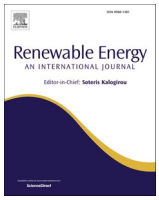
Renewable Energy 256 (2026) 124071

Contents lists available at ScienceDirect

Renewable Energy

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



Optimising biochar from agricultural residues: Predicting elemental composition with machine learning

Yao Fu *, Peter Cleall, Fei Jin **

School of Engineering, Cardiff University, Cardiff, CF24 3AA, UK

ARTICLE INFO

Keywords:
Agricultural residues
Biochar elemental composition
Feedstock characteristics
Machine learning
Pyrolysis parameters

ABSTRACT

Biochar, a material whose properties are critically defined by its elemental composition, has been promoted as a sustainable way to treat various biomass wastes, including agricultural residues. However, considerable variability in these compositions across studies necessitates precise predictive techniques. This research followed the PRISMA rules for data collection and study selection, compiling data on feedstock properties and pyrolysis parameters from 38 published studies. A novel Feature-oriented Imputation method was established and employed, utilizing K-Nearest Neighbours (KNN) or Random Forest (RF) imputer to fill in missing values for features with differing characteristics. The reprocessed data were then fed into six distinct datasets and analyzed using a Gradient Boosting Regression model to predict the contents of carbon (C), hydrogen (H), oxygen (O), nitrogen (N), phosphorus (P), and potassium (K) in biochar. The rigorous machine learning process yielded excellent accuracy rates: C (R^2 = 0.9088, RMSE = 4.0614), H (R^2 = 0.9068, RMSE = 0.4180), O (R^2 = 0.9172, RMSE = 2.6475), N (R^2 = 0.8950, RMSE = 0.3416), P (R^2 = 0.9699, RMSE = 0.0244), and K (R^2 = 0.9464, RMSE = 0.3842). A comprehensive analysis of feature importance revealed that feedstock properties generally hold more significance in determining the elemental composition of biochar compared to pyrolysis parameters. The highest heating temperature (HHT) emerged as the most influential parameter for the content of H and O, while the contents of N, P, and K were predominantly determined by their respective levels in the feedstock. From these insights, optimal pyrolysis parameters were derived to tailor biochar with different elemental compositions for various applications. The developed models offer a robust framework for predicting the elemental compositions of biochar derived from various agricultural biomass, thereby eliminating the need for complex and resource-intensive laboratory trials.

Abbreviation and Nomenclature

Food and Agriculture Organisation: FAO	Isolation Forest: IF
carbon: C	feedstock fixed carbon content: FC_F
hydrogen: H	feedstock volatile matter content: VM_F
oxygen: O	feedstock ash content: AC_F
nitrogen: N	feedstock carbon content: C_F
phosphorus: P	feedstock hydrogen content: H_F
potassium: K	feedstock oxygen content: O_F
cation exchange capacity: CEC	feedstock nitrogen content: N_F
electrical conductivity: EC	feedstock phosphorous content: P_F
highest heating temperature: HHT	feedstock potassium content: K_F
residence time: RT	feedstock cellulose content: Cel_F
heating rate: HR	feedstock hemicellulose content: Hem_F
volatile matter: VM	feedstock lignin content: Lig_F
ash content: AC	feedstock particle size: PS_F
fixed carbon: FC	Random Forest Imputer: RFI
Machine learning: ML	K-Nearest Neighbours Imputer: KNNI
artificial intelligence: AI	SHapley Additive exPlanations: SHAP
Gradient Boosting Regression: GBR	Partial Dependency Plots: PDP
SHapley Additive exPlanations: SHAP	Pearson correlation coefficient: PCC

(*continued on next column*)

(*continued*)

(*continued on next page*)

* Corresponding author.
** Corresponding author.
E-mail addresses: fuy49@cardiff.ac.uk (Y. Fu), jinf2@cardiff.ac.uk (F. Jin).

https://doi.org/10.1016/j.renene.2025.124071
Received 12 January 2025; Received in revised form 21 July 2025; Accepted 25 July 2025
Available online 26 July 2025

(*continued*)

Random Forest: RF	Gradient Boosting Decision Tree: GBDT
Artificial Neural Networks: ANN	eXtreme Gradient Boosting: XGBoost
Preferred Reporting Items for Systematic Reviews and Meta-Analyses: PRISMA	Multiple Linear Regression: MLR
dataset predicting biochar C content: Dt_C	Decision Tree: DT
dataset predicting biochar H content: Dt_H	Support Vector Machine: SVM
dataset predicting biochar O content: Dt_O	k-Nearest Neighbours: KNN
dataset predicting biochar N content: Dt_N	Rough Set Machine Learning: RSML
dataset predicting biochar S content: Dt_S	Linear Regression: LR
dataset predicting biochar P content: Dt_P	Multilayer Perceptron Neural Network: MLP-NN
dataset predicting biochar K content: Dt_K	Adaptive Neuro-Fuzzy Inference System: ANFIS

1. Introduction

The global production of agricultural residues, e.g. sugarcane bagasse, rice stalks, corn stalks, wood waste, is substantial each year, and addressing the reuse and appropriate disposal of these is of paramount importance [1]. For instance, the Food and Agriculture Organisation (FAO) reports a global production of 1208 million tons of corn in 2021, of which 205.87 million tons were burned, resulting in the release of 14.14 kilotons of N_2O and 545.36 kilotons of CH_4 to the atmosphere [2]. This practice contributes significantly to climate change, underscoring the need for sustainable waste management solutions to support global net-zero targets. Consequently, the production of biochar using agricultural residues has become a promising solution to promoting carbon neutrality and waste utilisation [3].

Biochar is a black solid product produced through pyrolysis in an oxygen-limited or anaerobic environment. Under high-temperature conditions above 250 °C [4], agricultural residues undergo a process where water and volatile contents escape, while the lignocellulosic composition gradually decomposes, degrades, and carbonises. This process ultimately yields biochar characterised by rich carbon (C) content and high porosity [5]. Alongside C, biochar comprises primarily hydrogen (H), oxygen (O), and nitrogen (N), and some minor elements such as phosphorus (P) and potassium (K). These elements contribute to the formation of biochar structure and diverse functional groups, thereby influencing various physical and chemical properties of biochar, including aromaticity, cation exchange capacity (CEC), electrical conductivity (EC), and alkalinity [6]. Leveraging these characteristics, agricultural residues derived biochar has gained widespread adoption across multiple environmental domains. Its capacity for carbon sequestration has been demonstrated through both structural characterisation [6] and CO_2 adsorption studies [7]. In the soil system, biochar can improve soil quality [8] and influence soil biota and nutrient dynamics [9]. In addition, it has been used to enhance soil productivity [10] and demonstrated sustainable performance in wastewater treatment applications [11].The elemental composition of biochar plays a critical role in determining its properties and effectiveness for various applications. As the most abundant element in biochar, C forms the backbone of biochar structure, linking with other elements to create functional groups. Inorganic C contributes to biochar's alkalinity and buffering capacity [12], thereby regulating soil pH, while the aromatic structures formed by C are highly effective in adsorbing soil and water pollutants [13]. H in biochar provides active functional groups and plays a critical bridging role in the adsorption of ionisable molecules [6]. A high O content is associated with greater hydrophilicity and porosity, which enhances soil water retention and nutrient holding capacity, neutralises soil acidity, promotes microbial activity, and effectively adsorbs pollutants [14]. Other elements in biochar like N, P, and K directly or indirectly impact soil nutrient levels [15], and activities of soil microorganisms and enzymes [9] in soil health improvement. Therefore, quantifying and optimising the elemental composition of biochar has practical significance in biochar applications. In this context, six elemental components - C, H, O, N, P, and K - were selected for investigation, as they are among the most functionally important and widely reported elements in biochar. C, H, and O form the primary framework of biochar's physical and chemical structure, whereas N, P, and K are essential nutrients influencing soil fertility and microbial activity. This selection enables a holistic evaluation of both structural and agronomic value of biochar across applications such as soil amendment, carbon sequestration, and pollution mitigation.

Studies increasingly reveal significant variations in biochar's elemental composition. It has been reported the elemental composition, such as major (C, H, O) [8] and nutrient elements (N, P, K, Ca, Mg) [16], of biochar derived from different biomass sources vary remarkably under identical pyrolysis conditions. Even within the same plants, Intani et al. [17] observed that maize stalk and cob produced biochar with different properties, while Liu et al. [18] confirmed this variability among different parts of pecan residues.Such differences directly influence the efficacy of biochar applications across various fields. To elucidate the factors underlying these variations in biochar' elemental composition, researchers have examined the influence of pyrolysis conditions, such as highest heating temperature (HHT), residence time (RT), and heating rate (HR). Generally, higher HHT promotes a more complete carbonisation reaction in lignocellulosic biomass, resulting in increased C content and ash, with more volatile materials lost as biogas [19]. Consequently, H, O, and N contents in biochar exhibit negative correlations with HHT, whereas C content shows a positive correlation [20]. Zhang et al. [21] further validated this trend by analysing biochar produced from multiple crop residues at 300–600 °C. Additionally, RT, HR, and biomass particle size exert varying impacts on biochar properties, though drawing precise conclusions remains challenging. However, beyond these controllable pyrolysis conditions, the inherent differences in agricultural residues' type entail relatively uncontrollable disparities in biomass components and compositions, significantly influencing biochar properties. Agricultural residues predominantly exhibit a pronounced lignocellulosic structure, distinguishing them from other feedstocks like manure, human waste and sludge. Wijitkosum [22] observed that Krachid tree, with higher lignin content, exhibited elevated C content (85.78 %), while rice husk, containing lower lignin content, displayed a much lower C content (47.67 %) in their respective biochar products. The reason is that lignin has a complex structure comprising benzene rings and polymers with high thermal stability. Moreover, the original proximate (i.e., volatile matter VM, ash content AC, and fixed carbon FC) and elemental composition of biomass influence biochar's elemental content to varying degrees. Therefore, drawing precise conclusions on how various factors and their interactions influence biochar's elemental composition remains challenging, despite a large number of experimental studies having been reported.

