Bioresource Technology 456 (2026) 134978

Contents lists available at ScienceDirect

Bioresource Technology

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

Insights into distinct biocarbon assembly and graphitization derived from large-scale bioresource by integrating machine learning with Raman and XRD spectroscopic profiling

Junsheng Yu [a,1], Jing Hong [a,1], Yujia Fang [a], Jiacheng Kang [a], Zongtai Tang [a], Lei Zhu [a], Bing Gao [a], Ling Luo [a], Dan Sun [b], Yanting Wang [a], Liangcai Peng [a], Hua Yu [a,*]

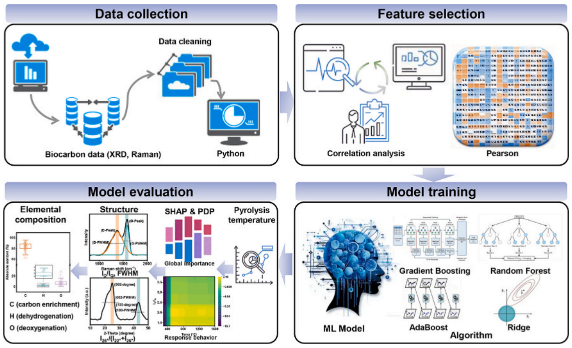
[a] *Key Laboratory of Fermentation Engineering (Ministry of Education), Hubei Key Laboratory of Industrial Microbiology, National "111" Center for Cellular Regulation & Molecular Pharmaceutics, Biomass & Bioenergy Research Center, School of Life & Health Sciences, Hubei University of Technology, Wuhan 430068, China*

[b] *School of Material Science & Chemical Engineering, Hubei University of Technology, Wuhan 430068, China*

HIGHLIGHTS

- Raman and XRD features serve as effective predictors of compositions.
- Gradient Boosting accurately predicts C, H, and O contents of biocarbon.
- SHAP and PDP reveal temperature–structure–element coupling mechanisms.
- Structural effects are most pronounced at intermediate pyrolysis temperatures.

GRAPHICAL ABSTRACT



ARTICLE INFO

Keywords:
Biocarbon
Lignocellulose
Carbon graphitization
Machine learning
Raman profiling
XRD spectroscopy

ABSTRACT

Understanding and predicting the elemental composition and assembly of diverse biocarbon materials are critical for quickly tailoring their physicochemical properties and functional applications. However, it remains challenging to sort out the complex coupling between pyrolysis conditions and carbon structural evolutions. By collecting all major sugars and lignocellulose resources, this study obtained a comprehensive dataset comprising 240 biocarbon samples derived from 17 precursors pyrolysis. Raman spectroscopy and X-ray diffraction profiling were employed to characterize all biocarbon samples, and multiple machine learning models were integrated to predict carbon/C, hydrogen/H, and oxygen/O contents. Among eleven regression algorithms, Gradient Boosting model exhibited the best performance, achieving robust predictions for C, H, and O contents and outperforming linear models in capturing nonlinear structure and composition relationships. Pearson correlation, SHAP interpretation, and partial dependence analysis could consistently sort out the dominant factor of pyrolysis temperature for governing carbon enrichment and H/O depletion, reflecting progressive aromatization, dehydrogenation and deoxygenation during pyrolysis. Structural descriptors including I_D/I_G, $I_{26°}/(I_{22°} + I_{26°})$,

---

* Correspondence author.
E-mail address: yh201906@foxmail.com (H. Yu).

[1] These authors contributed equally.

https://doi.org/10.1016/j.biortech.2026.134978
Received 2 February 2026; Received in revised form 15 April 2026; Accepted 23 May 2026

and interlayer spacing d_{002}, provided complementary information by linking defect density and graphitic ordering to elemental composition. Overall, this work has established an interpretable modeling framework that combines spectroscopic characterization with machine learning to enable rapid and non-destructive prediction of biocarbon composition and property, thereby providing mechanistic insights into biocarbon evolution for specific biocarbon design and assembly.

1. Introduction

Biocarbon is a versatile, cost-effective, and renewable carbonaceous material derived primarily from the pyrolysis of diverse biomass feedstocks, including agricultural and forestry residues, energy crops, municipal solid wastes, and algae (Wang et al., 2025b; Yang et al., 2025). Its structure originates mainly from the thermal transformation of cellulose, hemicellulose, lignin, and carbohydrates (monosaccharides, disaccharides, and polysaccharides) (Li et al., 2024c; Wang et al., 2024a). Owing to its high specific surface area, well-developed pore hierarchy, and abundant surface functional groups, biocarbon has attracted significant attention in environmental and energy-related applications (Hu et al., 2023; Li et al., 2024b). In environmental applications, it serves as an efficient adsorbent and soil amendment, while it also contributes to long-term carbon sequestration and greenhouse gas mitigation (Zhang et al., 2023a). In materials-oriented applications, the biocarbon produced under controlled pyrolysis conditions is increasingly utilized as a biomass-derived carbon material for energy storage and catalytic systems (Ai et al., 2024; Yu et al., 2022). Therefore, biocarbon offers broad potential for addressing resource depletion, environmental pollution, and energy sustainability challenges (Zhang et al., 2023b).

Biocarbon is primarily composed of carbon (C), hydrogen (H) and oxygen (O) with minor contributions from nitrogen (N), sulfur (S) and phosphorus (P) (Fu et al., 2026), and its elemental composition critically determines structure and functionality (Zhao et al., 2025). In particular, the relative contents of C, H, and O directly regulate aromaticity, polarity, stability, and chemical reactivity, thereby linking biomass characteristics to biocarbon performance (Pan et al., 2023). Carbon forms the structural backbone and enhances conductivity, stability, and carbon sequestration capacity with increasing content.(Luo et al., 2025), whereas hydrogen and oxygen contents are associated with calorific value, surface polarity, and functional group density, influencing adsorption and interfacial behavior (Hongwei et al., 2025; Liang et al., 2024b; Mariuzza et al., 2022). Hence, the variations in C/H/O contents fundamentally govern application-specific performance in environmental and energy systems (Murtaza et al., 2023).

The conventional approaches for predicting the elemental composition of biomass-derived carbon materials typically rely on kinetic modeling of pyrolysis reactions (Ranzi et al., 2017), which requires resolving multi-step reaction pathways and estimating numerous kinetic parameters (Fan et al., 2024). However, the high dimensionality and strong nonlinearity of biocarbon formation processes limit the applicability of such approaches across diverse conditions. In contrast, data-driven machine learning (ML) provides an efficient and robust alternative for capturing complex structure-composition relationships (Abdeldayem et al., 2025; Buratti et al., 2020). In particular, ML models can integrate pyrolysis conditions (e.g., temperature) with structural descriptors derived from Raman spectroscopy (e.g., I_D/I_G ratio), XRD (e.g., crystallinity and interlayer spacing), FTIR functional groups, and pore structure parameters (Wang et al., 2024b). Through non-linear interactive modeling, ML can accurately characterize the relationships of “temperature-structure-element” complex, such as the interaction between pyrolysis temperature and structural disorder or the synergistic effects of crystallinity and oxygen content (Zhang et al., 2025). This provides a predictive and interpretable framework for understanding of elemental regulation mechanisms in biocarbon systems (Song et al., 2024b; Haider Jaffari et al., 2023).

In this study, we established an integrated framework encompassing biocarbon preparation, model training and performance evaluation, based on a dataset of 240 samples derived from 17 distinct precursors. Eleven machine learning algorithms were then developed to predict the C, H, and O contents for systematically revealing the coupled associations between elemental composition and pyrolysis temperature, as well as Raman spectroscopy, and XRD characteristic parameters. Model performance was evaluated using the coefficient of determination (R^2) and root mean square error (RMSE) for both training and testing sets. Furthermore, SHapley Additive exPlanations (SHAP) and Partial Dependence Plots (PDP) were employed to interpret the feature contributions and elucidate the influence mechanisms of key variables. This enables mechanistic insights into structure–composition relationships and provides a foundation for the rational design and optimization of biocarbon properties.

2. Materials and methods

2.1. Biocarbon precursors

Seventeen representative biocarbon precursors were selected to high chemical diversity, categorized into pentoses (arabinose, xylose), hexoses (glucose, galactose), disaccharides (maltose, sucrose, lactose), polysaccharides (xylan, amylose, amylopectin), cellulose (microcrystalline cellulose), and lignin (alkali lignin, dealkaline lignin). Additionally, lignocellulosic biomass was represented by two *Miscanthus* germplasm resources (Mfl69 and Mfl148) (Li et al., 2016). All precursors were dried at 55 °C, sieved through a 40-mesh screen, and stored in sealed desiccators prior to pyrolysis (Gao et al., 2026; Yu et al., 2025).

2.2. Biocarbon preparation and characterization

For biocarbon production, 2.0 g of each precursor was placed in a tubular furnace (GSL-1800X) and pyrolyzed under a nitrogen atmosphere (Bakr et al., 2025). The samples were heated at a rate of 5 °C min^{-1} to target temperatures ranging from 300 to 1600 °C and held for 2 h. Subsequently, the furnace was cooled to 300 °C at a rate of 10 °C min^{-1}, followed by naturally cooling to room temperature. The resulting char residues were collected and treated with 1 mol L^{-1} HCl for 6 h to remove inorganic impurities. After acid washing, the samples were repeatedly rinsed with distilled water until neutral pH was reached, dried at 60 °C to constant weight, and stored for subsequent characterization and analysis. In total, 240 biocarbon samples were prepared (Yu et al., 2026). The biocarbon samples were characterized by Raman spectroscopy (ATR8300 series) to evaluate structural disorder and graphitization features (Yuan et al., 2025). Crystalline characteristics were identified using X-ray diffraction (XRD, FRINGE CLASS) (Wang et al., 2025a). Elemental compositions (C, H, and O contents) were determined using an elemental analyzer (EMA502) (Dong et al., 2019).

