

Machine Learning Accelerates the Directional Construction of the Specific Surface Area of Biochar

Jian Kang, Huamu Chen, Rui Hou,* Peng Sun, Fawang Li, Yanzhong Han, Guanyu Lei, and Jian Li

Biochar exhibits application potential for treating wastes and reducing carbon emissions, thereby improving efficiency in the petrochemical industry. Application prospects are particularly prominent in the remediation of petroleum-contaminated soil and carbon dioxide capture. Specific surface area of biochar serves as a key parameter governing its environmental application performance. However, complexity of biomass precursors and pyrolysis processes poses significant challenges to targeted design and prediction of biochar specific surface area *via* conventional experimental approaches. In this work four models were constructed and compared. The Random Forest model exhibited the best generalization ability, with a coefficient of determination of 0.79 and a root mean square error of 57.88 on the test set, thus being identified as the optimal prediction model. Pyrolysis process parameters were more dominant than elemental composition of biochar, and pyrolysis temperature was the most critical feature. Recommended pyrolysis parameters include temperatures above 700 °C and time exceeding 3 h, while elemental composition of biochar should favor a chemical composition with high carbon content (>40%) and high nitrogen content (>3%). These findings significantly reduce trial-and-error costs and accelerate the targeted development of biochar-based environmental materials, thereby advancing the practical application of biochar in pollution control and climate change mitigation.

DOI: 10.15376/biores.21.3.6218-6233

Keywords: Biochar; Machine learning; Random Forest; Specific surface area

Contact information: The Oil Production Technology Research Institute of No. 10 Oil Production Plant, Changqing Oilfield Company, 745100, Qingyang, Gansu, China;

* Corresponding author: gysh2001@126.com R.H

INTRODUCTION

Biochar refers to a carbon-rich, stable solid material produced from organic waste *via* pyrolysis under oxygen-limited or anaerobic conditions (Kumar *et al.* 2023; Senadheera *et al.* 2025). In recent years, it has sparked extensive research interest in global sustainable development and environmental remediation fields. Biochar raw materials mainly comprise agricultural waste (*e.g.*, rice straw, corn stover), forestry wastes (*e.g.*, wood chips, leaves), municipal organic waste, and other biomass materials (Kumar *et al.* 2023; Bano *et al.* 2025). Diverse raw material sources endow biochar with low-cost and environmental friendly characteristics, while also making it a key carrier for carbon sequestration and resource recycling. Biochar production typically relies on pyrolysis technology, where organic matter in raw materials is converted into high-carbon biochar under controlled temperature and residence time. Biochar encompasses characteristics such as large specific

surface area (SSA), strong adsorption capacity, good thermal stability, and environmental friendliness. These characteristics give biochar broad application prospects in wastewater treatment, soil amendment, gas adsorption, and other fields (Xiao *et al.* 2018; Ringsby *et al.* 2024; Bano *et al.* 2025).

Among the diverse applications of biochar, adsorption occupies a central position due to its high efficiency in environmental pollution remediation (Lu *et al.* 2025; Yi *et al.* 2025). For instance, in the petroleum industry, large volumes of oily wastewater and refining wastewater generated during production processes contain recalcitrant hydrocarbon pollutants and toxic substances. These pose a severe threat to aquatic ecosystems. Biochar has a hydrophobic surface, a well-developed porous structure, and tunable surface chemical properties. The biochar can effectively adsorb such oily pollutants and dissolved organic fractions, providing an economically feasible solution for advanced purification of water bodies in the petroleum industry (Liu *et al.* 2023; Zhou *et al.* 2023; Chen *et al.* 2024). As the primary physical parameter governing adsorption performance, SSA directly influences biochar adsorption capacity, catalytic activity, and soil amendment efficiency. Biochar adsorption capacity finds wide applications in water purification (*e.g.*, removal of heavy metal ions and organic pollutants) (Bano *et al.* 2025), soil remediation (*e.g.*, immobilization of organic toxins) (Yi *et al.* 2025), and gas capture (*e.g.*, adsorption of CO₂ and VOCs) (Xiao *et al.* 2018), with mechanisms relying on physical and chemical adsorption processes on the material surface. Among relevant parameters, SSA serves as the core indicator for evaluating biochar porosity development. A high SSA indicates a greater abundance of microporous and mesoporous structures, which significantly enhances adsorption capacity and rate. Additionally, SSA indirectly affects other key biochar properties, such as surface charge distribution (governing selectivity for ion adsorption), thermal stability, and bioavailability (Cui *et al.* 2024). Materials with low SSA often suffer from pore blockage, leading to slow adsorption kinetics and poor performance (Beryani *et al.* 2025). However, current production of biochar with SSA mainly relies on conventional trial-and-error approaches. Empirical optimization is conducted by repeatedly adjusting pyrolysis conditions (*e.g.*, temperature gradient experiments) or raw material composition and proportions. This approach is not only time-consuming and labor-intensive but also causes resource wastage, while struggling to capture nonlinear interactive effects among multiple parameters. More critically, environmental governance standards are becoming increasingly stringent. The industry urgently demands efficient and scalable preparation strategies for customized production of biochar with high SSA. Therefore, developing a data-driven prediction tool to achieve rapid and accurate estimation of SSA has become an urgent need to break through bottlenecks in biochar applications.

In recent years, rapid advancements in data science and artificial intelligence have driven the mature application of machine learning (ML) across diverse fields. Machine learning analyzes and learns from large datasets to uncover hidden laws and patterns, thus demonstrating remarkable advantages in prediction, classification, optimization and other tasks (Madika *et al.* 2025; Mortazavi 2025). In materials science and environmental science, ML has been widely applied. Typical applications include predicting material properties and environmental pollution (Palansooriya *et al.* 2022; Gao *et al.* 2025). Key strengths of ML include high efficiency, accuracy and the capability to handle complex nonlinear relationships. Leveraging existing experimental data, machine learning models automatically acquire mapping relationships between inputs and outputs to enable prediction of unknown data. For example, Liu *et al.* (2025) proposed a comprehensive ML-

based method to predict ammonia nitrogen adsorption capacity of biochar and determine optimal adsorption conditions. Zhang *et al.* (2024b) utilized machine learning to predict CO₂ adsorption capacity of biochar, providing a reliable approach for biochar adsorption performance evaluation. Thus, applying ML to predict biochar SSA reduces time and costs associated with traditional trial-and-error methods. This approach also establishes precise correlations between SSA and feedstock properties as well as pyrolysis conditions. It further delivers more accurate and efficient solutions for the design and preparation of biochar.

