Li et al. Biochar (2025) 7:20
https://doi.org/10.1007/s42773-024-00424-0

Biochar

PERSPECTIVE Open Access

Optimizing biochar for carbon sequestration: a synergistic approach using machine learning and natural language processing

Jiayi Li[1,2], Yixuan Chen[1,3], Chaojie Wang[2], Hanbo Chen[4], Yurong Gao[5], Jun Meng[1], Zhongyuan Han[6], Lukas Van Zwieten[7], Yi He[8*], Caibin Li[8], Gerard Cornelissen[9,10] and Hailong Wang[1,2,11*]

Abstract
Biochar is a promising technology for carbon storage and greenhouse gas (GHG) reduction, but optimizing it is challenging due to the complexity of natural systems. Machine learning (ML) and natural language processing (NLP) offer solutions through enhanced data analysis and pattern recognition, ushering in a new era of biochar research.

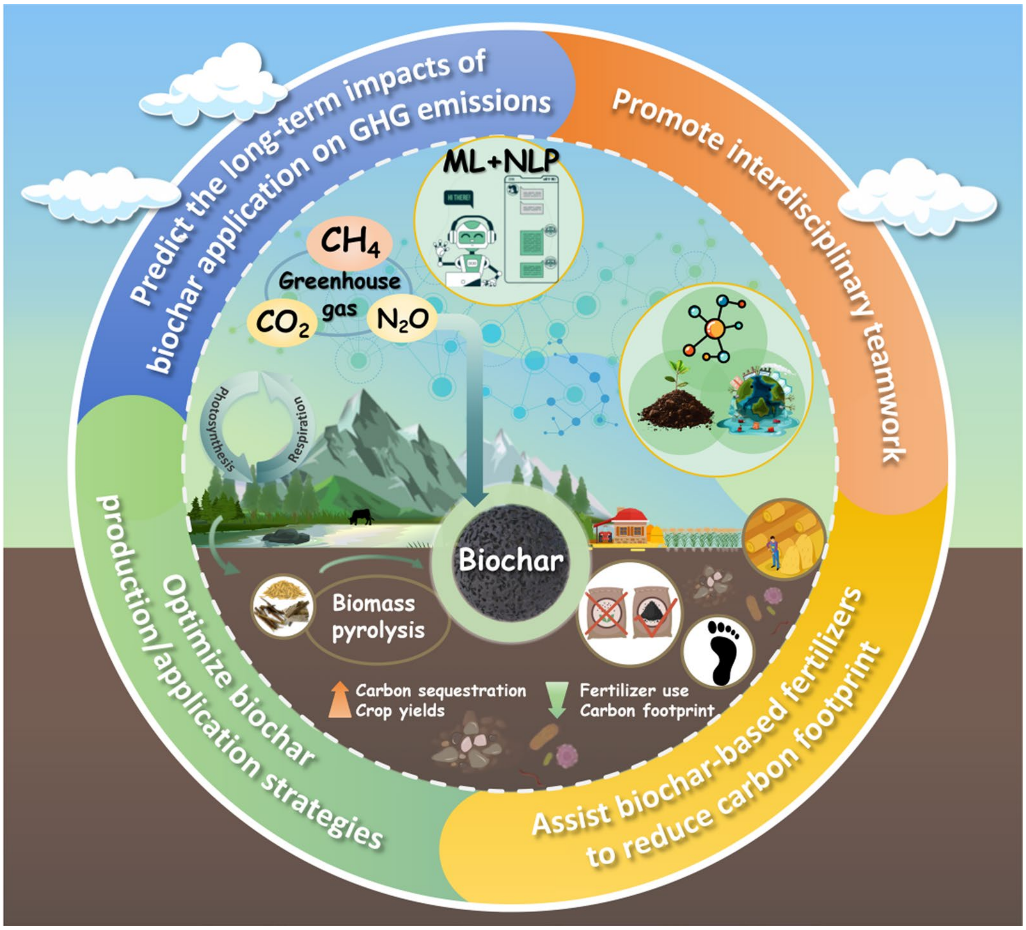
Highlights

- ML enhances biochar production by optimizing carbon sequestration and predicting long-term environmental impacts.
- NLP facilitates biochar data analysis, accelerates discoveries, and improves research productivity.
- Integrating biochar, ML, and NLP offers new pathways for achieving carbon neutrality and climate change mitigation.