2.3. Dataset preparation and variable selection

Based on Raman and XRD characterization, structural and process-related features were extracted for all 240 biocarbon samples. The input variables included pyrolysis temperature, Raman spectral parameters (I_D/I_G, D-band position, G-band position, D-band full width at half maximum (D-FWHM), and G-band full width at half maximum (G-FWHM)), and XRD-derived parameters (interlayer spacing d_{002}, 002 peak position, 100 peak position, 002-FWHM, 100-FWHM, and the intensity ratio $I_{26^\circ}/(I_{22^\circ} + I_{26^\circ})$). The elemental contents of C, H, and O were defined as the primary output variables, while elemental ratio (O/C, H/C, and O/H) were additionally recorded for descriptive analysis. Pyrolysis temperature and structural parameters were used as model inputs, and C, H, and O contents served as prediction targets to construct the machine learning dataset (Chen et al., 2025a). Based on the processed dataset, a complete machine learning framework was established, including data collection, feature selection, model training, and performance evaluation (Liang et al., 2024a).

2.4. *Machine learning models*

To reduce the risk of local overfitting and to comprehensively assess predictive performance, eleven regression algorithms were implemented, including Gradient Boosting (GB), Random Forest (RF), AdaBoost, Ridge Regression, Bayesian Ridge Regression (BRR), k-Nearest Neighbors (KNN), Histogram Gradient Boosting (HGB), Elastic Net (EN), Lasso Regression, Support Vector Regression (SVR), and a Multilayer Perceptron (MLP) neural network (Pakzad et al., 2024). These models span linear, nonlinear, ensemble-based, instance-based, and neural-network approaches, enabling a systematic comparison of learning strategies (Wen & Yang 2021). Ensemble methods (GB, RF, AdaBoost, and Histogram Gradient Boosting) capture complex nonlinear relationships through the aggregation of multiple weak learners, while linear models with regularization (Ridge, Bayesian Ridge, Lasso, and Elastic Net) address multicollinearity and enhance model interpretability (Kandpal et al., 2024). KNN predicts target values based on local similarity in feature space, SVR provides robustness against outliers by enforcing margin-based error constraints, and MLP enables flexible nonlinear mapping via layered neural representations (Elish 2014). Together, this diverse model set ensures a robust evaluation of the relationships between pyrolysis conditions, structural descriptors, and biocarbon elemental composition (Zhang et al., 2025).

2.5. *Model evaluation and interpretation*

Pearson correlation coefficient (PCC): To preliminarily assess linear relationships between input variables and output targets, the Pearson correlation coefficient (PCC) was calculated. An absolute PCC value greater than 0.5 ($|r| > 0.5$) indicates a strong linear correlation, while $r = 0$ indicates no linear correlation (Kravchenko et al., 2025). The PCC is defined as:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^n(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 \sum_{i=1}^{n}(y_i - \overline{y})^2}} \tag{1}$$

where x_i and y_i are the observed values of input and output variables, respectively, $\overline{x}$ and $\overline{y}$ are their mean values, and n represents the number of samples (Chen et al., 2025b).

Model performance evaluation: The dataset was randomly divided into training (80 %) and testing (20 %) subsets (Lu et al., 2025). Five-fold cross-validation was applied to reduce overfitting and ensure model robustness (Wei et al., 2024). Model performance was evaluated using the coefficient of determination (R^2) and the root mean square error (RMSE), calculated as:

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

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\boldsymbol{y}_i^{exp} - \boldsymbol{y}_i^{pred}\right)^2} \tag{3}$$

Where Y_i^{exp} and $Y_j^{\ pred}$ represent experimental and predicted values, respectively, $\overline{Y}^{exp}$ is the mean experimental value, and N is the total number of samples (Yin et al., 2025).

2.6. *SHAP analysis*

SHapley Additive exPlanations (SHAP) is a model-agnostic interpretability method rooted in cooperative game theory, which quantifies the contributions of each input feature to model predictions based on its marginal contribution (Palansooriya et al., 2022). The SHAP value (ϕ) represents the unique contribution of given feature to the outcome, and its additive property enables quantitative interpretation of individual features and feature combinations (Li et al., 2024a). The SHAP value for feature x_j is defined as:

$$\phi_j = \sum S \subseteq N\backslash\{x_i\}\frac{|S|!(P - |S| - 1)!}{N!}(f_x(S \cup \{x_i\}) - f_x(S)) \tag{4}$$

where S represents a subset of features excluding feature x_j, N denotes the complete set input features, and $f_x(S)$ is the model prediction obtained using features in subset S (Sun et al., 2024).

2.7. *Partial dependence plots (PDP)*

Partial dependence plots (PDPs) are visualization tools used to quantify the average marginal effect of one or two input features on model predictions, while averaging out the influence of all other variables (Leng et al., 2022). PDPs are particularly useful for interpreting nonlinear models and identifying feature–response relationships and interaction patterns (Song et al., 2024a). For a single feature x_i and a pair of features (x_i, x_j), the partial dependence functions are defined as:

$$f_\alpha(x_i) = \frac{1}{n}\sum_{i=1}^{n} f(x_i, x_i^0) \tag{5}$$

$$f_\alpha(x_i, x_j) = \frac{1}{n}\sum_{l=1}^{n} f(x_i, x_{j,} x_C^l) \tag{6}$$

where f_α represents the trained model, x_i and x_j are the features of interest, and x_c denotes the set of remaining features excluding x_i (or x_i, x_j), and n is the total number of samples (Huang et al., 2026).

3. Results and discussions

3.1. *Raman and XRD characteristics of biocarbon*

A total of 240 biocarbon samples derived from 17 different precursors were prepared over a wide pyrolysis temperature range of 300–1600 °C. Systematic analysis of their Raman and XRD parameters (Fig. 1a) revealed that the extracted features approximately showed normal distributions, validating the use of mean values as representative descriptors for subsequent analysis.

The statistical distributions of Raman spectral parameters are shown in Fig. 1b–d. The I_D/I_G ratio ranged from 1.16 to 6.51, while the D- and G-band positions varied from 1329.3 to 1432.2 cm^{-1}, and from 1572.9 to 1608.6 cm^{-1}, respectively. The full width at half maximum (D-FWHM) ranged from 48.6 to 420.0 cm^{-1}, and that of the G-FWHM from 54.9 to 197.6 cm^{-1}. The mean I_D/I_G value (2.24) indicated a highly defective carbon structure with a low degree of graphitization. Meanwhile, the average G-band position (1589.25 cm^{-1}) was notably higher than that of ideal graphite (~1580 cm^{-1}), suggesting the predominance of disordered carbon. The large median values of D-FWHM (202.60 cm^{-1}) and G-FWHM (103.06 cm^{-1}) further reflect short structural coherence lengths and pronounced structural disorder across the biocarbon dataset. The broad pyrolysis temperature range (300–1600 °C) encompasses low- (300–500 °C), intermediate- (500–1000 °C), and high-temperature

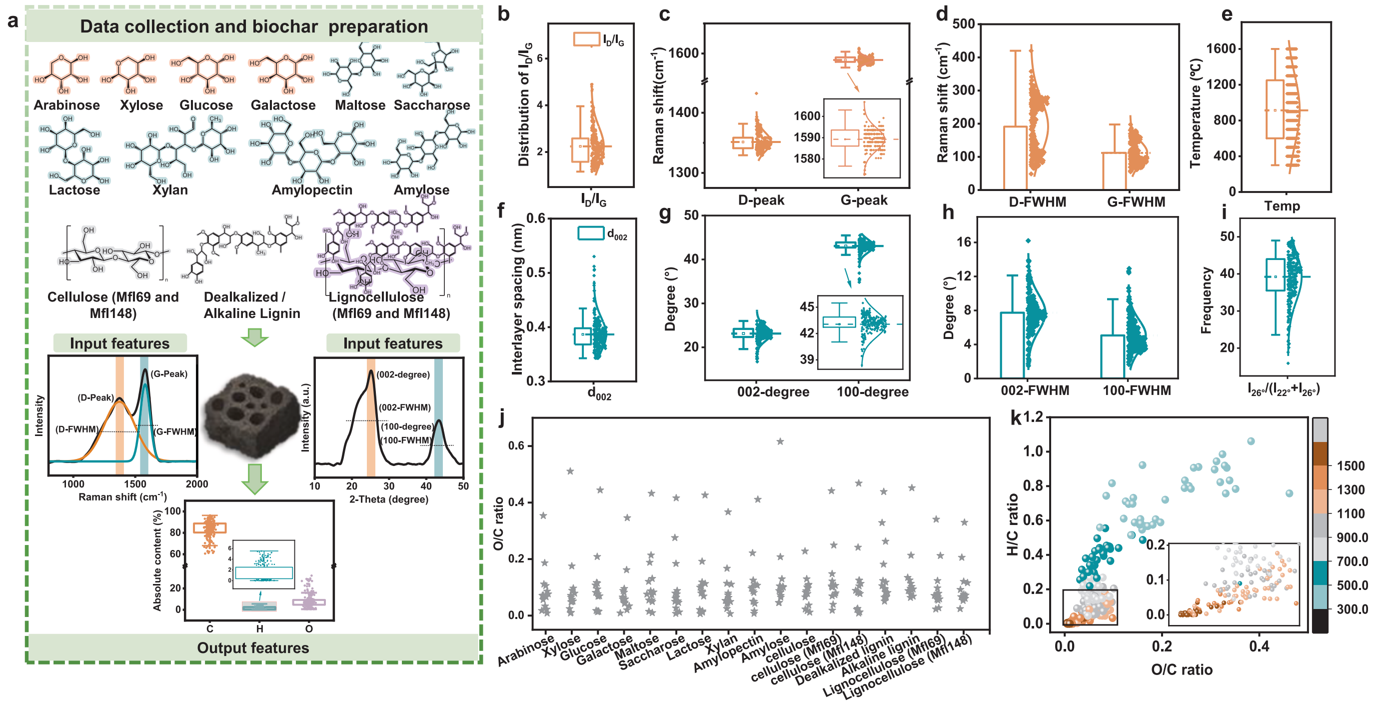


Fig. 1. Preparation workflow and statistical distributions of structural and compositional parameters of diverse biocarbon samples. (a) Schematic illustration of the biocarbon preparation process. (b–d) Statistical distributions of Raman spectroscopic parameters, including the I_D/I_G ratio (b), positions of the D and G bands (c), and the full width at half maximum (FWHM) of the D and G bands (d).(e) Distribution of pyrolysis temperatures. (f–i) Statistical analysis of X-ray diffraction (XRD) parameters, including interlayer spacing d_{002}, diffraction angles and FWHM of the (002) and (100) planes, and the degree of structural ordering.(j) Atomic O/C ratios of biocharbon samples derived from different precursor materials. (k) Temperature-dependent molar O/C ratios of biocarbon samples (n = 240).