This study aims to develop a SSA prediction model for biochar based on machine learning. The model integrates C, H, O, N elemental compositions (carbon, hydrogen, oxygen, nitrogen contents) of biochar and key pyrolysis parameters (*e.g.*, pyrolysis time, pyrolysis temperature, heating rate). Through collecting and analyzing extensive experimental data, multiple ML algorithms (*e.g.*, random forest (RF), neural network (NN)) will be employed. These algorithms will explore the influence of different input features on SSA. The core objective is to optimize biochar preparation processes and improve SSA prediction accuracy *via* machine learning. This research ultimately seeks to provide more effective biochar materials for wastewater treatment, soil improvement and other applications. It also aims to promote further development and application of ML in materials science.

EXPERIMENTAL

Feature Selection

The SSA of biochar is primarily influenced by reaction conditions (pyrolysis parameters) and elemental composition of biochar. Reaction conditions determine the process of biomass carbonization. They serve as external factors that directly affect the SSA of biochar. Key reaction conditions include three characteristics: pyrolysis time, pyrolysis temperature, and heating rate. Elemental composition of biochar typically includes major elements such as C, H, O, N, along with minor elements such as sulfur (S) and phosphorus (P). Among these, C, H, O, and N are common to the vast majority of biomass. In contrast, S, P, and other trace elements occur at very low rates. Thus, C, H, O, and N contents are selected as input features. In summary, the C, H, O, N contents of biochar, together with three reaction conditions: pyrolysis time, pyrolysis temperature, and heating rate are considered as input features. These features are used to predict the SSA of biochar.

Data Collection

A biochar SSA (BC-SSA) dataset was constructed. Searches for biochar-related studies on Web of Science were performed with keywords comprising BC, SSA, pyrolysis process, elemental analyzer, and others. Also, about 44 eligible peer-reviewed articles in total were considered. There were 181 data points retrieved and utilized to construct an initial BC-SSA dataset. Information on the utilized data points is presented in Table S1. During data collection, the following strategies was adopted to ensure the completeness of the dataset: (1) Based on expert knowledge of BC, preliminary analysis of the data was performed; (2) Heating rate within input features had partial missing values. Such missing values were input using the mean of the data.

Model Construction and Evaluation

The dataset was randomly split into two parts: an 80% training set and a 20% testing set. Then a five-fold cross-validation was performed on the training set to identify optimal hyperparameters. Optimal hyperparameters for the output features across different models are presented in Table S2. After hyperparameter tuning, four models were retrained: RF, NN, extreme gradient boosting (XGBoost), and gradient boosting regression (GBR). The predictive performance of the four models was evaluated on the testing set. Prediction accuracies of the models were then assessed using the coefficient of determination (R^2) and root mean square error (RMSE). Calculation methods for R^2 and RMSE are as follows,

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_p^i - y_t^i)^2}{\sum_{i=1}^N (y_t^i - y_m)^2} \quad (1)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_t^i - y_p^i)^2}{N}} \quad (2)$$

where y_p^i is the predicted value for the output feature, y_t^i is the true value, y_m is the mean of all output features, and N is the number of samples in the training or test set.

The Shapley Additive exPlanations (SHAP) method was employed to analyze features. Feature importance and the impact on biochar active sites was assessed. The Pearson correlation coefficient (PCC) was used to investigate correlations between features.

RESULTS AND DISCUSSION

Model Construction and Evaluation

To accurately predict the SSA of biochar, four elemental compositions (C, H, O, N) of biochar and three key pyrolysis process parameters (pyrolysis time, pyrolysis temperature, heating rate (HR)) as input features were used. The procedure was used to systematically construct and optimize four typical machine learning models: RF, NN, XGBoost, and GBR. Prior to modeling, to evaluate the strength of linear associations between input features and their potential relationships with the target variable (SSA), the Pearson correlation coefficient matrix was first calculated for all variables. Figure 1 shows the results from Pearson analysis, which indicated that the absolute values of correlation coefficients between all pairs of features were below 0.8. This outcome demonstrates the absence of severe multicollinearity among features. Sufficiently low correlations between features ensure provision of relatively independent information to the models. This establishes a solid foundation for subsequent development of robust prediction models.

The dataset was split into training and testing sets. The performance of each model was evaluated in terms of R^2 and RMSE. Results clearly demonstrated the performance variations among different algorithms for this prediction task.