Machine learning (ML) is a subcategory of artificial intelligence (AI) that utilises algorithms to efficiently learn and analyse relationships and patterns within multidimensional data. ML is instrumental in processing both classification and regression problems. Due to the intricate nature of biochar research, the application of ML offers novel perspectives and insights into comprehending and addressing biochar-related challenges [23]. In contrast to traditional statistical methods, ML enables the simultaneous consideration of crucial factors associated with the target variable while uncovering complex linear and non-linear correlations. Various ML models have been developed to predict biochar properties across different feedstocks and pyrolysis conditions. For instance, ensemble models such as Random Forest (RF), Gradient Boosting Decision Tree (GBDT), eXtreme Gradient Boosting (XGBoost), and AdaBoost [24], along with traditional models like Multiple Linear Regression (MLR), Decision Tree (DT), Support Vector Machine (SVM), and k-Nearest Neighbours (KNN) [25], have been applied to predict biochar yield. Zhu et al. [26] specifically employed RF to predict biochar yield

and carbon content. For surface characteristics prediction, Rough Set Machine Learning (RSML) was applied by Ang et al. [27], while Leng et al. [28] used RF and Gradient Boosting Regression (GBR). In predicting the biochar proximate composition, studies have used models such as Linear Regression (LR), AdaBoost, and Boosted Regression Trees (BRT) [29], as well as more complex ones like the Multilayer Perceptron Neural Network (MLP-NN) and Adaptive Neuro-Fuzzy Inference System (ANFIS) [30]. In addition, the prediction of prediction biochar heating value has been addressed using RSML [31], and a variety of regression models [32]. However, there is still limited understanding of specifically how agricultural residues influence the elemental composition of biochar under different pyrolysis conditions and how ML can help to optimise specific elements in biochar for more effective uses. For example, Li et al. [30] used neural networks to predict the C, H, O and N content of biochar, yet incorporated agriculture residues, food waste and poultry litter into the dataset, overlooking the inherent differences between feedstock types. Furthermore, as key components that affect soil fertility, the P and K contents cannot be disregarded. Similarly, Jiang et al. [33] explored deep neural network and light gradient boosting machine for biochar surface area prediction, without incorporating elemental composition as the target. While these studies demonstrated the feasibility of ML techniques, they often lacked broad feedstock diversity, detailed process variables, or interpretable model outputs. Moreover, they typically focused on a single or few output properties. In contrast, interpretable models were developed in this study to predict six key elemental components (C, H, O, N, P, K), using agricultural residues targeted and curated datasets, thus improving both prediction accuracy and practical relevance. Moreover, many previous studies have relied on conventional data preprocessing pipelines, which may overlook the impact of missing values on model performance and generalisability. To address this, a novel imputation strategy was proposed that is tailored to the characteristics of pyrolysis-related datasets, improving data completeness and model robustness. This methodological refinement contributes to a more rigorous application of ML in the biochar domain, ensuring that predictions are based on a more reliable and representative dataset.

In this work, characteristics of agricultural residues and pyrolysis conditions were examined as inputs and developed the GBR models to predict biochar's elemental (C, H, O, N, P, and K) contents. GBR models are adopted in this study due to its superior predictive performance, robustness to multicollinearity, and ability to capture complex nonlinear interactions while maintaining good interpretability through feature importance and SHapley Additive exPlanations (SHAP) analysis (supplementary material). Compared to RF, GBR yields better accuracy by minimizing bias through sequential model improvements [34]. Furthermore, it avoids the "black-box" nature of Artificial Neural Networks (ANN), allowing for clearer interpretation of model outputs—an essential consideration for practical biochar optimisation.

The aim of this study is to quantify the contributions of various factors affecting biochar elemental contents, identify their influencing patterns, and offer guidelines for pyrolysis parameters to optimise biochar production. This comprehensive analysis provides a robust framework for the development of tailored biochar prototypes, advancing understanding of the interplay between feedstock characteristics, pyrolysis conditions, ad biochar elemental composition, ultimately optimising its application across various environmental and agricultural domains. The findings of this study can significantly contribute to improving biochar design for specific applications and enhancing sustainability in agricultural waste management. Additionally, the use of ML models offers a valuable tool for refining biochar production processes, leading to more efficient resource utilisation. While this study primarily focused on lignocellulosic agricultural feedstocks, incorporating a broader spectrum of biomass sources - including industrial and forestry residues - could further enhance model generalisability in future applications.

2. Methodology

2.1. Data collection and preprocessing

The dataset for this study was gathered the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines by searching from Web of Science, Scopus and PubMed with "biochar" or "bio-char," "pyrolysis," "feedstock" or "biomass" in titles, abstracts, and keywords as search terms. After deleting duplicates (Fig. 1), article titles and abstracts were screened based on following criteria: (1) Only studies that employed agricultural residues biomass as the feedstock were included to focus on the most relevant biomass sources in biochar production. Studies using non-agricultural residues feedstocks (e.g., manure, food waste) were excluded; (2) Studies that employed pyrolysis as the biochar production method were selected. Studies using alternative methods (e.g., gasification, hydrothermal carbonisation) were excluded; (3) Studies that focused on biochar that had undergone modifications such as activation or loading of additional chemicals were excluded. This ensured that the biochar in the dataset reflected its natural characteristics from pyrolysis, without external alterations that could affect its elemental composition. Papers that met the criteria were selected as the source of the initial datasets after a rigorous screening process of the full text. Data from text were recorded manually, while those from figures were collected by plotdigitizer. Ultimately, 38 articles were selected with 332 instances in the initial dataset.

The initial dataset comprised 34 variables in the data collection process, including those pertaining to the paper meta information, feedstock characteristics, pyrolysis parameters, and experimental conditions (Table 1), as well as elemental composition of the biochar produced. The values and information of these variations are collected as complete as possible by going through all the content in the 38 study papers.

However, not all the variables listed in Table 1 are used as features for further ML. First, variables related to paper meta information were excluded, as they are not applicable for targeted numerical regression modeling. Second, features with a high proportion of missing values were filtered out to maintain dataset quality and model reliability. Specifically, any variable with more than 30 % missing data in sub-datasets was discarded. This threshold is commonly adopted in related studies [30]. Features discarded for this reason are indicated with an asterisk in Table S1.

High missingness in some features is likely attributable to practical and methodological limitations in biochar studies. For instance, extractives content is rarely reported independently, as it refer to the rest soluble part in biomass except for lignin, cellulose, and hemicellulose [35,36]. So, it is often omitted in studies that report those primary constituents. S content in biomass is often very low (e.g., 0.18–0.85 % in wood chips; [37]), requiring specialized instrumentation for accurate measurement, which may explain its absence in many studies. While these features may hold potential value for understanding biochar composition, their limited availability in the literature constrains their current utility in data-driven modeling.

Despite the need to exclude some features, their scientific relevance is not necessarily diminished. For example, carrier gas flow rate has been shown to affect biochar yield, particularly in fluidized bed and TGA pyrolysis systems [38]. Future expansions of the dataset with more comprehensive reporting may enable the inclusion of such features, enhancing model interpretability and generalisability.

Following feature selection, the left features in the dataset were further divided into two categories: (1) feedstock characteristics (2) pyrolysis process parameters. Subsequently, six separate datasets were created for each element of biochar (Dt_C, Dt_H, Dt_O, Dt_N, Dt_P and Dt_K) (Table 2). Features with a missing rate of less than 30 % were retained in each dataset to maintain the quality of the datasets. To reduce the negative impact of outliers on the model robustness, the Isolation Forest (IF, contamination = 0.1) in the scikit-learn library was

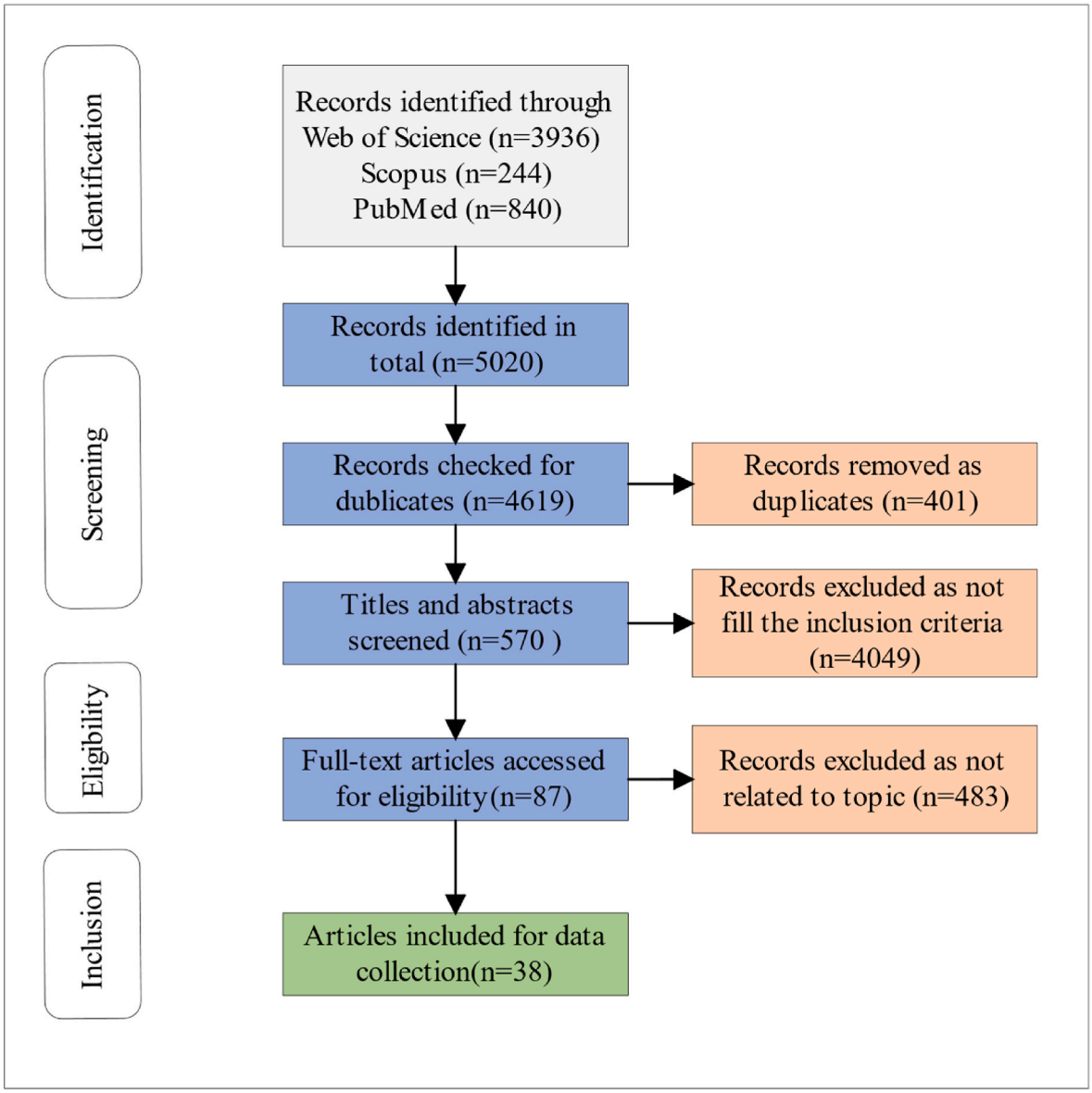


Fig. 1. Literature searching and data collection flowchart.

Table 1
Features and targets in the initial dataset with units and abbreviations.

Features		Targets
Feedstock characteristics (%)	Pyrolysis parameters	Biochar's elemental content (%)
Moisture[a], Fixed Carbon (FC_F), Volatile Matter (VM_F), Ash Content (AC_F), C (C_F), H (H_F), O (O_F), N (N_F), P (P_F), K (K_F), Cellulose (Cel_F), Hemicellulose (Hem_F), Lignin (Lig_F), Extractive[a]	Drying Temperature (°C), carrier gas[a], carrier gas flow rate (m^3/min)[a], Residence Time (min) (RT), Heating Rate (°C/min) (HR), Highest Heating Temperature (°C) (HHT), Feedstock Particle Size (mm) (PS_F), Vessel Dimensions (cm)[a], Vessel Volume (L)[a]	C (C_B), H (H_B), O (O_B), N (N_B), P (P_B), K (K_B)

[a] Features not included in the further 6 datasets because of data missing proportion higher than 30 % in all datasets.

employed to remove them from the datasets.