Keywords Artificial intelligence, Biochar optimization, Carbon storage, Carbon neutrality, Bibliometric analysis

*Correspondence:
Yi He
yancaoheyi@163.com
Hailong Wang
hailong.wang@fosu.edu.cn
Full list of author information is available at the end of the article

Graphical Abstract



1 Introduction

In the wake of the United Nations' Paris Agreement in 2015, dedicated to combating global warming, numerous nations have pledged to achieve net-zero emissions (Wang et al. 2023). China, as one of the world leading producers of agricultural commodities and a significant contributor to greenhouse gas (GHG) emissions, has committed to peaking its emissions by 2030 and achieving carbon neutrality by 2060 (Xia et al. 2023). This commitment underscores the pivotal role of China in the global transition toward a sustainable and low-carbon future.

Biochar, a carbon-rich byproduct of biomass pyrolysis (Chen et al. 2022), enhances soil quality and contributes significantly to carbon management (Liu et al. 2022; Deng et al. 2024), aligning with key United Nations Sustainable Development Goals (SDGs) (He et al. 2022).

The evolution of artificial intelligence (AI), particularly machine learning (ML), has facilitated the prediction of geological sequestration safety and the evaluation of carbon capture and storage (CCS) project efficacy (Yao et al. 2023). Researchers have also employed ML algorithms to optimize biochar synthesis, enhancing its carbon dioxide (CO_2) adsorption capabilities (Yuan et al. 2024). Natural language processing (NLP), a subset of AI, offers new insights and methodologies to biochar research. For instance, integrating language models with ML algorithms can provide precise experimental planning guidance for biochar studies (Yang et al. 2025). As AI technologies progress, there has been a noticeable surge in AI-associated research on carbon neutrality, illustrated by the upward trend in relevant studies (Fig. 1a). Despite the increasing interest in the synergistic application of biochar for carbon sequestration, the exploration of integrating NLP and ML within this field remains limited.

