Gul et al. Biochar (2026) 8:8
https://doi.org/10.1007/s42773-025-00511-w

Biochar

ORIGINAL RESEARCH

Open Access

Machine learning optimization for algal biochar yield: integrating experimental validation and sensitivity analysis

Jawad Gul[1], Muhammad Nouman Aslam Khan[1*], Umair Sikander[1], Asif Hussain Khoja[2], Melanie Kah[3] and Salman Raza Naqvi[4]

Abstract

Pyrolysis requires extensive experimentation to achieve optimum thermochemical conversion, which can be addressed by integrating machine learning (ML) predictive solutions. The abundant availability of algae with low volume footprint makes it viable green biomass to achieve thermochemical products. For optimum algal biochar (BC) yield production, the relation of ultimate, proximate analysis with process conditions is critical. This study's objective is twofold: It aims to develop a robust ML model, trained on diverse literature data and optimized using particle swarm optimization and genetic algorithm, that predicts BC yield across various feedstocks and conditions. Secondly, the optimum process parameters are derived to maximize BC yield with the experimental validation for the collected samples at their respective chemical and structural compositions. Limited data points for algal biomass induce a comparative analysis of ML models, including Gaussian process regression, ensembled tree (ET), decision tree and support vector machine. The predictive capability of ET enhanced through optimization performed exceptionally well for BC yield prediction with testing $R^2 = 0.77993$ and RMSE = 6.9792. 2D and 3D partial dependence plots imply that BC yield is primarily influenced by pyrolysis temperature, volatile matter, and heating rate with SHAP values of 1.2785, 0.3972, and 0.2949, respectively. Monte Carlo simulation and Sobol sensitivity analysis substantiate statistically the impact of selected features on algal BC yield. Inverse optimization of ET model suggests that the maximum BC yield production is 76.33% at a temperature of 500 °C, a heating rate of 10 °C/min, a residence time of 60 min, a N_2 flow rate of 0.5 L/min, and particle size of 1.5mm.

Highlights

- Developed the optimized ML model for algal biochar yield prediction on robust dataset
- Enhanced biochar production through experimental validated ML model
- Performed uncertainty and Sobol sensitivity analysis to quantify the impact of features on biochar yield

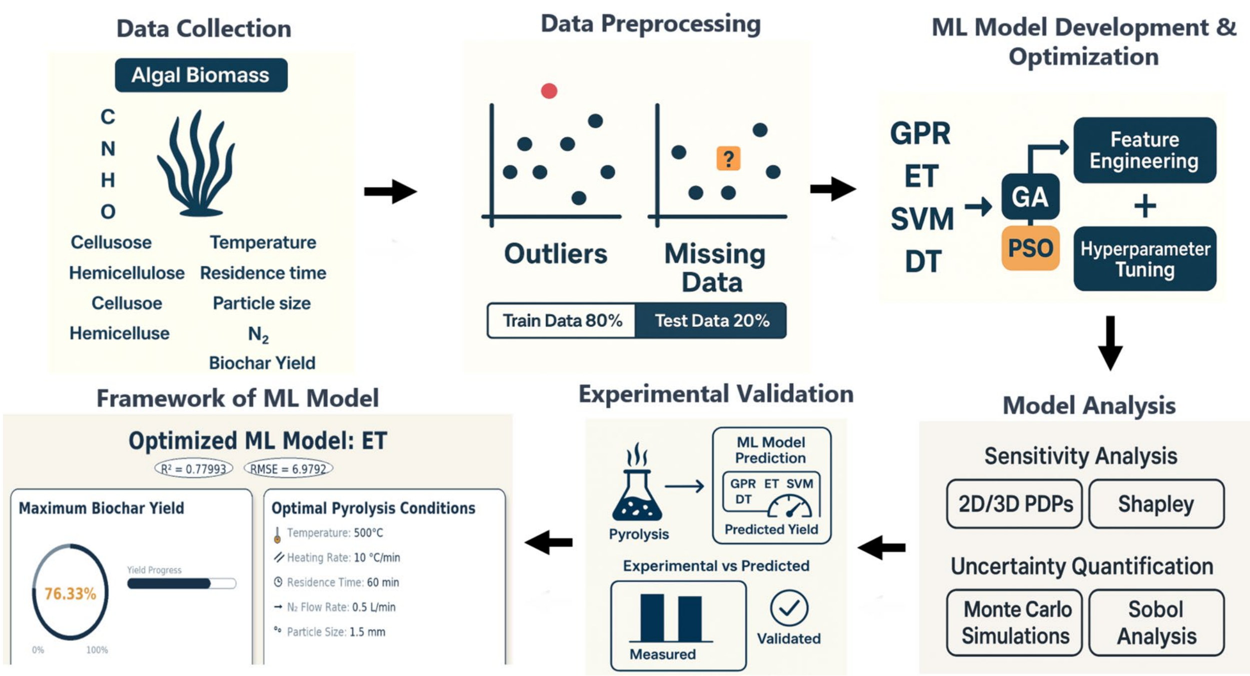
Keywords Artificial intelligence, Pyrolysis, Algae bi0omass, Optimization, Sensitivity analysis, Biochar

*Correspondence:
Muhammad Nouman Aslam Khan
mnouman@scme.nust.edu.pk
Full list of author information is available at the end of the article

Springer

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

Graphical Abstract



1 Introduction

Biochar, a carbon-rich material derived from the pyrolysis of biomass, holds considerable potential for carbon sequestration (Zhang et al. 2022), soil enhancement (Adhikari et al. 2022), and renewable energy production (Hoang et al. 2022) with various industrial and commercial applications. Traditionally, lignocellulosic biomass has been the predominant feedstock for biochar production, whereas algae represent an emerging, abundant marine resource with minimal land footprint (Sun et al. 2022). One of the significant challenges associated with producing BC from algal biomass is the substantial quantity of raw material needed to obtain a desirable yield. Due to the high moisture content and low density of algae, the conversion process results in a relatively low solid yield compared to the initial biomass input. As a result, achieving an optimal yield of BC is constrained by the requirement for large volumes of algae, which can be resource-intensive and limit the overall efficiency and scalability of the process. Furthermore, achieving optimal BC yields from algae poses a complex challenge due to the intricate interplay of processing conditions, the composition of algae biomass, and the desired BC properties. Traditional experimental techniques to unravel these relationships are often resource-intensive, time-consuming, and financially burdensome. Consequently, there is a compelling need to develop advanced and economically viable models capable of accurately predicting BC yields based on multiple variables. Data collected from the literature and shown in Table 1 provide an overview of different algae biomasses used for BC production and the equipment and temperatures employed. This specific data explicitly elaborates the range of BC yields reported across these conditions, offering insights into how biomass type and processing parameters influence yield outcomes.

Despite numerous experimental investigations into the factors affecting BC yield, it remains very challenging to identify the optimal set of production parameters that promote high yield. Since experiments are time-consuming and expensive, they can only assess the influence of a few parameters at a time. Integrating all available data enables a more holistic assessment of the influencing factors and provides a basis for developing prediction tools that consider both the feedstock and production factors. Therefore, establishing the most favourable and optimal parameters for algae-based BC production remains a significant challenge (Khan et al. 2022a). An adequate examination of algae-based BC yield can be achieved through the utilization of advanced tools of data modelling, machine learning, artificial intelligence, and big data, enabling a comprehensive analysis of algae pyrolysis

Table 1 A review of laboratory research focusing biochar yields

Algal feedstock	Type of apparatus	Type of pyrolysis	Temperature (°C)	Yield (%)	Reference
Macroalgae	Microwave reactor	Slow	240–400	54.8	Shuttleworth et al. (2012)
Brown *Laminaria japonica*	Horizontal electric furnace	Slow	200	78.34	Jung et al. (2016)
			400	63.64	
			600	37.96	
			800	27.05	
Algal *chlorella* residue	Thermogravimetric system	Slow	300	56.3	Chang et al. (2015)
			500	66.2	
			700	65	
Mix macroalgae species (*Fucus serratus, Laminaria digitata*)	Fluidized bed	Slow	500	29–36	Yanik et al. (2013)
Chlorella vulgaris	Fluidized bed	Fast	500	31	Wang et al. (2013)
Seawood	Fixed bed	Slow	250	93.95	Tag et al. (2016)
			600	61.5	
Cladophora glomerata	Fixed bed	Slow	400	44	Tag et al. (2016)
			500	40	
			600	39	
Chlamydomonas reinhardtii	Fixed bed	Slow	350	44	Yuan et al. (2015)
Chlorella vulgaris	Fixed bed	Slow	300	19.3	Yuan et al. (2015)
			900	43.46	

alongside various pyrolysis conditions and the properties of the feed (Khan et al. 2022a).

The application of machine learning (ML) is instrumental in predicting and validating the optimal yield of BC from algae (Khan et al. 2022b). ML algorithms comprehensively predict and evaluate relationships among input parameters and output variables specific to algae-based BC production (Pathy et al. 2020). For instance, the development of the regression-based Support Vector Machine (SVM) model ($R^2 = 0.96$) enabled the analysis of BC yield (Cao et al. 2016) but was based on only a dataset of 33 points. Previous studies (Ullah et al. 2021) developed an optimized ML model for general biomass for BC production using the metaheuristic algorithms without cross-validation, resulting in model overfitting. Random Forest, employed on just 150 data points, reported a test R^2 of 0.97 (Leng et al. 2024) while Gradient Boosted Decision Trees, trained on a dataset expanded from 93 to 341 points, achieved an R^2 regarding 0.998 (Zhou et al. 2024). These results, while seemingly strong, raise concerns of overfitting due to limited data. Small datasets restrict the robustness and generalizability of ML models.

A comprehensive dataset for algae-based BC is essential, along with an experimentally validated, optimized ML model that explicitly captures algae's complex structural and chemical composition. Algae's potential as a biomass source for BC is promising, but its chemical variability complicates consistent yield prediction (Nguyen et al. 2024). To address this challenge, the ML model was developed on a robust dataset to accurately predict BC yield based on the structural and chemical composition of several algal strains. This approach reduces the need for extensive experimental trials, making biomass conversion more efficient and thus economical. First, key pyrolysis reaction conditions—temperature, residence time, nitrogen flow, particle size, and heating rate—along with biomass properties such as structural analysis, proximate, and ultimate composition, were consolidated as input features. The relationship between pyrolysis reaction conditions and biomass properties was investigated with BC yield using different ML algorithms, and model performance was improved through Genetic Algorithm (GA) and Particle Swarm Optimization (PSO).

