitting in this study was evidenced by the fact that training R^2 for all the models was in the range of ≥0.9, whereas test R^2 was comparatively lower at a value of 0.8. It is

https://doi.org/10.1021/acs.est.1c01849
Environ. Sci. Technol. 2021, 55, 11925−11936

worth mentioning that this was not a case of acute overfitting, but rather a minor one as the difference in training and test R^2 for all the models was off by values of ∼0.12−0.15.

To address this issue, the k-fold CV was effectively used, where the training data were divided into k subsets (k = 5 in this study); at each training iteration, one of the k subsets was used as the validation set and the other $k - 1$ subsets formed the training set. The prediction accuracy was averaged over all k trials to obtain the total effectiveness of the model. This significantly reduced the bias, as much of the data were used for fitting during the training process, and simultaneously reduced the variance, as most of the data were also being used in the validation set. Evidently, the overfitting was effectively countered as the results of the k-fold CV (k = 5) in terms of R^2 closely matched that of the test data, thereby improving the generalization of the model (Table 1).

Table 1. Comparative Evaluation of Tree-Based ML Models Using the Master Data Set

	GBDT	LGB	XGB
training R^2	0.98	0.92	0.96
5-fold CV R^2	0.81	0.78	0.81
test R^2	0.84	0.78	0.83
RMSE (mmol/g)	0.66	0.77	0.68

The entire data set was additionally categorized into RPC and HDPC subdata sets. The entire process of model development and tuning of the respective training sets and evaluation of the test data set was performed for the RPC and HDPC subdata sets. Figure 5a−f presents the joint scatter plots of the actual versus predicted values of CO_2 adsorption on RPCs and HDPCs, as determined by the three models. Although the GBDT and XGB exhibited comparable performances in the training and CV, the former marginally outperformed the latter in terms of higher test R^2 at 0.86 and 0.79, and least RMSE at 0.61 and 0.69 in the RPC and HDPC subdata sets, respectively. These results suggested that all the models evaluated here had comparable performances and served their intended purposes. In general, the GBDT model slightly outperformed the XGB and LGB models when encountering unlabeled data and was subjected to less overfitting given the minimum difference in R^2 between the training and test, thus, proving its generalization capabilities (Table 2).

In addition to the comparative evaluation of the models, several important observations were made. First, despite the fact that the GBDT model performed better than XGB on the RPC and HDPC subdata sets, the fivefold CV R^2 and test R^2 of the XGB model attained were similar to those of the GBDT model. This condition might have been due to the training of the XGB model on more data (e.g., the master data set had 527 data points compared with the RPCs and HDPCs, which had 193 and 334 data points, respectively), allowing models to learn better and make more accurate and generalized predictions. Intuitively, boosting trees and ML models in general perform better with substantial amounts of data (data quality is a significant factor not discussed here), which could account for the aforementioned observation. Another interesting observation was that the LGB model, which had the worst prediction performance in the master data set and HDPC subdata set, yielded the best prediction performance in the RPC subdata set. Although it could be argued that all three models provided high prediction accuracy in the RPC subdata set, this implicitly indicated that the quality of data on the RPC subdata set was not sufficiently enriched (e.g., most of the data spread for the input features would have been centered around certain mean values), thus failing to significantly contribute to the improvement of the model performance on the master data set.

Feature Importance Analysis. The agnostic ML model explainer, for example, the permutation importance or MDA, was used to determine the effects of the set of input variables, which included the elemental compositions and textural properties of BWDPCs and the adsorption parameters on the target variable (e.g., CO_2 uptake). This study was specifically conducted for the GBDT model, which was identified as the best-performing model in this research. Figure 6a presents the overall influence of each input feature on the target variable. The figure depicts the absolute MDA values for each input feature on the variable. A high MDA value for the input feature intuitively indicates a decrease in the prediction performance of the model when the feature is discarded. Thus, an input feature with a high MDA indicates a significant effect on the model prediction.

The permutation importance plots for the entire data set yielded interesting observations. The pressure and temperature at which the adsorption experiments were carried out were the most significant parameters influencing the model prediction. This finding was followed by the textural properties of SA, TPV, and MPV in the order of decreasing precedence and finally by compositional features in the decreasing order of N > O > H > C. Based on the ranking of the MDA values, the importance of input features was distinctively observed and classified into three categories, with the adsorption process parameters being the most significant, followed by the textural and compositional properties of biochar.

The feature importance test was also conducted on the RPC and HDPC subdata sets. Figure 6b,c shows the results, respectively. A similar trend was observed for the RPC subdata set, in which the adsorption pressure and temperature were the most significant, followed by the textural properties of BWDPCs. Specifically, the adsorption pressure, temperature, and MPV were the top three important input features for the model prediction on the RPC data set. Similarly, in the HDPC data set, the adsorption temperature, pressure, and SA were the most dominant factors. The physisorption driven by weak van der Waals forces dominated the porous carbon-based CO_2 capture process.[6,11,45] This finding implied that adsorption parameters (e.g., temperature and pressure) dramatically affected the CO_2 adsorption performance, such as the decrease in the CO_2 uptake with the increase in adsorption temperature and/or decrease in adsorption pressure. In addition, the MPVs (especially those of the narrow pores under 0.6 or 0.8 nm in diameter) demonstrated a linear relationship with CO_2 uptake at specific adsorption parameters,[6,45−47] for example, the effective pore size must be at least twice the kinetic diameter of the CO_2 molecule (0.33 nm).