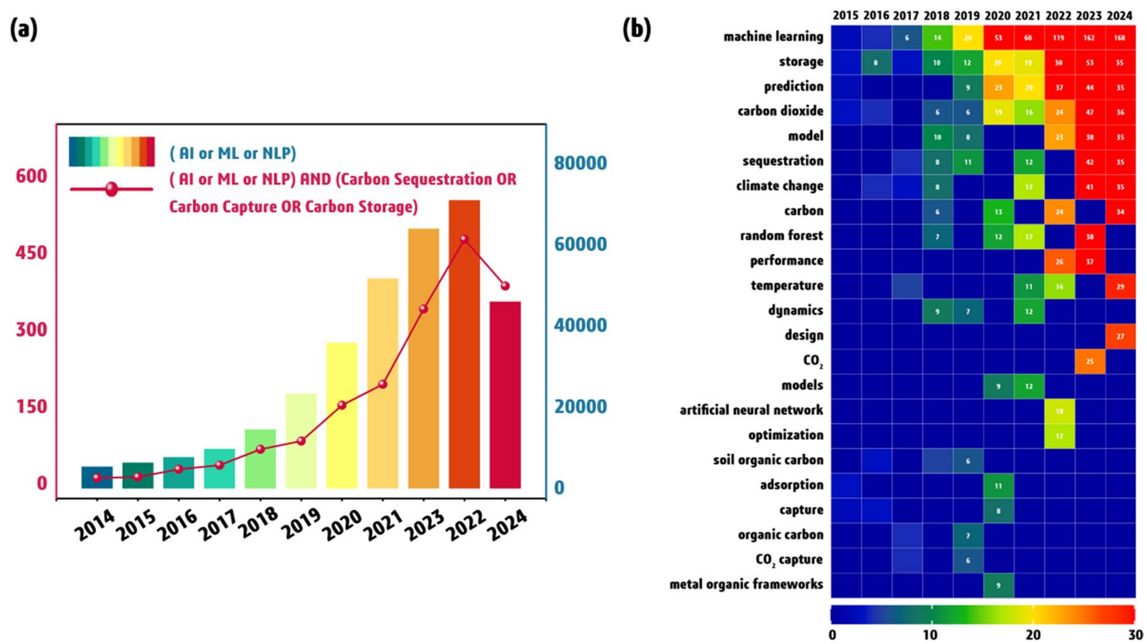


Fig. 1 **a** The growth in publication counts for the past decade (2014–2024) (Web of Science). Boolean search string: (TS = (AI) OR TS = (Machine Learning) OR TS = (Natural Language Processing)) AND (TS = (Carbon Sequestration) OR TS = (Carbon Capture) OR TS = (Carbon Storage)). **b** Temporal trend of emerging keywords in the past decade. Data analysis was based on annual frequency. Keywords are ordered by the cumulative frequency, and any frequencies above five are labeled

The full potential of utilizing NLP and ML to optimize biochar production processes and further enhance its carbon sequestration capabilities has yet to be fully harnessed and realized.

In this study, a bibliometric analysis on "AI and carbon sequestration" was conducted in the "Scopus core collection". The keywords in 1782 research articles and reviews from the Web of Science™ Core Collection, spanning the last decade (2014–2024), have been meticulously examined. Keyword analysis revealed "machine learning", "storage", and "prediction" as dominant subdomains (Fig. 1b), with "biochar" emerging as the top-trending keyword in the past year (Fig. 2a). This surge is likely driven by the transformative potential of biochar in waste management and carbon sequestration. Furthermore, a detailed evaluation of the interconnections among 19 high-frequency keywords related to biochar over the past decade (Fig. 2b) highlights its central role in advancing carbon-neutral technologies. The co-occurrence results indicate a high frequency of mention between biochar and several key terms, including "random forest" (a category of machine learning algorithms), "climate change", and "carbon sequestration". Notably, the strong correlation between biochar and "random forest" suggests a preference for ensemble-based ML models in optimizing biochar properties. This finding aligns

(See figure on next page.)

Fig. 2 **a** The research interest in keywords within the field of AI and carbon sequestration. Boolean search string: (TS = (AI) OR TS = (Machine Learning) OR TS = (Natural Language Processing)) AND (TS = (Carbon Sequestration) OR TS = (Carbon Capture) OR TS = (Carbon Storage)) from 2014 to 2024 (n = 1782) is visualized. Both the X and Y axes utilize a logarithmic scale. To provide a clear view of keyword trends, the dashed diagonal line in the figure serves as a boundary for changes in keyword trends. Keywords positioned above the diagonal line indicate an upward trend in research interest in the current period (2014–2022), while those below the line suggest a declining trend. When a keyword is above the diagonal line, the higher its coordinate value and the closer it is to the Y-axis, the more frequently it appears in the current period (2023–2024), and the more attention it has received compared to the previous period (2014–2022), with its color approaching yellow. When a keyword has large values for both coordinates and is above the diagonal line, it indicates that these keywords have consistently had a high frequency of occurrence over the past 10 years, and their popularity is still rising in the current period (2023–2024). **b** The 19 keywords most frequently co-occur with biochar emerging in recent ten years. The keywords (nodes) are ordered by their overall frequencies from steel blue to yellow color. Edge width and color are used to represent the co-occurrence between keywords; a thicker edge with a darker color means that the two keywords have a higher co-occurrence