To address the missing values, a novel feature-oriented imputation method was applied in this study, which strategically aligns imputation techniques with the inherent characteristics of different feature types. Specifically, we categorized the features into two groups based on their statistical information and practical implications: (i) feedstock properties, includes FC_F, VM_F, AC_F, C_F, H_F, O_F, N_F, P_F, K_F, Cel_F, Hem_F, Lig_F, which are continuous variables generally following a normal distribution, and (ii) pyrolysis parameters, includes RT, HR, HHT, PS_F, which are discrete by nature and often constrained by experimental design. For features reflecting the nature of the feedstock, Random Forest Imputer (RFI) was used. RFI is an integrated learning technique based on multiple decision trees through bootstrapping and random feature selection and aggregates their outputs to predict missing values. This algorithm can effectively deal with complex nonlinear relationships among features [39]. Here, the applied RFI utilised the correlation between feedstock features to predict missing values, which maintained the intrinsic structure and dependencies of the data, thereby improving the prediction accuracy of the subsequent models. On the other hand, for discrete features such as pyrolysis parameters artificially

Table 2
Pre-processed datasets information. The meaning of the abbreviations are shown in Table 1.

Dataset	Number of Instances	Number of Features	Feedstock Features	Pyrolysis Features	Target
Dt_C	229	13	FC_F, VM_F, AC_F, C_F, H_F, O_F, N_F, Cel_F, Hem_F, Lig_F	RT, HR, HHT	C_B
Dt_H	284	14	FC_F, VM_F, AC_F, C_F, H_F, O_F, N_F, Cel_F, Hem_F, Lig_F	RT, HR, HHT, PS_F	H_B
Dt_O	270	14	FC_F, VM_F, AC_F, C_F, H_F, O_F, N_F, Cel_F, Hem_F, Lig_F	RT, HR, HHT, PS_F	O_B
Dt_N	258	14	FC_F, VM_F, AC_F, C_F, H_F, O_F, N_F, Cel_F, Hem_F, Lig_F	RT, HR, HHT, PS_F	N_B
Dt_P	75	14	FC_F, VM_F, AC_F, C_F, H_F, O_F, N_F, P_F, K_F, Cel_F, Hem_F, Lig_F	RT, HHT	P_B
Dt_K	98	15	FC_F, VM_F, AC_F, C_F, H_F, O_F, N_F, P_F, K_F, Cel_F, Hem_F, Lig_F	RT, HR, HHT	K_B

set based on the experimental design, the K-Nearest Neighbours Imputer (KNNI) was selected for data imputation. The KNNI works by identifying the K nearest (n_neighbors = 5) known points to the missing values in the data space, and then estimating the missing values based on the values of these neighbours. This method is well-suited for analysing discrete and relatively independent features [40]. This approach is particularly effective for preserving realistic, experimentally observed values, and avoids the introduction of nonphysical or excessively granular data points that could arise from inappropriate imputation of inherently discrete variables.

Previous studies in biochar research have often overlooked the heterogeneity of feature types during imputation. For example, several works employed a single imputation algorithm - such as RFI [41] or KNNI [42] - across all variables without differentiating between continuous and discrete data. Such approaches could risk introducing bias or reducing the interpretability of model outputs.

This novel feature-oriented imputation strategy employed in this study offers a more tailored, statistically grounded strategy that enhances both the fidelity and quality of the imputed dataset and the robustness of the subsequent ML models. This method ensures that the data preprocessing process aligns with both the statistical nature and practical constraints of the input features, thereby laying a more solid foundation for downstream prediction tasks.

Finally, each dataset was randomly split into training and testing datasets in a ratio of 8:2. To ensure that each model received the same dataset, each dataset was split only once and then used in turn to train the models. The StandardScaler from the scikit-learn library was employed to perform feature standardisation for all the datasets, facilitating better interpretation and visualisation of ML models.

2.2. Model development and interpretation

In this study, GBR was employed to predict the elemental composition of biochar. GBR is a tree-based model refers to a class of models based on decision trees in ML models. This type of model is suitable for small-scale datasets, has stronger interpretability, and is suitable for high-dimensional data [43]. GBR is an integrated learning algorithm that employs decision trees as the base learners. Each of these decision trees is constructed based on a correction of the residuals of the previous decision trees to minimise the loss function (Eq. (1)) [44]. This approach improves the model's generalisation ability, reducing overfitting and boosting its performance on unseen data.

$$\sum_{i=1}^{N} L((y_i, \hat{y}_{m-1}) + f_m(X_i)) \tag{1}$$

where N is the instance number of the training dataset, y_i is the actual output of the instance x_i, $\hat{y}$ is the predicted output from $(m-1)^{th}$ base learner, $f_m(x_i)$ refers to the predicted value of the sample X_i from the m^{th} learner.

A grid search was employed with 5-fold cross validation to identify the optimal hyperparameters combination for each model. The n_estimators was set from 1 to 50 for all the GBR models, which regulates the number of decision trees. The learning rate from 0.01 to 0.3 was used which controls the contribution of each decision tree in an iteration.

After ML model development, the performance of models was assessed by various techniques including Feature Importance Analysis, SHapley Additive exPlanations (SHAP), 1-way Partial Dependency Plots (1-way PDP) and 2-way Partial Dependency Plots (2-way PDP). Feature Importance Analysis assesses the importance of each feature by calculating its contribution to the prediction, providing an intuitive view for identifying the features that have the greatest impact on the prediction results. SHAP value analysis is an advanced feature interpretation method based on game theory, aiming to quantify the contribution of each feature to each prediction [45]. It provides a detailed view of the influence and direction of action of each input feature to the model. 1-way PDP is used to visualise the impact of a feature on the predicted target while other features are held constant, while 2-way PDP assesses the impact of two interacting features on the predicted results. The combined use of these methods enhances the explainability and transparency of the model, which is crucial to ensure the impartiality and scientific validity of the model. And the use of these interpretation methods will identify key features, elucidate the relationship between each feature and the target, and optimise biochar pyrolysis parameters.

2.3. Performance evaluation

After constructing all the models, their training and generalisation performance is evaluated by the coefficient of determination (R^2) and the root mean squared error (RMSE). The R^2 statistic (Eq. (2)) is a measure relative to the model (Y = X), which explains the degree of error between predicted and true values. It ranges from 0 to 1, with values closer to 1 indicating less error between the predicted and true values and so a more valid model. The RMSE is a measure of the prediction error and how well the model fits the observations (Eq. (3)), so the smaller the RMSE, the better the model performs.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{2}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{3}$$

where y_i is the real target value of the ith sample, $\hat{y}_i$ is the predicted target value, and $\bar{y}$ is the average of the actual value of the target in all n samples.

3. Results and discussion

3.1. Dataset imputation and statistical analysis

For each dataset (see Table 2), features with missing values are imputed by the feature-oriented imputation method as detailed in Section 2.1 (Fig. 2). For instance, in Dt_C, features FC_F, VM_F, Cel_F, Hem_F, Lig_F, O_F, RT and HR have missing data with proportions of

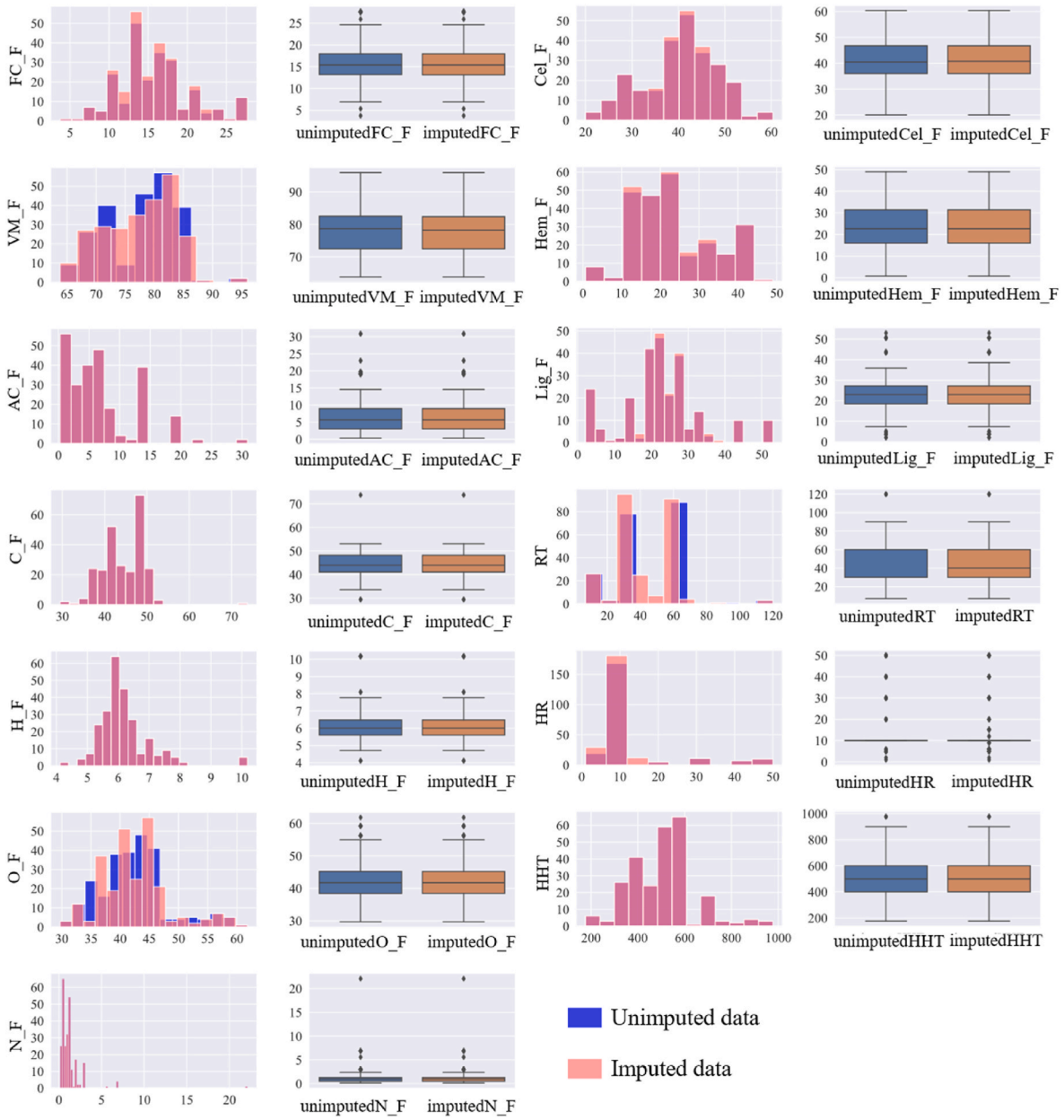


Fig. 2. Comparison of statistical distributions of features in the dataset Dt_C before and after imputation. Dt_C: dataset for predicting the biochar's carbon content. The back layer blue bars show the statistical distributions of the original unimputed data, and transparent light red bars visualise those of the imputed features. The overlapped regions are shown as pink colour.