This study integrates machine learning with advanced optimization techniques to establish a robust predictive framework. ML algorithms—Decision Trees (DT), Ensemble Trees (ET), Gaussian Process Regression (GPR), and SVM—were evaluated to determine the most effective model for BC yield prediction. To enhance model performance, advanced optimization techniques were employed for hyperparameter tuning and feature selection. The testing and training datasets were appropriately segregated to ensure that different ranges of BC yield (17–79%) are well represented in both sets, which is essential for developing an effective ML model. Using fivefold cross-validation ensures that the model is optimally tuned to perform well on unseen data. The ET model demonstrated superior predictive capability with a testing R^2 value of 0.7799 and a RMSE value of 6.9792. This model identified key process parameters influencing BC yield, including temperature, volatile matter and heating rate, with SHAP values of 1.2785, 0.3972 and 0.2949, respectively. The predicted algal yield through the ET model was experimentally validated with three different algal biomass samples. Using inverse optimization, the maximum BC yield was reported as 76% with a temperature of 500 °C, a heating rate of 10 °C/min, a resident time of 60 min, an N_2 flow rate of 0.5 L/min, and a particle size of 1.5 mm. An error of 5% was recorded in experimental validation with the experimental algal yield of 78%. The proposed framework not only refines yield prediction but also provides an objective understanding of parameters governing algal BC production, surpassing conventional experimental methodologies. This approach provides a sustainable and cost-effective solution for BC production, leveraging ML and optimization techniques to overcome the challenges of chemical variability in algae biomass. The advanced AI-based ET architecture exhibits robust predictive performance, establishing it as a reliable tool for guiding pre-experimental designs in algal-based studies. Figure 1 shows the successive stages of prediction and validation of the algal BC yield.

2 Methodology

2.1 Data sourcing and preprocessing

A comprehensive dataset was collected from existing literature using Boolean connectors (AND, OR) with keywords such as 'BC yield,' 'algae,' 'pyrolysis,' 'biomass conversion,' 'algal conversion,' 'proximate analysis,' and 'ultimate analysis.' Searches were conducted in Web of Science, ScienceDirect, and Google Scholar for studies published between 2015 and 2025. Inclusion criteria focused on studies providing detailed experimental data on BC yield from algae pyrolysis. Studies with incomplete data, non-peer-reviewed sources, or lacking relevant parameters were excluded. A total of 48 research papers meeting these quality criteria were selected, contributing 373 data points specifically for algal biomass for ML model development and assessment. Outliers in the data were identified and rectified using smoothing techniques, ensuring the integrity and accuracy of the dataset (Khan et al. 2022b). The updated dataset was partitioned randomly, allocating 80% (298 data points) for training and 20% (75 data points) for testing. ML was trained to predict the BC yield with tunned k-fold cross-validation. Input parameters include Carbon (C%), Nitrogen (N%), Hydrogen (H%), Oxygen (O%), Fixed carbon (FC%), Volatile matter (VM%), Moisture Content (MC%), Ash Content (Ash%), Lignin (%), Cellulose (%), Hemicellulose (%), Pyrolysis temperature (°C), Heating Rate (°C/min),

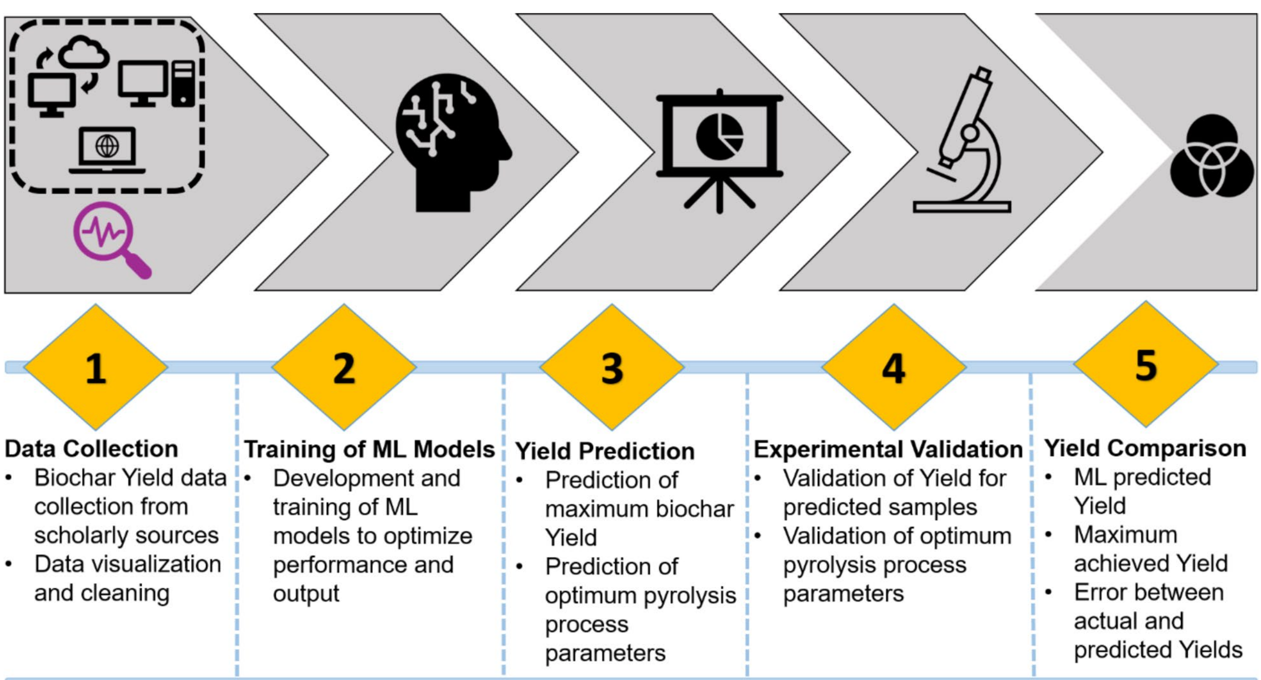


Fig. 1 Research scheme of successive steps for algal biochar yield prediction

Residence time (min), Particle size (mm), and Nitrogen (N_2) flowrate (L/min).

The violin plots combined with box plots effectively showcase the distribution patterns and central tendencies of the features (Potter et al. 2006) affecting BC yield. The violin plots represent kernel density estimations, giving a clear view of skewness and kurtosis to understand asymmetry and tail behaviour, while the box plots inside indicate the median, interquartile range (Critchley and Jones 2008), and any outliers present as represented in Fig. 2. C showed a relatively symmetric distribution (44.85%), ensuring stable availability of energy content in BC. In contrast, N is right-skewed (1.51%), indicating limited incorporation. H (5.93%) and O (41.95%) influence VM and BC's combustion properties (Hadey et al. 2022), impacting energy release and reactivity. MC varies considerably with a right-skewed distribution (5.32%), affecting pyrolysis energy efficiency (Westerhof et al. 2007). Lignin is left-skewed (21.81%), contributing to aromaticity and long-term stability. Cellulose (34.55%) and hemicellulose (24.28%) show moderate distributions, influencing thermal degradation and volatile release. Among process parameters, pyrolysis temperature shows bimodal distribution (526.12 °C), reflecting varied thermal regimes impacting BC properties. The heating rate (20.79 °C/min) and residence time (46.66 min) are right-skewed, influencing reaction kinetics and pore structure. Particle size is symmetrically distributed (2.03 mm), affecting heat transfer and mass diffusion during pyrolysis. Analyzing skewness and kurtosis ensures reliable statistical analysis and predictive modelling.

Pearson's correlation with the dendrogram between the variables is represented in Fig. 3. The structural analysis (encompassing cellulose and hemicellulose contents), ultimate analysis (H, O, C and N as elemental composition of biomass), proximate analysis, and