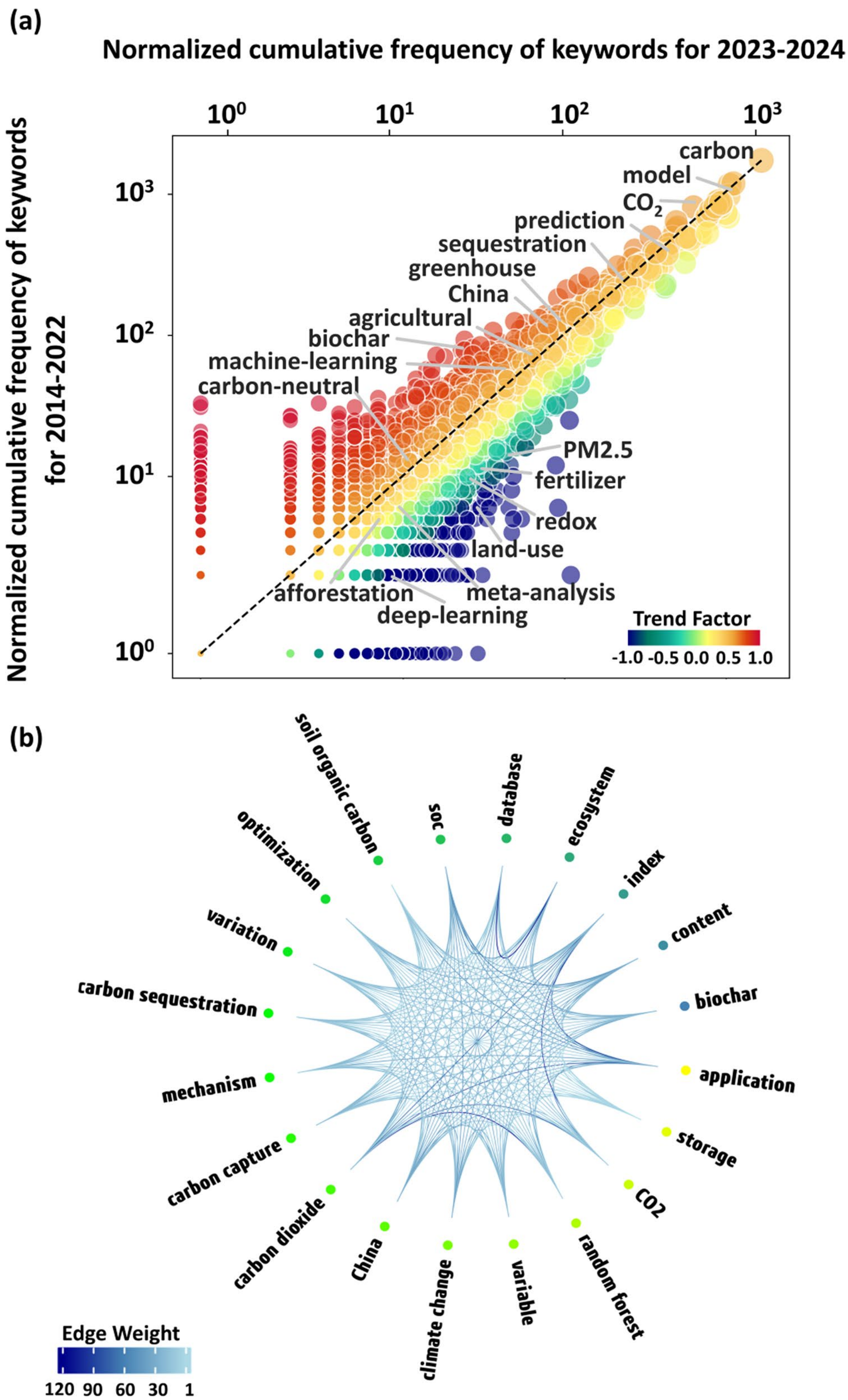


Fig. 2 (See legend on previous page.)

with the recent studies emphasizing the effectiveness of ensemble models in optimizing biochar properties for carbon sequestration (Yuan et al. 2021; Yuan et al. 2024). The bibliometric analysis of the past decade has underscored the increasing significance of AI, especially in the form of ML, in enhancing biochar research by optimizing its production and application, as well as in predicting and synthesizing knowledge within the field, thereby positioning biochar as a multifaceted tool in the fight against climate change. This paper presents a novel approach integrating ML and NLP to enhance the carbon sequestration potential of biochar. ML optimizes production and CO_2 adsorption, while NLP accelerates data analysis, boosting research efficiency. Combining these technologies opens new pathways for optimizing the role of biochar in climate change mitigation and contributing to global carbon neutrality efforts.