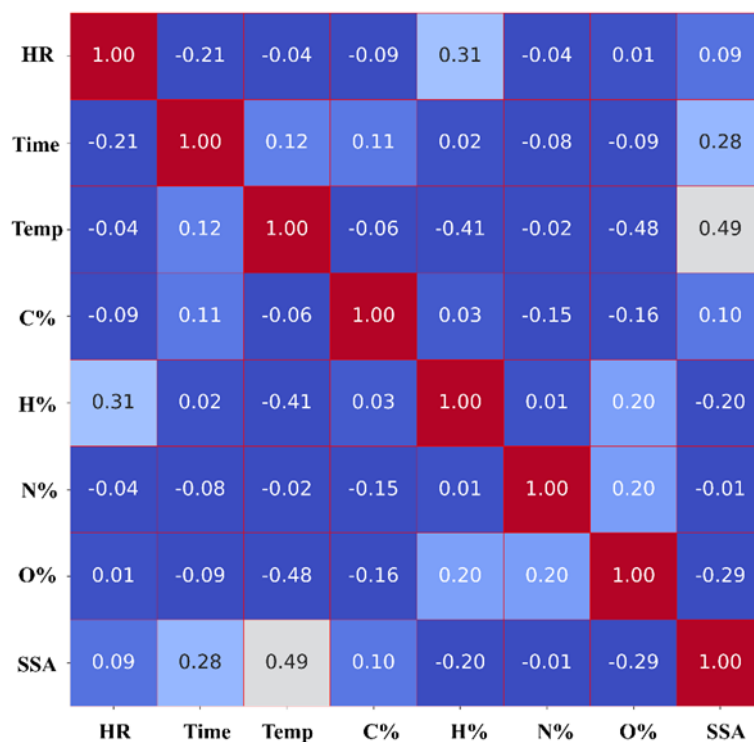


Fig. 1. Pearson correlation coefficient between features

As shown in Fig. 2 and Table 1, the four models exhibited strong predictive capability. However, generalization performance of the models on the testing set showed significant differences. Among them, the RF model performed best. The value of R^2 for the training set reached 0.88, and R^2 for the testing set attained as high as 0.79, which was much higher than that of other models. Meanwhile, RMSE of the testing set for the RF model (57.88) was the lowest. In contrast, gradient boosting models such as XGB and GBR, despite achieving R^2 values of 0.87 and 0.84 for the training set, indicated strong data fitting capability and showed a certain degree of performance decline on the tested set (with testing set R^2 values of 0.76 and 0.64, respectively). The NN model performed between the RF model and gradient boosting models, with a testing set R^2 of 0.78. Overall, while maintaining strong learning capability for training data, the RF model exhibited the strongest generalization capability. The RF model can provide relatively accurate and stable predictions of SSA for new biochar samples. This provides a reliable model foundation for subsequent feature importance analysis and practical applications.

Table 1. Comparative Evaluation with R^2 and RMSE of Various ML Models Trained with the SSA

Target		RF	XGB	GBR	NN
k	Train R^2	0.88	0.87	0.84	0.83
	Test R^2	0.79	0.76	0.64	0.78
	Train RMSE	6.3	62	42	71
	Test RMSE	82	95	102	92

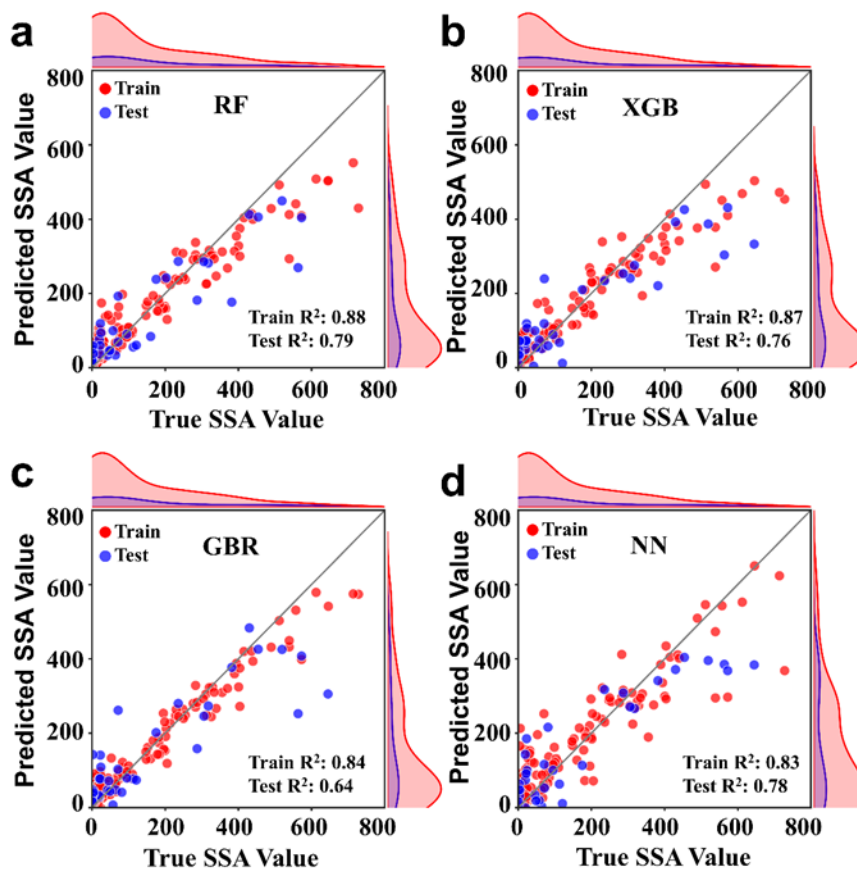


Fig. 2. Performance of machine learning models: a: RF; b: XGB; c: GBR; and d: NN

Feature Importance Analysis

To deeply explain the internal decision-making, the mechanism of the optimal RF model the SHAP method was employed. SHAP analysis quantitatively assesses the contribution and impact direction of each input feature on biochar SSA predictions. SHAP values, rooted in game theory, assign a unified and comparable value to the contribution of each feature in each prediction sample. This approach provides clear explanations at both global and local levels.

The SHAP feature plot in Fig. 3a intuitively illustrates the global relative importance of the seven input features in model predictions. Analysis results indicate that pyrolysis temperature is the most critical factor affecting biochar SSA. The average absolute SHAP value of temperature accounted for 39.0% of the sum of the average absolute SHAP values across all features, making it the most influential factor in the model. This indicates that, within the ML framework, pyrolysis temperature dominates the prediction of biochar properties—primarily by controlling the carbonization degree and the development of pore structures (Senadheera *et al.* 2024). Pyrolysis time ranks second, with a contribution of 14.3%. Sufficient pyrolysis time ensures adequate progression of pyrolysis reactions. Adequate time facilitates the formation of stable pore structures. Notably, elemental compositions of biomass also exhibit significant influence. Among these, the importance of N content (11.3%) and C content (11.2%) is very close. Such a finding indicates that chemical compositions of raw materials play a decisive role in pore properties of the final product. As the skeleton of biochar, carbon content may correlate with the proportion of fixed carbon, directly affecting the construction of carbon-based

skeletons. The importance of nitrogen may stem from its involvement in forming nitrogen-containing functional groups and heterocyclic structures during pyrolysis; these structures may affect the reactivity of carbon layers and the arrangement of graphite microcrystals, further influencing pore-forming behavior (Liu *et al.* 2019).