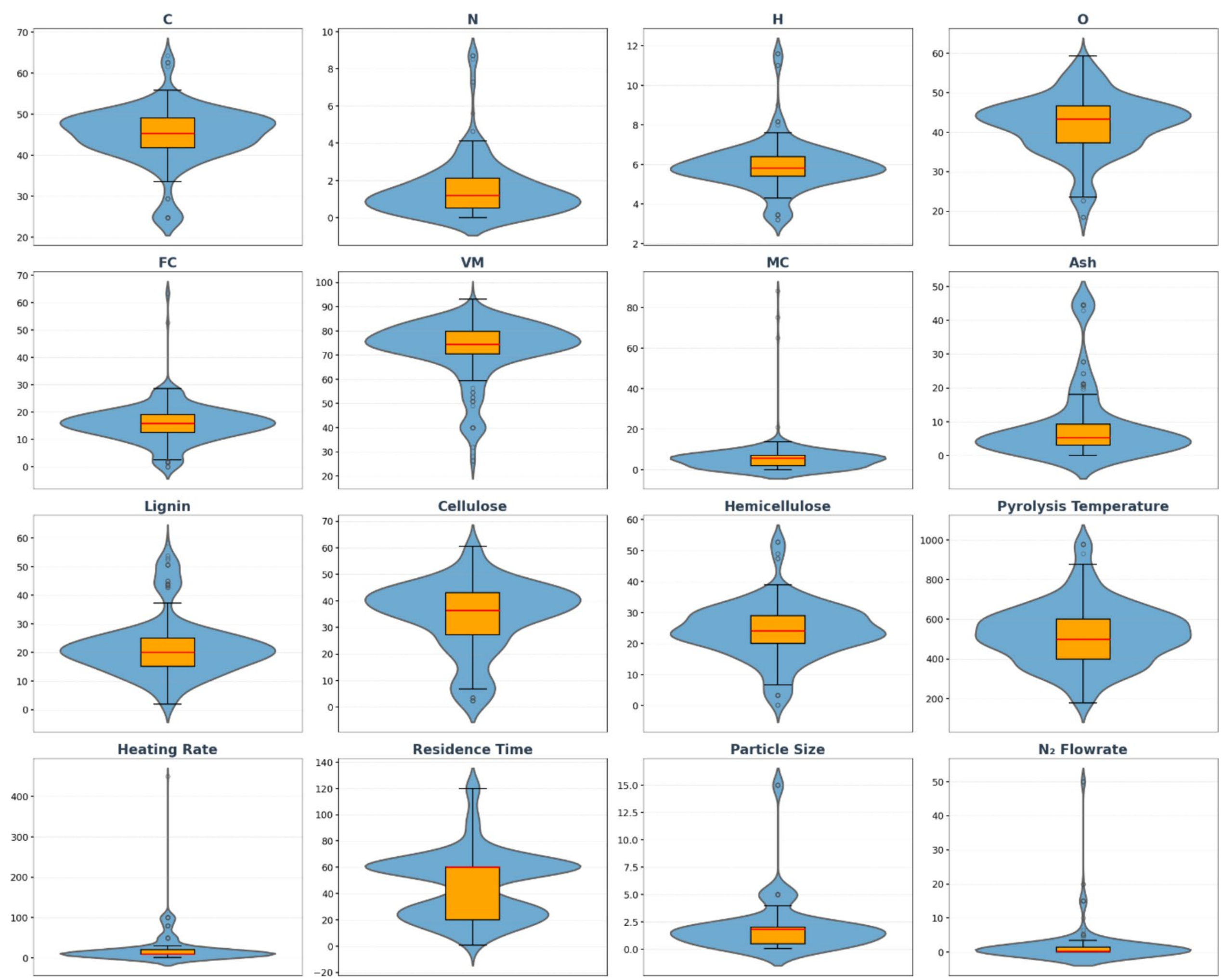


Fig. 2 The violin plots for data distribution of input parameters

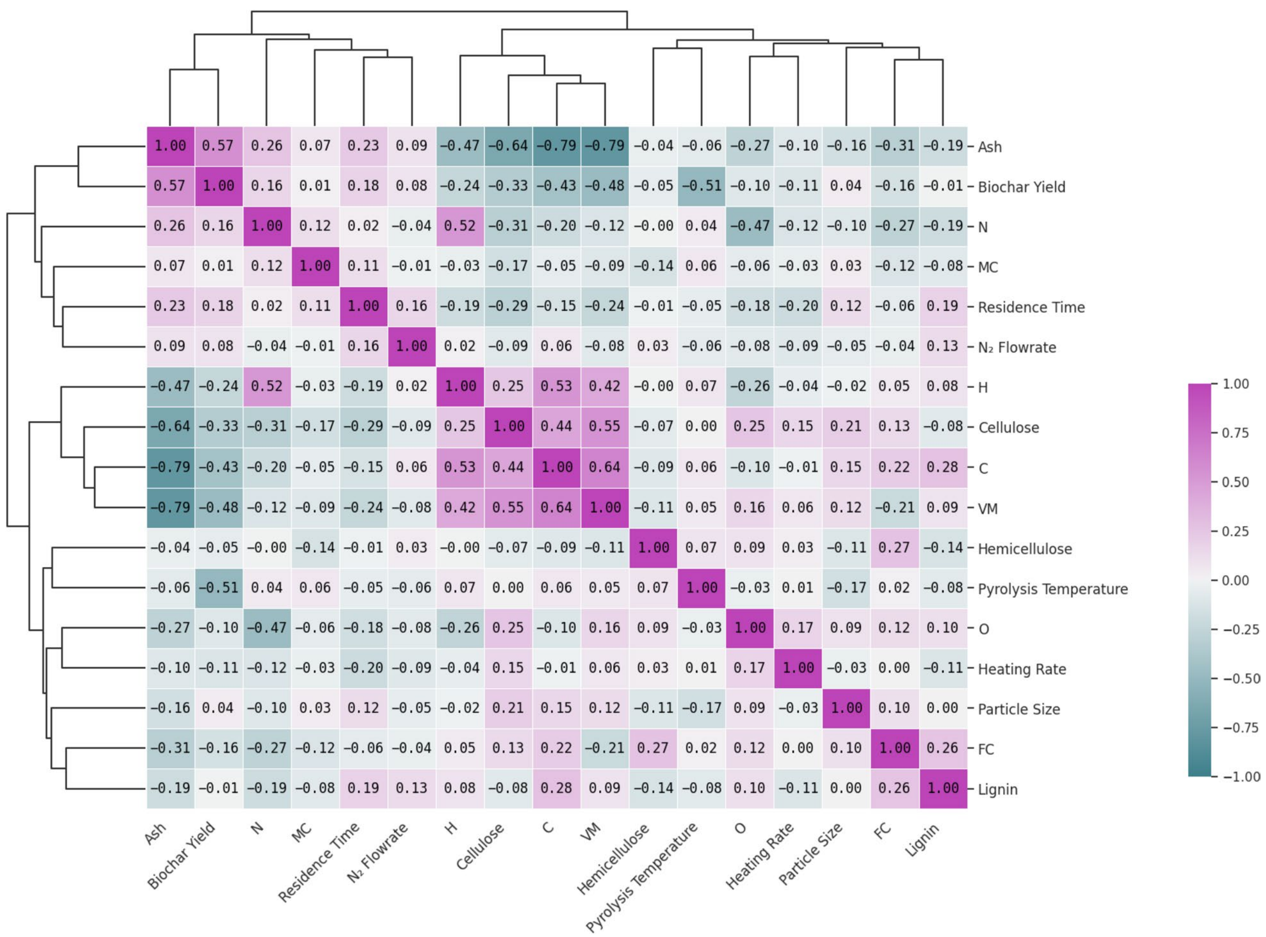


Fig. 3 Relationship of variables using Pearson's correlation coefficient

experimental conditions (residence time, particle size, temperature, heating rate, N_2 flowrate and temperature) are all assigned a PCC value. C and VM are moderately correlated ($r = 0.64$), indicating similar thermal degradation. Ash negatively correlates with VM ($r = -0.79$), reflecting its role in reducing volatiles. C directly impacts H ($r = 0.53$) and cellulose ($r = 0.44$) but indirectly BC ($r = -0.43$) and N ($r = -0.20$). The clustering groups cellulose, hemicellulose, and VM highlight their influence on energy content. Ash, MC, and N are clustered, indicating their impact on BC. The analysis makes the data set favourable for modelling a robust data set as reported in previous literature (Ullah et al. 2023).

2.2 Machine learning methods

Due to the challenges associated with the availability of experimental data, it is necessary to find the best performing ML model for the pyrolysis of algae. This leads to a comparative analysis of various ML models—ET, SVM, GPR, and DT, to develop a robust ML model for predicting BC yield derived from the pyrolysis of algae with experimental validation. These models were constructed based on a comprehensive analysis of proximate and ultimate analysis parameters, encompassing factors from algae composition to processing conditions. The training and testing of the ML model ensure reliable execution to develop patterns and trends in predicting BC yield from algae under various experimental conditions. Learning from the dataset would enable ML models to accurately predict BC yield for a range of algae strains with reduced experimentation. Statistical parameters, including the coefficient of correlation (R^2) and root mean square error (RMSE), were used to evaluate the performance of ML models, given in Eqs. 1 and 2 (Haq et al. 2022).

$$R^2 = \frac{\left[\sum_{i=1}^{n}\left(y_{p,i} - \bar{y}_p\right) \times \left(y_{a,i} - \bar{y}_a\right)\right]^2}{\sum_{i=1}^{n}\left(y_{p,i} - \bar{y}_p\right)^2 \times \sum_{i=1}^{n}\left(y_{a,i} - \bar{y}_a\right)^2} \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{p,i} - y_{a,i}\right)^2} \tag{2}$$

where $\bar{y}_a$ = observed output mean, $y_{a,i}$ = observed output, $\bar{y}_p$ = predicted output mean, $y_{p,i}$ = Predicted output, and n = total data points.

The SVM algorithm used in this study identifies complex patterns within the dataset to make precise algal BC yield predictions. Due to the varying resource and feedstock conditions for pyrolysis, SVM can accurately predict a wide range of production scenarios. Decision Tree plays a key role in deciphering the complexities of BC yield prediction from algae pyrolysis. DT handles complex datasets to achieve precise outcomes, and the generated tree structure reveals the influential factors affecting BC yield.

Through probabilistic modelling principles, GPR accurately predicts BC yield and estimates the uncertainty associated with those predictions. GPR's flexibility effectively manages the involvement of numerous variables, including algae composition, processing conditions, and environmental factors. ET combines multiple decision trees to generate a robust predictive model. DT makes the most accurate predictions by splitting the dataset into smaller parts and deciding based on the leading features within each subset. Ensemble methods improve predictive performance by combining the BC yield of multiple individual models. This method represents intricate relationships in the dataset while adapting to multiple factors affecting BC yield. ET's yield estimation for BC is more accurate and comprehensive due to the collective knowledge of multiple models.

2.3 Optimization techniques

To maximize ML models' predictive power and accuracy, feature selection and hyperparameter tuning through optimization is a key step. The detailed approach finetuned the models, enabling their proficiency in BC yield prediction. Optimization techniques, GA and PSO, systematically search and identify the optimal configuration of hyperparameters and key features through the exploration of multiple parameter combinations (Zhong et al. 2021). GA functions as a population-based feature selection algorithm, differing from the traditional approach of choosing a single solution. In the beginning, both genetic and binary data are considered as chromosomes carrying problem solutions. Each value within the population possesses distinctive fitness, collectively yielding diverse potential solutions (Huang et al. 2020). For BC yield, a population is generated for parameters such as proximate and ultimate analysis data and process conditions. At a specific time, this combination of solutions (populations) is termed a generation. The pivotal component in GA is the fitness function, serving as a fundamental parameter that defines the optimization problem at hand (Naveed et al. 2024). PSO is an optimization algorithm that has a multidimensional application. Its applications are widespread, especially in solving optimization problems characterized by high-dimensional search spaces and intricate objective functions. The initiation of PSO involves randomly placing a swarm of particles within the search space. Each particle subsequently adjusts its position by relying on its own experience and that of its neighbours (Papadakis and Markaki 2019). Through iterative updates to their velocities and positions, particles diligently explore the search space in pursuit of optimal solutions.

2.4 Materials and methods

Algae samples were collected from three different locations: Rawal Dam, Simli Dam and Shahpur Dam in Pakistan. Various strains of algal consortia are present throughout freshwater bodies in Pakistan. The selected three samples are the most readily and easily harvestable. The samples were initially washed and sun-dried. After rinsing with DI water, they were dried in a oven for 12 h at 60 °C. Upon drying, the samples were ground to a particle size of 60–120 µm, which is the most commonly reported particle size for obtaining optimal algal BC yields. Moreover, this range lies in between the predicted conditions by ML methods for testing via experimentation. The prepared samples were then placed in a vacuum oven at 105 °C for 12 h to dry the samples thoroughly. The samples were characterized by proximate and ultimate analysis since ML models require the input values in the form of the elemental compositions to predict the BC yield. The various algal biomass samples were denoted as Sample 1, Sample 2, and Sample 3, respectively. Table 2 provides the values of ultimate and proximate analysis for each sample.