(1000–1600 °C) regimes (Fig. 1d), corresponding to distinct carbonization stages associated with substantial changes in elemental composition (C/H/O).

The distributions of XRD-derived structural parameters are presented in Fig. 1f–i, including interlayer spacing d_{002} (0.3427–0.5302 nm), 002 peak position (16.7°–26.0°), 100 peak position (1.49°–16.2°), 002-FWHM (34.8°–45.4°), 100-FWHM (1.65°–12.9°), and the intensity ratio $I_{26°}/(I_{22°} + I_{26°})$ (15.9 %–49.0 %). The mean d_{002} value (0.3866 nm) is substantially larger than that of ideal graphite (0.3354 nm), indicating a predominantly turbostratic or highly disordered carbon structure. Consistently, the relatively low average 002 and 100 peak positions (23.1° and 43.0°), coupled with the broad peak widths (002-FWHM = 7.73°, 100-FWHM = 5.06°), confirm the small crystallite size and low crystallinity of the biocarbon samples. Overall, Raman and XRD analyses consistently demonstrate that the biocarbon samples exhibit a high degree of structural disorder and low graphitization, which are characteristic features of biocarbon materials produced over broad temperature ranges and from diverse precursors. The wide distributions of structural parameters reflect both the feedstock heterogeneity and the progressive structural evolution driven by increasing pyrolysis temperature.

As shown in Fig. 1j and Fig. S1, the elemental compositions of biocarbon vary substantially across different precursors. The O/C, H/C, and H/O ratios range from 0.63 to 61.5 %, 0–8.84 %, and 0–55.6 %, respectively, with average values of 10.6 %, 2.01 %, and 16.6 %. Both O/C and H/C ratios decrease systematically with increasing pyrolysis temperature (Fig. 1k), reflecting thermally induced devolatilization. As temperature increases, oxygen- and hydrogen-containing species are progressively removed, leading to relative carbon enrichment in the solid residue.

3.2. *Pearson correlation analysis*

The Pearson correlation coefficient (PCC) was employed to quantify the linear relationships between the input feature and elemental compositions. The resulting correlation matrix (Fig. 2a) and direction of correlations between the 12 input features and the three target variables (C, H, and O contents), where color intensity and ellipse size represent correlation magnitude and sign. The majority of correlations was statistically significant ($p \leq 0.05$), confirming strong statistical associations between biocarbon structural characteristics and its elemental composition.

Pyrolysis temperature exhibited a moderate positive correlation with carbon content (r = 0.42) where it had strongly negative correlations with oxygen and hydrogen contents (r = − 0.71 and − 0.89, respectively). This trend aligns with the progressive thermal decomposition of aliphatic groups ($–CH_2$, $–CH_3$) and oxygen-containing functional groups (–OH, –COOH) as temperature increases from 300 to 1600 °C. During pyrolysis, hydrogen and oxygen were released as volatile species (CO_2, H_2O, CH_4, and H_2), resulting in a relative enrichment of carbon from approximately 50–60 % at low temperatures to 70–80 % at intermediate temperatures. At higher temperatures, enhanced aromatic condensation and polymerization promoted the formation of graphite-like carbon frameworks, being coupled with the decomposition of residual ether and carbonyl functionalities to enable the carbon content at more than 95 %.

Raman-derived parameters are correlated with elemental composition by reflecting variations in structural disorder and defect density. Specifically, the I_D/I_G ratio, D-FWHM, G-FWHM, and D-band position are negatively correlated with carbon content (r = − 0.52 to − 0.17), but positively associated with hydrogen (r = 0.22–0.75) and oxygen contents (r = 0.23–0.64). Conversely, the G-band position scales positively with carbon content (r = 0.50) and negatively correlated with hydrogen and oxygen contents (r = − 0.57). These relationships suggest that biocarbon with higher defect density and broader Raman features tend to retain more hydrogen- and oxygen-containing functionalities, whereas increased structural ordering is linked to carbon enrichment. XRD-derived parameters primarily reflect crystalline ordering and interlayer structural evolution and therefore exhibit systematic

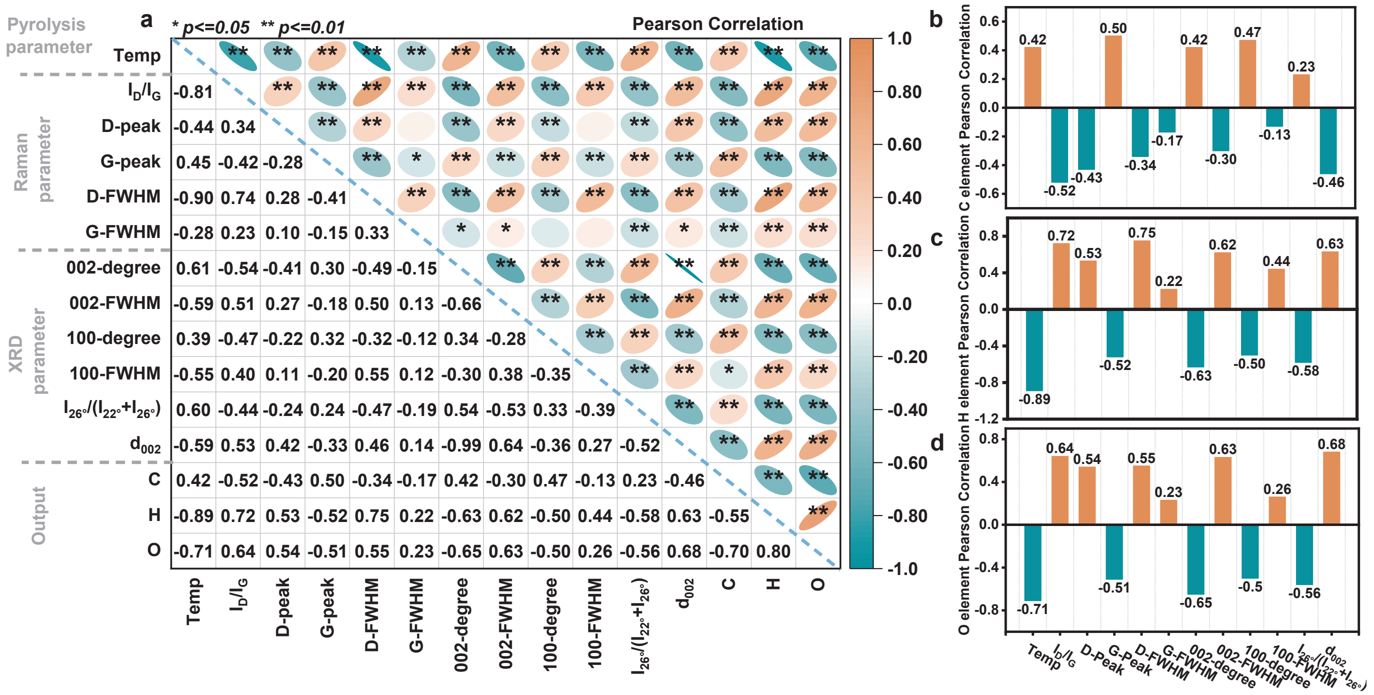


Fig. 2. Pearson correlation analysis between structural descriptors and elemental composition of biocarbon based on the complete dataset (n = 240). (a) Pearson correlation coefficient (PCC) matrix illustrating pairwise correlations among Raman, XRD, pyrolysis temperature, and elemental composition variables for biocarbon samples derived from 17 precursors over a temperature range of 300–1600 °C. (b–d) Correlation coefficients between representative Raman and XRD input features and carbon (C), hydrogen (H), and oxygen (O) contents, respectively (*$p < 0.05$, **$p < 0.01$).

correlations with elemental composition. The interlayer spacing d_{002}, 002-FWHM, and 100-FWHM show negative correlations with carbon content (r = − 0.46, −0.30, and − 0.13) and positive correlation with hydrogen (r = 0.63, 0.62, and 0.44) and oxygen contents (r = 0.68, 0.63, and 0.26). Conversely, the 002 peak position, 100 peak position, along with the intensity ratio $I_{26°}/(I_{22°} + I_{26°})$, being positively correlated with carbon content (r = 0.42, 0.47, and 0.44) and negatively correlated with hydrogen (r = − 0.63, −0.50, and − 0.58) and oxygen (r = − 0.65, −0.50, and − 0.56). These trends are consistent with the progressive transition from turbostratic carbon toward more ordered structures during pyrolysis, accompanied by heteroatom elimination and contraction of the interlayer spacing toward the graphite value (0.3354 nm). These results summarize the correlations between representative input features: pyrolysis temperature, Raman parameters (e.g., I_D/I_G, G-