Notably, in all three data sets, N contributed the most among the compositional features. The critical role of the N content in the porous carbon-based CO_2 capture process was verified in this study, consistent with the findings of previous research.[11,15,48,49] This suggested that most of the N content could be considered as effective N-functional groups for enhancing the CO_2 adsorption capacity, based on the ML model developed here. Therefore, effective N-doping treat-

https://doi.org/10.1021/acs.est.1c01849
Environ. Sci. Technol. 2021, 55, 11925−11936

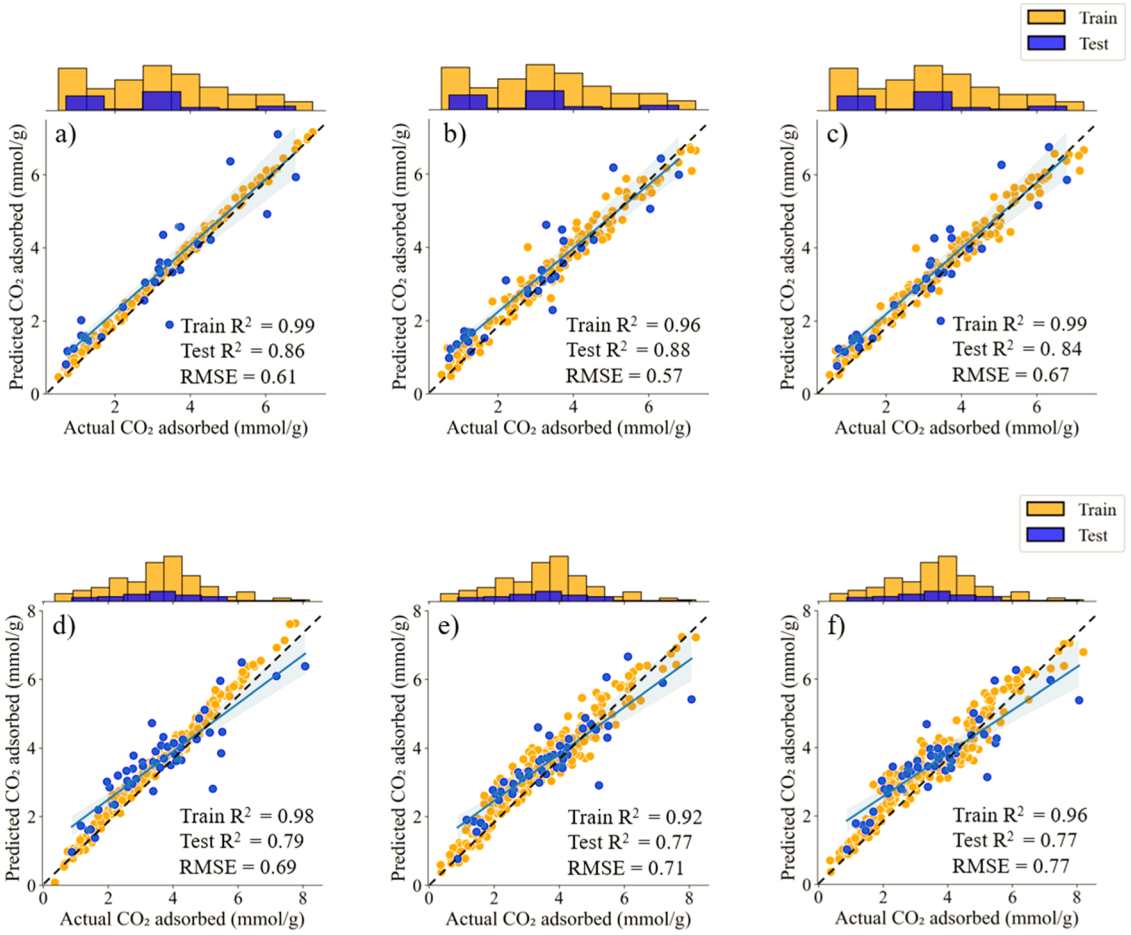


Figure 5. Model prediction on the RPC and HDPC data set. (a−c) GBDT, LGB, and XGB for the RPC subdata set, respectively. (d−f) GBDT, LGB, and XGB for the HDPC subdata set, respectively. The blue shades represent 95% confidence intervals of the regression line on the test points. The black dashed lines represent the line of equality ($y = x$). The prediction performance of the various models on the HDPC subdata set followed similar trends to that of those in the master data set, but with less accuracy. Interestingly, in the RPC subdata set, the LGB model showed the highest accuracy on the test data set, despite having the least performance in all other cases—which could be partly attributed to the RPC subdata set having a smaller dispersion of data, thus allowing the LGB to learn and predict effectively.