2.4.1 Pyrolysis

A tube furnace was used for the validation of ML predicted pyrolysis conditions, with a nitrogen flow for inert conditions. The biomass was reduced in size and then pyrolyzed in the presence of nitrogen flow. Size-reduced algal samples were placed in the quartz tube in a boat crucible. Algal samples were weighed before and after pyrolysis to obtain BC yield. S2 in the supplementary material shows the experimental setup of algal pyrolysis.

3 Results and discussion

3.1 Optimization based feature selection

ML models predict BC yield based on the inter-dependance of proximate and ultimate analysis along with process conditions. Removal of a single parameter from the model prediction may significantly improve or decrease the accuracy and efficiency of the model (Meyer et al. 2019). Parameters are essentially called features,

Table 2 Elemental compositions of obtained three algae samples

Technique	Sample	Simli (1)	Shahpur (2)	Rawal (3)
Ultimate (wt%, daf)	C	44.15	46.9	49.92
	H	6.7	11.6	9.9
	N	1.35	8.7	1.55
	O[a]	43.6	23.5	31.5
Proximate (wt%, ar)	VM	82.45	70.5	82.12
	MC	3.91	6.7	2.98
	Ash	1.4	9.3	1.99
	FC	13.15	13.5	11.43
Lignocellulosic (wt%, daf)	Cellulose	45.12	44.71	28.45
	Hemicellolose	31.42	32.33	45.14
	Legnin	12.01	11.92	11.45

ar: as received; daf: dry ash free; a: calculated by difference

Table 3 Feature selection by GA and PSO for respective ML models

Optimization	Model	Selected features
GA	DT	H, O, FC, Ash, Pyrolysis Temperature, Heating Rate, Residence Time, Particle Size, N_2 Flowrate
	SVM	N, H, Ash, Lignin, Pyrolysis Temperature, Residence Time, Particle Size, N_2 Flowrate
	ET	N, O, FC, Ash, Cellulose, Pyrolysis Temperature, Heating Rate, Residence Time, Particle Size, N_2 Flowrate
	GPR	N, H, O, FC, MC, Lignin, Hemicellulose, Pyrolysis Temperature, Residence Time, Particle Size, N_2 Flowrate
PSO	DT	C, H, O, FC, MC, Ash, Pyrolysis Temperature, Heating Rate, Residence Time, Particle Size
	SVM	C, H, O, FC, VM, MC, Ash, Cellulose, Pyrolysis Temperature, Heating Rate, N_2 Flowrate
	ET	C, H, FC, VM, Cellulose, Hemicellulose, Pyrolysis Temperature, Heating Rate, Residence Time, Particle Size, N_2 Flowrate
	GPR	FC, Hemicellulose, Pyrolysis Temperature, Heating Rate, Residence Time, Particle Size, N_2 Flowrate

and feature selection is a crucial part of ML modelling. GA and PSO are the optimization techniques that select the features for training of respective ML models (Ghosh et al. 2019). It is important to mention that selected features are indicators of the extent of their contribution to the prediction of BC yield rather than the scientific relation of the feature with the BC yield. This step is a way towards increasing the predictive capability of model for BC yield prediction. The features selected by GA and PSO for each of the models are shown in Table 3.

3.2 Hyperparameters tuning for optimum biochar yield

The hyperparameters of ML models were optimized to increase the predictive capability of ML models. The parameters assigned for tuning across various ML models were identified using fivefold cross-validation on MATLAB. Hyperparameter selections, their corresponding ranges, and the resulting optimized values are presented in Fig. S1 in the supplementary material. The tuning and optimization of these hyperparameters were conducted utilizing both GA and PSO for ET, SVM, DT and GPR.

For ET refinement, optimum hyperparameter values included several learning cycles set at 29, a learning rate of 0.334, and the chosen method as LSBoost. The application of PSO for SVM yielded optimal values, including a box constraint value of 49.747, a kernel scale value of 38.120, an epsilon of 0.0095256, and a Gaussian kernel function. For GPR, the optimized hyperparameters comprised a sigma of 31.222, an Ardexponential exponential kernel function, and no basic function. DT showed a minimum leaf size of 4.160 and Surrogate 'on' for the optimization via PSO. After GA refinement for the ET, the reported optimum hyperparameter values included several learning cycles set at 10, a learning rate of 1, and the chosen method as LSBoost. For SVM, the application of the GA yielded optimal values, including a box constraint value of 989.5, a kernel scale value of 142.044, an epsilon of 11.662, and a Gaussian kernel function. In the case of GPR, the optimized hyperparameters, after GA application, comprised a sigma of 32.665, a Ardexponential exponential kernel function, and no basic function. DT showed a minimum leaf size of 2 and Surrogate 'off' for the optimization via GA.

3.3 Prediction performance

BC yield predictions were performed using various ML models, including SVM, DT, ET, and GPR. Utilizing GA-based feature selection, all four models exhibited improved prediction accuracy. After GA-based feature selection, the testing R^2 value increased to 0.7023, while the testing RMSE decreased to 8.1173, demonstrating the removal of irrelevant features. The GA-optimized ET, GPR, DT, and SVM models achieved testing R^2 values of 0.7023, 0.69537, 0.68499, and 0.36254, with corresponding testing RMSE values of 8.1173, 8.2113, 8.35, and 11.878, respectively. Figure 4 illustrates the performance of these models. PSO-optimized ML models demonstrated superior performance compared to GA, with testing R^2 and testing RMSE values for ET were 0.77993 and 6.9792, respectively, indicating its strong predictive capability. GPR followed with a testing R^2 of 0.76521 and an RMSE of 7.2088, while SVM and DT achieved testing R^2 values of 0.74817 and 0.5663 respectively.

ET emerged as the best-performing model overall as shown in Fig. 4f, with a testing R^2 value of 0.77993 and a RMSE value of 6.9792, followed by GPR with a testing R^2 value of 0.76521 and a RMSE value of 7.2088. SVM and DT had testing R^2 values of 0.74817 and 0.5663, with RMSE values of 7.4658 and 9.7976, respectively. The consolidated result of each model is given in Fig. S3 in the supplementary material.

3.4 Uncertainty and sensitivity analysis

3.4.1 Shapley analysis

The ET model, enhanced by the Shapley method, effectively delineates the intricate relationship between input parameters and BC yield. Operating on the principle of feature attribution magnitude, the Shapley method serves as a valuable tool for assessing the relative importance of various input parameters in the context of composition and pyrolysis conditions under BC yield. The analysis in Fig. 5a highlights pyrolysis temperature as the most critical factor (1.2785), confirming its dominant role in determining BC yield. Higher temperatures enhance devolatilization, reducing solid BC (Haykiri-Acma et al. 2017). Volatile matter (0.3972) also significantly affects yield, as higher VM leads to increased gas release and lower char retention (Lang et al. 2005). The heating rate (0.2949) influences char formation, with rapid heating promoting volatilization and reducing BC, while slower rates favor retention (Lalaymia et al. 2025). Ash content (0.2771) contributes to char stability, as it remains after pyrolysis (Ganesan et al. 2025). Lignin (0.2521), being thermally stable, enhances BC retention (Kudelytė et al. 2025). FC (0.1828) supports stability, whereas O (0.1655) promotes oxidative decomposition, lowering yield. Cellulose (0.1049) decomposes rapidly, producing less char, while moisture (0.0975) affects heating but not final yield (Zhang et al. 2021). Particle size (0.0315), N content (0.0249), hemicellulose (0.0180), and H (0.0047) have minimal effects. Hemicellulose decomposes early, forming more volatiles than char. The results reinforce that temperature, heating rate, and feedstock composition are key determinants of BC production, crucial for optimizing pyrolysis in applications like carbon sequestration and wastewater treatment.

The SHAP analysis in Fig. 5b illustrates the impact of different parameters on the model's BC yield predictions. Pyrolysis temperature exhibits the most significant influence, with higher values (pink) leading to negative SHAP values, indicating a reduction in BC yield, while lower temperatures (blue) contribute positively. Pyrolysis temperature is the most influential factor, with higher values reducing BC yield due to increased volatile release. Ash content positively affects yield, as its inorganic fraction remains after pyrolysis. High VM and MC reduce BC formation by promoting gas production. A higher heating rate follows the same trend, accelerating volatilization and lowering char retention. Lignin supports BC retention due to its thermal stability, while cellulose and hemicellulose decompose rapidly, reducing yield. O and H contents contribute to decomposition, further lowering char formation. Particle size and N have minimal impact; though larger particles slightly favor retention by slowing heat transfer. This analysis confirms that temperature, heating rate, and feedstock composition are key factors in optimizing BC production.

3.4.2 Monte carlo simulations and sobol sensitivity analysis

Monte Carlo simulations are a valuable tool for assessing uncertainty in predictive modelling (Janssen 2013), making them well-suited for studying algae BC yield. Since the input parameters can vary significantly, Monte Carlo simulations help to assess the variability by generating a range of possible outcomes rather than a single predicted value. Figure 6a represents the distribution of predicted BC obtained from Monte Carlo simulations, where random variations were introduced to selected input features. The mean predicted BC, indicated by the blue dashed line, is 35.01, representing the expected value of the model's output after accounting for input uncertainties. This serves as the central tendency of predictions. The 95% confidence interval is marked by grey dashed lines, with the lower bound at 22.93 (2.5% of predictions fall below this value) and the upper bound at 55.35 (97.5% of predictions fall below this value). This implies that the BC is 95% likely to be between 23.93 and 55.35, reflecting the uncertainty associated with input variations. The histogram itself exhibits a right-skewed distribution, indicating that while most predictions are concentrated