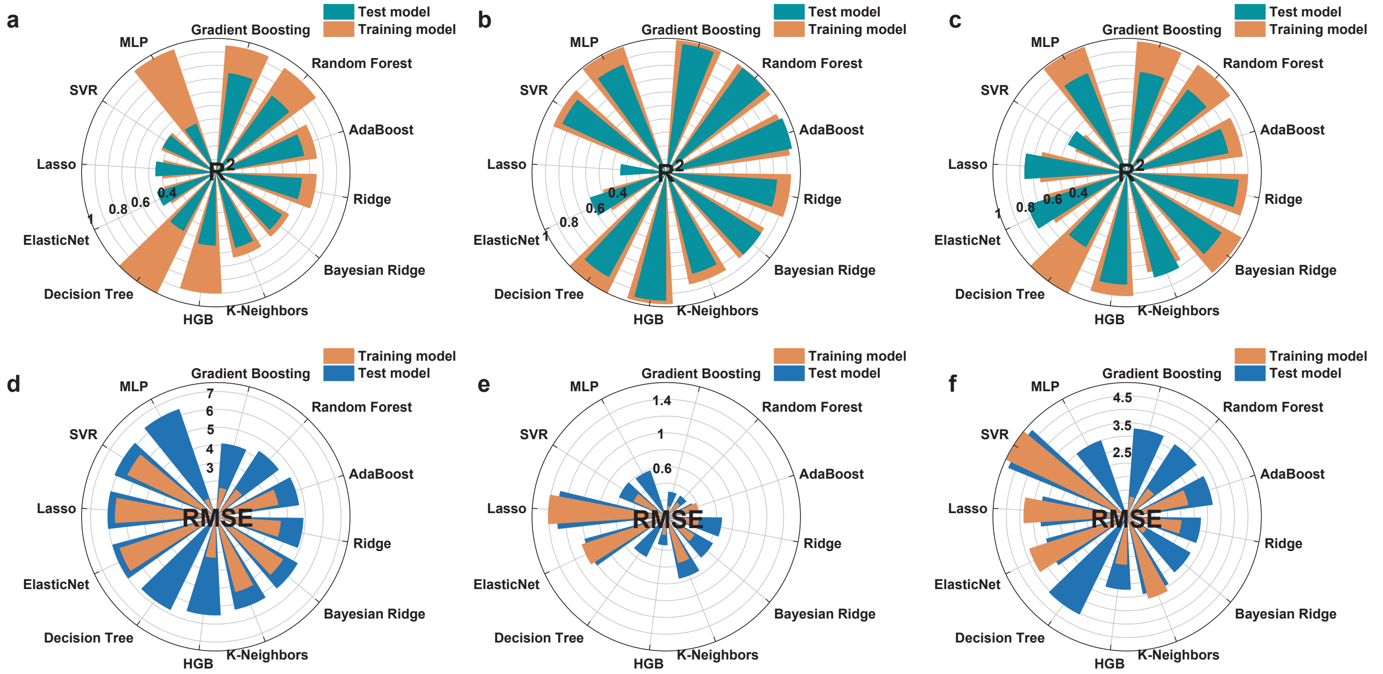


Fig. 3. Performance comparison of eleven machine learning models for predicting the elemental composition of biocarbon samples. (a–c) Coefficient of determination (R^2) for carbon (C), hydrogen (H), and oxygen (O) contents. (d–f) Root mean square error (RMSE) for carbon (C), hydrogen (H), and oxygen (O) contents.

band position), and XRD parameters (e.g., d_{002}, 002 peak position) and the output variables C, H, and O contents (Fig. 2b–d). A quantitative comparison of correlation coefficients highlights distinct differences in both the strength and direction of associations across different categories of structural descriptors, providing a solid foundation for subsequent machine learning modeling. Notably, these correlations do not imply direct causality, but rather reflect the co-evolution of elemental composition and carbon structure during temperature-driven pyrolysis. To further disentangle the relative contributions of temperature and structural parameters and to capture potential nonlinear effects, machine learning models and interpretable analysis methods are employed in the following sections.

3.3. *Model training and prediction*

To evaluate efficacy of machine learning (ML) in predicting the elemental composition of biocarbon, eleven regression models were implemented to predict C, H, and O contents. Model performance was evaluated using the coefficient of determination (R^2) and root mean square error (RMSE), allowing direct and quantitative comparison among different models (Fig. 3, Table S1). Hyperparameter tuning for all models was performed using randomized search with 5-fold cross-validation to ensure a fair comparison. Furthermore, this study compared the R^2 and RMSE of the optimal machine learning model with the previously-reported predictions (Table S2) (Deng et al., 2024; Fu et al., 2026; Leng et al., 2022; Li et al., 2022; Liao et al., 2025; Zhu et al., 2019). The optimal hyperparameter combinations of the final predictive models were summarized in (Tables S3–S5). For carbon prediction, the overall model performance followed the descending order: Gradient Boosting (GB) > Random Forest (RF) > AdaBoost > Ridge > Bayesian Ridge > k-Neighbors > Histogram Gradient Boosting (HGB) > Elastic Net (EN) > Lasso > Support Vector Regression (SVR) > ANN. A similar ranking trend was observed for hydrogen prediction, whereas Ridge performed best for oxygen, followed by AdaBoost, RF, and GB. Given that carbon constitutes the dominant fraction of biocarbon and is the predominant element and highly sensitive to structural evolution during pyrolysis, subsequent analyses focused on the four top-performing models for carbon: GB, RF, AdaBoost, and Ridge.

Prediction accuracy was further assessed using scatter plots of predicted versus experimental values, supplemented by marginal histograms, confidence bands, and prediction intervals for all 240 biocarbon samples (Fig. 4). For carbon content, the GB model (Fig. 4a) showed excellent agreement with the ideal y = x line, particularly in the 80–95 % range corresponding to high-temperature biocarbon samples. Only slight underestimation is observed for a limited number of low-carbon samples (50–70 %). GB achieved the highest predictive performance (R^2 = 0.747, RMSE = 4.11), effectively capturing the two-stage nonlinear behavior of carbon enrichment: a gradual increase at low pyrolysis temperatures followed by rapid graphitization at higher temperatures. Notably, the moderate R^2 value can be attributed to the high heterogeneity of the dataset, which includes 17 chemically diverse precursors and a wide temperature range (300-1600 °C) spanning multiple nonlinear carbonization regimes. This complexity introduces a substantial variability in carbon content, and increases a prediction difficulty compared to relatively more homogeneous datasets. The RF model (Fig. 4b) reproduced the overall trend but shows larger deviations in the intermediate carbon range (70–80 %) (R^2 = 0.710, RMSE = 4.40), likely reflecting feedstock heterogeneity and variance introduced by ensemble averaging across decision trees. The AdaBoost model (Fig. 4c) tends to overestimate carbon content in the 60–70 % range (R^2 = 0.669, RMSE = 4.70), suggesting sensitivity to noisy samples that are more difficult to predict. In contrast, the Ridge model (Fig. 4d) systematically underestimates high-carbon (>90 %) samples and overestimates low-carbon (<60 %) samples (R^2 = 0.641, RMSE = 4.89), highlighting the inherent limitations of linear models in capturing nonlinear carbon enrichment processes during pyrolysis.

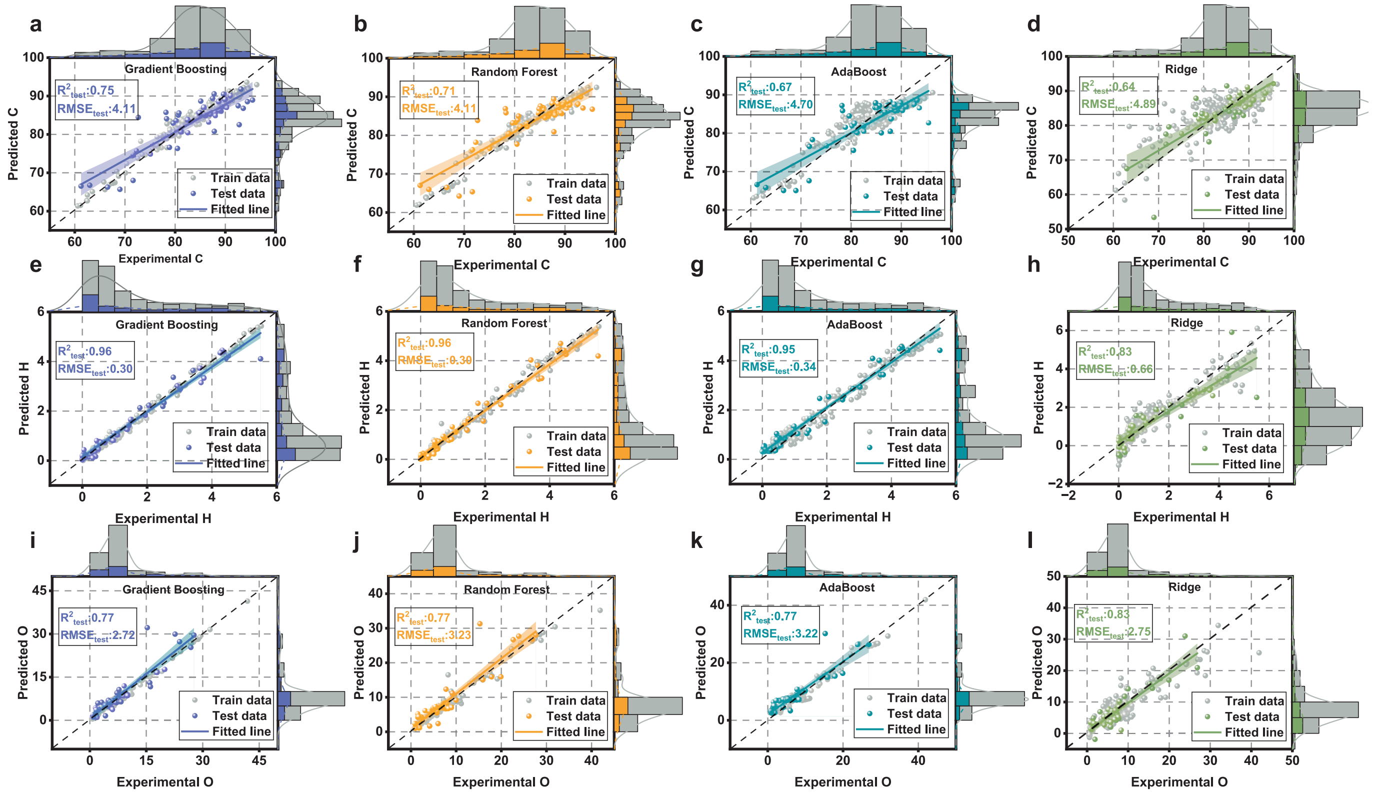


Fig. 4. Comparison between predicted and experimental elemental contents obtained using different machine learning models, including Gradient Boosting, Random Forest, AdaBoost, and Ridge regression. A total of 192 data points were used for model development. (a–d) Carbon content; (e–h) hydrogen content; (i–l) oxygen content. RMSE denotes the root mean square error.