Table 2. Comparative Evaluation of the Tree-Based ML Models on the RPC and HDPC Subdata Sets

	RPCs			HDPCs		
	GBDT	LGB	XGB	GBDT	LGB	XGB
training R^2	0.99	0.96	0.99	0.98	0.92	0.96
5-fold CV R^2	0.88	0.86	0.88	0.76	0.67	0.73
test R^2	0.86	0.88	0.84	0.79	0.77	0.77
RMSE (mmol/g)	0.61	0.57	0.67	0.69	0.71	0.71

ments have been widely investigated to enhance the CO_2 uptake and selectivity over other gases through a simple acid−base interaction.[2,48−51] Lastly, the overall feature importance followed similar trends in each of the three data sets (e.g., adsorption parameters > textural properties > compositional properties), whereas the individual parameters in these categories showed subtle changes. Through the average of the contributions for each of the feature classes in all the three data sets used here, it was realized that adsorption parameters, textural properties, and compositional properties contributed 62, 27, and 11%, respectively, toward the target prediction.

In addition to permutation importance analysis, PDPs were devised to better understand how each input feature affected CO_2 adsorption. The CO_2 uptake could be effectively enhanced by increasing the SA of the BWDPCs to 1800 m^2/g. However, a further increase in the SA to 2000−3000 m^2/g did not significantly increase the CO_2 uptake (Figure 7a). Figure 7b exhibits an optimal TPV of 0.6−0.9 cm^3/g for the maximum CO_2 uptake, and Figure 7c displays that the CO_2 uptake linearly increased with an MPV up to 0.6 cm^3/g, after which no significant increase was observed. These findings were mainly due to the pore sizes of 0.6 and 0.8 nm, which were verified as critical indicators for CO_2 capture at 0.15 and 1 bar, respectively.[47] Moreover, a linear relationship was detected between the MPV of narrow pores less than 0.6 or 0.8 nm and the CO_2 uptake at different adsorption conditions.[6,47] Thus, CO_2 adsorption capacity could continually increase to a certain value when the SA reaches 1800 m^2/g or the MPV reaches 0.6 cm^3/g; it might not increase significantly to higher levels with the further increase in the SA and MPV. As shown in Figures 7d,e, a nitrogen content of 7−10 wt % and oxygen content of 10−14 wt % could effectively enhance the CO_2 uptake, mainly due to nitrogen- and oxygen-based basic functional groups (e.g., pyrrolic-N, pyridinic-N, −OH, etc.) via the Lewis basic theory.[6,15] In addition, several biomass wastes that are naturally rich in nitrogen content are widely considered as promising candidates for synthesizing porous carbons with high CO_2 capture performance.[11] Figures 7f,g

https://doi.org/10.1021/acs.est.1c01849
Environ. Sci. Technol. 2021, 55, 11925−11936

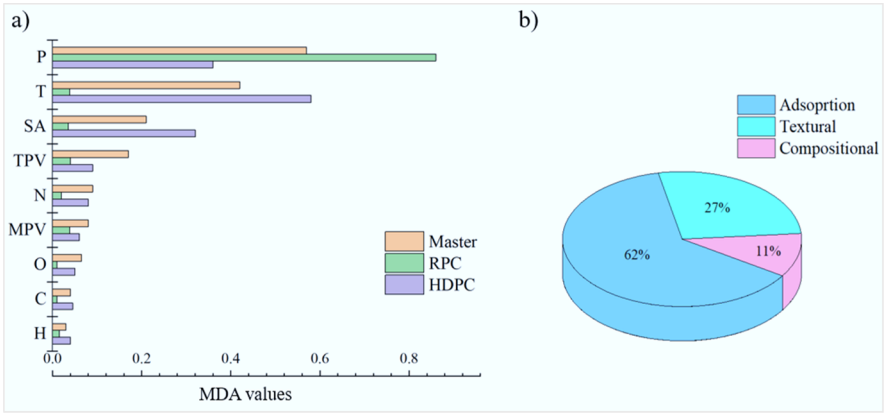


Figure 6. Feature importance analysis represented in the form of (a) 3D bar graphs for the MDA values in the master data set, RPC subdata set, and HDPC subdata set. (b)Pie chart representing the average of the contributions for each of the feature classes (adsorption parameters, textural properties, and compositional properties) toward the target prediction. A recurring trend was observed in all the data sets that led to the conclusion that adsorption parameters > textural properties > compositional properties for CO_2 adsorption by the BWDPCs. The abbreviated features with their units are as follows—*T*: temperature (°C), *P*: pressure (bar), SA: surface area (m^2/g), TPV: total pore volume (cm^3/g), MPV: micropore volume (cm^3/g), N: nitrogen content (wt %), O: oxygen content (wt %), C: carbon content (wt %), and H: hydrogen (wt %).