9.22 %, 9.22 %, 2.84 %, 2.84 %, 2.84, 1.42 %, 20.21 % and 13.12 %, respectively (Table S1). Statistical analysis illustrated in Fig. 2 reveals some degree of skewness in features such as RT and HR, and the data for these features tend to cluster around a few values. For example, the HR of 10 °C/min is commonly used in pyrolysis studies, and RT is frequently adopted as 30 or 60 min. This occurs because these parameters are experimental variables that researchers chose empirically or followed from previous studies. Moreover, these features typically consist of integer data. Consequently, employing RF for imputation is unsuitable for these parameters as it could introduce noise and generate unrealistic data points, including decimal values. Therefore, KNN imputation method was employed as it effectively preserved the original range and distribution patterns of the data, ensuring that no additional skewness was introduced post-imputation (Fig. 2). For features such as Cel_F, Hem_F, and Lig_F, the bar charts display significant overlaps between the original and imputed data, indicating that RF imputation is effective without introduction of biases or distortions. This is because these feedstock properties are inherently more variable and less influenced by experimental settings, and the RF method captured the correlations between these features successfully.

Fig. 2 also illustrates the range of properties inherent to agricultural residues, including elemental composition, proximate components, and lignocellulosic constituents. Specifically, it demonstrates that agricultural residues, rich in cellulose, hemicellulose and lignin but with minimal non-combustible minerals, typically exhibit lower ash content and higher volatile matter content, which is also noted by Aller [46]. The high volatile matter content is due to the propensity of cellulose and hemicellulose to degrade and decompose readily during the pyrolysis process [47].

The violin plots in Fig. 3(a) depicts the distributions of the prediction targets in the 7 datasets, The shaded part shows the probability density distribution of the data. The wide part indicates that the data points are

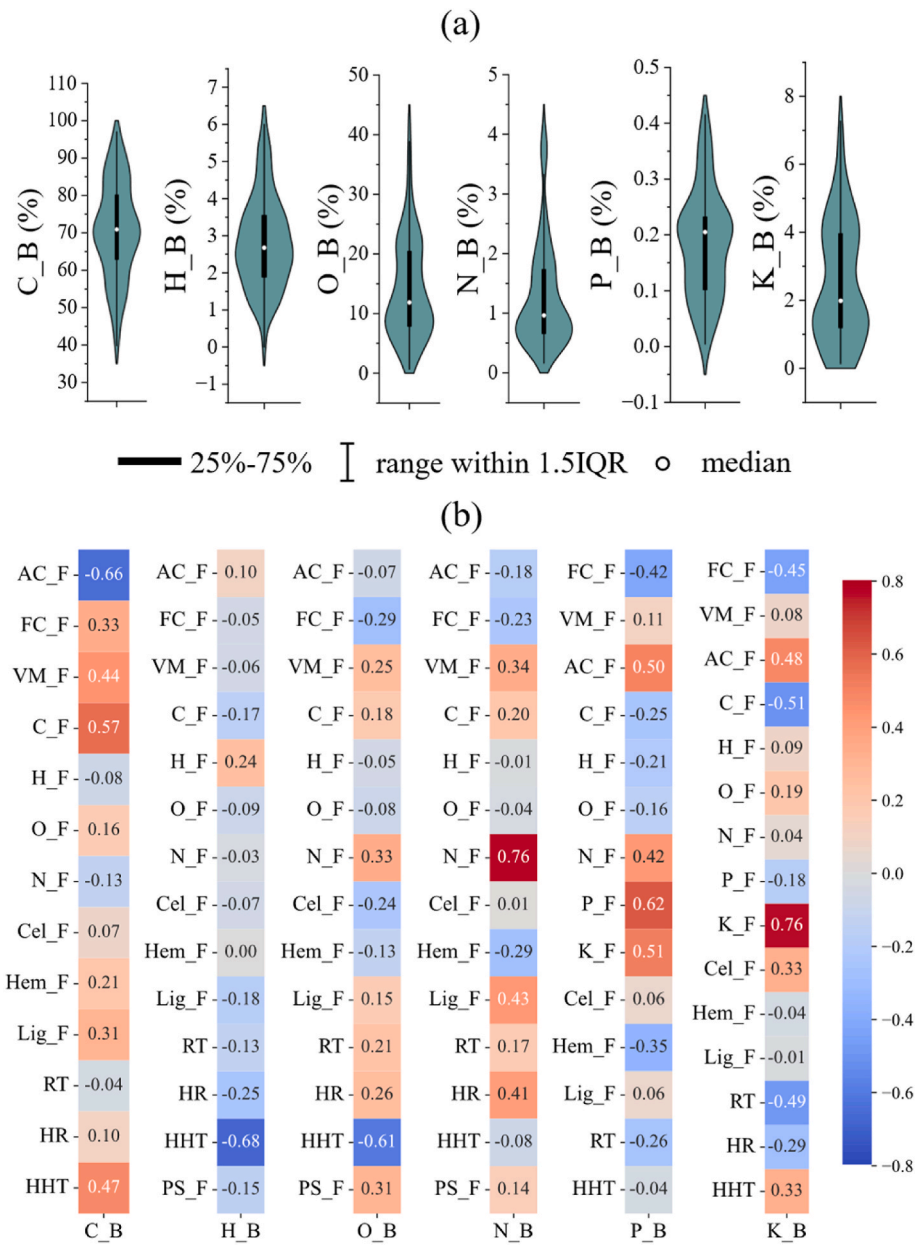


Fig. 3. Distribution of the targets and Pearson correlation matrix between targets and features in each dataset.

concentrated, and the narrow part indicates that the data points are sparse. The black bars in the middle are box plots, showing the mean and the first and third quartiles of the data. It shows that the contents of C, H, O, N, P, and K are 39.78 %–97.04 %, 0 %–6.35 %, 0.65 %–42.56 %, 0.16 %–4.21 %, 0.0043 %–0.4152 %, and 0.14 %–0.27 %, respectively. The C content shows a wide distribution with significant occurrences at extremely low or high percentages, indicating its large variability across different biochar samples. For O, N, and K, a pronounced peak near the lower range suggests their low contents in the majority of samples. The distributions of N and O are clearly skewed towards lower concentrations with a long tail towards higher concentrations, reflecting the heterogeneity of the data.

The linear correlations between biochar's elemental composition and the influencing factor in each dataset are shown in Fig. 3 (b) by the Pearson correlation matrix. The carbon content in the feedstock (C_F) shows the strongest positive correlation (Pearson correlation coefficient (PCC) = 0.57) with the biochar's C content (C_B), whereas the ash content of the feedstock (AC_F) negatively correlated with it (PCC = −0.66). HHT exhibits a significant negative correlation with both H (PCC = −0.68) and O (PCC = −0.61) contents of biochar. The elements of N, P, and K in biochar are primarily derived from those in feedstocks and hence significantly correlated with N_F (PCC = 0.76), P_F (PCC = 0.62), and K_F (PCC = 0.76), respectively.

These observations suggest complex interactions and variability between the composition of agricultural residue-derived biochar, the characteristics of the feedstock and the pyrolysis parameters. Understanding these complexities and predicting the elemental composition of biochar is essential for optimising its application in the environment.

3.2. *Model development and performance evaluation*

GBR models were developed for each pre-processed dataset. In the hyperparameter tunning process, the grid search with a 5-fold cross-validation method was used to find the optimal model parameters. Finally, the optimised models achieving highest R^2 in this grid search process were chosen for subsequent prediction and interpretations.

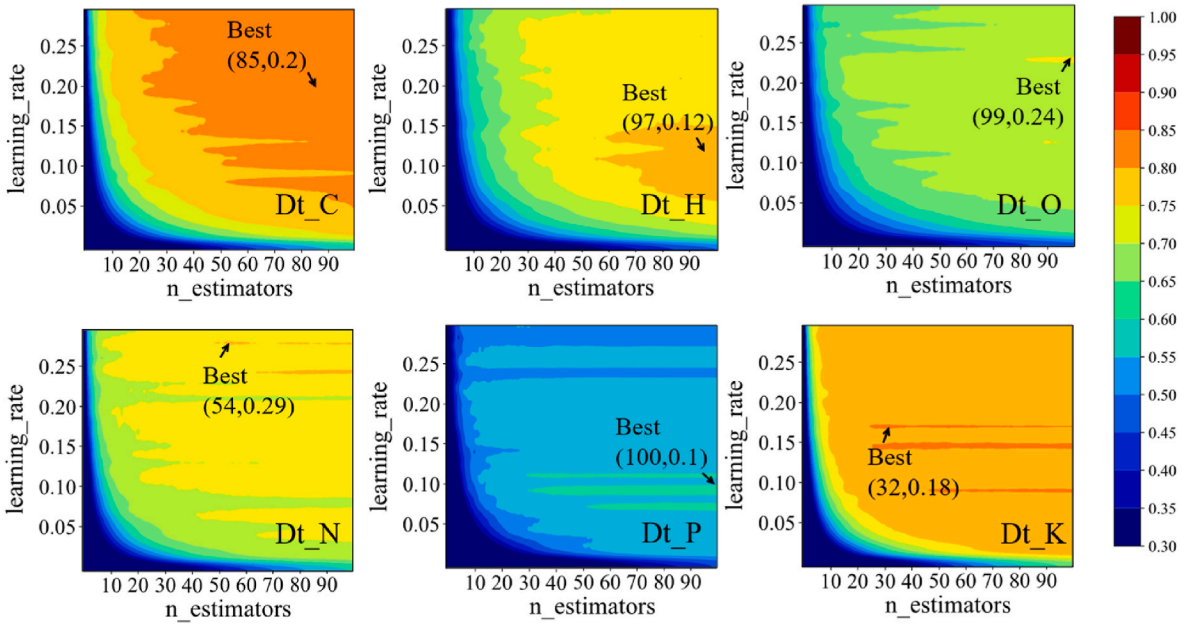


Fig. 4. Hyperparameter tuning heatmap of GBR model on each dataset for predicting biochar's C, H, O, N, P and K contents.

Fig. 4 illustrates the heat map of each dataset in the grid search, the arrows point to the locations of best-performing model parameters.

Fig. 5 presents the scatter plots showing the predicting performance of the GBR models for each element's content. Each panel displays the scatter plots of the predicted versus actual values, with separate markers denoting training and testing data. The linear fits, along with the 95 % confidence and prediction intervals, are depicted to assess the model accuracy and prediction reliability. The detailed R^2 and RMSE values for each model are listed in each plot. The training scores reflect the models' ability to learn from the training datasets, whereas testing scores assess their generalisation capabilities on the testing datasets. For instance, the GBR model demonstrated excellent predictive performance for biochar C content with R^2 of 0.9713 in training and 0.9088 in testing. The relatively low RMSE of 2.1572 in training and 4.0614 in testing suggests a robust model with strong predictive capabilities and reasonable generalisability to unseen data. The overall performance of the GBR model across all elemental content predictions, i.e., C, H, O, N, P, K, demonstrates its strong robustness and reliability as a predictive tool. In particular, the high R^2 values and relatively low RMSE across both training and testing datasets indicate that the model is highly accurate and generalises well to unseen data. This suggests that such models can serve as valuable estimators for researchers seeking to predict biochar composition before conducting actual experiments. By providing high-accuracy predictions, these models have the potential to guide experimental designs, optimise conditions, and reduce the need for extensive trial-and-error in laboratory settings.