2 Biochar and carbon sequestration

Carbon capture, utilization, and storage (CCUS) has emerged as a pivotal strategy in the quest for carbon neutrality (Wei et al. 2021). In this context, biochar plays a vital role due to its aromatic carbon structure, which ensures its stability in soil. Research shows that biochar not only enhances soil carbon content but also reduces GHG emissions and improves soil quality (Lehmann et al. 2021; Bolan et al. 2022). Its unique porous structure enables the adsorption of atmospheric CO_2 within soil pores, positioning biochar as a promising agent in CCUS efforts (Cao et al. 2022; Dissanayake et al. 2020). Furthermore, integrating biochar into agricultural practices significantly improves soil structure, enhances cation exchange capacity, and increases the soil's ability to retain water and nutrients (Chen et al. 2022). These, in turn, promote crop growth and boost plants' CO_2 assimilation capacity, contributing to a more sustainable agricultural ecosystem (Fahad et al. 2016).

Biochar application has the potential to mitigate up to 0.92 Gt CO_2 yr^{-1} under sustainable scenarios (Deng et al. 2024). Additionally, it has been shown to influence the GHG flux within the soil environment by altering microbial activities, resulting in annual mitigated emissions of 2.97×10^7 t CO_2e for CO_2, 1.53×10^8 t CO_2e for CH_4, and 1.27×10^8 t CO_2e for N_2O (Wang et al. 2023). By linking microbial and mineral carbon processes, biochar carbon pump (BCP) promotes long-term carbon sequestration and supports climate change mitigation (Chen et al. 2024). In this paper, we present an overview of a novel approach that leverages ML techniques to optimize the carbon sequestration efficiency of biochar. This approach applies ML models to refine biochar synthesis processes, potentially enhancing its CO_2 adsorption capacity and predicting its long-term impact on GHG emissions. The integration of ML in biochar research offers the possibility of accelerating the development of high-performance biochar and providing precise, data-driven guidance for biochar production and application, contributing to global efforts toward carbon neutrality.

3 ML for enhancing biochar synthesis and carbon sequestration efficiency

Recent research advancements have shown that ML techniques play a significant role in optimizing the synthesis of biochar, enhancing its CO_2 adsorption performance and predicting its carbon sequestration potential. For instance, Yuan et al. (2021) utilized various boosting algorithms, including Gradient Boosting Decision Trees (GBDTs), Light Gradient Boosting machines (LGBs), and Extreme Gradient Boost (XGB). They combined these algorithms with feature importance analysis techniques such as permutation importance and mean decrease in accuracy to assess the significance of each input feature. By employing partial dependence plots, the authors demonstrated the impact of individual features on the target variable, thereby providing valuable insights into both the CO_2 adsorption process and the optimization of biomass waste-derived porous carbon synthesis. Building on this work, Yuan et al. (2024) developed an advanced ML-driven framework to extract key insights from experimental data and fine-tune biochar synthesis parameters, significantly improving its CO_2 adsorption capacity. By leveraging the predictive capabilities of Random Forest (RF) and Gaussian Process (GP) regression, alongside optimization techniques like Particle Swarm Optimization (PSO) and Shapley Additive exPlanations (SHAP) value analysis, they successfully predicted the microporous volume ($V_{0.8}$) of biochar. Through three active learning cycles, they effectively doubled their CO_2 adsorption capacity, accelerating the development of biochar for this application. In addition to these studies, Tee et al. (2022) employed an Artificial Neural Network (ANN) model to predict the yield and Brunauer-Emmett-Teller (BET) surface area of biochar based on pyrolysis conditions and biomass composition. The study demonstrated that the precision of the ANN model in predicting carbon sequestration potential of biochar offers a valuable predictive tool for screening biomass feedstocks and assessing carbon sequestration potential. This is especially critical because the surface area of biochar directly affects its CO_2 adsorption performance, while its yield determines the feasibility and efficiency under different production conditions. Furthermore, Ding et al. (2018) used Boosted Regression Trees (BRTs) to evaluate biochar's effects on soil organic carbon (SOC) decomposition, identifying incubation time as the key driver. Their findings highlighted critical factors influencing biochar's impact on native SOC cycling, providing insights for better predicting soil carbon sequestration following biochar amendment. In summary, ML profoundly enhances the research and development of biochar technology by accurately predicting key properties, refining synthesis methodologies, and assessing environmental impacts. This not only augments the efficiency of biochar innovation but also provides robust technical support for the achievement of carbon neutrality objectives.