For hydrogen prediction, both GB (Fig. 4e) and RF (Fig. 4f) closely follow the y = x line across the 1–5 % H range, with slight underestimation for extremely low-hydrogen samples (<1 %). These models achieve the highest accuracy (R^2 = 0.964, RMSE = 0.301–0.303). AdaBoost (Fig. 4g) also demonstrates strong predictive performance (R^2 = 0.954, RMSE = 0.342), whereas Ridge (Fig. 4h) shows increased dispersion (R^2 = 0.827, RMSE = 0.665), again reflecting the limited ability of linear regression to capture nonlinear devolatilization behavior. For oxygen content, the Ridge model (Fig. 4l) achieves the best overall performance (R^2 = 0.832, RMSE = 2.75), although it tends to underestimate samples with high oxygen contents (>25 %). The RF (Fig. 4j) and AdaBoost (Fig. 4k) models provide reliable predictions for low-to-medium oxygen contents (R^2 = 0.767–0.768, RMSE = 3.22–3.23), while the GB model (Fig. 4i) performs particularly well in the 5–10 % oxygen range but slightly overestimates samples between 15 % and 20 % (R^2 = 0.756, RMSE = 3.30). This discrepancy suggests a reduced sensitivity to oxygen removal at intermediate temperatures under the current parameter configuration.

Based on the strongly predictive performance achieved across multiple models, it confirmed that pyrolysis temperature combined with Raman- and XRD-derived structural descriptors encoded sufficient information to reconstruct the elemental composition of biocarbon. We then generated learning curves for the GB model to assess whether the current dataset size was sufficient for ensemble methods. The curves show that both the training score and the cross–validation score plateau after approximately 150–180 samples, indicating that further increased sample size (without increased precursor diversity) would only yield marginal gains (Fig. S2). It suggests that the dataset covers multiple precursor categories with distinct chemical characteristics, leading to inherently different feature distributions in the spectral descriptors. Among all tested algorithms, the GB model offers the most balanced performance, delivering the highest accuracy for carbon prediction, matching RF for hydrogen, and providing reliable oxygen predictions without pronounced systematic bias. Although Ridge regression achieved a slightly higher accuracy for oxygen prediction, Gradient Boosting was selected as the representative model for interpretability analysis to ensure consistency across all target elements and to avoid element-specific model bias in mechanistic interpretation. Its strong adaptability to complex, nonlinear structure–composition relationships justified its selection for subsequent interpretability and mechanism analyses.

3.4. Feature importance of input variables

The relative importance of input features for predicting the elemental composition of biocarbon was evaluated using the Gradient Boosting (GB) model integrated with SHAP (SHapley Additive exPlanations) analysis (Fig. 5). Each point represents a single sample, with its horizontal position corresponding to the SHAP value and the color indicating the feature magnitude. For carbon prediction (Fig. 5a–c), pyrolysis temperature (Temp) emerges as the most influential feature, accounting for 40.4 % of the total contribution (Fig. 5b). Higher temperatures consistently yield positive SHAP values, reflecting the progressive removal of hydrogen- and oxygen-containing volatiles and subsequent carbon enrichment. The I_D/I_G ratio shows predominantly negative SHAP values (down to −2.18), indicating that highly defective carbon structures are associated with lower predicted carbon content, whereas increased structural ordering (higher $I_{26^\circ}/(I_{22^\circ}+I_{26^\circ})$) provides

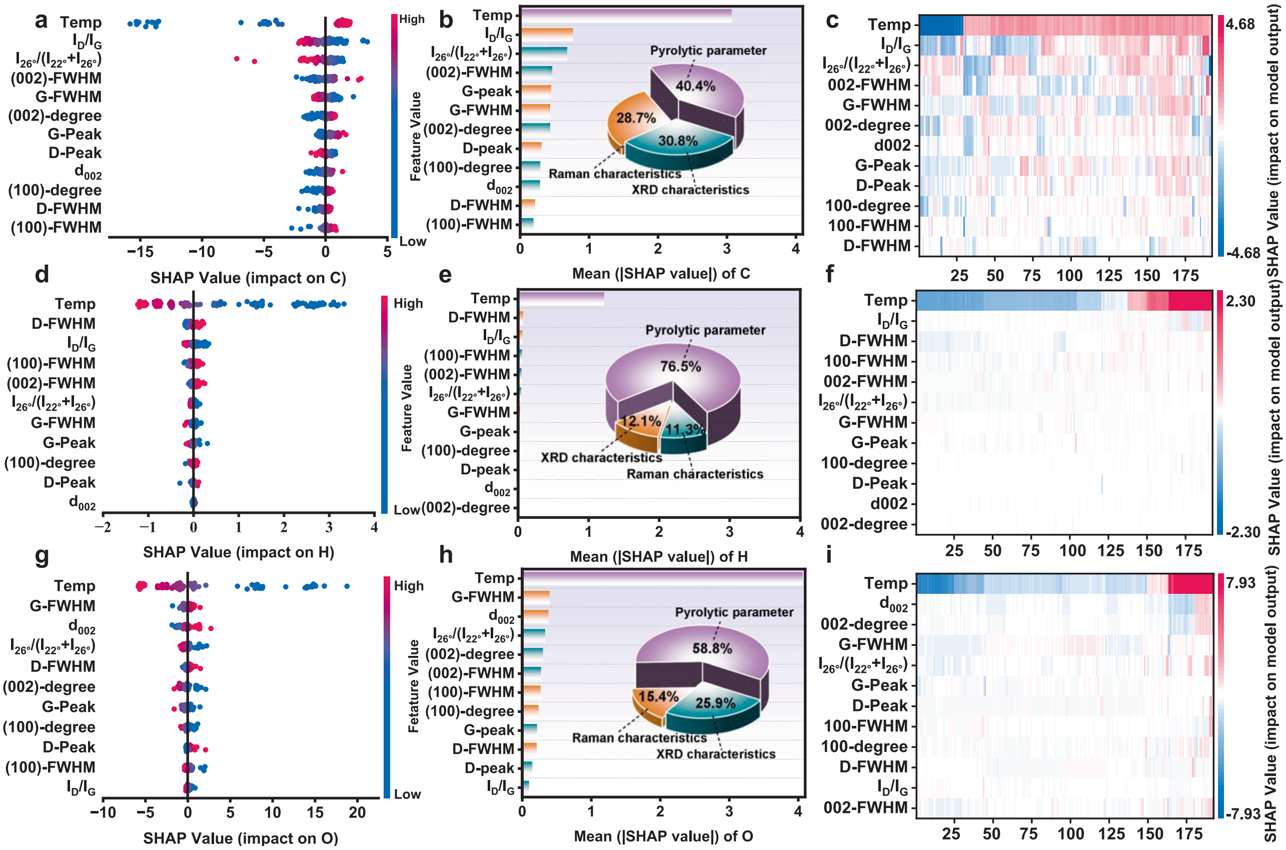


Fig. 5. Interpretation of the Gradient Boosting model based on the feature-selected dataset. (a, d, g) SHAP analysis illustrating the relative contributions of key variables;(b, e, h) feature importance rankings of the Gradient Boosting model;(c, f, i) combined heatmaps for carbon, hydrogen, and oxygen contents (n = 192). Temp denotes pyrolysis temperature; D-/G- peak refers to the positions of the D and G bands; D/G FWHM represents the full width at half maximum of the Raman peaks; (002)-, (100)- FWHM indicates the full width at half maximum of the (002) and (100) diffraction peaks; 002/100 angle denotes the corresponding diffraction angles.

positive contributions. XRD-derived features collectively account for 30.8 % of the total importance, highlighting the significance of interlayer stacking and crystallite ordering in modulating carbon content. For hydrogen prediction (Fig. 5d–f), temperature again dominates, contributing 76.5 % of the total feature importance. Higher temperatures correlate with negative SHAP values, consistent with the rapid volatilization of hydrogen-containing groups. Among structural descriptors, I_D/I_G and D-FWHM are the most influential, underscoring that defect density and structural disorder facilitate hydrogen retention. XRD parameters contribute 12.1 %, reflecting the indirect crystallinity impact on hydrogen retention. For oxygen prediction (Fig. 5g–i), temperature remains the dominant factor (58.8 % contribution), with increasing temperature associated with progressively more negative SHAP values, aligning with accelerated thermal decomposition of oxygen-containing functional groups ($-COOH \rightarrow CO_2$, $-OH \rightarrow H_2O$). XRD features contribute 25.9 % of the total importance, with d_{002} and 002 peak position being the most influential structural variables. Raman-derived descriptors such as I_D/I_G exhibit positive SHAP values, indicating greater oxygen retention in more defective, disordered carbon structures.

Overall, SHAP analysis quantifies both the relative contributions and the directionality of key features in determining biocarbon elemental composition. While SHAP reveals feature influence at both global and local levels, it may not fully capture the nonlinear dependencies across feature ranges. Consequently, Partial Dependence Plots (PDPs) were employed in the subsequent section to visualize the marginal and interaction effects of key variables on C, H, and O contents, offering complementary mechanistic insights into the structure-composition relationships.