3.3. Feature importance and SHAP analysis

The feature importance values were calculated from the attributes of GBR models shown above to evaluate the contribution of each feature in predicting the target. Fig. 6 presents a model-based feature importance analysis for predicting the elemental composition of biochar. It comprises 6 sub-figures, each corresponding to a different target predicted from the best performing GBR model and highlighting the relative importance of pyrolysis parameters and feedstock characteristics.

The results in Fig. 6 show that in most cases, feedstock characteristics have a more substantial impact in determining the elemental composition of biochar. This could be attributed to the fact that the most of elements, particularly N, P and K, are directly derived from the biomass and are incorporated into the biochar in forms that are relatively stable and less affected by the pyrolysis process. These elements are typically bound in the feedstock and do not undergo significant loss or transformation during pyrolysis, making feedstock characteristics the dominant factor in predicting their presence in the final biochar [39]. In fact, elements such as N, P, and K show exceptionally high feature importance values, with their contributions being 93.5 %, 95.3 %, and 97.0 %, respectively, as indicated by our models. However, the contents of H exhibit a different pattern where pyrolysis parameters, particularly HHT, play a more influential role with an overall 67 % contribution, which is consistent with the PCC values shown in Fig. 2. This aligns with the findings in Li et al. [30] that thermal conditions can alter the chemical structure and composition of the resulting biochar, particularly affecting more volatile elements such as H and O under high temperatures. At higher temperatures, the decomposition of biomass components releases volatile compounds, leading to a reduction in H and O content.

This comparison between the influences of feedstock and pyrolysis illustrates the complex interplay between the inherent characteristics of feedstocks and the pyrolysis treatment during biochar production. Feedstock's characteristics play a key role in determining the concentration of essential elements such as N, P, and K, which are critical for biochar applications in soil fertility and nutrient cycling. On the other hand, pyrolysis conditions, especially temperature, are crucial for adjusting the biochar's properties related to volatile elements, such as H and O, that influence biochar's role in carbon sequestration and pollutant adsorption.

These findings underscore the importance of selecting the appropriate agricultural residues for producing biochar with specific elemental compositions. The control of pyrolysis conditions can further fine-tune biochar's characteristics, depending on the desired application, such as nutrient-rich soil amendments or high-carbon biochar for carbon sequestration.

Fig. 7 illustrates the results of the SHAP analysis, providing a more nuanced perspective on how various features influence the model's predictions, the detailed mathematical mechanisms of SHAP can be found in supplementary material. The magnitude of the SHAP value signifies the extent of the feature's influence on the target, with larger absolute values indicating a greater impact. The sign of the SHAP value indicates whether the influence is positive or negative. Each data point

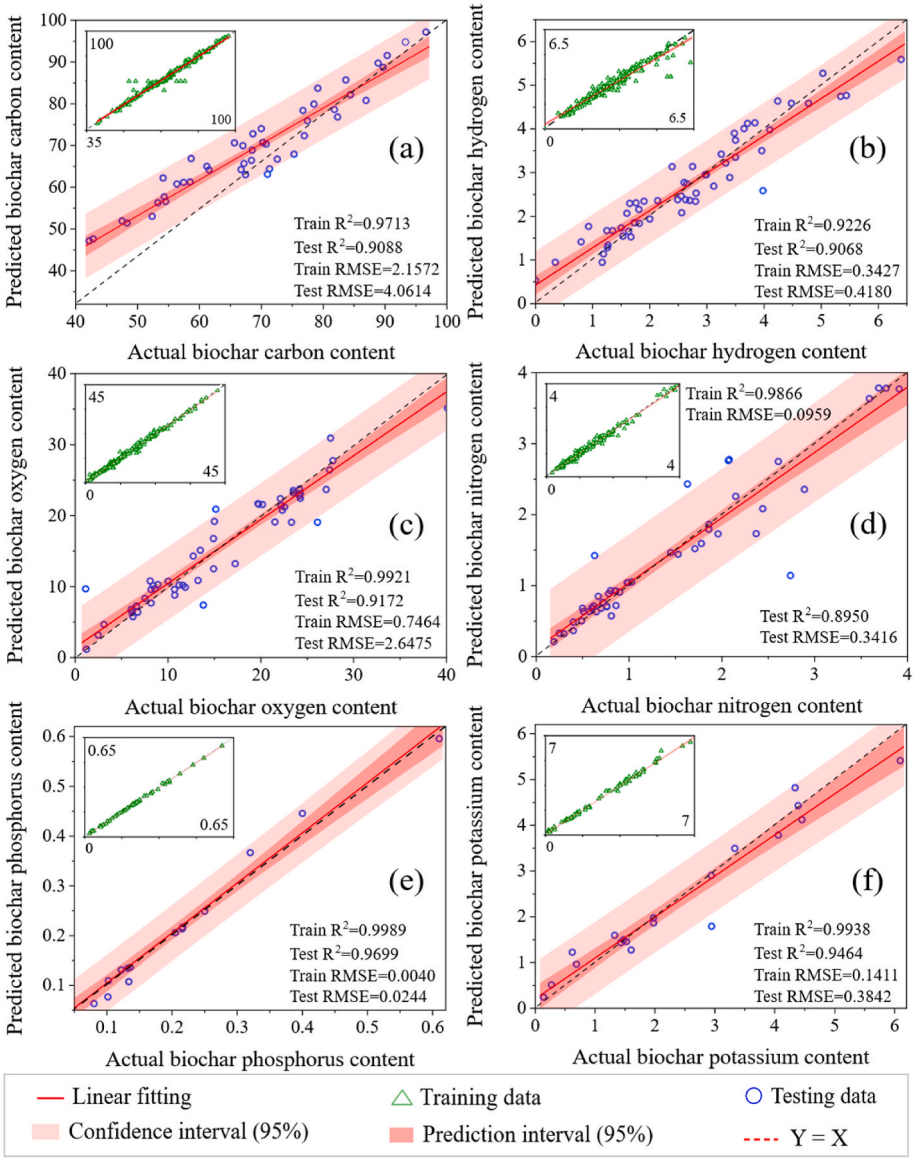


Fig. 5. Best-performing GBR models for each dataset in predicting biochar elemental compositions. (a) C_B; (b) H_B; (c) O_B; (d) N_B; (e) P_B; (f) K_B.

in the dataset is represented by the horizontally distributed point, where the color denotes the magnitude of the data point's value. A dense clustering of points at varying locations on the SHAP value scale reflects the variability in the influence that specific features exert on the model output. Tightly clustered points suggest that a feature consistently affects predictions, whereas widely spread points indicate significant variability in impacts across observations.

The feature importance analysis illustrated in Fig. 6 compares the overall importance of each feature, while the SHAP analysis provides deeper insights by revealing how individual data points contribute to the predicted elemental composition of biochar. For example, features such as higher AC_F exhibit lower SHAP values (Fig. 7(a)) when it has the highest feature importance as a whole (Fig. 6(a)) in predicting biochar C content. This quantification allows for a deeper understanding of which features are most influential in shaping the model's outputs and which features may have marginal effects. These insights are crucial for practitioners who aim to optimise pyrolysis conditions for specific biochar properties.

Additionally, the variability in SHAP values across the dataset suggests that certain features may have differing effects depending on the input data or experimental conditions. For example, higher HHT will negatively decrease the H content in biochar, and lower HHT will promote higher H retention in biochar. Combining Figs. 6 and 7 offers a comprehensive analysis of the influencing factors and their contributions in predicting the elemental composition of biochar and in tailoring the expected biochar in future experiments.

Fig. 6(a) suggests that HR and RT contribute minimally to predicting the biochar C content. This is because the lignocelluloses especially hemicellulose and cellulose in agricultural biomass are easily degraded in relatively low pyrolysis temperatures. Keiluweit et al. [5] studied the dynamic change of plant biomass in pyrolysis and found that the cellulose and hemicellulose begin to volatilise and decompose at ~200 °C, with the carbonisation process proceeding more rapidly than other types of feedstocks, such as manure and sludge. The top three important influencing factors are AC_F, HHT, and C_F (Figs. 6(a) and 7(a)), showing that higher temperatures and higher feedstock C content lead to higher C content in the biochar, while feedstock ash content is negatively correlated.

Similarly, the HHT has dominating contributions in both biochar H and O content predictions (Fig. 6(b) and (c)), and there are clear negative correlations between HHT and H or O (Fig. 7(b) and (c)). The other two most important factors for biochar H content retention are H_F and

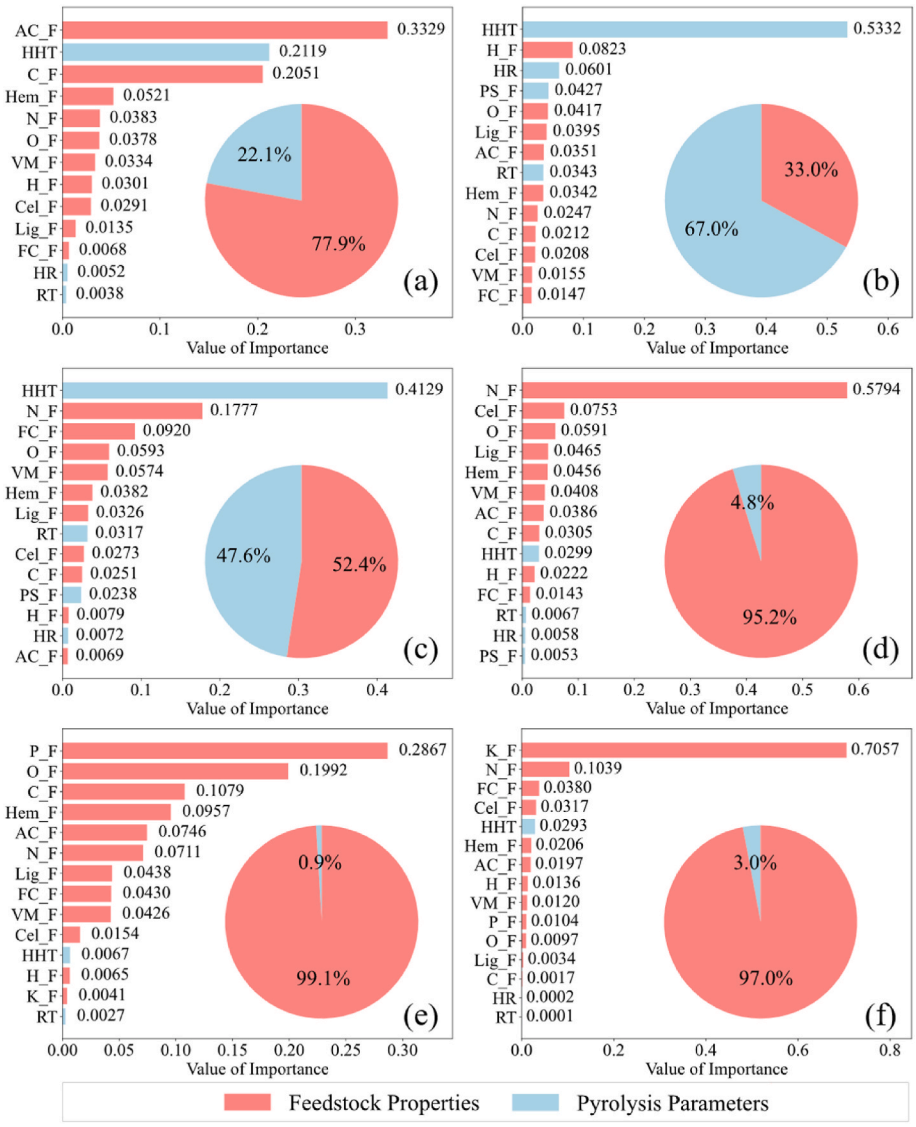


Fig. 6. GBR model-based feature importance analysis in predicting biochar's elemental composition. (a) Dt_C; (b) Dt_H; (c) Dt_O; (d) Dt_N; (e) Dt_P; (f) Dt_K.