4 Integrating ML and NLP for cross-disciplinary advancements in biochar research

The integration of ML and NLP can play a pivotal role in facilitating interdisciplinary research and fostering collaborative efforts across multiple disciplines. For instance, this synergy enhances knowledge extraction and literature analysis in the field of Earth sciences (Lin et al. 2024), expediting the discovery of novel materials and enabling accurate predictions of chemical reactions in materials science (Maik Jablonka et al. 2023). In ecology, such integration advances our comprehension of species' roles within ecosystems and their ecological contributions, offering valuable insights for conservation strategies and climate change adaptation (Domazetoski et al. 2024; Wang et al. 2024). Against this backdrop, Qiao et al. (2024) revealed the research trends and focal points in the field of microbial carbon capture spanning the past three decades. Their bibliometric analysis highlighted the emergence of biochar as a promising carbon sequestration medium, with its research prominence experiencing a sustained and escalating interest. In another study, Paula et al. (2022) demonstrated the great potential of NLP in systematically analyzing large amounts of scientific literature. The study processed 10,975 pieces of relevant literature through automatic reading-interpreting-extracting computational routines (namely, the a.RIX engine) and identified the most commonly used biomass precursors and optimal production parameters through ML algorithms. The researchers used cluster and network analysis to extract key parameters related to the synthesis and post-processing of carbon functional materials (CFM). This data-driven approach demonstrates the powerful promise of ML and NLP in experimental design and materials optimization, significantly accelerating the process of extracting valuable information from massive data and driving the efficient development of CFM materials. The latest research has shown that integrating ChatGPT with ML-based Bayesian algorithms enables researchers to meticulously design experiments and provide tailored solutions for specific soil conditions, enhancing the precision and efficiency of biochar synthesis (Yang et al. 2025). By embracing the sophisticated integration of NLP and ML, researchers can fine-tune biochar preparation processes, achieving targeted soil amelioration with enhanced precision. In summary, the application of NLP technology in carbon sequestration has greatly improved the ability to extract valuable insights from scientific literature. Combined with ML algorithms, it offers intelligent guidance for biochar synthesis and carbon sequestration, advancing the development of carbon capture methodologies and providing a more effective approach to addressing climate change.

5 Conclusions and perspectives

Biochar has emerged as a promising agent in the realms of carbon capture, utilization, and storage (CCUS), demonstrating significant potential. The integration of ML has advanced biochar research by optimizing synthesis processes, enhancing CO_2 adsorption, and predicting carbon sequestration potential. Additionally, NLP enables intelligent analysis of scientific literature, revealing the complex relationships between the synthesis, properties, and applications of biochar. The combination of ML and NLP introduces a new approach to biochar research, offering innovative tools and perspectives for advancing carbon sequestration technologies. Future research should focus on the following key areas:

- Model selection and optimization: Selecting appropriate ML algorithms, such as Random Forest (RF), Gaussian Process (GP) regression, and Artificial Neural Networks (ANN), is essential for specific applications. Performance can be further enhanced through parameter tuning and algorithm adjustments to improve the efficiency and accuracy of biochar optimization.
- Multi-model integration: Combining ML models with ensemble learning techniques, such as Boosting or Stacking, can improve the accuracy and robustness of predictions, providing precise guidance for biochar performance under varying production conditions. Future research should develop hybrid ML models integrating supervised and unsupervised learning for comprehensive biochar optimization.
- Interactive platform development: Real-time interactive platforms can dynamically adjust experimental data and model predictions, offering more precise control over biochar synthesis processes and enabling personalized optimization.
- Biochar supply chain optimization: ML can optimize the entire biochar supply chain, from production to application. Additionally, NLP can evaluate the economic and environmental impacts from existing literature, offering data-driven support for industrial decision-making.

- Biochar policy and regulatory framework: NLP can analyze global policies and regulations related to biochar and integrate this information into ML models. This will help predict biochar adoption and carbon sequestration potential under different policy scenarios, facilitating the formulation of effective policies and broader biochar adoption.
- Industrial-scale application of AI-driven biochar optimization: As ML technologies continue to advance, future research should prioritize the scaling of AI-driven optimization methods for industrial applications in biochar production. The development of scalable ML frameworks, combined with comprehensive big data analytics, will effectively streamline and enhance the efficiency of biochar production and application processes. By conducting thorough life cycle assessments and economic analyses on a larger scale, the environmental and economic benefits of these methods can be better assessed, thereby aligning with global carbon neutrality goals.