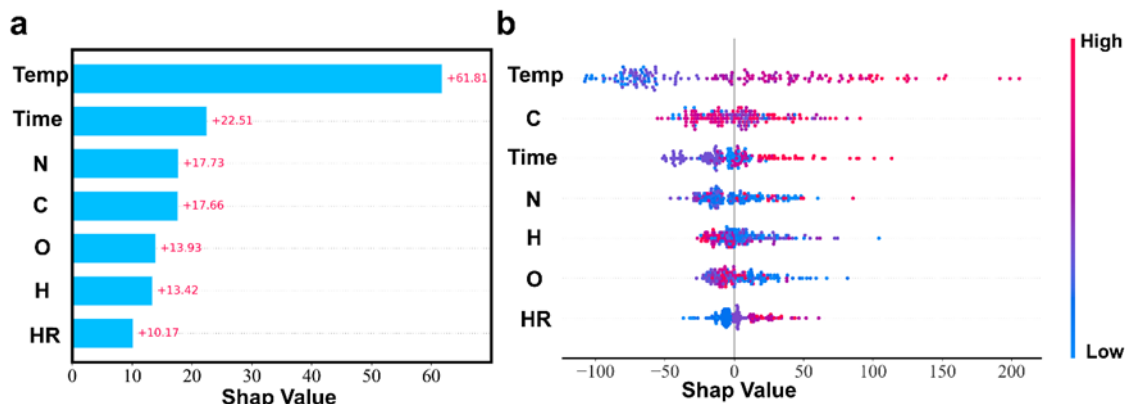


Fig. 3. SHAP feature analysis: a: Feature importance analysis based on SHAP mean absolute values; and b: Relationship between feature values and SSA based on SHAP point plots (Positive SHAP values correspond to increased predicted SSA, and negative values to decreased predicted SSA).

Next in order are O and H contents. The ratios of H/C and O/C are generally regarded as key indicators of biomass chemical structure and pyrolysis processes. Such ratios correlate strongly with volatile matter content (Iswardi *et al.* 2025). Levels of H and O contents directly affect the amount and rate of volatile product release during pyrolysis. This effect regulates the formation and collapse of pores. After aggregating features by category, total contribution of pyrolysis process parameters reaches 60.1%. In contrast, total contribution of biochar's own elemental compositions is 39.9%. This indicates that while chemical properties of raw materials form the foundation, post-treatment pyrolysis process conditions play a more critical dominant role in the controllable synthesis of biochar SSA. This provides clear guidance for practical production: precise regulation of pyrolysis temperature and time enables targeted design and optimization of biochar SSA over a wide range.

To further uncover how each feature affects specific surface area, SHAP feature dependence dot plots (Fig. 3b) were used. SHAP values for pyrolysis temperature generally show a positive correlation with feature values. In other words, higher pyrolysis temperatures correspond to positive SHAP values. This is fully consistent with the well-known rule that increasing temperature within a certain range promotes increases in SSA (Atinafu *et al.* 2025; Loc and Phuong 2025). Pyrolysis time and heating rate (HR) also exhibit positive correlations. Longer pyrolysis times promote increases in SSA. Higher HR values similarly correspond to larger SSA. Dot plot distributions for N and C contents may exhibit more complex nonlinear relationships. The distributions suggest a potential optimal range where contributions to increasing SSA are maximized. Excessively high or low contents may instead reduce such contributions. Dot plots for H and O contents show a negative correlation trend: higher H or O contents often correspond to negative SHAP values. This may result from higher H and O contents indicating higher aliphatic and volatile components in biomass during pyrolysis. Fixed carbon skeletons formed after pyrolysis may thus be less stable and less developed (Loc and Phuong 2025).

Feature Influence Mechanism

To further uncover complex nonlinear relationships between key features and biochar SSA, PDP plots were used. PDP analysis quantifies the marginal effect of each feature on model predictions after removing interference from other features. The four most important features from SHAP analysis were selected: pyrolysis temperature, pyrolysis time, carbon content, and nitrogen content. The PDP clearly depict dynamic variation processes. Each feature affects SSA within specific value ranges. Such plots provide valuable insights for understanding underlying mechanisms.

Figure 4a displays the partial dependence of the predicted SSA on pyrolysis temperature. The curve reveals a nonlinear relationship. In the low-temperature range of 200 to 500 °C, the curve remains flat. This indicates that SSA changes minimally during this stage. This aligns with characteristics of the initial pyrolysis stage. This stage primarily involves biomass dehydration, depolymerization, and partial hemicellulose decomposition. Limited pore development has been observed at this point (Yu *et al.* 2022). However, when temperature exceeds 500 °C, the curve begins to rise sharply, continuing until 900 °C. This phenomenon is highly consistent with the basic principles of biochar formation (Jin *et al.* 2025).

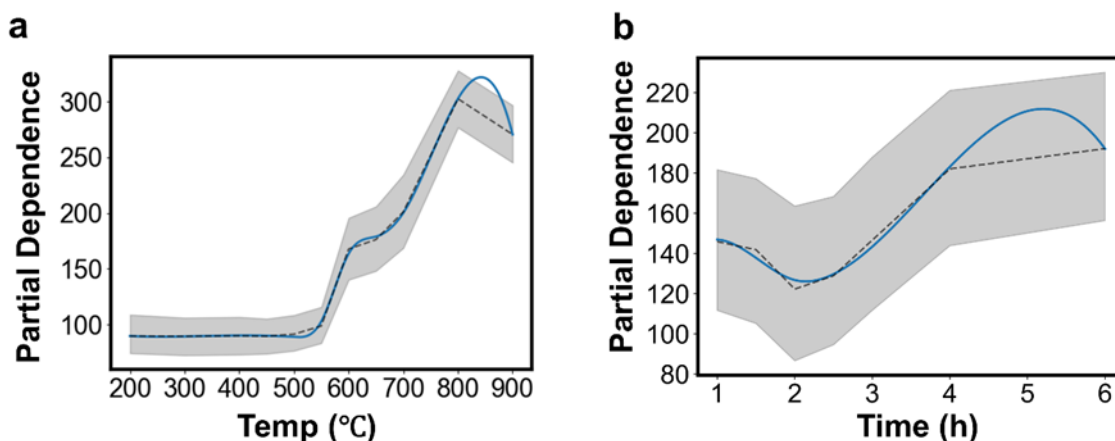


Fig. 4. Partial dependence plots showing the marginal effect of (a) pyrolysis temperature and (b) pyrolysis time on the model-predicted SSA of biochar. Note: The y-axis represents the partial dependence value (mean predicted SSA after averaging out the influence of all other features).