HR, which show positive and negative influences individually. For biochar O content, the N content and fixed carbon content in feedstocks are the other two critical factors. N_F has a positive influence on biochar O content, while FC_F has a negative influence. Normally, the H element in biomass has relatively lower stabilisation than the O element and escapes in the gas phase in pyrolysis [48]. Hence, lower HR is better for retaining more H in biochar. The strong correlation between N_F and biochar O content may be related to the formation of the nitrate functional groups [49] which is supported by the strong correlation between in O_F and biochar N content in Fig. 6(d). Unfortunately, there are no clear trends detected in O_F in Fig. 7(d) may be because of limited data points and narrow data distribution.

For predicting more N, P, and K content in biochar, the results from Fig. 6(d)–(f) and Fig. 7(d)–(f) show that the N, P, and K content in biomass feedstocks are most important factors, and pyrolysis parameters have relatively lower influence and hence choosing the appropriate feedstock is key for developing nutrient-rich biochar for soil applications.

By analysing the characteristics of the feedstock and the conditions of the pyrolysis process, these predictive models will enable more precise control of biochar's elemental composition and more efficient use of biochar in different applications. This approach not only maximises the environmental benefits of biochar, but also improves the efficiency of biochar production and application processes.

3.4. *Pyrolysis parameters optimisation*

Fig. 8 presents the 1-way PDP for HHT which is the most influential pyrolysis parameter across different models and for most elements (see Fig. 6). It should be noted that the trends observed could be related to the data density (shown as the ticks on the x-axis). In some instances, despite the absence of data, changes in trends are still observed because the model predictions are based on the overall data distribution learned during the training phase. Therefore, the model is capable of making informed estimates for data points that were not directly observed during training, allowing it to provide predictions for intervals within the feature values where data might be sparse or completely absent. For biochar's C content, an increase in HHT initially has a positive effect, peaking at ~600 °C, suggesting thermal decomposition or volatilisation at higher temperatures. In contrast, increasing HHT from 300 °C to 800 °C significantly reduces the H content. Fig. 8 (c) and 8 (d) exhibit many peaks and troughs, indicating that O and N contents are highly sensitive to HHT, with certain temperatures favoring O and N retention or loss. Similar patterns also exist for P, and K contents, thought the

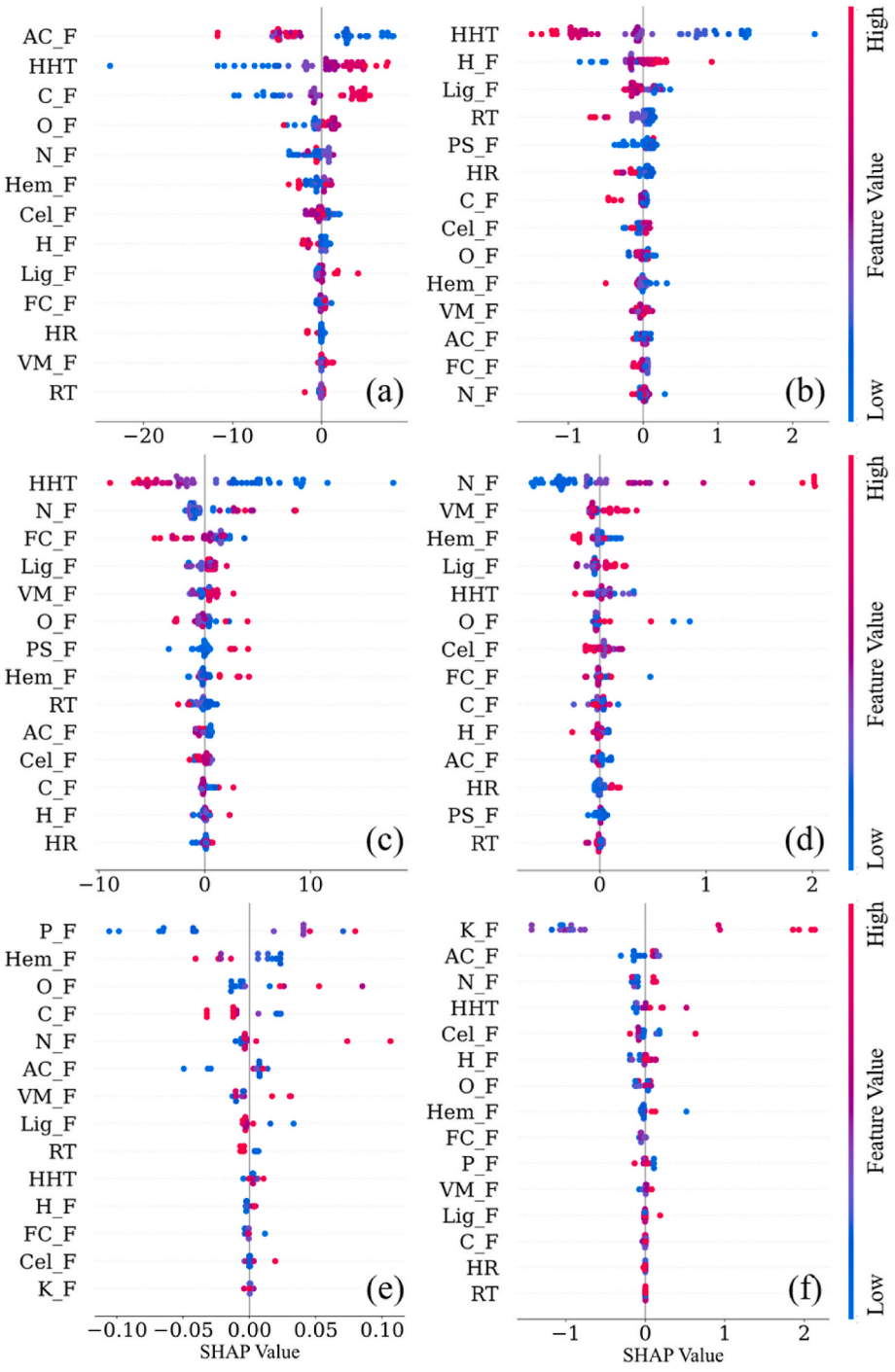


Fig. 7. GBR Model-based SHAP analysis in predicting biochar's elemental composition. (a) Dt_C; (b) Dt_H; (c) Dt_O; (d) Dt_N; (e) Dt_P; (f) Dt_K.

trendlines are smoother indicating their lower sensitivity to HHT.

The selected 2-way PDPs presented in Fig. 9 explore the interactions between pyrolysis parameters and reveal the variations of biochar's element contents with two different pyrolysis parameter variables. These plots utilize contour lines to demarcate zones of varying colors for enhanced clarity. The numbers marked on these contour lines represent the threshold values of the predicted target variables, which aid in identifying the conditions favorable for maximising specific elemental contents in biochar. For C_B, at shorter RT, an increase in HHT results in higher C content. As RT increases, the positive effect of HHT on C_B becomes more pronounced. This indicates that to maximize C_B, a shorter RT should be paired with a higher HHT, whereas with a longer RT, the required HHT can be reduced. Conversely, for H_B, the interaction between HHT and RT shows the opposite trend. For P_B, an increase in RT consistently decreases P_B. Thus, at appropriate temperatures, a shorter RT is preferable to maximize biochar phosphorus content.

The optimal conditions were derived to obtain the maximum contents of various elements in biochar (Table 3). The selection and range of these parameters were determined based on the heatmap and contour lines depicted in the 2-way PDP in Fig. 9. It should be noted that Table 3 only listed the optimal ranges of parameters that exert consistent impacts on the targeted elements from the PDP plots to ensure clarity and precision in the recommendations.

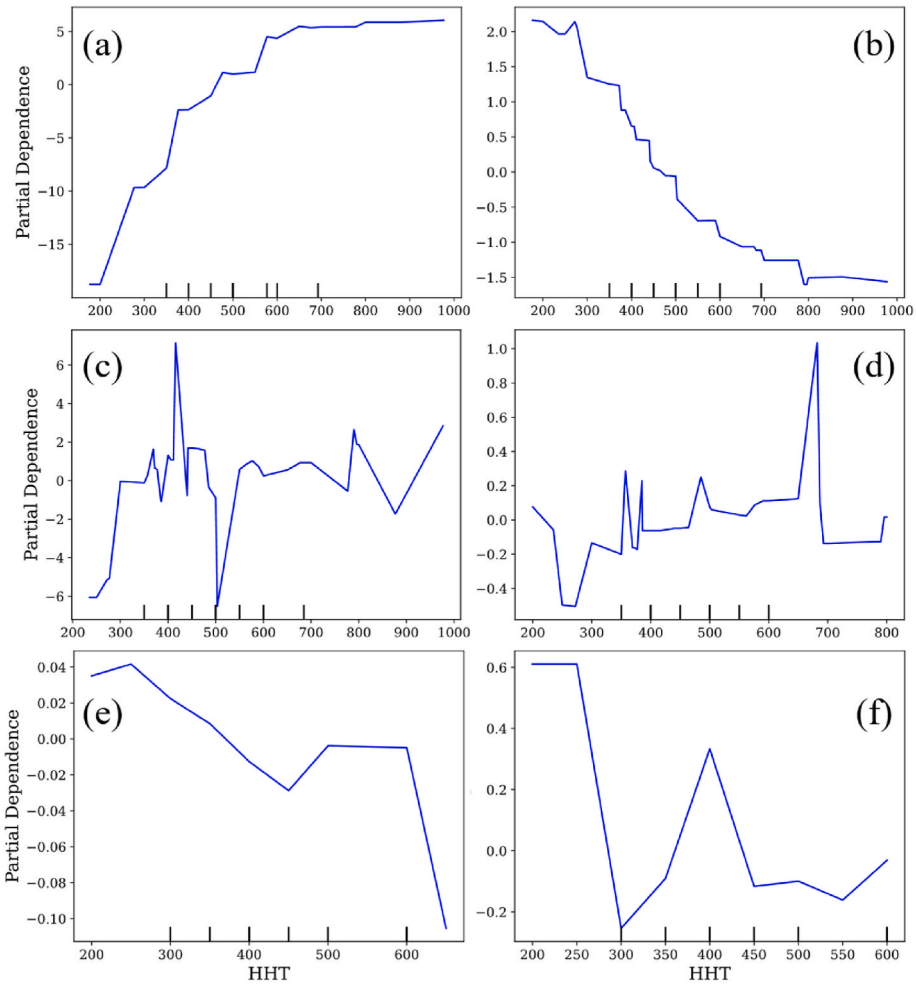


Fig. 8. One-way PDP for HHT in each dataset in predicting biochar elemental composition content. (a) Dt_C; (b) Dt_H; (c) Dt_O; (d) Dt_N; (e) Dt_P; (f) Dt_K.