3.5. Univariate partial dependence analysis

Based on the SHAP feature importance analysis, the four most influential input variables (temperature, I_D/I_G, $I_{26^\circ}/(I_{22^\circ} + I_{26^\circ})$, and 002-FWHM) were selected to construct one-dimensional (1D) and two-dimensional (2D) partial dependence plots (PDPs). These plots elucidate the marginal effects, nonlinear responses, and potential interactions of key variables across the C, H and O contents of biocarbon samples (Fig. S2). Temperature exhibits the strongest effect on carbon content (Fig. S3a). As temperature rising, the C content initially increases rapidly before plateauing, reflecting an accelerated devolatilization of H- and O-containing species and progressive enrichment of the carbon skeleton, ultimately reaching a high-temperature carbonization limit. The I_D/I_G ratio shows a weak negative effect on C content, as higher values indicate greater defect density and lower graphitization, which are associated with reduced carbon enrichment. In contrast, $I_{26^\circ}/(I_{22^\circ} + I_{26^\circ})$ exerts a positive influence, as higher values correspond to an increased proportion of ordered, graphitized carbon. The 002-FWHM displays a weak negative relationship with C content, consistent with smaller ordered domains and higher structural disorder. For hydrogen content (Fig. S3b), temperature drives a pronounced and nearly linear decrease. Hydrogen is mainly associated with thermally labile functional groups and aliphatic structures, which are readily cleaved and released as H_2 and CH_4 upon heating. Consequently, H content decreases rapidly and continuously with increasing temperature. In contrast, H content exhibits a slight increase with higher I_D/I_G, a slight decrease with increasing $I_{26^\circ}/(I_{22^\circ} + I_{26^\circ})$, and a gradual increase with larger 002-FWHM values. Oxygen content follows a temperature-dependent decay (Fig. S3c), characterized by a rapid decline at low temperatures followed by a more gradual decrease at higher temperatures. This trend reflects the sequential thermal decomposition of oxygen-containing functional groups, including carboxyl, carbonyl, and hydroxyl moieties, which are progressively removed as CO and CO_2. Oxygen content increases with higher I_D/I_G, decreases slightly with increasing $I_{26^\circ}/(I_{22^\circ} + I_{26^\circ})$, and increases modestly with larger 002-FWHM values.

3.6. Bivariate partial dependence analysis

To further explore the synergistic interplay between pyrolysis temperatures and structural features, bivariate partial dependence analysis was performed (Fig. 6), revealing nonlinear interactions between process conditions and structural parameters. For carbon content, the combined effect of temperature and I_D/I_G (Fig. 6a–c) is minimal in the low-temperature range (300–600 °C), where dehydration and the release of light volatiles dominate and the carbon framework remains largely undeveloped. In the intermediate temperature regime (600–1000 °C), regions characterized by lower defect density (lower I_D/I_G) exhibit more rapid carbon enrichment, indicating that increasing structural ordering facilitates carbon accumulation during active carbonization. At high temperatures (1000–1600 °C), carbon content approaches saturation as biocarbon becomes progressively graphitized, rendering the influence of structural parameters negligible. For hydrogen content (Fig. 6d–f), temperature induces a continuous decline across the entire range. At low temperatures, higher D-FWHM, I_D/I_G, and 100-FWHM values are associated with enhanced retention of hydrogen-containing functional groups. During the intermediate temperature stage, extensive C—H bond cleavage and structural rearrangements accelerate hydrogen removal. At high temperatures (>1000 °C), hydrogen content in all samples decreases to below 1 %, and the contribution of structural features becomes insignificant. Oxygen content exhibited a similar temperature-dependent trajectory (Fig. 6g–i). At low temperatures, higher G-FWHM, larger interlayer spacing d_{002}, or lower $I_{26^\circ}/(I_{22^\circ} + I_{26^\circ})$ ratios reflect greater structural disorder and are associated with increased oxygen retention. As temperature entered the intermediate range, oxygen content decreases rapidly, accompanied by a progressive weakening of structural effects. At high temperatures, oxygen-containing functional groups were almost entirely decomposed, and structural parameters no longer exert a discernible negligible influence.

The bivariate PDP results demonstrate that temperature is the dominant driver of C, H, and O evolution in biocarbon, whereas structural features modulate elemental contents only within specific temperature windows, primarily at intermediate temperatures. The temperature-driven aromatization, dehydrogenation, and deoxygenation processes determine the principal trends in elemental composition, while structural parameters regulate the rate and stage-dependent variations of these transformations.

Combined with the univariate PDP analysis, these results establish a clear temperature–structure–element coupling mechanism. Pyrolysis temperature determines the overall trajectories of C, H, and O contents, whereas structural parameters related to defect density and graphitization degree exert secondary control that is most pronounced at intermediate temperatures (600–1000 °C), where carbon skeleton rearrangement and functional group elimination are most active. At low temperatures, structural disorder promotes retention of H- and O-containing functionalities, while at high temperatures (>1000 °C), extensive graphitization leads to near-complete heteroatom removal and suppresses structural control. This window-dependent synergistic effect explains the observed nonlinear relationships between temperature, structure, and elemental composition, and provides a mechanistic framework for rational regulation of biocarbon properties through controlled pyrolysis.

Overall, SHAP and PDP analyses collectively establish a unified temperature-structure-element coupling mechanism. Pyrolysis temperature governs the global evolution of elemental composition, while structural descriptors modulate elemental retention in a temperature-dependent manner, with the strongest influence occurring in the intermediate temperature regime where structural rearrangement and functional group decomposition are most active. This integrated interpretation links feature importance, marginal effects and interaction effects into a coherent mechanistic understanding of biocarbon formation.

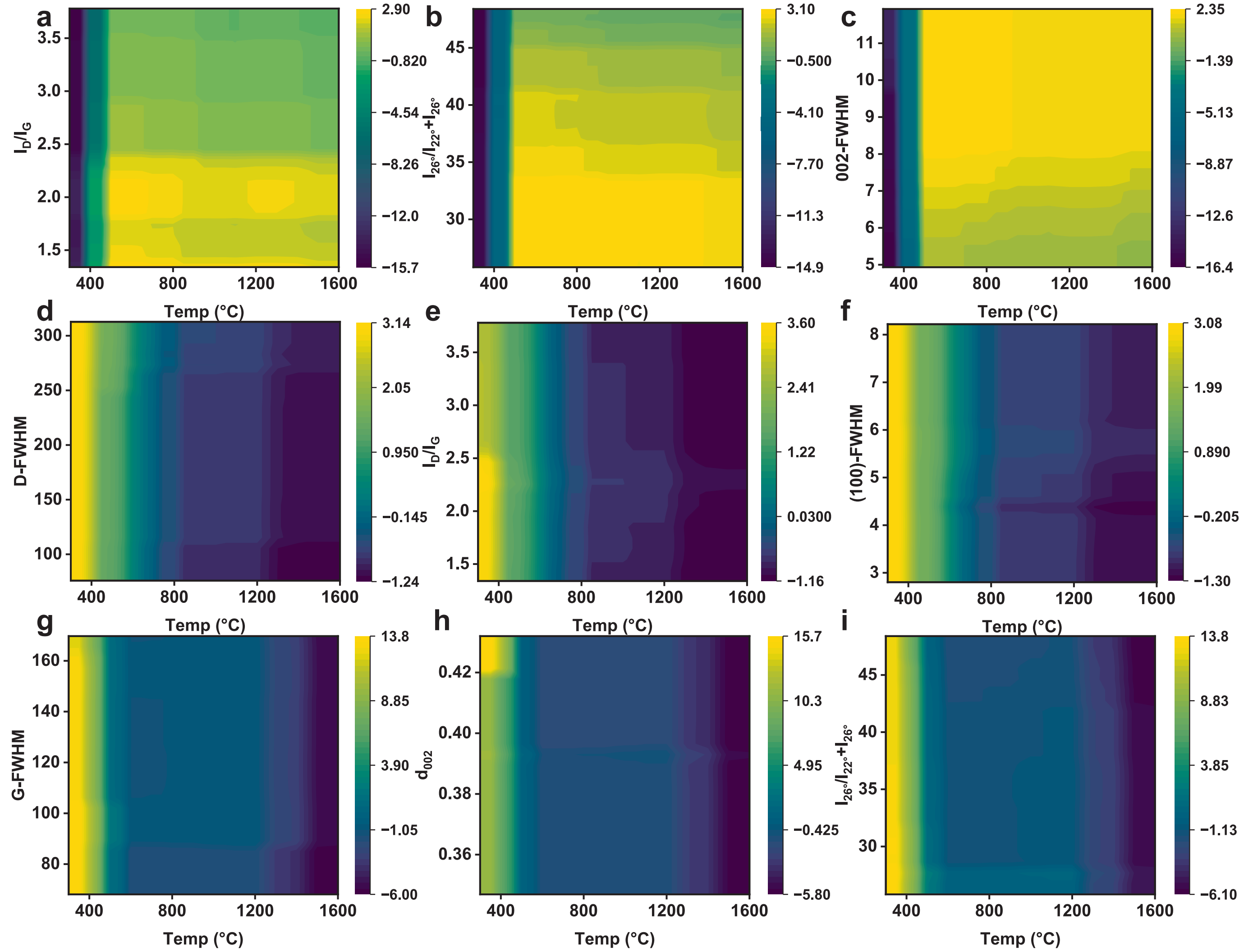


Fig. 6. Two-dimensional partial dependence plots illustrating the combined effects of pyrolysis temperature (Temp) and structural features on biocarbon elemental contents. (a–c) Carbon content for Temp versus I_D/I_G, Temp versus $I_{26°}/(I_{22°} + I_{26°})$, and Temp versus 002-FWHM; (d–f) Hydrogen content for Temp versus D-FWHM, Temp versus I_D/I_G, and Temp versus 100-FWHM; (g–i) Oxygen content for Temp versus G-FWHM, Temp versus d_{002}, and Temp versus $I_{26°}/(I_{22°} + I_{26°})$.

4. Conclusions

This study explores that the elemental composition of biocarbon is primarily governed by pyrolysis temperature and the resultant structural evolution, which can be effectively quantified using Raman- and XRD-derived descriptors and machine learning models. Among the evaluated models, Gradient Boosting exhibits a superior predictive performance for C, H, and O elements, revealing the robustness and suitability of nonlinear approaches for describing complex structure-composition relationships. Correlation analysis, SHAP interpretation and partial dependence analyses consistently identify the pyrosis temperature as the dominant factor driving carbon enrichment and depletion of hydrogen and oxygen, being attributable to progressive aromatization, dehydrogenation and deoxygenation. Furthermore, structural parameters provide complementary insights by linking defect density and graphitic ordering to elemental composition. Ultimately, this study has demonstrated a rapid and non-destructive strategy for accurately predicting large-scale biocarbon compositions and properties applicable for specific biocarbon design and assembly.