Acknowledgements
The authors thank the valuable comments of anonymous reviewers and editor.

Author contributions
Jiayi Li: Conceptualization, Data curation, Investigation, Visualization, Writing-original draft. Yixuan Chen: Data curation, Software. Chaojie Wang: Data curation, Software. Hanbo Chen: Data curation, Writing-review & editing. Yurong Gao: Data curation, Software. Jun Meng: Writing-review & editing. Zhongyuan Han: Writing-review & editing. Lukas Van Zwieten: Writing-review & editing. Yi He: Writing-review & editing, Supervision. Caibin Li: Writing-review & editing. Gerard Cornelissen: Writing-review & editing. Hailong Wang: Conceptualization, Supervision, Funding acquisition. The authors read and approved the final manuscript.

Funding
This study was funded by the the National Key Research and Development Program of China (2023YFD1700800), the National Natural Science Foundation of China (22276031, 42107031), the Key Research Platform and Project for Higher Education Institutions in Guangdong Province in 2022 (2022KCXTD049), China National Tobacco Corporation Guizhou Provincial Company Science and Technology Program (2022XM11), and the Guangdong Foundation for Program of Science and Technology Research, China (2023B1212060044).

Availability of data and materials
The authors confirm that all relevant data are included in the article.

Declarations

Competing interests
Wang Hailong and Meng Jun are Executive Editors of the journal Biochar, and they were not involved in the peer-review or handling of the manuscript. Lukas Van Zwieten is an EBM of the journal Biochar, and he was not involved in the peer-review or handling of the manuscript. The authors have no other competing interests to disclose.

Author details
[1]Agronomy College, Shenyang Agricultural University, Shenyang 110866, China. [2]School of Environment and Chemical Engineering, Foshan University, Foshan 528000, China. [3]State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-Products, Key Laboratory of Biotechnology in Plant Protection of Ministry of Agriculture and Zhejiang Province, Institute of Plant Virology, Ningbo University, Ningbo 315211, China. [4]Key Laboratory of Recycling and Eco-Treatment of Waste Biomass of Zhejiang Province, School of Environment and Natural Resources, Zhejiang University of Science & Technology, Hangzhou 310023, China. [5]School of Environmental Science and Engineering, Guangzhou University, Guangzhou 510006, China. [6]School of Electronic Information Engineering, Foshan University, Foshan 528000, China. [7]NSW Department of Primary Industries, Wollongbar Primary Industries Institute, 1243 Bruxner Highway, Wollongbar, NSW 2477, Australia. [8]Yancao Production Technology Center, Bijie Yancao Company of Guizhou Province, Bijie 551700, China. [9]Norwegian Geotechnical Institute (NGI), 0484 Oslo, Norway. [10]Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences (NMBU), 1430 Ås, Norway. [11]Guangdong Provincial Key Laboratory of Integrated Agro-Environmental Pollution Control and Management, Institute of Eco-Environmental and Soil Sciences, Guangdong Academy of Sciences, Guangzhou 510650, China.

Received: 2 September 2024 Revised: 4 December 2024 Accepted: 24 December 2024
Published online: 20 January 2025