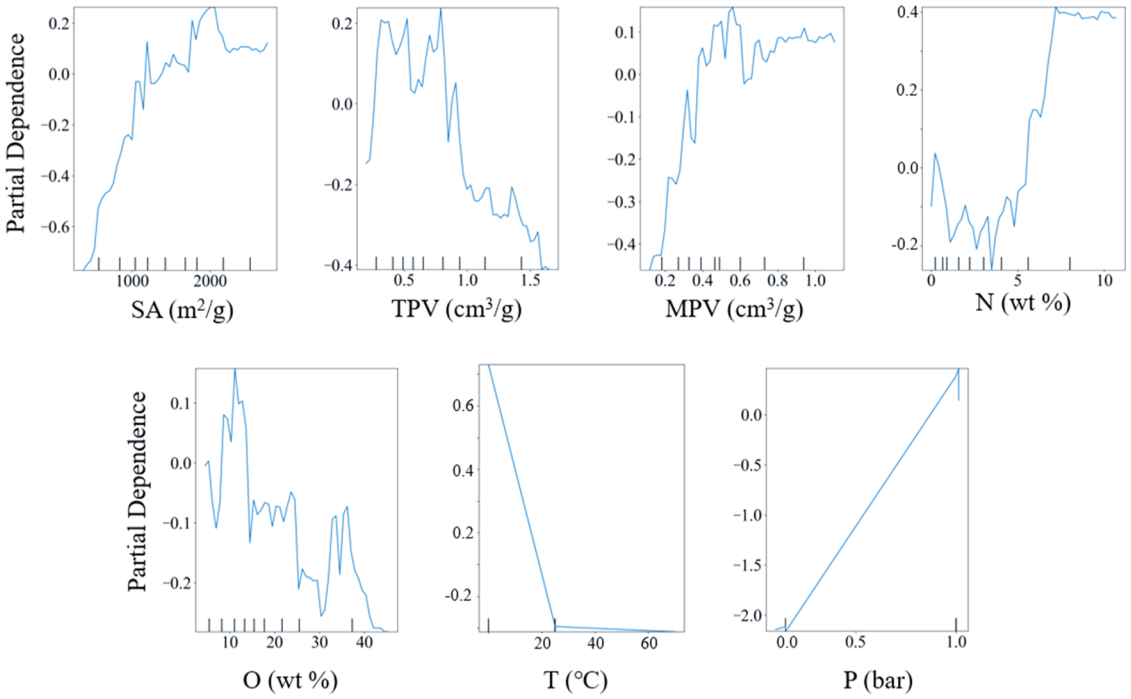


Figure 7. PDPs to understand how individual input features impacted the CO_2 adsorption on BWPDCs. They revealed that the PDP CO_2 adsorption capacity could keep increasing to a certain value when the SA reaches 1800 m^2/g or MPV reaches 0.6 cm^3/g and may not increase significantly to higher levels with further increase. Moreover, low temperatures and high pressures are preferred for CO_2 adsorption. Though strong evidence for the compositional features was not effectively realized, the nitrogen content in the range of 7−10 wt % and oxygen content of 10−14 wt % could effectively enhance CO_2 adsorption. The abbreviated features with their units are as follows—SA: surface area (m^2/g), TPV: total pore volume (cm^3/g), MPV: micropore volume (cm^3/g), N: nitrogen content (wt %), O: oxygen content (wt %), *T*: temperature (°C), and *P*: pressure (bar).

shows that low temperatures and high pressures were preferred for CO_2 adsorption. This result depended mainly on the natural CO_2 adsorption mechanism using BWDPCs. To date, it has been concluded that physisorption mainly dominates CO_2 adsorption on carbon-based porous carbon, considering that (a) the CO_2 adsorption capacity decreases with increasing adsorption temperature or decreasing adsorption pressure, and (b) the isosteric heat of adsorption (Q_{st}) generally ranges from 20 to 50 kJ/mol.[3,6,16]

The permutation importance and PDPs contributed to revealing essential insights into the adsorption of CO_2 by BWDPCs, which were also supported by the existing literature.[6,11,15,16,47] Although the permutation importance plot revealed how each input feature contributed to model prediction, the PDPs helped in realizing how CO_2 adsorption varied when the adsorption parameters or textural properties of the BWDPCs varied. Through the combined inference derived from the permutation importance plots and PDPs, the priority for achieving high CO_2 adsorption using BWDPCs could be

https://doi.org/10.1021/acs.est.1c01849
Environ. Sci. Technol. 2021, 55, 11925−11936

given as follows: adsorption parameters > textural properties > compositional properties. Given that physisorption dominated the CO_2 capture process using BWDPCs, the CO_2 adsorption on the surface of BWDPCs was mainly driven by van der Waals forces, which were drastically affected by the operating temperature and pressure. In addition to adsorption parameters, compared with nitrogen and oxygen functional groups, textural properties (especially the limited MPV under 0.6 or 0.8 nm pore diameter) played a more critical role in the high-performance CO_2 capture process.