Typically, in the high-temperature region above 500 °C, cellulose and lignin undergo deep aromatization and rearrangement of graphite microcrystals. Large amounts of volatile substances escape. This creates abundant microporous and mesoporous structures. Such structures cause SSA to increase exponentially (Zhao *et al.* 2025; Xu *et al.* 2026). This result clearly identifies a critical threshold temperature of 500 °C for preparing biochar with high SSA. It holds significant value for process guidance. Figure 4b shows the partial dependence of the predicted SSA on pyrolysis time. The profile exhibits an initial decline followed by a gradual increase. During early pyrolysis (1 to 2 h), SSA decreases slightly. A possible explanation is the collapse or blockage of some initially formed micropores, which is likely due to secondary tar deposition or temporary pore clogging (Di *et al.* 2025). As pyrolysis continues between 2 h and 6 h, the curve stabilizes and rises steadily. Prolonged heating promotes more complete pyrolysis, decomposing intermediate deposits and allowing original pore structures to develop and stabilize. New

pores may also form. These results suggest that insufficient pyrolysis duration hinders the development of stable porous frameworks.

Figure 5a presents the PDP plot for carbon content. Overall, biochar SSA increases gradually with rising carbon content, reflecting the fundamental role of the carbon skeleton as the structural support for porosity. Once carbon content exceeds 70%, the slope becomes steeper. This indicates that pyrolysis becomes exceptionally effective in converting highly carbonized precursors into porous carbon materials with ultra-high SSA.

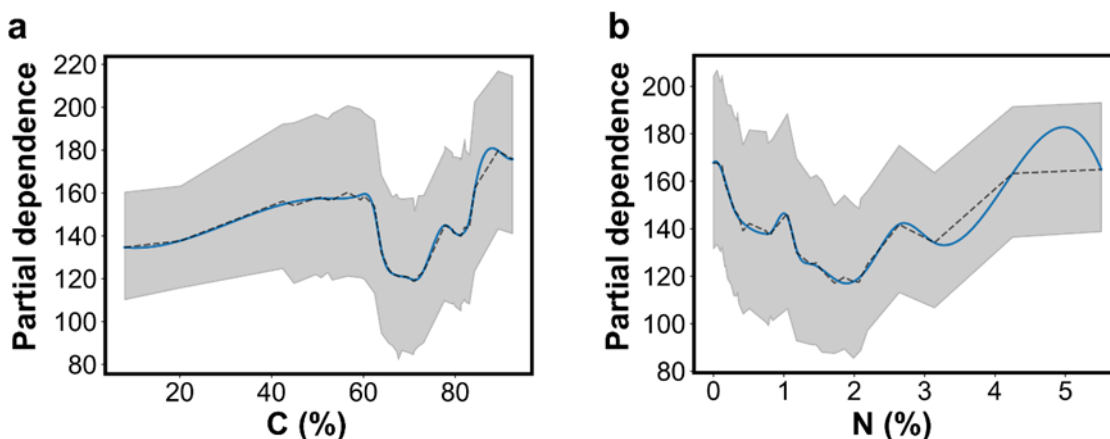


Fig. 5. Partial dependence plots showing the marginal effect of (a) the content of C element and (b) the content of N element on the model-predicted SSA of biochar. Note: The y-axis represents the partial dependence value (mean predicted SSA after averaging out the influence of all other features).

The PDP plot for nitrogen content (Fig. 5b) reveals a distinct inflection point in its influence on SSA. At low concentrations (0 to 2%), the curve declines, suggesting that limited nitrogen incorporation may promote non-porous nitrogen-containing structures or hinder graphitization, thereby suppressing pore development. As nitrogen content rises above 2% and reaches 5%, the curve shifts upward, indicating enhanced SSA. Two mechanisms may explain this trend: first, decomposition of nitrogen-containing functional groups releases gases that etch the carbon framework, creating new pores (Huang *et al.* 2025); second, nitrogen doping introduces defect sites, which can function as micropores (Zhang *et al.* 2024a). These findings highlight nitrogen's dual role in biochar preparation either inhibitory or promotive depending on its initial concentration in the feedstock.

While PDP effectively is able to illustrate the individual marginal effects of specific features on biochar SSA, real-world pyrolysis involves a complex interplay among these factors. To investigate potential synergistic or antagonistic effects, a two-dimensional interaction analysis was performed for two critical feature pairs. These pairs are pyrolysis temperature with pyrolysis time, and carbon content with nitrogen content. Results revealed that optimized parameter combinations for producing high-surface-area biochar, offering actionable, formula-like guidance for tailored synthesis.

Figure 6a presents a two-dimensional contour plot of the interaction between pyrolysis temperature and time. The SSA contours exhibit strong gradients across the parameter space. The region with temperatures below 600 °C and durations shorter than 2 h corresponds to the lowest predicted surface areas. These conditions represent insufficient pyrolysis, failing to develop a porous structure effectively. A distinct high-performance

“golden zone” is clearly identified. Maximum predicted SSA is achieved when pyrolysis temperature exceeds 700 °C and duration extends beyond 3 hours. This pattern demonstrates a significant synergistic effect between temperature and time. High temperature provides the essential energy for micropore formation. Sufficient duration ensures complete heat and mass transfer, allowing carbon framework stabilization (Leng *et al.* 2022; Loc and Phuong 2025). That finding clearly identifies a reliable strategy for efficient production of high-surface-area biochar.

Figure 6b shows the interaction contour plot for carbon and nitrogen contents. Achieving high SSA requires a specific compositional window, rather than independent carbon and nitrogen thresholds. The model predicts high surface areas when carbon content exceeds 40% and nitrogen content simultaneously surpasses 3%. This finding carries important physicochemical implications. A carbon content above 40% serves as a fundamental prerequisite, ensuring sufficient carbon for constructing porous networks. Given adequate carbon, higher nitrogen content (>3%) plays a beneficial role. Nitrogen incorporation into the carbon lattice likely induces electronic structure modifications and lattice distortions. Those modifications generate numerous defects. Those defects function as micropores. Furthermore, the decomposition of nitrogen-functional groups may release gaseous species. Released gases cause *in-situ* etching of the carbon matrix. Such etching creates additional porosity (Jiang *et al.* 2022; Li *et al.* 2025). This nitrogen-assisted pore formation aligns perfectly with the positive influence of higher nitrogen levels observed in the single-variable PDP analysis.