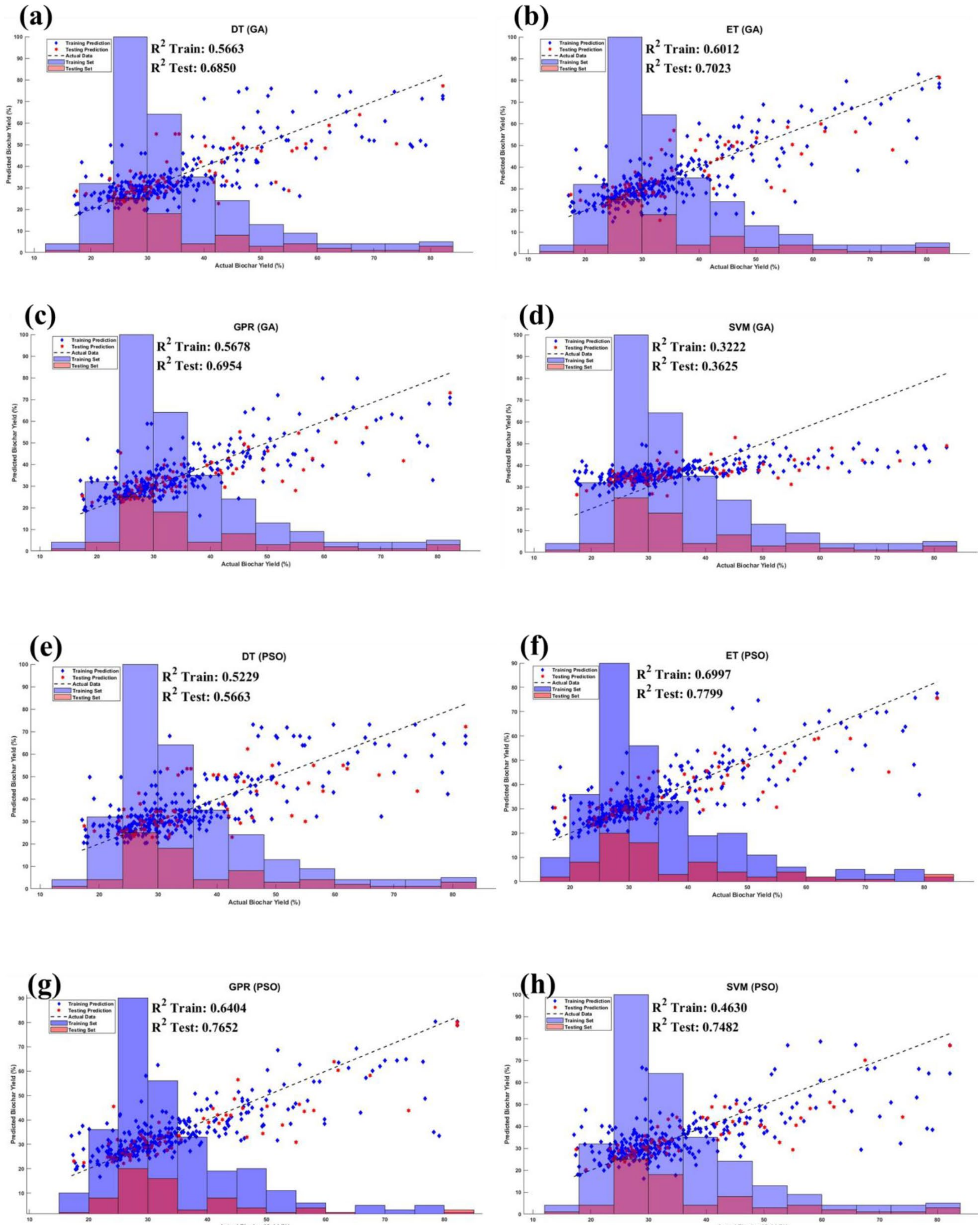


Fig. 4 Enhanced biochar prediction performance using optimization techniques

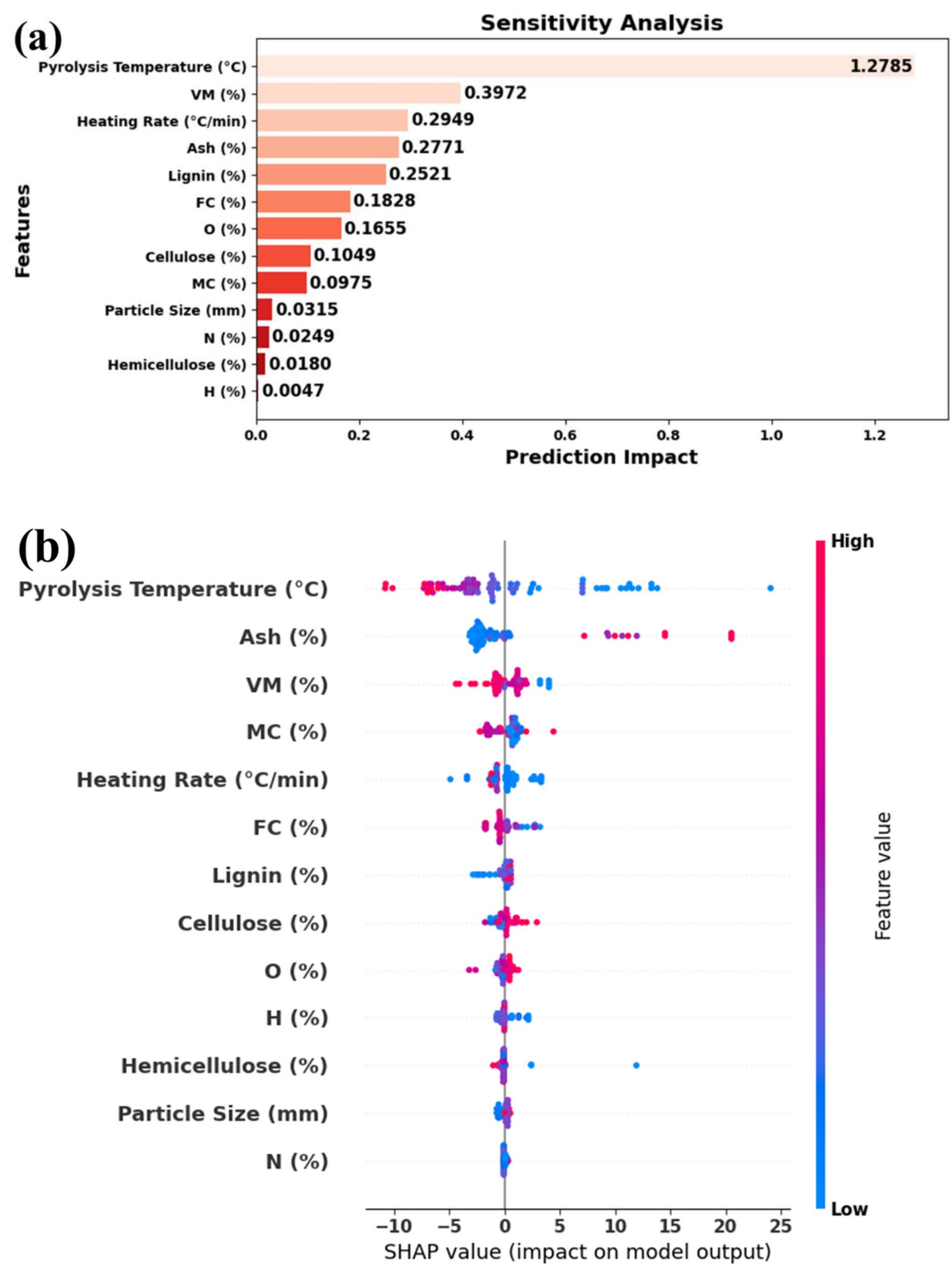


Fig. 5 Shapley analysis of input features for biochar yield prediction

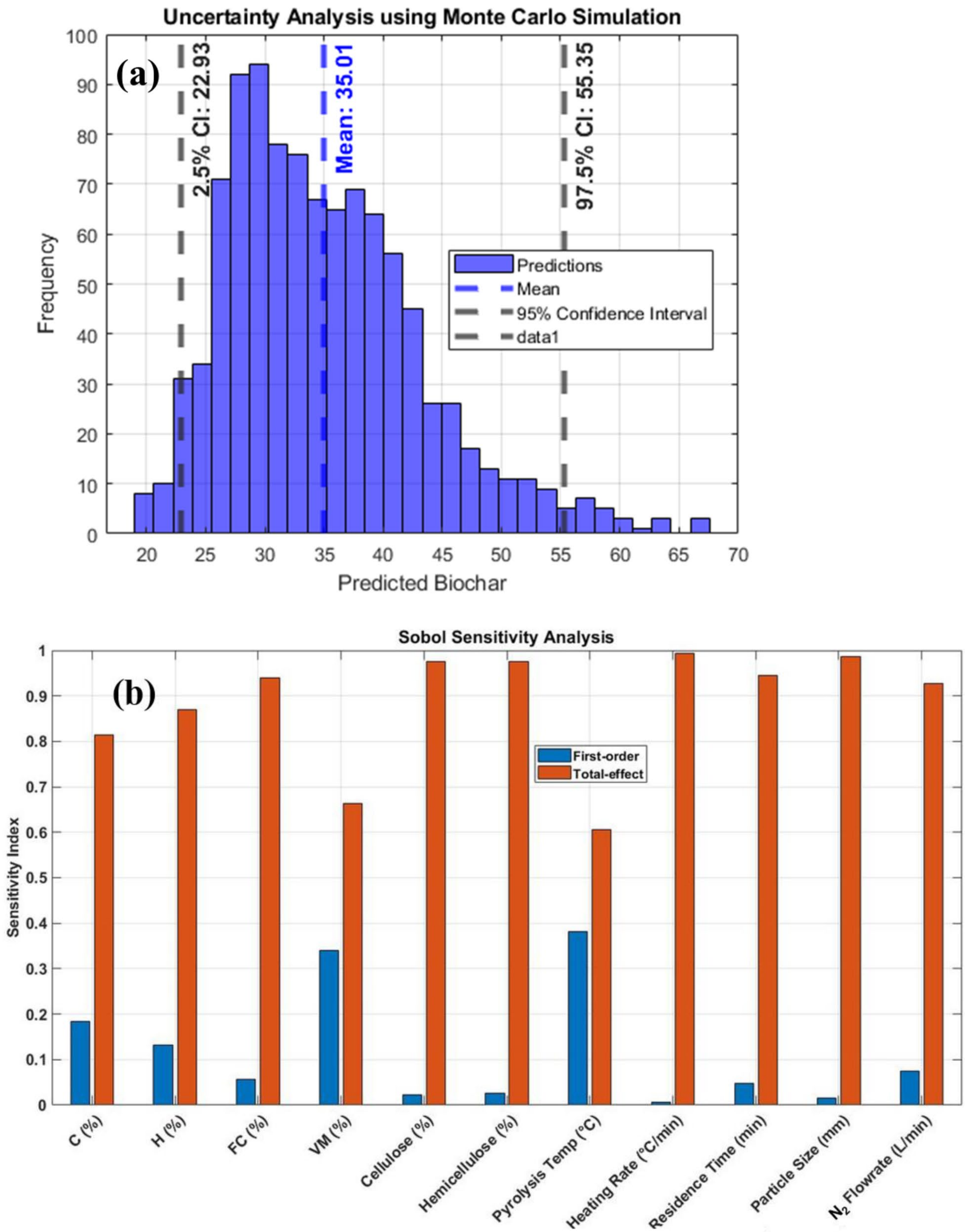


Fig. 6 **a** Uncertainty analysis using Monte Carlo simulations **b** Sobol sensitivity analysis

in the range of 30–40, some predictions exceed 60. This suggests that certain input features lead to significantly higher values of BC, likely due to nonlinear effects that need investigation through Sobol sensitivity analysis. The highest frequency of BC is observed within the 30–40 range, while predictions within the 50–70 range occur less frequently. Although the mean prediction is 35.01, the considerable variation, with a 95% confidence range from 22.93 to 55.35, underlines the importance of incorporating uncertainty analysis in decision-making. Understanding this range helps assess risk and variability, ensuring more informed and robust conclusions based on the model's BC predictions.