Insights from feature analysis after the development of ML models facilitated the derivation of vital information required to deduce how the ML model works. This information is relevant in the field of chemical and material science, where such knowledge could be leveraged for process optimization in the design of experiments or for guiding material synthesis compared with conventional intuition and trial-and-error-based synthesis. Specific to the application of BWDPCs for CO_2 adsorption, a field that has rapidly progressed in the last decade, challenges still persist, with the published literature presenting subtle yet contrasting views on the critical factors that influence CO_2 adsorption on BWDPCs.[11,13,16,52] These findings were limited to a given set of experiments (most likely from the respective research groups) and could lead to bias in the interpretation. By contrast, a purely data-driven approach was used here, in which a wide variety of published experimental data were used for predictive analytics, followed by feature importance, presenting a generalized approach and objective analysis. Moreover, this modeling approach is cross-deployable and can be used to investigate other porous carbons (e.g., zeolites, metal−organic frameworks, etc.) for CO_2 adsorption and not just those derived from biomass waste. In addition, given that the findings and inferences of these data-driven models were substantiated by the existing literature, these models offered certain mechanistic knowledge regarding the actual CO_2 capture process, which increased the confidence in their acceptance. Thus, based on the insights from this study, our next endeavor is twofold: (1) to devise a guided synthesis strategy to maximize CO_2 adsorption, which will be rationalized by optimization of the adsorption parameters and textural properties of BWDPCs; and (2) experimental validation of the optimizer, followed by iterative addition of the new experimental data points to the existing database and making the database open source to allow other researchers in the community to leverage from it.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.est.1c01849.

BWDPCs without heteroatom-doping treatment, BWDPCs with heteroatom-doping treatment, descriptive statistics of the textural properties prior to data imputation, comparative evaluation of the linear regression and RF model to map TPV and MPV as a function of SA, and PCC plot of the textural properties (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Xiaonan Wang − *Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117585, Singapore;* orcid.org/0000-0001-9775-2417; Phone: +65 6601 6221; Email: chewxia@nus.edu.sg

Yong Sik Ok − *Korea Biochar Research Center, APRU Sustainable Waste Management Program & Division of Environmental Science and Ecological Engineering, Korea University, Seoul 02841, Republic of Korea;* orcid.org/0000-0003-3401-0912; Phone: +82 2 3290 3044; Email: yongsikok@korea.ac.kr

Authors

Xiangzhou Yuan − *Korea Biochar Research Center, APRU Sustainable Waste Management Program & Division of Environmental Science and Ecological Engineering, Korea University, Seoul 02841, Republic of Korea; R&D Centre, Sun Brand Industrial Inc., Jeollanam-do 57248, Republic of Korea*

Manu Suvarna − *Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117585, Singapore*

Sean Low − *Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117585, Singapore*

Pavani Dulanja Dissanayake − *Korea Biochar Research Center, APRU Sustainable Waste Management Program & Division of Environmental Science and Ecological Engineering, Korea University, Seoul 02841, Republic of Korea;* orcid.org/0000-0002-6934-5037

Ki Bong Lee − *Department of Chemical & Biological Engineering, Korea University, Seoul 02841, Republic of Korea;* orcid.org/0000-0001-9020-8646

Jie Li − *Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117585, Singapore*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.est.1c01849

Author Contributions

X.Y. and M.S. contributed equally to this work. X.Y.: data collection, writing (review and editing), and visualization; M.S.: modeling, writing (review and editing), and visualization; S.L.: modeling; P.D.D.: data collection; K.B.L.: writing (review and editing); J.L.: visualization and writing; X.W.: conceptualization, writing (review and editing), and supervision; and Y.S.O.: conceptualization, writing (review and editing), and supervision.

Notes

The authors declare no competing financial interest.

We present the ML pipeline developed in this study in the form of an open-source IPython notebook hosted on GitHub (https://github.com/ssuvarnamanu/CO2-adsorption-by-biowaste-based-porous-carbon). A clean Excel datasheet with the essential data is also attached to this repository, which can be directly imported into the IPython notebook. The pipeline is generic and cross-deployable in nature and can be used as a template for the analysis of other related CO_2 adsorbents.

■ ACKNOWLEDGMENTS

This work was supported by the Cooperative Research Program for Agriculture Science and Technology Development (Project no. PJ01475801), Rural Development Administration, Republic of Korea. This work was also supported by the National Research Foundation of Korea(NRF) grant funded

by the Korea government(MSIT) (No. 2021R1A2C2011734). Y.S.O, X.Y., and P.D.D. were partly supported by the KU Future Research Grant (KU FRG) Fund, Korea Biochar Research Center (KBRC) Fund, and the Association of Pacific Rim Universities (APRU) Sustainable Waste Management Program from the Korea University, Republic of Korea. X.W., M.S., S.L., and J.L. were supported by the Singaporean RIE2020 Advanced Manufacturing and Engineering (AME) IAF-PP grant "Cyber-physical production system (CPPS) toward contextual and intelligent response" by the Agency for Science, Technology and Research under grant no. A19C1a0018 and the model factory at SIMTech.