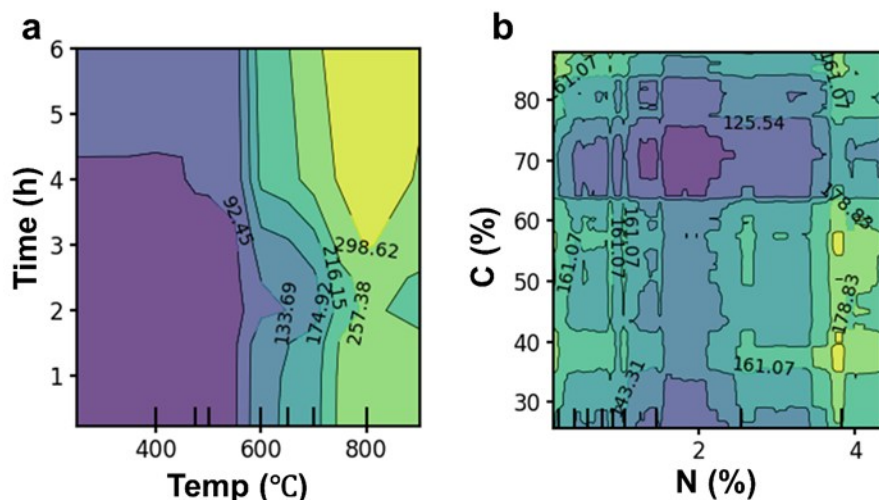


Fig. 6. Interaction between features: a: Pyrolysis temperature and pyrolysis time; and b: The content of elements C and the content of element N

The study provides robust methodological support and practical guidance for reducing experimental trial-and-error costs and enabling precise design of biochar SSA. Data-driven design paradigm and well-defined optimized parameter combinations provided in this study can directly guide applications in specific environmental remediation scenarios. Such scenarios include oil and gas field wastewater treatment and petrochemical VOCs abatement. This study offers a feasible material technical pathway for efficient treatment of "three wastes" pollution from the petrochemical industry.

Limitations

Despite the promising predictive performance and practical guidance provided by our machine learning framework, several limitations should be acknowledged. First, the dataset used in this study was compiled from diverse published literature, which inevitably introduces heterogeneity arising from differences in experimental conditions across studies. While such diversity enhances the generalizability of our models to some extent, it also introduces uncontrolled confounding factors that are not explicitly captured by the input features. Second, certain critical parameters (*e.g.*, exact gas flow rate, pressure, or the degree of volatile–char interactions) were often unreported in the original studies and therefore could not be included as predictive features. The absence of these variables may lead to an underestimation of prediction uncertainty. Consequently, the accuracy of model predictions for a specific experimental setup may be lower than the cross-validation metrics reported here. Future work should focus on expanding the dataset to enhance model generalizability and applying this predictive paradigm to other critical biochar properties.

CONCLUSIONS

1. A machine learning model based on the random forest (RF) algorithm was successfully developed to accurately predict biochar specific surface area from elemental composition and pyrolysis parameters. RF modeling, combined with the Shapley Additive ExPlanations (SHAP) system and partial dependence analysis, revealed that pyrolysis parameters dominate biochar specific surface area, with a greater influence than elemental composition. This finding empirically confirms that post-processing conditions can effectively modulate pore structure. Pyrolysis temperature emerged as the most critical variable: temperatures below 500 °C exhibited minimal effect, while above 500 °C a pronounced promotional effect was observed.
2. Regarding feedstock composition, carbon and nitrogen contents played the most significant and complex roles. Nitrogen exhibited a dual effect: inhibiting pore development at lower concentrations but promoting porosity at higher levels, possibly through synergistic interaction with carbon. Specific combinations were identified as optimal strategies for achieving high specific surface area: high temperature (>700 °C) with long duration (>3 h), and high carbon content (>40%) with high nitrogen content (>3%).
3. In summary, a systematic analytical framework encompassing model development, performance evaluation, mechanistic interpretation, and pathway optimization was established. This framework not only achieved reliable predictive performance within the scope of the collected data, but also extracted underlying physicochemical principles, transforming the machine learning model into valuable scientific knowledge. A practical application prospect of this framework is to guide the production of biochar with tailored specific surface area for environmental remediation, particularly in wastewater treatment and petroleum-contaminated soil rehabilitation, where biochar's surface area serves as a key parameter governing its performance.

Conflict of Interest

The authors declare no competing financial interest.

REFERENCES CITED

- Atinafu, D. G., Choi, J. Y., Nam, J., Kang, Y. J., and Kim, S. (2025). "Insights into the effects of biomass feedstock and pyrolysis conditions on the energy storage capacity and durability of standard biochar-based phase-change composites," *Biochar* 7(1), 18. <https://doi.org/10.1007/s42773-024-00396-1>
- Bano, A., Aziz, M. K., Prasad, B., Ravi, R., Shah, M. P., Lins, P. V. D., Meili, L., and Prasad, K. S. (2025). "The multifaceted power of biochar: A review on its role in pollution control, sustainable agriculture, and circular economy," *Environmental Chemistry and Ecotoxicology* 7, 286-304. <https://doi.org/10.1016/j.enceco.2025.01.004>
- Beryani, A., Flanagan, K., You, S. J., Forsberg, F., Viklander, M., and Blecken, G. T. (2025). "Critical field evaluations of biochar-amended stormwater biofilters for PFAS and other organic micropollutant removals," *Water Research* 281, article 123547. <https://doi.org/10.1016/j.watres.2025.123547>
- Chen, Y. H., Wang, F. M., Gao, J. Q., He, X. H., Liu, Q. L., and Liu, L. (2024). "Enhancing bioremediation of petroleum-contaminated soil by sophorolipids-modified biochar: Combined metagenomic and metabolomic analyses," *Science of the Total Environment* 951, article 175772. <https://doi.org/10.1016/j.scitotenv.2024.175772>
- Cui, S., Lv, J. L., Hough, R., Fu, Q., An, L. H., Zhang, Z. L., Ke, Y. X., Liu, Z. K., and Li, Y. F. (2024). "Recent advances and prospects of neonicotinoid insecticides removal from aquatic environments using biochar: Adsorption and degradation mechanisms," *Science of the Total Environment* 939, article 173509. <https://doi.org/10.1016/j.scitotenv.2024.173509>
- Di, H. K., Zhang, T., Ma, H. R., Zhang, G. Z., Li, K. J., Han, J. Z., and Yang, J. S. (2025). "Biomass tar modified biochar with enhanced micropores and oxygen content for high efficiency CO capture," *Separation and Purification Technology* 378, article 134678. <https://doi.org/10.1016/j.seppur.2025.134678>
- Gao, W. J., Xu, Y. S., Chang, X. L., Xu, X., Li, N., Yan, B. B., Chen, G. Y., and Duan, X. G. (2025). "Machine learning-driven global optimization of single-atom catalyst-mediated advanced oxidation processes," *Environmental Science and Technology* 59(44), 24044-24054. <https://doi.org/10.1021/acs.est.5c07237>
- Huang, T., Wu, Y., Liu, G. S., Chen, Y. Y., Huang, G. S., Fu, F. B., Zu, X. H., Li, H., Liu, Q. Y., and Zhang, W. L. (2025). "Modulation of micropores to remold the rate capability of lignin-derived porous carbon cathodes in zinc-ion hybrid supercapacitors," *Journal of Power Sources* 646, article 237282. <https://doi.org/10.1016/j.jpowsour.2025.237282>
- Iswardi, A. H., Hubble, A. H., Lehmann, J., and Goldfarb, J. L. (2025). "Can batch lab-scale studies of slow pyrolysis describe biochar yields and properties upon upscaling to a continuous auger kiln?" *Journal of Cleaner Production* 496, article 145150. <https://doi.org/10.1016/j.jclepro.2025.145150>
- Jiang, G. S., Senthil, R. A., Sun, Y. Z., Kumar, T. R., and Pan, J. Q. (2022). "Recent progress on porous carbon and its derivatives from plants as advanced electrode materials for supercapacitors," *Journal of Power Sources* 520, article 230886. <https://doi.org/10.1016/j.jpowsour.2021.230886>