CRediT authorship contribution statement

Junsheng Yu: Writing – original draft, Visualization, Software, Methodology, Investigation. **Jing Hong:** Writing – original draft, Software, Methodology, Investigation. **Yujia Fang:** Writing – original draft, Software, Methodology, Investigation. **Jiacheng Kang:** Writing – original draft, Software, Methodology, Investigation. **Zongtai Tang:** Writing – review & editing, Investigation. **Lei Zhu:** Writing – review & editing, Investigation. **Bing Gao:** Writing – review & editing, Investigation. **Ling Luo:** Software, Methodology, Investigation. **Dan Sun:** Software, Methodology, Investigation. **Yanting Wang:** Writing – review & editing, Funding acquisition. **Liangcai Peng:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Hua Yu:** Writing – review & editing, Supervision, Methodology, Funding acquisition.

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

This work was supported by the National Natural Science Foundation of China (32500325 to H. Y), the Initiative Grant of Hubei University of Technology for High-level Talents (GCC20230001 to L. P).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.biortech.2026.134978.

Data availability

The data that has been used is confidential.

References

Abdeldayem, O.M., Al-Sakkari, E.G., Ortiz, D., Dupont, C., Ferras, D., Ouali, M.-S., Ragab, A., Kennedy, M., 2025. Human-centric ensemble AI for hydrothermal carbonization modeling and hydrochar properties prediction. J. Environ. Chem. Eng. 13 (5), 117826. https://doi.org/10.1016/j.jece.2025.117826.

Ai, Y., Wang, H., Liu, P., Yu, H., Sun, M., Zhang, R., Tang, J., Wang, Y., Feng, S., Peng, L., 2024. Insights into contrastive cellulose nanofibrils assembly and nanocrystals catalysis from dual regulations of plant cell walls. Sci. Bull. 69 (24), 3815–3819. https://doi.org/10.1016/j.scib.2024.06.013.

Bakr, M.M.A., Wang, Y., Hao, P., Dawoud, M.M.A., Peng, L., Wang, Y., 2025. Biochar from Al (OH)$_3$-activated sugarcane bagasse for efficient removal of cadmium: Optimization using the response surface method (RSM). Biomass Convers. Bior. 15 (12), 19487–19504. https://doi.org/10.1007/s13399-025-06594-6.

Buratti, Y., Le Gia, Q.T., Dick, J., Zhu, Y., Hameiri, Z., 2020. Extracting bulk defect parameters in silicon wafers using machine learning models. npj Comput. Mater. 6 (1), 142. https://doi.org/10.1038/s41524-020-00410-7.

Chen, Z., Gong, F., Song, J., Zhang, K., 2025b. Interpretable prediction of biomass-derived biochar characteristics: combining machine learning with shapley additive explanation. Energy 335, 137877. https://doi.org/10.1016/j.energy.2025.137877.

Chen, Q., Xue, P., Yang, Z., Wang, C., Yang, H., Zhang, S., 2025a. Machine learning for product distribution prediction of one-step pyrolysis model of coal. J. Energy Inst. 121, 102152. https://doi.org/10.1016/j.joei.2025.102152.

Deng, Y., Pu, B., Tang, X., Liu, X., Tan, X., Yang, Q., Wang, D., Fan, C., Li, X., 2024. Machine learning prediction of fundamental sewage sludge biochar properties based on sludge characteristics and pyrolysis conditions. Chemosphere 369, 143812. https://doi.org/10.1016/j.chemosphere.2024.143812.

Dong, X., Wang, F., Fan, X., Zhao, Y.-P., Wei, X.-Y., Wang, R.-Y., Ma, F.-Y., Liu, J.-M., Li, B., 2019. Evaluation of elemental composition obtained by using mass spectrometer and elemental analyzer: a case study on model compound mixtures and a coal-derived liquid. Fuel 245, 392–397. https://doi.org/10.1016/j.fuel.2019.02.089.

Elish, M.O., 2014. A comparative study of fault density prediction in aspect-oriented systems using MLP, RBF, KNN, RT, DENFIS and SVR models. Artif. Intell. Rev. 42 (4), 695–703. https://doi.org/10.1007/s10462-012-9348-9.

Fan, S., Zhang, W., Fan, X., Wang, S., Fang, X., Zhou, N., Xu, H., 2024. Hydrothermal pretreatment and pyrolytic conversion of biogas residue into biochar for efficient adsorption of tetracycline. Fuel 358, 130244. https://doi.org/10.1016/j.fuel.2023.130244.

Fu, Y., Cleall, P., Jin, F., 2026. Optimising biochar from agricultural residues: predicting elemental composition with machine learning. Renew. Energ. 256, 124071. https://doi.org/10.1016/j.renene.2025.124071.

Gao, B., Yu, J., Luo, L., Zhu, L., Tang, Z., Wang, Y., Kang, J., Sun, D., Peng, H., Wang, Y., Peng, L., Yu, H., 2026. Cascading valorization of *Miscanthus* biomass for bioethanol, lactic acid and nanomaterials via green-like pretreatment with reusable fertilizers. Renew. Energy 264, 125555. https://doi.org/10.1016/j.renene.2026.125555.

Haider Jaffari, Z., Jeong, H., Shin, J., Kwak, J., Son, C., Lee, Y.-G., Kim, S., Chon, K., Hwa Cho, K., 2023. Machine-learning-based prediction and optimization of emerging contaminants' adsorption capacity on biochar materials. Chem. Eng. J. 466, 143073. https://doi.org/10.1016/j.cej.2023.143073.

Hongwei, Z., Hongbao, Z., Haibin, G., Rupeng, Z., Li, W., Bokai, B., 2025. Regulation mechanisms and kinetic characterization of nonionic surfactants on bituminous coal wettability across multiple scales. Surf. Interf. 79, 108193. https://doi.org/10.1016/j.surfin.2025.108193.

Hu, Z., Li, Q., Chen, Y., Li, T., Wang, Y., Zhang, R., Peng, H., Wang, H., Wang, Y., Tang, J., Nauman Aftab, M., Peng, L., 2023. Intermittent ultrasound retains cellulases unlock for enhanced cellulosic ethanol with high-porosity biochar for dye adsorption using desirable rice mutant straw. Bioresour. Technol. 369, 128437. https://doi.org/10.1016/j.biortech.2022.128437.

Huang, W., Li, K., Miao, S., 2026. Real-time process design enabled by an interpretable GBDT model for high-fidelity prediction of pyrolysis products from crop residues. Bioresour. Technol. 442, 133742. https://doi.org/10.1016/j.biortech.2025.133742.

Kandpal, S., Tagade, A., Sawarkar, A.N., 2024. Critical insights into ensemble learning with decision trees for the prediction of biochar yield and higher heating value from pyrolysis of biomass. Bioresour. Technol. 411, 131321. https://doi.org/10.1016/j.biortech.2024.131321.

Kravchenko, E., Minkina, T., Privizentseva, D., Zhelnin, M., Bauer, T., Zhao, Y., Srivastava, S., Shan, S., 2025. Machine learning prediction of biochar properties derived from food waste. J. Environ. Chem. Eng. 13 (3), 116791. https://doi.org/10.1016/j.jece.2025.116791.

Leng, L., Yang, L., Lei, X., Zhang, W., Ai, Z., Yang, Z., Zhan, H., Yang, J., Yuan, X., Peng, H., Li, H., 2022. Machine learning predicting and engineering the yield, N content, and specific surface area of biochar derived from pyrolysis of biomass. Biochar 4 (1), 63. https://doi.org/10.1007/s42773-022-00183-w.

Li, Y., Gupta, R., You, S., 2022. Machine learning assisted prediction of biochar yield and composition via pyrolysis of biomass. Bioresour. Technol. 359, 127511. https://doi.org/10.1016/j.biortech.2022.127511.

Li, Y., He, B., Zhang, H., Liu, J., Li, S., Wang, H., Peng, H., Wang, Y., Dai, J., Wang, Y., Peng, L., Kang, H., 2024c. Enriched extensin and cellulose for non-collapse biochar assembly to maximize carbon porosity and dye adsorption with high bioethanol production. Ind. Crop Prod. 222, 119924. https://doi.org/10.1016/j.indcrop.2024.119924.

Li, X., Liao, H., Fan, C., Hu, H., Li, Y., Li, J., Yi, Z., Cai, X., Peng, L., Tu, Y., 2016. Distinct geographical distribution of the *Miscanthus* accessions with varied biomass enzymatic saccharification. PLoS One 11 (8), e0160026. https://doi.org/10.1371/journal.pone.0160026.

Li, J., Pan, L., Huang, Y., Liu, X., Ye, Z., Wang, Y., 2024a. Biochar design for antibiotics adsorption via a hybrid machine-learning-based optimization framework. Sep. Purif. Technol. 348, 127666. https://doi.org/10.1016/j.seppur.2024.127666.

Li, T., Peng, H., He, B., Hu, C., Zhang, H., Li, Y., Yang, Y., Wang, Y., Bakr, M.M.A., Zhou, M., Peng, L., Kang, H., 2024b. Cellulose de-polymerization is selective for bioethanol refinery and multi-functional biochar assembly using brittle stalk of corn mutant. Int. J. Biol. Macromol. 264, 130448. https://doi.org/10.1016/j.ijbiomac.2024.130448.

Liang, J., Chen, X., Duan, X., Gu, X., Zhao, X., Zha, S., Chen, X., 2024b. Natural aging and adsorption/desorption behaviors of polyethylene mulch films: Roles of film types and exposure patterns. J. Hazard. Mater. 466, 133588. https://doi.org/10.1016/j.jhazmat.2024.133588.

Liang, H., Li, W., Li, Y., Li, Y., 2024a. Machine learning-based multi-objective optimization and physical-geometrical competitive mechanisms for 3D woven thermal protection composites. Int. J. Heat Mass Transf. 232, 125920. https://doi.org/10.1016/j.ijheatmasstransfer.2024.125920.

Liao, H., Fan, S., Xie, Y., Yang, X., Wang, L., Chen, W., 2025. Machine learning prediction of pore structure and nitrogen content of N-doped biochar derived from biomass pyrolysis: Effects of biomass elemental compositions, preparation processes, and preparation methods. Fuel 398, 135559. https://doi.org/10.1016/j.fuel.2025.135559.