REFERENCES

(1) Creamer, A. E.; Gao, B. Carbon-based adsorbents for postcombustion CO_2 capture: A Critical Review. *Environ. Sci. Technol.* **2016**, *50*, 7276−7289.
(2) Yang, M.; Guo, L.; Hu, G.; Hu, X.; Xu, L.; Chen, J.; Dai, W.; Fan, M. Highly cost-effective nitrogen-doped porous coconut shell-based CO_2 sorbent synthesized by combining ammoxidation with KOH activation. *Environ. Sci. Technol.* **2015**, *49*, 7063−7070.
(3) Wang, J.; Yuan, X.; Deng, S.; Zeng, X.; Yu, Z.; Li, S.; Li, K. Waste polyethylene terephthalate (PET) plastics-derived activated carbon for CO_2 capture: a route to a closed carbon loop. *Green Chem.* **2020**, *22*, 6836−6845.
(4) Dissanayake, P. D.; Choi, S. W.; Igalavithana, A. D.; Yang, X.; Tsang, D. C. W.; Wang, C.-H.; Kua, H. W.; Lee, K. B.; Ok, Y. S. Sustainable gasification biochar as a high efficiency adsorbent for CO_2 capture: A facile method to designer biochar fabrication. *Renew. Sustain. Energ. Rev.* **2020**, *124*, 109785.
(5) Global Monitoring Laboratory. National Oceanic and Atmospheric Administration Trends in atmospheric carbon dioxide. https://www.esrl.noaa.gov/gmd/ccgg/trends/ (assessed May 2021).
(6) Yuan, X.; Lee, J. G.; Yun, H.; Deng, S.; Kim, Y. J.; Lee, J. E.; Kwak, S. K.; Lee, K. B. Solving two environmental issues simultaneously: Waste polyethylene terephthalate plastic bottle-derived microporous carbons for capturing CO_2. *Chem. Eng. J.* **2020**, *397*, 125350.
(7) Igalavithana, A. D.; Choi, S. W.; Shang, J.; Hanif, A.; Dissanayake, P. D.; Tsang, D. C. W.; Kwon, J.-H.; Lee, K. B.; Ok, Y. S. Carbon dioxide capture in biochar produced from pine sawdust and paper mill sludge: Effect of porous structure and surface chemistry. *Sci. Total Environ.* **2020**, *739*, 139845.
(8) Jahandar Lashaki, M.; Khiavi, S.; Sayari, A. Stability of amine-functionalized CO_2 adsorbents: a multifaceted puzzle. *Chem. Soc. Rev.* **2019**, *48*, 3320−3405.
(9) Osamn, A. I.; Hefny, M.; Abdel Maksoud, M. I. A.; Elgarahy, A. M.; Rooney, D. W. Recent advances in carbon capture storage and utilisation technologies: a review. *Environ. Chem. Lett.* **2021**, *19*, 797−849.
(10) Wang, J.; Huang, L.; Yang, R.; Zhang, Z.; Wu, J.; Gao, Y.; Wang, Q.; O'Hare, D.; Zhong, Z. Recent advances in solid sorbents for CO_2 capture and new development trends. *Energ. Environ. Sci.* **2014**, *7*, 3478−3518.
(11) Singh, G.; Lakhi, K. S.; Sil, S.; Bhosale, S. V.; Kim, I.; Albahily, K.; Vinu, A. Biomass derived porous carbon for CO_2 capture. *Carbon* **2019**, *148*, 164−186.
(12) Singh, V. K.; Anil Kumar, E. Measurement and analysis of adsorption isotherms of CO_2 on activated carbon. *Appl. Therm. Eng.* **2016**, *97*, 77−86.
(13) Danish, M.; Ahmad, T. A review on utilization of wood biomass as a sustainable precursor for activated carbon production and application. *Renew. Sustain. Energ. Rev.* **2018**, *87*, 1−21.
(14) Liu, W.-J.; Jiang, H.; Yu, H.-Q. Emerging applications of biochar-based materials for energy storage and conversion. *Energ. Environ. Sci.* **2019**, *12*, 1751−1779.
(15) Yuan, X.; Li, S.; Jeon, S.; Deng, S.; Zhao, L.; Lee, K. B. Valorization of waste polyethylene terephthalate plastic into N-doped microporous carbon for CO_2 capture through a one-pot synthesis. *J. Hazard. Mater.* **2020**, *399*, 123010.
(16) Singh, G.; Lee, J.; Karakoti, A.; Bahadur, R.; Yi, J.; Zhao, D.; AlBahily, K.; Vinu, A. Emerging trends in porous materials for CO_2 capture and conversion. *Chem. Soc. Rev.* **2020**, *49*, 4360−4404.