- Jin, J. K., Ma, H. X., Liang, H. R., and Zhang, Y. Y. (2025). "Biopolymer-derived carbon materials for wearable electronics," *Advanced Materials* 37(22), article 202414620. <https://doi.org/10.1002/adma.202414620>
- Kumar, A., Bhattacharya, T., Shaikh, W. A., Roy, A., Chakraborty, S., Vithanage, M., and Biswas, J. K. (2023). "Multifaceted applications of biochar in environmental management: A bibliometric profile," *Biochar* 5(1), 11. <https://doi.org/10.1007/s42773-023-00207-z>
- Leng, L. J., Yang, L. H., Lei, X. N., Zhang, W. J., Ai, Z. J., Yang, Z. Q., Zhan, H., Yang, J. P., Yuan, X. Z., Peng, H. Y., and Li, H. L. (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, J. B., Wang, R. X., Han, L. L., Wang, T. Y., Asakura, Y., Wang, C. Y., Wang, G. X., Xu, X. T., and Yamauchi, Y. (2025). "Unveiling the neglected role of oxygen doping in nitrogen-doped carbon for enhanced capacitive deionization performance," *Nature Communications* 16(1), article 1996. <https://doi.org/10.1038/s41467-025-56694-0>
- Liu, C., Balasubramanian, P., An, J. X., and Li, F. Y. (2025). "Machine learning prediction of ammonia nitrogen adsorption on biochar with model evaluation and optimization," *NPJ Clean Water* 8(1), 13. <https://doi.org/10.1038/s41545-024-00429-z>
- Liu, X. R., Luo, Z. Y., Yu, C. J., and Xie, G. L. (2019). "Conversion mechanism of fuel-N during pyrolysis of biomass wastes," *Fuel* 246, 42-50. <https://doi.org/10.1016/j.fuel.2019.02.042>
- Liu, Z., Li, Z., Chen, S. G., and Zhou, W. Z. (2023). "Enhanced phytoremediation of petroleum-contaminated soil by biochar and urea," *Journal of Hazardous Materials* 453, article 131404. <https://doi.org/10.1016/j.jhazmat.2023.131404>
- Loc, N. X., and Phuong, D. T. M. (2025). "Optimizing biochar production: a review of recent progress in lignocellulosic biomass pyrolysis," *Frontiers of Agricultural Science and Engineering* 12(1), 148-172. <https://doi.org/10.15302/J-Fase-2024597>
- Lu, J. K., Li, Y., Wang, B., Zhao, T., Wang, M., and Si, H. Y. (2025). "Analysis of monomeric and competitive adsorption mechanisms of nutrient ions on biochar surfaces based on molecular dynamics simulations," *Bioresource Technology* 416, article 131746. <https://doi.org/10.1016/j.biortech.2024.131746>
- Madika, B., Saha, A., Kang, C., Buyantogtokh, B., Agar, J., Wolverton, C. M., Voorhees, P., Littlewood, P., Kalinin, S., and Hong, S. B. (2025). "Artificial intelligence for materials discovery, development, and optimization," *ACS Nano* 19(30), 27116-27158. <https://doi.org/10.1021/acsnano.5c04200>
- Mortazavi, B. (2025). "Recent advances in machine learning-assisted multiscale design of energy materials," *Advanced Energy Materials* 15(9), article 202403876. <https://doi.org/10.1002/aenm.202403876>
- Palansooriya, K. N., Li, J., Dissanayake, P. D., Suvarna, M., Li, L. Y., Yuan, X. Z., Sarkar, B., Tsang, D. C. W., Rinklebe, J., Wang, X. N., and Ok, Y. S. (2022). "Prediction of soil heavy metal immobilization by biochar using machine learning," *Environmental Science and Technology* 56(7), 4187-4198. <https://doi.org/10.1021/acs.est.1c08302>
- Ringsby, A. J., Ross, C. M., and Maher, K. (2024). "Sorption of soil carbon dioxide by biochar and engineered porous carbons," *Environmental Science and Technology* 58(19), 8313-8325. <https://doi.org/10.1021/acs.est.4c02015>