The identification of optimal parameter ranges not only provides valuable guidance for producing biochar with desired elemental profiles but also aids biochar producers in achieving target properties without the need for extensive experimental trials. This approach ensures that biochar production is both efficient and aligned with specific environmental and agricultural needs. For instance, biochar with enhanced C content can sequester and store more carbon in soils, contributing to soil health and climate change mitigation. Biochar with higher H and O contents could possess more surface functional groups, which can improve soil fertility by enhancing nutrient retention and water holding capacity [14]. Biochar rich in N, P, and K is particularly beneficial for enriching the nutrient content of agricultural soils, supporting the growth of nutrient-demanding crops such as corns, onions, beans, and potatoes [6].

PDP analysis elucidates the relationships between pyrolysis parameters (HHT, RT, HR) and the elemental composition (C, H, O, N, P, K) of biochar derived from agricultural residues. These insights not only enhance the understanding of process–property linkages but also provide a practical foundation for optimising pyrolysis conditions to target specific biochar compositions.

By integrating feedstock properties and controllable pyrolysis settings, the developed GBR models can serve as predictive tools to guide experimental design prior to empirical testing. Practitioners can utilize the model to simulate expected elemental compositions of biochar produced from specific agricultural residues under varying pyrolysis conditions, thereby enabling more efficient planning of experiments and resource allocation.

A pertinent case is the producing biochar from Brewer's spent grain (BSG) reported by Ref. [50]. Their study demonstrated that pyrolysis temperature significantly influences the biochar's properties, affecting its suitability for energy applications and plant growth enhancement. Applying our model to such feedstocks allows researchers to predict the elemental composition outcomes across different pyrolysis regimes, supporting the selection of conditions that optimise energy content (via high C concentration) or plant growth potential (via elevated N, P and K levels).

In this way, the proposed GBR-based prediction framework not only facilitates data-driven insights but also provides a decision-support mechanism for tailoring biochar production to specific application needs - whether for energy recovery, carbon sequestration, or soil fertility enhancement.

4. Conclusions

This study evaluates the efficacy of GBR machine learning model in predicting the elemental composition of biochar derived from agricultural residues, specifically focusing on C, H, O, N, P, and K contents. The results shows that the GBR model achieved robust preformation in both learning and generalisation stages with training R^2 from 0.9226 to 0.9989 and testing R^2 from 0.8950 to 0.9699. The interactions between input features and each target were extensively analyzed using feature importance analysis and SHAP methods. Based on the insights derived from 1-way and 2-way PDPs, optimal pyrolysis parameters were identified, facilitating production of biochar with desirable elemental

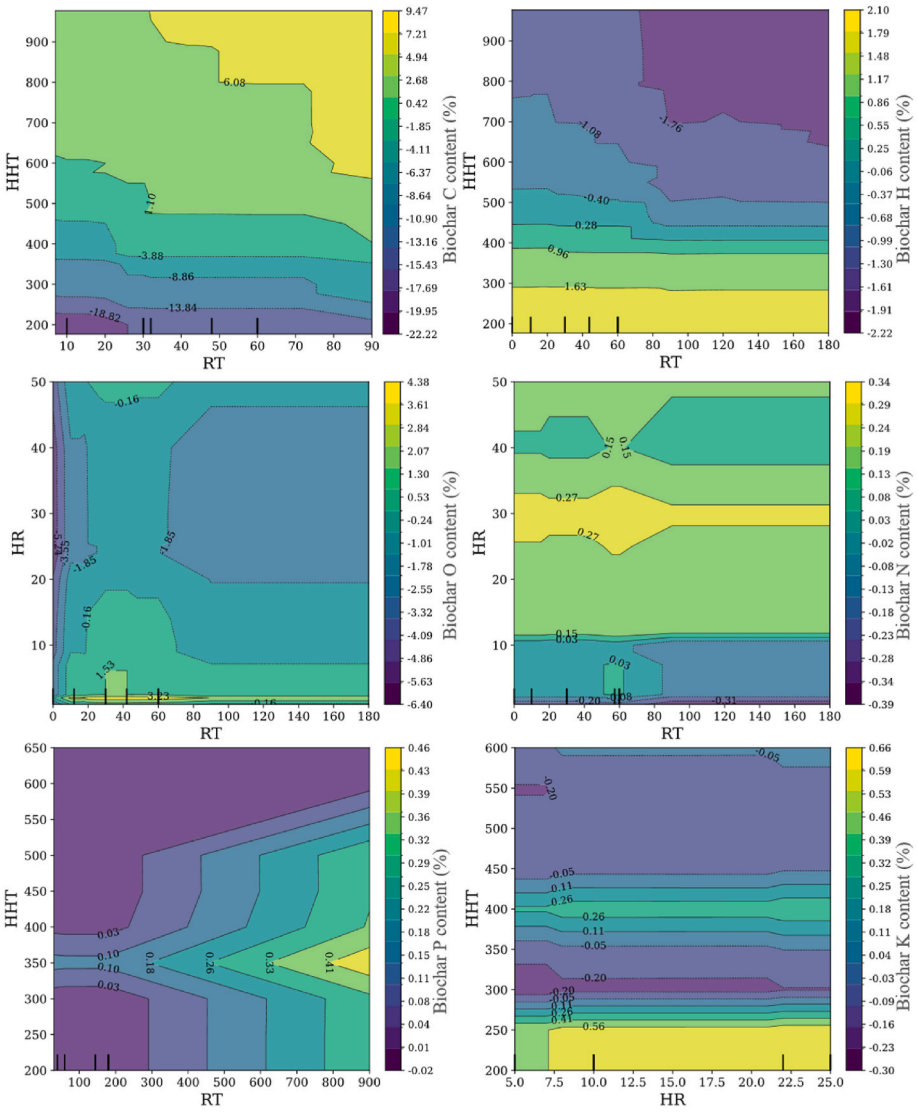


Fig. 9. Selected two-way PDPs showing the optimal pyrolysis parameters to maximize specific elements in biochar.

Table 3
Optimal pyrolysis parameters from 2-way PDP analysis to produce biochar with the maximum elemental content.

Target element	Optimal pyrolysis parameters
Carbon	600 °C < HHT <1000 °C, 30min < RT < 90 min
Hydrogen	150 °C < HHT <300 °C, RT < 40min, HR < 15 °C/min
Oxygen	300 °C < HHT <800 °C, 20 min < RT < 70min, HR < 10 °C/min, PS_F < 5 mm
Nitrogen	600 °C < HHT <700 °C, 25 °C/min < HR < 35 °C/min
Phosphorus	200 °C < HHT <500 °C, 500min < RT
Potassium	200 °C < HHT <250 °C, 50min < RT < 100min, 7.5 °C/min < HR < 25 °C/min

profiles. This comprehensive analysis not only assists researchers and practitioners in developing prototypes of optimally tailored biochar for different applications but also advanced our understanding of the interrelationships among feedstock characteristics, pyrolysis parameters, and the resultant biochar's elemental composition. This study, therefore, provides a robust framework for advancing biochar technology and optimising its application in various environmental and agricultural domains.

5. Limitations and future research

While this study demonstrates the potential of machine learning in predicting biochar elemental composition, several limitations remain.

Firstly, the dataset, although substantial, was constrained by inconsistent reporting across the literature. Features such as feedstock particle size, feedstock P content, and gas flow rate were often excluded due to high missing rates of the data, limiting the model's comprehensiveness. Standardised data reporting in future studies will be essential to address this gap.

Secondly, the study focused on common lignocellulosic agricultural residues feedstocks, whereas alternative sources such as aquatic biomass, food waste, and industrial by-products - now gaining interest in biochar research - were underrepresented. Including these in future work would enhance model generalisability across feedstock types.

Lastly, environmental and post-pyrolysis conditions (e.g., humidity, storage time) were not considered due to data scarcity. These factors can significantly influence biochar properties and should be integrated into future models for more realistic predictions.

Future research should therefore focus on improving data quality, expanding feedstock diversity, and incorporating environmental variables to strengthen model robustness and applicability.

CRediT authorship contribution statement

Yao Fu: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Peter Cleall:** Writing – review & editing, Supervision. **Fei Jin:** Writing – review & editing, Validation, Supervision, Investigation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.renene.2025.124071.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] A. Sheer, et al., Trends and social aspects in the management and conversion of agricultural residues into valuable resources: a comprehensive approach to counter environmental degradation, food security, and climate change, Bioresour. Technol. 394 (Feb. 2024) 130258, https://doi.org/10.1016/j.biortech.2023.130258.

[2] Faostat. https://www.fao.org/faostat/en/#data. (Accessed 16 November 2024).

[3] J. Lehmann, et al., Biochar in climate change mitigation, Nat. Geosci. 14 (12) (Dec. 2021) 883–892, https://doi.org/10.1038/s41561-021-00852-8.

[4] J. Lehmann, S. Joseph, Biochar for environmental management: an introduction, in: Biochar for Environmental Management, second ed., Routledge, 2015.

[5] M. Keiluweit, P.S. Nico, M.G. Johnson, M. Kleber, Dynamic molecular structure of plant biomass-derived black carbon (Biochar), Environ. Sci. Technol. 44 (4) (Feb. 2010) 1247–1253, https://doi.org/10.1021/es9031419.

[6] X. Xiao, B. Chen, Z. Chen, L. Zhu, J.L. Schnoor, Insight into multiple and multilevel structures of biochars and their potential environmental applications: a critical review, Environ. Sci. Technol. 52 (9) (May 2018) 5027–5047, https://doi.org/10.1021/acs.est.7b06487.

[7] A. Creamer, B. Gao, M. Zhang, Carbon dioxide capture using biochar produced from sugarcane bagasse and hickory wood, Chem. Eng. J. 249 (Aug. 2014) 174–179, https://doi.org/10.1016/j.cej.2014.03.105.

[8] J.H. Windeatt, A.B. Ross, P.T. Williams, P.M. Forster, M.A. Nahil, S. Singh, Characteristics of biochars from crop residues: potential for carbon sequestration and soil amendment, J. Environ. Manag. 146 (Dec. 2014) 189–197, https://doi.org/10.1016/j.jenvman.2014.08.003.