(17) Ma, X.; Yang, Y.; Wu, Q.; Liu, B.; Li, D.; Chen, R.; Wang, C.; Li, H.; Zeng, Z.; Li, L. Underlying mechanism of CO_2 uptake onto biomass-based porous carbons: Do adsorbents capture CO_2 chiefly through narrow micropores? *Fuel* **2020**, *282*, 118727.
(18) Osman, A. I. Mass spectrometry study of lignocellulosic biomass combustion and pyrolysis with NOx removal. *Renew. Energ.* **2020**, *146*, 484−496.
(19) Li, J.; Pan, L.; Suvarna, M.; Tong, Y. W.; Wang, X. Fuel properties of hydrochar and pyrochar: Prediction and exploration with machine learning. *Appl. Energ.* **2020**, *269*, 115166.
(20) Li, J.; Zhu, X.; Li, Y.; Tong, Y. W.; Ok, Y. S.; Wang, X. Multi-task prediction and optimization of hydrochar properties from high-moisture municipal solid waste: Application of machine learning on waste-to-resource. *J. Clean. Prod.* **2021**, *278*, 123928.
(21) Zhu, X.; Wang, X.; Ok, Y. S. The application of machine learning methods for prediction of metal sorption onto biochars. *J. Hazard. Mater.* **2019**, *378*, 120727.
(22) Zhang, K.; Zhong, S.; Zhang, H. Predicting Aqueous Adsorption of Organic Compounds onto Biochars, Carbon Nanotubes, Granular Activated Carbons, and Resins with Machine Learning. *Environ. Sci. Technol.* **2020**, *54*, 7008−7018.
(23) Guo, H.-n.; Wu, S.-b.; Tian, Y.-j.; Zhang, J.; Liu, H.-t. Application of machine learning methods for the prediction of organic solid waste treatment and recycling processes: A review. *Bioresour. Technol.* **2021**, *319*, 124114.
(24) Cha, D.; Park, S.; Kim, M. S.; Kim, T.; Hong, S. W.; Cho, K. H.; Lee, C. Prediction of Oxidant Exposures and Micropollutant Abatement during Ozonation Using a Machine Learning Method. *Environ. Sci. Technol.* **2021**, *55*, 709−718.
(25) Zorn, K. M.; Foil, D. H.; Lane, T. R.; Hillwalker, W.; Feifarek, D. J.; Jones, F.; Klaren, W. D.; Brinkman, A. M.; Ekins, S. Comparing Machine Learning Models for Aromatase (P450 19A1). *Environ. Sci. Technol.* **2020**, *54*, 15546−15555.
(26) Michel, A. P. M.; Morrison, A. E.; Preston, V. L.; Marx, C. T.; Colson, B. C.; White, H. K. Rapid Identification of Marine Plastic Debris via Spectroscopic Techniques and Machine Learning Classifiers. *Environ. Sci. Technol.* **2020**, *54*, 10630−10637.
(27) Zhu, X.; Tsang, D. C. W.; Wang, L.; Su, Z.; Hou, D.; Li, L.; Shang, J. Machine learning exploration of the critical factors for CO_2 adsorption capacity on porous carbon materials at different pressures. *J. Clean. Prod.* **2020**, *273*, 122915.
(28) Golden, C. E.; Rothrock, M. J., Jr.; Mishra, A. Comparison between random forest and gradient boosting machine methods for predicting Listeria spp. prevalence in the environment of pastured poultry farms. *Food Res. Int.* **2019**, *122*, 47−55.
(29) Konstantinov, A. V.; Utkin, L. V. Interpretable machine learning with an ensemble of gradient boosting machines. *Knowl. Base Syst.* **2021**, *222*, 106993.
(30) Yin, X.; Fallah-Shorshani, M.; McConnell, R.; Fruin, S.; Franklin, M. Predicting Fine Spatial Scale Traffic Noise Using Mobile Measurements and Machine Learning. *Environ. Sci. Technol.* **2020**, *54*, 12860−12869.
(31) Tan, D.; Suvarna, M.; Shee Tan, Y.; Li, J.; Wang, X. A three-step machine learning framework for energy profiling, machine-state prediction and production estimation in smart process manufacturing. *Appl. Energ.* **2021**, *291*, 116808.
(32) Pathy, A.; Meher, S.; P, B. Predicting algal biochar yield using eXtreme Gradient Boosting (XGB) algorithm of machine learning methods. *Algal Res.* **2020**, *50*, 102006.
(33) Fan, J.; Ma, X.; Wu, L.; Zhang, F.; Yu, X.; Zeng, W. Light Gradient Boosting Machine: An efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agric. Water Manag.* **2019**, *225*, 105758.