- Senadheera, S. S., Withana, P. A., Lim, J. Y., You, S. M., Chang, S. X., Wang, F., Rhee, J. H., and Ok, Y. S. (2024). "Carbon negative biochar systems contribute to sustainable urban green infrastructure: A critical review," *Green Chemistry* 26(21), 10634-10660. <https://doi.org/10.1039/d4gc03071k>
- Senadheera, S. S., Withana, P. A., You, S., Tsang, D. C. W., Hwang, S. Y., and Ok, Y. S. (2025). "Sustainable biochar: Market development and commercialization to achieve ESG goals," *Renewable and Sustainable Energy Reviews* 217, article 115744. <https://doi.org/10.1016/j.rser.2025.115744>
- Xiao, X., Chen, B. L., Chen, Z. M., Zhu, L. Z., and Schnoor, J. L. (2018). "Insight into multiple and multilevel structures of biochars and their potential environmental applications: A critical review," *Environmental Science and Technology* 52(9), 5027-5047. <https://doi.org/10.1021/acs.est.7b06487>
- Xu, T., Chen, L. Y., Chen, J., and Sadvakasova, A. K. (2026). "Evolution in functional groups of agricultural straw during pyrolysis at elevated temperature: Synchrotron radiation infrared spectroscopy (vol 437, 133125, 2025)," *Bioresource Technology* 439, article 133335. <https://doi.org/10.1016/j.biortech.2025.133335>
- Yi, B. J., Guo, F., Wei, Q., Huang, F., Senadheera, S. S., Masek, O., Sarmah, A. K., Rinklebe, J., and Ok, Y. S. (2025). "Functionalized biochar: Mechanisms of oxygen-containing functional groups in water pollutant remediation," *ACS ES&T Water*, 5(10), 5764-5784. <https://doi.org/10.1021/acsestwater.5c00416>
- Yu, S. J., Wang, L. Z., Li, Q. H., Zhang, Y. G., and Zhou, H. (2022). "Sustainable carbon materials from the pyrolysis of lignocellulosic biomass," *Materials Today Sustainability* 19, article 100209. <https://doi.org/10.1016/j.mtsust.2022.100209>
- Zhang, G. Q., Liu, F. Q., Zhu, Q. L., Qian, H., Zhong, S. C., Tan, J. Z., Zheng, A. M., Liu, F. J., and Jiang, L. L. (2024a). "Triple templates directed synthesis of nitrogen-doped hierarchically porous carbons from pyridine rich monomer as efficient and reversible SO adsorbents," *Small* 20(46), article 202404548. <https://doi.org/10.1002/smll.202404548>
- Zhang, J. J., Zhang, X., Li, X. Q., Song, Z. T., Shao, J. G., Zhang, S. H., Yang, H. P., and Chen, H. P. (2024b). "Prediction of CO adsorption of biochar under KOH activation via machine learning," *Carbon Capture Science and Technology* 13, article 100309. <https://doi.org/10.1016/j.ccst.2024.100309>
- Zhao, J. Q., Jiang, Y. Q., Chen, X. Y., Wang, C. Q., and Nan, H. Y. (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>
- Zhou, H. H., Huang, X. M., Jiang, L. J., Shen, Q., Sun, H., Yi, M., Wang, X., Yao, X. H., Wu, Y. F., Zhang, C. F., and Tang, J. W. (2023). "Improved degradation of petroleum contaminants in hydraulic fracturing flowback and produced water using laccase immobilized on functionalized biochar," *Environmental Technology and Innovation* 32, article 103280. <https://doi.org/10.1016/j.eti.2023.103280>

Article submitted: February 26, 2026; Peer review completed: May 3, 2026; Revised version received and accepted: May 8, 2026; Published: May 21, 2026.

DOI: 10.15376/biores.21.3.6218-6233

APPENDIX**Table S1.** Article Information of the Source of Data Set

1	DOI:10.1016/j.apcatb.2021.120926
2	DOI:10.1016/j.scitotenv.2020.144281
3	DOI:10.1016/j.energy.2019.05.214
4	DOI:10.1016/j.chemosphere.2021.130456
5	DOI:10.1016/j.jhazmat.2021.127225
6	DOI:10.1016/j.jhazmat.2021.127692
7	DOI:10.1016/j.jhazmat.2021.125294
8	DOI:10.1016/j.jaap.2018.06.018
9	DOI:10.1016/j.cej.2021.130370
10	DOI:10.1016/j.chemosphere.2019.125007
11	DOI:10.1016/j.cej.2022.135552
12	DOI:10.1016/j.jwpe.2022.102626
13	DOI:10.1016/j.chemosphere.2019.05.087
14	DOI:10.1016/j.cej.2019.121908
15	DOI:10.1016/j.cej.2019.06.006
16	DOI:10.1016/j.jaap.2017.07.015
17	DOI:10.1016/j.scitotenv.2020.137015
18	DOI:10.1016/j.jhazmat.2019.121980
19	DOI:10.1016/j.jenvman.2023.118076
20	DOI:10.1016/j.jhazmat.2023.130999
21	DOI:10.1016/j.watres.2023.119659
22	DOI:10.1016/j.chemosphere.2023.138690
23	DOI:10.1016/j.cej.2023.142908
24	DOI:10.1016/j.cej.2022.135669
25	DOI:10.1016/j.seppur.2022.121525
26	DOI:10.1016/j.scitotenv.2022.158532
27	DOI:10.1021/acs.energyfuels.9b01105
28	DOI:10.1016/j.scitotenv.2020.138672
29	DOI:10.1016/j.jece.2021.106049
30	DOI:10.1016/j.scitotenv.2020.137584
31	DOI:10.1016/j.jhazmat.2020.122048
32	DOI:10.1016/j.scitotenv.2020.136532
33	DOI:10.1016/j.cej.2021.130584
34	DOI:10.1016/j.envpol.2022.120644
35	DOI:10.1016/j.jenvman.2019.03.082
36	DOI:10.1016/j.molliq.2023.121257
37	DOI:10.1016/j.jclepro.2017.12.090
38	DOI:10.1016/j.envpol.2022.120832
39	DOI:10.1016/j.scitotenv.2020.136550
40	DOI:10.1016/j.ecmx.2021.100176
41	DOI:10.1016/j.scitotenv.2019.134767
42	DOI:10.1016/j.jenvman.2019.03.110
43	DOI:10.1016/j.ecoenv.2021.113107
44	DOI:10.1016/j.jece.2022.107328

Table S2. The Tuned Hyper-parameters of RF, GBR and XGB Model for the Prediction of SSA

Target	Model	Hyper-Parameters (Three Vital Parameters)		
SSA	RF	n_estimators	max_depth	min_samples_split
		58	25	1
	GBR	n_estimators	max_depth	learning_rate
		200	3	0.03
	XGB	n_estimators	max_depth	learning_rate
		300	8	0.01
	NN	hidden_layer_sizes	batch_size	solver
		100	16	adam