References
Bolan N et al (2022) Multifunctional applications of biochar beyond carbon storage. Int Mater Rev 67:150–200. https://doi.org/10.1080/09506608.2021.1922047
Cao L et al (2022) Straw and wood based biochar for CO_2 capture: adsorption performance and governing mechanisms. Sep Purif Technol 287:120592. https://doi.org/10.1016/j.seppur.2022.120592
Chen H et al (2022) Engineered biochar for environmental decontamination in aquatic and soil systems: a review. Carbon Res 1:1–25. https://doi.org/10.1007/s44246-022-00005-5
Chen Y et al (2024) Biochar as a green solution to drive the soil carbon pump. Carbon Res 3:44. https://doi.org/10.1007/s44246-024-00132-1
Deng X et al (2024) Exploring negative emission potential of biochar to achieve carbon neutrality goal in China. Nat Commun 15:1085. https://doi.org/10.1038/s41467-024-45314-y
Ding F et al (2018) A meta-analysis and critical evaluation of influencing factors on soil carbon priming following biochar amendment. J Soils Sediments 18:1507–1517. https://doi.org/10.1007/s11368-017-1899-6
Dissanayake PD et al (2020) Sustainable gasification biochar as a high efficiency adsorbent for CO_2 capture: a facile method to designer biochar fabrication. Renew Sustain Energy Rev 124:109785. https://doi.org/10.1016/j.rser.2020.109785
Domazetoski, V. 2024. Enhancing Ecological Knowledge Discovery Using Large Language Models. Master's Thesis, Georg-August-Universität Göttingen.
Fahad S et al (2016) A combined application of biochar and phosphorus alleviates heat-induced adversities on physiological, agronomical and quality attributes of rice. Plant Physiol Biochem 103:191–198. https://doi.org/10.1016/j.plaphy.2016.03.001
He M et al (2022) Waste-derived biochar for water pollution control and sustainable development. Nat Rev Earth Environ 3:444–460. https://doi.org/10.1038/s43017-022-00306-8
Lehmann J et al (2021) Biochar in climate change mitigation. Nat Geosci 14:883–892. https://doi.org/10.1038/s41561-021-00852-8
Lin Z et al (2024) GeoGalactica: a scientific large language model in geoscience. arXiv Preprint. https://doi.org/10.4855/arXiv.2401.00434
Liu Z et al (2022) Challenges and opportunities for carbon neutrality in China. Nat Rev Earth Environ 3:141–155. https://doi.org/10.1038/s43017-021-00244-x
Maik Jablonka K et al (2023) 14 examples of how LLMs can transform materials science and chemistry: a reflection on a large language model hackathon. Digit Discov 2:1233–1250. https://doi.org/10.1039/D3DD00113J
Paula AJ et al (2022) Machine learning and natural language processing enable a data-oriented experimental design approach for producing biochar and hydrochar from biomass. Chem Mater 34:979–990. https://doi.org/10.1021/acs.chemmater.1c02961

Qiao L et al (2024) Microbial carbon capture - evolving trends, interconnections, and recent spotlights of the past three decades. Chem Eng J 482:148970. https://doi.org/10.1016/j.cej.2024.148970
Tee JX et al (2022) Prediction of carbon sequestration of biochar produced from biomass pyrolysis by artificial neural network. J Environ Chem Eng 10:107640. https://doi.org/10.1016/j.jece.2022.107640
Wang L et al (2023) Role of biochar toward carbon neutrality. Carbon Res 2:2. https://doi.org/10.1007/s44246-023-00035-7
Wang C et al (2024) Interpretable machine learning for predicting heavy metal removal and optimizing biochar characteristics. J Water Process Eng 68:106484. https://doi.org/10.1016/j.jwpe.2024.106484
Wei YM et al (2021) A proposed global layout of carbon capture and storage in line with a 2℃ climate target. Nat Clim Change 11:112–118. https://doi.org/10.1038/s41558-020-00960-0
Xia L et al (2023) Integrated biochar solutions can achieve carbon-neutral staple crop production. Nature Food 4(3):236–246. https://doi.org/10.1038/s43016-023-00694-0
Yang H et al (2025) An intelligent approach: integrating ChatGPT for experiment planning in biochar immobilization of soil cadmium. Sep Purif Technol 352:128170. https://doi.org/10.1016/j.seppur.2024.128170
Yao P et al (2023) Application of machine learning in carbon capture and storage: an in-depth insight from the perspective of geoscience. Fuel 333:126296. https://doi.org/10.1016/j.fuel.2022.126296
Yuan X et al (2021) Applied machine learning for prediction of CO_2 adsorption on biomass waste-derived porous carbons. Environ Sci Technol 55:11925–11936. https://doi.org/10.1021/acs.est.1c01849
Yuan X et al (2024) Active learning-based guided synthesis of engineered biochar for CO_2 capture. Environ Sci Technol 58:6628–6636. https://doi.org/10.1021/acs.est.3c10922