Lu, X., Han, Y., Wang, L., Chen, H., Feng, Y., Chen, Y., Feng, Y., Xue, L., 2025. Machine learning-driven optimization of metal-modified biochar for phosphorus adsorption and wastewater remediation. J. Environ. Chem. Eng. 13 (5), 118839. https://doi.org/10.1016/j.jece.2025.118839.

Luo, W., Li, J., Yu, Y., Su, W., Ren, Z., Niu, B., Chao, Z., Fan, J., Yi, W., 2025. Synthesis of porous CNS@C@NC concave cubic nanobox composite as cathode material for aluminum batteries. Chem. Eng. J. 510, 161898. https://doi.org/10.1016/j.cej.2025.161898.

Mariuzza, D., Lin, J.-C., Volpe, M., Fiori, L., Ceylan, S., Goldfarb, J.L., 2022. Impact of Co-Hydrothermal carbonization of animal and agricultural waste on hydrochars' soil amendment and solid fuel properties. Biomass Bioenergy 157, 106329. https://doi.org/10.1016/j.biombioe.2021.106329.

Murtaza, G., Ahmed, Z., Usman, M., Li, Y., Tariq, A., Rizwan, M., 2023. Effects of biotic and abiotic aging techniques on physiochemical and molecular characteristics of biochar and their impacts on environment and agriculture: a review. J. Soil Sci. Plant Nutr. 23 (2), 1535–1564. https://doi.org/10.1007/s42729-023-01201-x.

Pakzad, S.S., Ghalehnovi, M., Ganjifar, A., 2024. A comprehensive comparison of various machine learning algorithms used for predicting the splitting tensile strength of steel fiber-reinforced concrete. Case Stud. Constr. Mater. 20, e03092. https://doi.org/10.1016/j.cscm.2024.e03092.

Palansooriya, K.N., Li, J., Dissanayake, P.D., Suvarna, M., Li, L., Yuan, X., Sarkar, B., Tsang, D.C.W., Rinklebe, J., Wang, X., Ok, Y.S., 2022. Prediction of soil heavy metal immobilization by biochar using machine learning. Environ. Sci. Technol. 56 (7), 4187–4198. https://doi.org/10.1021/acs.est.1c08302.

Pan, X., Wang, X., Liu, L., Kuang, S., Zheng, H., 2023. CO_2 and H_2O as sweep gases elevated carbon stability and decreased phytotoxicity of biochars. Chem. Eng. J. 472, 145035. https://doi.org/10.1016/j.cej.2023.145035.

Ranzi, E., Debiagi, P.E.A., Frassoldati, A., 2017. Mathematical modeling of fast biomass pyrolysis and bio-oil formation. note II: secondary gas-phase reactions and bio-oil formation. ACS Sustain. Chem. Eng. 5 (4), 2882–2896. https://doi.org/10.1021/acssuschemeng.6b03098.

Song, C., Cai, F., Yang, S., Wang, L., Liu, G., Chen, C., 2024a. Machine learning-based prediction of methane production from lignocellulosic wastes. Bioresour. Technol. 393, 129953. https://doi.org/10.1016/j.biortech.2023.129953.

Song, Y., Huang, Z., Jin, M., Liu, Z., Wang, X., Hou, C., Zhang, X., Shen, Z., Zhang, Y., 2024b. Machine learning prediction of biochar physicochemical properties based on biomass characteristics and pyrolysis conditions. J. Anal. Appl. Pyrol. 181, 106596. https://doi.org/10.1016/j.jaap.2024.106596.

Sun, Y., Sun, P., Jia, J., Liu, Z., Huo, L., Zhao, L., Zhao, Y., Niu, W., Yao, Z., 2024. Machine learning in clarifying complex relationships: biochar preparation procedures and capacitance characteristics. Chem. Eng. J. 485, 149975. https://doi.org/10.1016/j.cej.2024.149975.

Wang, W., Chang, J.-S., Lee, D.-J., 2024b. Machine learning applications for biochar studies: a mini-review. Bioresour. Technol. 394, 130291. https://doi.org/10.1016/j.biortech.2023.130291.

Wang, H., Li, S., Wu, L., Zou, W., Zhang, M., Wang, Y., Lv, Z., Chen, P., Liu, P., Yang, Y., Peng, L., Wang, Y., 2025a. Semi-overexpressed OsMYB86L2 specifically enhances cellulose biosynthesis to maximize bioethanol productivity by cascading lignocellulose depolymerization via integrated rapid-physical and recyclable-chemical processes. Green Chem. 27 (30), 9127–9143. https://doi.org/10.1039/D5GC00658A.

Wang, M., Wang, Y., Liu, J., Yu, H., Liu, P., Yang, Y., Sun, D., Kang, H., Wang, Y., Tang, J., Fu, C., Peng, L., 2024a. Integration of advanced biotechnology for green carbon. Green Carbon. 2 (2), 164–175. https://doi.org/10.1016/j.greenca.2024.02.006.

Wang, Y., Zhang, H., Li, Y., Yu, H., Sun, D., Yang, Y., Zhang, R., Yu, L., Ma, F., Aftab, M. N., Peng, L., Wang, Y., 2025b. Effective xylan integration for remodeling biochar uniformity and porosity to enhance chemical elimination and CO_2 adsorption. Int. J. Biol. Macromol. 291, 138865. https://doi.org/10.1016/j.ijbiomac.2024.138865.

Wei, X., Liu, Y., Shen, L., Lu, Z., Ai, Y., Wang, X., 2024. Machine learning insights in predicting heavy metals interaction with biochar. Biochar 6 (1), 10. https://doi.org/10.1007/s42773-024-00304-7.

Wen, S.-C., Yang, C.-H., 2021. Time series analysis and prediction of nonlinear systems with ensemble learning framework applied to deep learning neural networks. Inf. Sci. 572, 167–181. https://doi.org/10.1016/j.ins.2021.04.094.

Yang, W., He, B., Yu, H., Zhang, H., Li, Y., Liu, J., Peng, H., Wang, H., Liu, P., Wang, Y., Peng, L., Sun, D., 2025. An optimal conversion cascading of distinct lignocelluloses to maximize bioethanol productivity and bio-adsorption capacities of Cd^{2+}/Cr(VI) and red/blue dyes selective for complete biomass recycling. Int. J. Biol. Macromol. 321, 146214. https://doi.org/10.1016/j.ijbiomac.2025.146214.

Yin, Q., Wang, M., Zhu, X., Gao, X., Wang, R., Zhao, Z., 2025. Research on the production and characteristics of carbon materials with high CO_2 adsorption performance: based on machine learning and dung beetle optimizer methods. Carbon 239, 120323. https://doi.org/10.1016/j.carbon.2025.120323.

Yu, H., Hu, M., Hu, Z., Liu, F., Yu, H., Yang, Q., Gao, H., Xu, C., Wang, M., Zhang, G., Wang, Y., Xia, T., Peng, L., Wang, Y., 2022. Insights into pectin dominated enhancements for elimination of toxic Cd and dye coupled with ethanol production in desirable lignocelluloses. Carbohydr. Polym. 286, 119298. https://doi.org/10.1016/j.carbpol.2022.119298.

Yu, J., Tang, Z., Zhu, L., Gao, B., Hong, J., Fang, Y., Kang, J., Sun, D., Peng, H., He, B., Bakr, M.M.A., Wang, Y., Peng, L., Yu, H., 2026. Low-temperature $ZnCl_2$ activation of distinct *Miscanthus* lignin as highly-porous biochar assembly for efficient removal of organic dyes, tetracycline and Cr(VI). Ind. Crop Prod. 240, 122630. https://doi.org/10.1016/j.indcrop.2026.122630.

Yu, H., Zhang, G., Liu, J., Liu, P., Peng, H., Teng, Z., Li, Y., Ren, X., Fu, C., Tang, J., Li, M., Wang, Y., Wang, L., Peng, L., 2025. A functional cascading of lignin modification via repression of caffeic acid O-methyltransferase for bioproduction and anti-oxidation in rice. J. Adv. Res. 78, 1–9. https://doi.org/10.1016/j.jare.2025.01.048.

Yuan, R., Guo, Y., Gurgan, I., Siddique, N., Li, Y.-S., Jang, S., Noh, G.A., Kim, S.H., 2025. Raman spectroscopy analysis of disordered and amorphous carbon materials: a review of empirical correlations. Carbon 238, 120214. https://doi.org/10.1016/j.carbon.2025.120214.

Zhang, R., Gao, H., Wang, Y., He, B., Lu, J., Zhu, W., Peng, L., Wang, Y., 2023a. Challenges and perspectives of green-like lignocellulose pretreatments selectable for low-cost biofuels and high-value bioproduction. Bioresour. Technol. 369, 128315. https://doi.org/10.1016/j.biortech.2022.128315.

Zhang, R., Hu, Z., Wang, Y., Hu, H., Li, F., Li, M., Ragauskas, A., Xia, T., Han, H., Tang, J., Yu, H., Xu, B., Peng, L., 2023b. Single-molecular insights into the breakpoint of cellulose nanofibers assembly during saccharification. Nat. Commun. 14 (1), 1100. https://doi.org/10.1038/s41467-023-36856-8.

Zhang, P., Wang, B., Liu, H., Rui, Y., Li, G., Zhang, J., 2025. Biomass pyrolysis characterisation based on machine learning: identification of key factors affecting biochar stability. Biomass Bioenerg. 203, 108293. https://doi.org/10.1016/j.biombioe.2025.108293.

Zhao, J., Jiang, Y., Chen, X., Wang, C., Nan, H., 2025. Unlocking the potential of element-doped biochar: from tailored synthesis to multifunctional applications in environment and energy. Biochar 7 (1), 77. https://doi.org/10.1007/s42773-025-00467-x.

Zhu, X., Li, Y., Wang, X., 2019. Machine learning prediction of biochar yield and carbon contents in biochar based on biomass characteristics and pyrolysis conditions. Bioresour. Technol. 288, 121527. https://doi.org/10.1016/j.biortech.2019.121527.