[9] J. Lehmann, M.C. Rillig, J. Thies, C.A. Masiello, W.C. Hockaday, D. Crowley, Biochar effects on soil biota – a review, Soil Biol. Biochem. 43 (9) (Sep. 2011) 1812–1836, https://doi.org/10.1016/j.soilbio.2011.04.022.

[10] H. Singh, B.K. Northup, C.W. Rice, P.V.V. Prasad, Biochar applications influence soil physical and chemical properties, microbial diversity, and crop productivity: a meta-analysis, Biochar 4 (1) (Dec. 2022), https://doi.org/10.1007/s42773-022-00138-1.

[11] M. Kamali, L. Appels, E.E. Kwon, T.M. Aminabhavi, R. Dewil, Biochar in water and wastewater treatment - a sustainability assessment, Chem. Eng. J. 420 (Sep. 2021) 129946, https://doi.org/10.1016/j.cej.2021.129946.

[12] J.-H. Yuan, R.-K. Xu, H. Zhang, The forms of alkalis in the biochar produced from crop residues at different temperatures, Bioresour. Technol. 102 (3) (Feb. 2011) 3488–3497, https://doi.org/10.1016/j.biortech.2010.11.018.

[13] M. Qiu, et al., Biochar for the removal of contaminants from soil and water: a review, Biochar 4 (1) (Mar. 2022) 19, https://doi.org/10.1007/s42773-022-00146-1.

[14] Z. Chen, X. Xiao, B. Chen, L. Zhu, Quantification of chemical States, dissociation constants and contents of oxygen-containing groups on the surface of biochars produced at different temperatures, Environ. Sci. Technol. 49 (1) (Jan. 2015) 309–317, https://doi.org/10.1021/es5043468.

[15] Y. Ding, et al., Biochar to improve soil fertility. A review, Agron. Sustain. Dev. 36 (2) (May 2016) 36, https://doi.org/10.1007/s13593-016-0372-z.

[16] Y. Sun, et al., Effects of feedstock type, production method, and pyrolysis temperature on biochar and hydrochar properties, Chem. Eng. J. 240 (Mar. 2014) 574–578, https://doi.org/10.1016/j.cej.2013.10.081.

[17] K. Intani, S. Latif, Z. Cao, J. Müller, Characterisation of biochar from maize residues produced in a self-purging pyrolysis reactor, Bioresour. Technol. 265 (Oct. 2018) 224–235, https://doi.org/10.1016/j.biortech.2018.05.103.

[18] Z. Liu, et al., Comparative analysis of the properties of biochars produced from different pecan feedstocks and pyrolysis temperatures, Ind. Crop. Prod. 197 (2023), https://doi.org/10.1016/j.indcrop.2023.116638.

[19] S. Wang, H. Zhang, H. Huang, R. Xiao, R. Li, Z. Zhang, Influence of temperature and residence time on characteristics of biochars derived from agricultural residues: a comprehensive evaluation, Process Saf. Environ. Prot. 139 (Jul. 2020) 218–229, https://doi.org/10.1016/j.psep.2020.03.028.

[20] M. Rafiq, et al., Biochar amendment improves alpine meadows growth and soil health in Tibetan Plateau over a three year period, Sci. Total Environ. 717 (May 2020) 135296, https://doi.org/10.1016/j.scitotenv.2019.135296.

[21] X. Zhang, P. Zhang, X. Yuan, Y. Li, L. Han, Effect of pyrolysis temperature and correlation analysis on the yield and physicochemical properties of crop residue biochar, Bioresour. Technol. 296 (Jan. 2020) 122318, https://doi.org/10.1016/j.biortech.2019.122318.

[22] S. Wijitkosum, Biochar derived from agricultural wastes and wood residues for sustainable agricultural and environmental applications, Int. Soil Water Conserv. Res. 10 (2) (Jun. 2022) 335–341, https://doi.org/10.1016/j.iswcr.2021.09.006.

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

[24] Z. Dong, X. Bai, D. Xu, W. Li, Machine learning prediction of pyrolytic products of lignocellulosic biomass based on physicochemical characteristics and pyrolysis conditions, Bioresour. Technol. 367 (Jan. 2023) 128182, https://doi.org/10.1016/j.biortech.2022.128182.

[25] A. Hai, et al., Machine learning models for the prediction of total yield and specific surface area of biochar derived from agricultural biomass by pyrolysis, in: Environmental Technology & Innovation, vol. 30, ELSEVIER, May 2023, https://doi.org/10.1016/j.eti.2023.103071. Radarweg 29, 1043 NX Amsterdam, Netherlands.

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

[27] J.C. Ang, et al., Development of predictive model for biochar surface properties based on biomass attributes and pyrolysis conditions using rough set machine learning, in: Biomass & Bioenergy, vol 174, Pergamon-Elsevier Science LTD, ENGLAND, Jul. 2023, https://doi.org/10.1016/j.biombioe.2023.106820. The Boulevard, Langford Lane, Kidlington, Oxford Ox5 1gb.

[28] L. Leng, et al., Machine learning predicting and engineering the yield, N content, and specific surface area of biochar derived from pyrolysis of biomass, BIOCHAR 4 (1) (Dec. 2022), https://doi.org/10.1007/s42773-022-00183-w.

[29] A.T. Le, et al., Precise prediction of biochar yield and proximate analysis by modern machine learning and SHapley additive exPlanations, Energy Fuels 37 (22) (Nov. 2023) 17310–17327, https://doi.org/10.1021/acs.energyfuels.3c02868.

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

[31] J. Tang, et al., Prediction model for biochar energy potential based on biomass properties and pyrolysis conditions derived from rough set machine learning, Environ. Technol. (Mar. 2023) 1–15, https://doi.org/10.1080/09593330.2023.2192877.

[32] H. Yaka, M.A. Insel, O. Yucel, H. Sadikoglu, A comparison of machine learning algorithms for estimation of higher heating values of biomass and fossil fuels from ultimate analysis, in: FUEL, vol 320, Elsevier Sci Ltd, London, England, Jul. 15, 2022, https://doi.org/10.1016/j.fuel.2022.123971, 125 London Wall.

[33] Z. Jiang, et al., Machine learning prediction of biochar-specific surface area based on plant characterization information, Renew. Energy 243 (Apr. 2025) 122633, https://doi.org/10.1016/j.renene.2025.122633.

[34] Y. Song, et al., Machine learning prediction of biochar physicochemical properties based on biomass characteristics and pyrolysis conditions, J. Anal. Appl. Pyrolysis 181 (Aug. 2024) 106596, https://doi.org/10.1016/j.jaap.2024.106596.

[35] K.P. Shadangi, P.K. Sarangi, A.K. Behera, Chapter 3 - characterization techniques of biomass: physico-chemical, elemental, and biological, in: K.P. Shadangi, P. K. Sarangi, K. Mohanty, I. Deniz, A.R. Kiran Gollakota (Eds.), Bioenergy Engineering, Woodhead Publishing, 2023, pp. 51–66, https://doi.org/10.1016/B978-0-323-98363-1.00022-3.

[36] C. Whittaker, I. Shield, Factors affecting wood, energy grass and straw pellet durability – a review, Renew. Sustain. Energy Rev. 71 (May 2017) 1–11, https://doi.org/10.1016/j.rser.2016.12.119.

[37] G. Jomantas, K. Buinevičius, J. Šereika, Sulfur emission dependence on various factors during biomass combustion, Energies 18 (7) (Mar. 2025) 1701, https://doi.org/10.3390/en18071701.

[38] N. Toscano Miranda, I. Lopes Motta, R. Maciel Filho, M.R. Wolf Maciel, Sugarcane bagasse pyrolysis: a review of operating conditions and products properties, Renew. Sustain. Energy Rev. 149 (Oct. 2021) 111394, https://doi.org/10.1016/j.rser.2021.111394.

[39] L. Leng, et al., Machine learning predicting wastewater properties of the aqueous phase derived from hydrothermal treatment of biomass, Bioresour. Technol. 358 (Aug. 2022) 127348, https://doi.org/10.1016/j.biortech.2022.127348.

[40] X.C. Nguyen, Q.V. Ly, T.T.H. Nguyen, H.T.T. Ngo, Y. Hu, Z. Zhang, Potential application of machine learning for exploring adsorption mechanisms of pharmaceuticals onto biochars, Chemosphere 287 (Jan. 2022) 132203, https://doi.org/10.1016/j.chemosphere.2021.132203.

[41] J. Ma, S. Zhang, X. Liu, J. Wang, Machine learning prediction of biochar yield based on biomass characteristics, Bioresour. Technol. 389 (Dec. 2023) 129820, https://doi.org/10.1016/j.biortech.2023.129820.

[42] C. Liu, P. Balasubramanian, J. An, F. Li, Machine learning prediction of ammonia nitrogen adsorption on biochar with model evaluation and optimization, npj Clean Water 8 (1) (Feb. 2025) 1–12, https://doi.org/10.1038/s41545-024-00429-z.
[43] O. Rahmati, et al., Land subsidence modelling using tree-based machine learning algorithms, Sci. Total Environ. 672 (Jul. 2019) 239–252, https://doi.org/10.1016/j.scitotenv.2019.03.496.
[44] Y. Sun, et al., Machine learning in clarifying complex relationships: biochar preparation procedures and capacitance characteristics, Chem. Eng. J. 485 (Apr. 2024) 149975, https://doi.org/10.1016/j.cej.2024.149975.
[45] L. Leng, et al., Machine-learning-aided prediction and engineering of nitrogen-containing functional groups of biochar derived from biomass pyrolysis, Chem. Eng. J. 485 (Apr. 2024) 149862, https://doi.org/10.1016/j.cej.2024.149862.
[46] M.F. Aller, Biochar properties: transport, fate, and impact, Crit. Rev. Environ. Sci. Technol. 46 (14–15) (Aug. 2016) 1183–1296, https://doi.org/10.1080/10643389.2016.1212368.
[47] J. Sun, F. He, Y. Pan, Z. Zhang, Effects of pyrolysis temperature and residence time on physicochemical properties of different biochar types, Acta Agric. Scand. Sect. B Soil Plant Sci 67 (1) (Jan. 2017) 12–22, https://doi.org/10.1080/09064710.2016.1214745.
[48] A.L. Sullivan, R. Ball, Thermal decomposition and combustion chemistry of cellulosic biomass, Atmos. Environ. 47 (Feb. 2012) 133–141, https://doi.org/10.1016/j.atmosenv.2011.11.022.
[49] L. Leng, et al., Nitrogen containing functional groups of biochar: an overview, Bioresour. Technol. 298 (Feb. 2020) 122286, https://doi.org/10.1016/j.biortech.2019.122286.
[50] R. Zabaleta, et al., Brewer's spent grain-based biochar as a renewable energy source and agriculture substrate, J. Mater. Cycles Waste Manag. 26 (6) (Nov. 2024) 3787–3801, https://doi.org/10.1007/s10163-024-02078-3.