The Sobol sensitivity analysis provides a detailed assessment of how various input parameters influence BC yield (Hao et al. 2024), capturing both their direct effects and interactions. The first-order sensitivity index (blue bars) quantifies the independent impact of each parameter, while the total-effect sensitivity index (orange bars) accounts for both direct effects and interactions with other variables. The Sobol sensitivity analysis plot in Fig. 6b provides a detailed breakdown of the influence of different input parameters on the predicted BC yield, considering both their direct effects and interactions. Figure 6b shows that heating rate, particle size, residence time, and N_2 flow rate have higher total order values of 0.99, 0.98, 0.93 and 0.92, respectively, as compared to their first order values. This indicates the presence of nonlinear behaviour and collective significant effects. Temperature has a moderate first-order value to influence BC yield, but it impacts the yield significantly when considered in connection with process variables such as heating rate, and particle size. This explains the observed values of BC towards the tail in Fig. 6a. Furthermore, for the Ensemble Tree model, this nonlinear behaviour and compounding effects result in higher BC yield. The skewed distribution is caused by nonlinear effects (Lu et al. 2021; Aghbashlo et al. 2021) and interactions between these influential variables. Since the model is highly sensitive to changes in these key factors, even small variations can lead to larger shifts in predicted values, creating the longer tail on the right side of the graph. This implies that most parameters contribute to BC yield variation primarily through interactions rather than through their individual influences. Pyrolysis temperature emerges as the dominant parameter, with high first-order and total-effect indices, confirming its critical role in BC yield determination. Additionally, C, FC, and VM exhibit moderate first-order effects but substantial total contributions, emphasizing their dual role in independent and interaction-driven variations. From a practical perspective, pyrolysis temperature should be carefully controlled and optimized due to its dominant impact. Parameters like heating rate, particle size, residence time and lignin primarily affect BC yield through interactions. This analysis aids in refining predictive models and optimizing pyrolysis conditions for improved BC production efficiency.

3.5 Effect of parameters

Conducting a PDPs analysis is instrumental in discerning the influence of input variables on BC yield. In Fig. 7a, the PDP illustrates the relationship between temperature and BC yield. At temperatures below 200 °C, BC yield remains high, around 60–70%, due to minimal thermal degradation. As temperature increases from 200 °C to 400 °C, yield decreases sharply, fluctuating between 30% and 50%, likely due to the decomposition of hemicellulose and cellulose (Faleeva et al. 2024). Around 400–600 °C, yield shows significant variations, dropping to nearly 20% at times, which may be attributed to the breakdown of lignin and secondary reactions. Beyond 600 °C, yield steadily declines, reaching around 10% at 900–1000 °C, as higher temperatures promote gasification and volatile release. The colour gradient reflects decomposition intensity, reinforcing temperature as a key factor in BC production. Figure 7b illustrates that at low flow rates (0–10), BC yield fluctuates between 30% and 55%, possibly due to inconsistent volatile removal and heat distribution. A peak around 10 suggests an optimal flow rate that balances inert conditions with controlled volatilization. As the flow rate increases from 10 to 25, BC yield drops sharply to nearly 25–30%, indicating enhanced volatile removal, which reduces solid residue. Beyond 25, BC yield rises again, stabilizing around 45% at higher flow rates (40–50), likely due to reduced residence time limiting secondary decomposition. Figure 7c depicts the effect of residence time on BC yield. Initially (0–20 min), BC yield fluctuates around 30–35%, likely due to incomplete thermal decomposition. As residence time increases to 60 min, BC yield stabilizes around 35–40%, with a sharp peak exceeding 60% at approximately 65 min. The spike is attributed to the transient formation of stable carbonaceous intermediates, such as aromatic clusters, which momentarily enhance char yield before undergoing secondary degradation. A similar pattern has been observed in slow pyrolysis of lignocellulosic biomass, where intermediate tar condensation or limited volatilization leads to a temporary increase in solid yield (Amalina et al. 2022). Following this peak, the yield declines due to further decomposition or thermal cracking of previously formed structures, aligning with established thermal degradation kinetics. Figure 7d illustrates that at smaller particle sizes (0–2 mm), BC yield fluctuates between 30% and 45%, likely due to rapid heat transfer and inconsistent pyrolysis behavior. As particle size increases to around 5

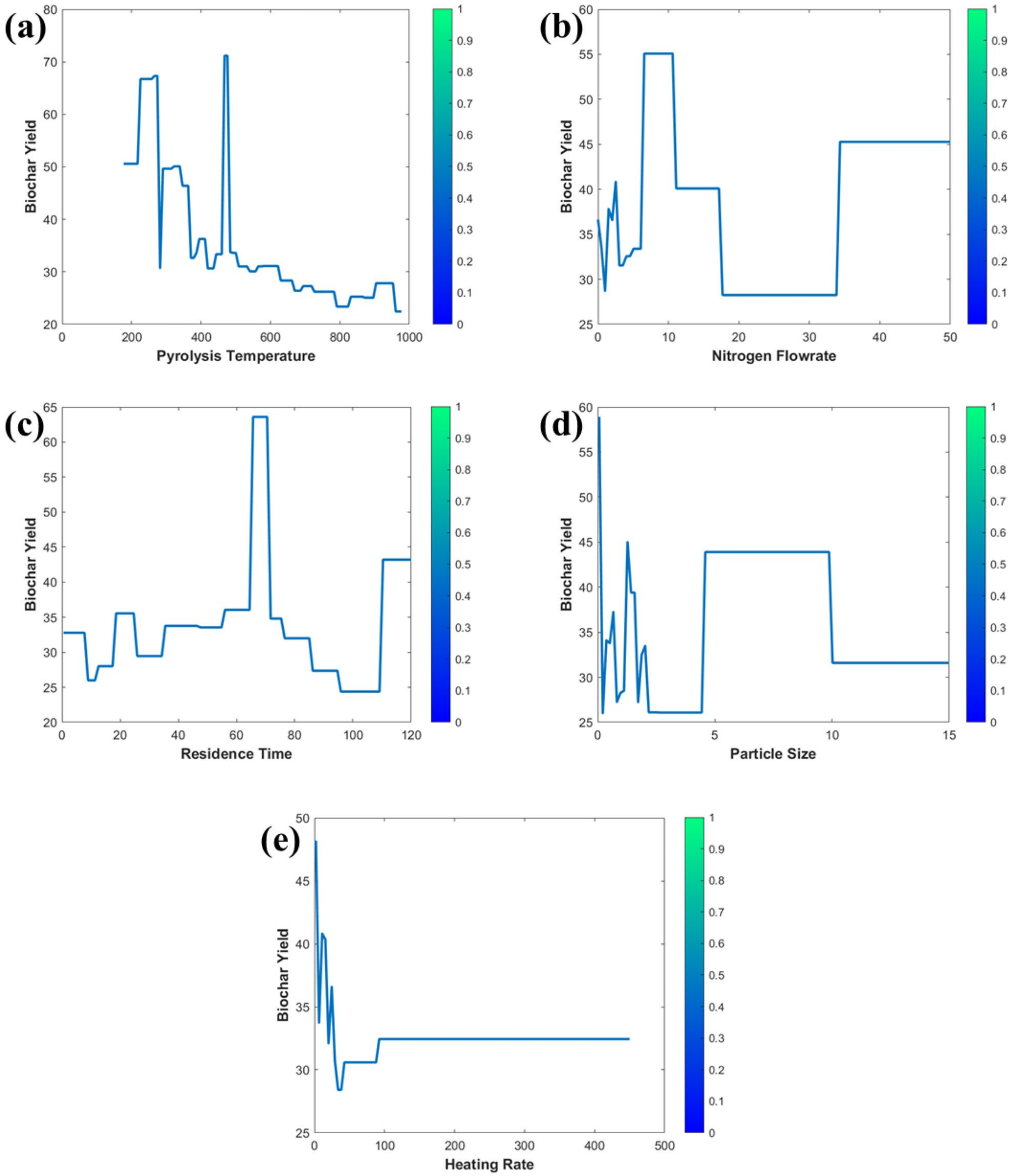


Fig. 7 2-D partial dependance plots of pyrolysis conditions against biochar yield

mm, BC yield stabilizes at approximately 45–50%, suggesting an optimal size for char formation where heat penetration and devolatilization are balanced. Figure 7e shows that at low heating rates (0–50 °C/min), BC yield fluctuates between 30% and 45%, likely due to varying decomposition rates of biomass components. As the heating rate increases beyond 50 °C/min, BC yield stabilizes at approximately 33–35%, suggesting that higher heating rates favor volatile release, reducing solid char formation.

The breakdown of algae for thermochemical conversion requires optimal pyrolysis conditions to achieve optimal BC yield. Ionic reaction mechanisms may become prominent under certain conditions. Examining the ET model predictions, it is evident that certain features have a marginal impact.