https://doi.org/10.1021/acs.est.1c01849
Environ. Sci. Technol. 2021, 55, 11925−11936

(34) Yuan, X.; Im, S. I.; Choi, S. W.; Lee, K. B. Removal of Cu(II) ions from aqueous solutions using petroleum coke-derived microporous carbon: Investigation of adsorption equilibrium and kinetics. *Adsorption* **2019**, *25*, 1205−1218.
(35) Altmann, A.; Toloşi, L.; Sander, O.; Lengauer, T. Permutation importance: A corrected feature importance measure. *Bioinformatics* **2010**, *26*, 1340−1347.
(36) Elshawi, R.; Al-Mallah, M. H.; Sakr, S. On the interpretability of machine learning-based model for predicting hypertension. *BMC Med. Inf. Decis. Making* **2019**, *19*, 146.
(37) Molnar, C. Interpretable Machine Learning - A Guide for Making Black Box Models Explainable. https://christophm.github.io/interpretable-ml-book/ (accessed May 2021).
(38) Liu, D.; Li, L.; Rostami-Hodjegan, A.; Bois, F. Y.; Jamei, M. Considerations and Caveats when Applying Global Sensitivity Analysis Methods to Physiologically Based Pharmacokinetic Models. *AAPS J.* **2020**, *22*, 93.
(39) Scikit − learn, Machine Learning in Python. https://scikit-learn.org/stable/ (accessed May 2021).
(40) Pennline, H. W. *Sorbent Research for the Capture of Carbon Dioxide*; U.S. Department of Energy, National Energy Technology Laboratory, 2016.
(41) Chen, Y.; Yang, H.; Yang, Q.; Hao, H.; Zhu, B.; Chen, H. Torrefaction of agriculture straws and its application on biomass pyrolysis poly-generation. *Bioresour. Technol.* **2014**, *156*, 70−77.
(42) Chen, D.; Mei, J.; Li, H.; Li, Y.; Lu, M.; Ma, T.; Ma, Z. Combined pretreatment with torrefaction and washing using torrefaction liquid products to yield upgraded biomass and pyrolysis products. *Bioresour. Technol.* **2017**, *228*, 62−68.
(43) Sher, F.; Pans, M. A.; Sun, C.; Snape, C.; Liu, H. Oxy-fuel combustion study of biomass fuels in a 20 kWth fluidized bed combustor. *Fuel* **2018**, *215*, 778−786.
(44) Purdy, A.; Pathare, P. B.; Wang, Y.; Roskilly, A. P.; Huang, Y. Towards sustainable farming: Feasibility study into energy recovery from bio-waste on a small-scale dairy farm. *J. Clean. Prod.* **2018**, *174*, 899−904.
(45) Kim, M.-J.; Choi, S. W.; Kim, H.; Mun, S.; Lee, K. B. Simple synthesis of spent coffee ground-based microporous carbons using K_2CO_3 as an activation agent and their application to CO_2 capture. *Chem. Eng. J.* **2020**, *397*, 125404.
(46) Jang, E.; Choi, S. W.; Lee, K. B. Effect of carbonization temperature on the physical properties and CO_2 adsorption behavior of petroleum coke-derived porous carbon. *Fuel* **2019**, *248*, 85−92.
(47) Choi, S. W.; Tang, J.; Pol, V. G.; Lee, K. B. Pollen-derived porous carbon by KOH activation: Effect of physicochemical structure on CO_2 adsorption. *J. CO2 Util.* **2019**, *29*, 146−155.
(48) Yang, Z.; Zhang, G.; Guo, X.; Xu, Y. Designing a novel N-doped adsorbent with ultrahigh selectivity for CO_2: Waste biomass pyrolysis and two-step activation. *Biomass Convers. Bior.* **2020**, DOI: 10.1007/s13399-020-00633-0.
(49) Yang, Z.; Zhang, G.; Xu, Y.; Zhao, P. One step N-doping and activation of biomass carbon at low temperature through $NaNH_2$. An effective approach to CO_2 adsorbents. *J. CO2 Util.* **2019**, *33*, 320−329.
(50) Yang, F.; Wang, J.; Liu, L.; Zhang, P.; Yu, W.; Deng, Q.; Zeng, Z.; Deng, S. Synthesis of porous carbons with high N-content from shrimp shells for efficient CO_2-capture and gas separation. *ACS Sustain. Chem. Eng.* **2018**, *6*, 15550−15559.
(51) Chen, J.; Yang, J.; Hu, G.; Hu, X.; Li, Z.; Shen, S.; Radosz, M.; Fan, M. Enhanced CO_2 capture capacity of nitrogen-doped biomass-derived porous carbons. *ACS Sustain. Chem. Eng.* **2016**, *4*, 1439−1445.
(52) Dissanayake, P. D.; You, S.; Igalavithana, A. D.; Xia, Y.; Bhatnagar, A.; Gupta, S.; Kua, H. W.; Kim, S.; Kwon, J.-H.; Tsang, D. C. W.; Ok, Y. S. Biochar-based adsorbents for carbon dioxide capture: A critical review. *Renew. Sustain. Energ. Rev.* **2020**, *119*, 109582.

https://doi.org/10.1021/acs.est.1c01849
Environ. Sci. Technol. 2021, 55, 11925−11936