In Fig. 8a, BC yield ranges from approximately 20% to over 70%, with the highest values observed at lower pyrolysis temperatures and moderate particle sizes. As temperature increases beyond 500 °C, BC yield declines, indicating greater thermal degradation and volatile release. At smaller particle sizes (below 5 mm), yield fluctuations are noticeable, likely due to rapid heat transfer affecting decomposition rates. The color gradient confirms this trend, with higher BC yields shown in green at lower temperatures and moderate particle sizes, while lower yields appear in blue at high temperatures. In Fig. 8b, BC yield varies from approximately 20% to over 70%. Higher BC yields are observed at lower pyrolysis temperatures and moderate N_2 flow rates, as indicated by the green regions in the plot. As pyrolysis temperature increases beyond 500 °C, BC yield declines, likely due to enhanced decomposition and volatile loss. Similarly, at N_2 flow rates above 40 L/min, BC yield decreases, suggesting an influence on heat transfer and reaction kinetics. In Fig. 8c the BC yield varies from approximately 20% to over 70%. Higher BC yields are observed at lower pyrolysis temperatures and moderate residence times. As the pyrolysis temperature increases beyond 500 °C, BC yield declines, likely due to increased thermal degradation and volatilization. Additionally, at very high residence times, BC yield tends to decrease, suggesting excessive degradation of solid carbon content. Figure. 8d shows the impact of particle size and residence time on BC yield. The yield varies between approximately 20% and 80%, with the highest values observed for smaller particle sizes and lower residence times, indicated by the green regions. As particle size increases beyond 5 mm, BC yield declines significantly, suggesting that larger particles experience incomplete pyrolysis due to heat transfer limitations. Similarly, extended residence times do not contribute to a substantial increase in BC yield, indicating possible secondary decomposition of BC at prolonged exposure. Figure 8e illustrates the relationship between N_2 flow rate, residence time, and BC yield. At low N_2 flow rates (around 20–30), BC yield reaches a peak of approximately 75–80%. Residence time also impacts BC yield, with shorter times (near 10–20) resulting in higher yields, whereas extended times (above 50) lead to a decline. This trend suggests that excessive N_2 flow and prolonged residence times promote volatile loss, reducing the overall BC yield.

3.6 Experimental validation: pyrolysis of algae biomass

Three pyrolysis experiments were conducted on the samples collected. Conditions for the experiment were fixed for the ML model prediction and experimental validation. This enabled real-time validation of the results produced by the ML models prediction and the experimental setup. ML models also provided the conditions in which the setup would yield maximum BC. Those conditions were also tested and reported as the result of the study. Table 4 shows the experimental conditions on which the obtained algal samples were tested. Conditions for maximum BC yield predicted by ML model are also listed at the end for experimental validation. Conditions for maximum BC yield predicted by ML model are also enlisted at the end for experimental validation with heating rate of 10 °C/min, residence time of 60 min, N^2 flow rate of 0.5 L/min and particle size of 1.5 mm.

Pyrolysis of algae samples was carried out in a tube furnace at various temperatures, heating rates, and residence time conditions to validate the results of the ML models. Various experimental conditions were selected to validate the predicted and experimental BC yield. The yields of three samples predicted by ML models were experimentally validated and reported. The error was recorded to be less than 6% for the actual algal samples. Similarly, the conditions for maximum yield provided by ML model were tested and reported for the error of 1.12% as shown in Table 4. ML predicted yield for the maximum conditions using sample 3 was predicted to be 76.334%, while the actual experimentally obtained yield was 75.42%. Sample 3 exhibited the highest predicted yield for a higher particle size. This observation is attributed to the particle size exhibiting a low first-order Sobol index but a high total order index of 0.98. This indicates that, while particle size alone has a limited impact on BC yield, its influence becomes markedly more significant when interacting with process parameters. The interactive effects between variables influence the yield outcome more significantly than individual variables alone.

3.7 Graphical user interface

A graphical user interface was developed in this research to predict BC yield in real time. An interactive environment takes input in the form of ultimate, proximate composition and feed characteristics to predict the BC yield based on the ET prediction model. Fig. S3 in the supplementary file represents the GUI interface.

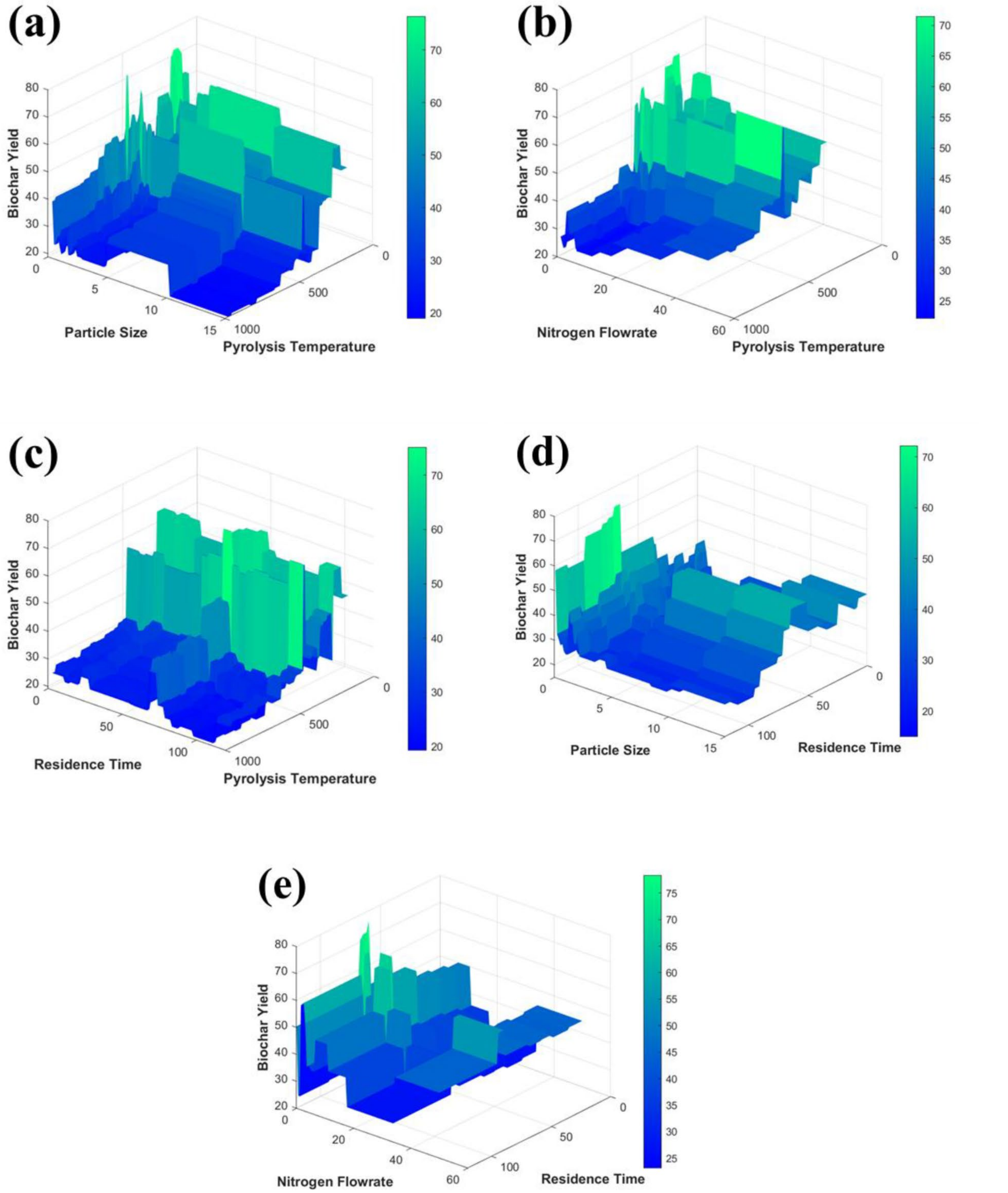


Fig. 8 3-D pdps of pyrolysis conditions against biochar yield

Table 4 Pyrolysis condition provided by ML model for experimental validation

Sample	Pyrolysis temperature (°C)	Heating Rate (°C/min)	Residence Time (min)	Particle Size (mm)	N_2 (L/min)	Prediction (%)	Experiment (%)	Error (%)
1	500	10	60	1.25	0.5	44.72	42.16	5.72
2	250	15	60	5	0.8	67.23	68.37	1.67
3	330	15	60	5	0.8	64.01	62.24	2.84
Max Yield	500	10	60	1.5	0.5	76.33	75.42	1.12

4 Conclusion

Diverse and varied profile of algal biomass challenges quantification of properties for optimum yield prediction. Data-driven optimised ML models integrate key predictors with algal profiles, enabling better predictions and offering resource conservation. ML models were employed to predict BC yield based on the proximate, ultimate, and pyrolysis conditions. The PSO-optimized training R^2 values of ET, GPR, DT, and SVM were 0.6997, 0.6456, 0.5229, and 0.4630, while testing R^2 values were 0.7799, 0.7652, 0.5663, and 0.7482, respectively. The ET model exhibited superior predictive performance ($R^2 = 0.7792$), and its inverse optimization capability uniquely enables researchers to utilize comprehensive model for prediction, significantly reducing traditional trial-and-error methods. In contrast to reported studies, a comprehensive dataset was employed, with tunned k-fold cross validation and the application of inverse optimization to identify the conditions for maximum BC yield. Also, for the stability of ML model, uncertainty analysis through Monte Carlo simulation and Sobol sensitivity analysis was performed to quantify the model performance with 1st and total order effects on algal BC yield. The crucial features influencing BC prediction using the ET model were identified as FC, MC, and pyrolysis temperature. The ET model underscores the potential for enhancing BC production through elevated temperature (350–750 °C) and heightened MC. The results obtained from BC yield predictions were tested and cross-verified for three different algal biomass samples with recorded error less than 6%. The pyrolysis conditions for maximum BC yield were obtained from inverse optimization and the experiment was concluded with an error of 1.12%.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1007/s42773-025-00511-w.

Additional file 1.

Acknowledgements
Jawad Gul, the first author, is thankful to National University of Sciences and Technology, Islamabad, Pakistan for PhD N-Centive scholarship.

Author contributions
Jawad Gul: Methodology, Software, Data curation, Formal analysis, Writing—original draft. Muhammad Nouman Aslam Khan: Conceptualization, Methodology, Investigation, Software, Supervision, Writing—review and editing. Umair Sikander: Formal analysis, Validation, Methodology, Writing—original draft. Asif Hussain: Investigation, Resources, Visualization, Writing—original draft. Melanie Kah: Formal analysis, Investigation, Writing—review and editing. Salman Raza Naqvi: Conceptualization, Methodology, Writing—review and editing.

Funding
No funding was received to assist with the preparation of this manuscript.

Data availability
Data sets used during the current study are available from the corresponding author on request.

Declarations

Competing interests
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.

Author details
[1]Alternative Fuels and Sustainability Laboratory, School of Chemical and Materials Engineering, National University of Sciences and Technology, Islamabad H-12, Pakistan. [2]Fossil Fuels Laboratory, Department of Thermal Energy Engineering, U.S.-Pakistan Centre for Advanced Studies in Energy, National University of Sciences & Technology, Islamabad H-12, Pakistan. [3]Faculty of Science, Environment, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand. [4]Department of Engineering and Chemical Sciences, Karlstad University, Karlstad, Sweden.

Received: 26 March 2025 Revised: 2 September 2025 Accepted: 8 September 2025
Published online: 07 January 2026

References

Adhikari S, Timms W, Mahmud MAP (2022) Optimising water holding capacity and hydrophobicity of biochar for soil amendment—a review. Sci Total Environ 851:158043

Aghbashlo M, Almasi F, Jafari A, Nadian MH, Soltanian S, Lam SS, Tabatabaei M (2021) Describing biomass pyrolysis kinetics using a generic hybrid intelligent model: a critical stage in sustainable waste-oriented biorefineries. Renew Energy 170:81–91

Amalina F, Abd Razak AS, Krishnan S, Sulaiman H, Zularisam AW, Nasrullah M (2022) Biochar production techniques utilizing biomass waste-derived materials and environmental applications–a review. J Hazard Mater Adv 7:100134

Cao H, Xin Y, Yuan Q (2016) Prediction of biochar yield from cattle manure pyrolysis via least squares support vector machine intelligent approach. Bioresour Technol 202:158–164

Chang Y-M, Tsai W-T, Li M-H (2015) Chemical characterization of char derived from slow pyrolysis of microalgal residue. J Anal Appl Pyrolysis 111:88–93

Critchley F, Jones MC (2008) Asymmetry and gradient asymmetry functions: density-based skewness and kurtosis. Scand J Stat 35:415–437

Faleeva YM, Lavrenov VA, Zaichenko VM (2024) Investigation of plant biomass two-stage pyrolysis based on three major components: cellulose, hemicellulose, and lignin. Biomass Convers Biorefin 14:14519–14529

Ganesan A, Rezazgui O, Langlois S, Boussabbeh C, Barnabé S (2025) Pyrolytic conversion of construction, renovation, and demolition (CRD) wood wastes in Québec to biochar: production, characterization, and identifying relevant stability indices for carbon sequestration. Sci Total Environ 965:178650

Ghosh M, Guha R, Alam I, Lohariwal P, Jalan D, Sarkar R (2019) Binary genetic swarm optimization: a combination of GA and PSO for feature selection. J Intell Syst 29:1598–1610

Hadey C, Allouch M, Alami M, Boukhlifi F, Loulidi I (2022) Preparation and characterization of biochars obtained from biomasses for combustible briquette applications. Sci World J 2022:2554475

Hao S, Ryu D, Western AW, Perry E, Bogena H, Franssen HJH (2024) Global sensitivity analysis of APSIM-wheat yield predictions to model parameters and inputs. Ecol Modell 487:110551

Haq ZU, Ullah H, Khan MNA, Naqvi SR, Ahsan M (2022) Hydrogen production optimization from sewage sludge supercritical gasification process using machine learning methods integrated with genetic algorithm. Chem Eng Res des 184:614–626

Haykiri-Acma H, Kurt G, Yaman S (2017) Properties of biochars obtained from RDF by carbonization: influences of devolatilization severity. Waste Biomass Valorizat 8:539–547

Hoang AT, Goldfarb JL, Foley AM, Lichtfouse E, Kumar M, Xiao L, Ahmed SF, Said Z, Luque R, Bui VG (2022) Production of biochar from crop residues and its application for anaerobic digestion. Bioresour Technol. https://doi.org/10.1016/j.biortech.2022.127970

Huang J, Zhu Y, Kelly JT, Jang C, Wang S, Xing J, Chiang P-C, Fan S, Zhao X, Yu L (2020) Large-scale optimization of multi-pollutant control strategies in the Pearl River Delta region of China using a genetic algorithm in machine learning. Sci Total Environ 722:137701

Janssen H (2013) Monte-Carlo based uncertainty analysis: sampling efficiency and sampling convergence. Reliab Eng Syst Saf 109:123–132

Jung K-W, Jeong T-U, Kang H-J, Ahn K-H (2016) Characteristics of biochar derived from marine macroalgae and fabrication of granular biochar by entrapment in calcium-alginate beads for phosphate removal from aqueous solution. Bioresour Technol 211:108–116

Khan M, Ullah Z, Mašek O, Naqvi SR, Khan MNA (2022a) Artificial neural networks for the prediction of biochar yield: a comparative study of metaheuristic algorithms. Bioresour Technol 355:127215

Khan AA, Gul J, Naqvi SR, Ali I, Farooq W, Liaqat R, AlMohamadi H, Štěpanec L, Juchelková D (2022b) Recent progress in microalgae-derived biochar for the treatment of textile industry wastewater. Chemosphere. https://doi.org/10.1016/j.chemosphere.2022.135565

Kudelytė V, Eimontas J, Paulauskas R, Striūgas N (2025) Co-pyrolysis of plastic waste and lignin: a pathway for enhanced hydrocarbon recovery. Energies 15:1–19

Lalaymia I, Belaadi A, Ghernaout D (2025) Studying Gaussian deconvolution and multicomponent kinetics models in *Agave* cellulosic fibers pyrolysis: application in sustainable bioenergy for cleaner production. Biomass Bioenerg 192:107488

Lang T, Jensen AD, Jensen PA (2005) Retention of organic elements during solid fuel pyrolysis with emphasis on the peculiar behavior of nitrogen. Energy Fuels 19:1631–1643

Leng L, Lei X, Al-Dhabi NA, Wu Z, Yang Z, Li T, Zhang W, Liu W, Zhan H, Peng H (2024) Machine-learning-aided prediction and engineering of nitrogen-containing functional groups of biochar derived from biomass pyrolysis. Chem Eng J 485:149862

Lu L, Gao X, Gel A, Wiggins GM, Crowley M, Pecha B, Shahnam M, Rogers WA, Parks J, Ciesielski PN (2021) Investigating biomass composition and size effects on fast pyrolysis using global sensitivity analysis and CFD simulations. Chem Eng J 421:127789

Meyer H, Reudenbach C, Wöllauer S, Nauss T (2019) Importance of spatial predictor variable selection in machine learning applications–moving from data reproduction to spatial prediction. Ecol Modell 411:108815

Naveed MH, Khan MNA, Mukarram M, Naqvi SR, Abdullah A, Haq ZU, Ullah H, Al Mohamadi H (2024) Cellulosic biomass fermentation for biofuel production: review of artificial intelligence approaches. Renew Sustain Energy Rev 189:113906

Nguyen VG, Sharma P, Ağbulut Ü, Le HS, Truong TH, Dzida M, Tran MH, Le HC, Tran VD (2024) Machine learning for the management of biochar yield and properties of biomass sources for sustainable energy. Biofuels Bioprod Biorefin 18:567–593

Papadakis S, Markaki M (2019) An in depth economic restructuring framework by using particle swarm optimization. J Clean Prod 215:329–342

Pathy A, Meher S, Balasubramanian P (2020) Predicting algal biochar yield using extreme gradient boosting (XGB) algorithm of machine learning methods. Algal Res 50:102006

Potter K, Hagen H, Kerren A, Dannenmann P. Methods for presenting statistical information: the box plot. 2006. p. 97–106.

Shuttleworth P, Budarin V, Gronnow M, Clark JH, Luque R (2012) Low temperature microwave-assisted vs conventional pyrolysis of various biomass feedstocks. J Nat Gas Chem 21:270–274

Sun J, Norouzi O, Mašek O (2022) A state-of-the-art review on algae pyrolysis for bioenergy and biochar production. Bioresour Technol 346:126258

Tag AT, Duman G, Ucar S, Yanik J (2016) Effects of feedstock type and pyrolysis temperature on potential applications of biochar. J Anal Appl Pyrolysis 120:200–206

Ullah Z, Naqvi SR, Farooq W, Yang H, Wang S, Vo D-VN (2021) A comparative study of machine learning methods for bio-oil yield prediction—a genetic algorithm-based features selection. Bioresour Technol 335:125292

Ullah H, Haq ZU, Naqvi SR, Khan MNA, Ahsan M, Wang J (2023) Optimization based comparative study of machine learning methods for the prediction of bio-oil produced from microalgae via pyrolysis. J Anal Appl Pyrolysis 170:105879

Wang K, Brown RC, Homsy S, Martinez L, Sidhu SS (2013) Fast pyrolysis of microalgae remnants in a fluidized bed reactor for bio-oil and biochar production. Bioresour Technol 127:494–499

Westerhof RJM, Kuipers NJM, Kersten SRA, Van Swaaij WPM (2007) Controlling the water content of biomass fast pyrolysis oil. Ind Eng Chem Res 46:9238–9247

Yanik J, Stahl R, Troeger N, Sinag A (2013) Pyrolysis of algal biomass. J Anal Appl Pyrolysis 103:134–141

Yuan T, Tahmasebi A, Yu J (2015) Comparative study on pyrolysis of lignocellulosic and algal biomass using a thermogravimetric and a fixed-bed reactor. Bioresour Technol 175:333–341

Zhang C, Chao L, Zhang Z, Zhang L, Li Q, Fan H, Zhang S, Liu Q, Qiao Y, Tian Y (2021) Pyrolysis of cellulose: evolution of functionalities and structure of bio-char versus temperature. Renew Sustain Energy Rev 135:110416

Zhang Y, Maierdan Y, Guo T, Chen B, Fang S, Zhao L (2022) Biochar as carbon sequestration material combines with sewage sludge incineration ash to prepare lightweight concrete. Constr Build Mater 343:128116

Zhong S, Zhang K, Bagheri M, Burken JG, Gu A, Li B, Ma X, Marrone BL, Ren ZJ, Schrier J (2021) Machine learning: new ideas and tools in environmental science and engineering. Environ Sci Technol 55:12741–12754

Zhou B, Li H, Wang Z, Huang H, Wang Y, Yang R, Huo R, Xu X, Zhou T, Dong X (2024) Prediction of phosphate adsorption amount, capacity and kinetics via machine learning: a generally physical-based process and proposed strategy of using descriptive text messages to enrich datasets. Chem Eng J 479:147503