Journal Pre-proof

Bridging Biochar Synthesis and Dye Wastewater Treatment via an Interpretable End-to-End Machine Learning Framework

Yan Gao , Kai Ran , Yunyi Yang , Pengjing Wang , Jingjing Yao , Haipu Li , Zuhong Lin

PII: S2772-4433(26)00007-3
DOI: https://doi.org/10.1016/j.recm.2026.100176
Reference: RECM 100176

To appear in: *Resources Chemicals and Materials*

Received date: 11 December 2025
Revised date: 14 February 2026
Accepted date: 27 February 2026

Please cite this article as: Yan Gao , Kai Ran , Yunyi Yang , Pengjing Wang , Jingjing Yao , Haipu Li , Zuhong Lin , Bridging Biochar Synthesis and Dye Wastewater Treatment via an Interpretable End-to-End Machine Learning Framework, *Resources Chemicals and Materials* (2026), doi: https://doi.org/10.1016/j.recm.2026.100176

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2026 The Author(s). Publishing services by Elsevier B.V. on behalf of KeAi Communications Co., Ltd.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Highlights

- An end-to-end ML framework bridges biochar synthesis and dye wastewater treatment.
- Framework achieved R2=0.97-0.99 and RMSE=0.01-0.04 with strong accuracy.
- Online platform enables rapid biochar design with real-time analysis.
- Optimized biochar showed 125.25 mg/g adsorption, 18.71% above literature.

Journal Pre-proof

Bridging Biochar Synthesis and Dye Wastewater Treatment via an Interpretable End-to-End Machine Learning Framework

Yan Gao[a,b,1], Kai Ran[a,b,1], Yunyi Yang[a], Pengjing Wang[a], Jingjing Yao[a,b], Haipu Li[a,b,*], Zuhong Lin[a,b,*]

[a]Center for Environment and Water Resources, College of Chemistry and Chemical Engineering, Central South University, Changsha 410083, China

[b]Key Laboratory of Hunan Province for Water Environment and Agriculture Product Safety, Changsha 410083, China

*Corresponding author. E-mail address: lihaipu@csu.edu.cn; 212301023@csu.edu.cn.

Journal Pre-proof

Abstract

Biochar has emerged as an essential material for wastewater remediation due to its tunable physicochemical properties. There is a pressing need to develop rapid and efficient strategies to guide the rational design of biochar, facilitating its transition from laboratory research to practical applications. In this work, we developed a sequentially coupled end-to-end machine learning framework (EEMLF) that integrates an online, user-friendly platform enabling rapid prediction of optimal biochar synthesis and its adsorption capacity without programming expertise. The EEMLF had high and robust predictive accuracy on dye removal end and biochar synthesis end (R_1^2/R_2^2 = 0.99/0.97; $RMSE_1/RMSE_2$ =0.01/0.04). Through EEMLF, pyrolysis temperature and the

activator-to-biomass ratio were identified as the dominant factors governing biochar properties, while specific surface area emerged as the primary determinant of adsorption performance. Guided by the outputs of the EEMLF, the waste reed straw was pyrolyzed to produce a biochar (RSCH-800), RSCH-800's methylene blue adsorption capacity reached 125.25 mg/g, nearly twice that of the unactivated one (RSC-800). For real wastewater ($TOC_{Initial}$ = 672.6 mg/L), RSCH-800 achieved 85% TOC removal at a solids-to-liquid ratio of 1.3 g/L within 2 h. This work delivers a generalizable data-driven framework that enables the rapid development of high-performance biochar for dye wastewater, even broader polluted wastewater remediation.

Keywords

Carbon material design; Adsorption process; Contaminant removal; Environmental remediation; Machine learning.

Journal Pre-proof

1. Introduction

The printing and dyeing industry annually releases 10-20% of its wastewater, causing severe environmental pollution[1-3]. The complex aromatic structures of dyes confer high chemical stability, rendering them resistant to biodegradation and often associated with toxic and carcinogenic effects[4, 5]. With increasing demands for environmental protection and goals such as carbon sequestration, adsorption has become a promising choice due to its operational simplicity, the renewability of adsorbent, and the effective removal of pollutants[6]. Biochar, produced from waste biomass, is particularly attractive due to its adjustable specific surface area, well-developed pore structure, and abundant active functional sites[7, 8].

The adsorption performance of biochar is directly influenced by its intrinsic physicochemical properties and the adsorption process conditions[9, 10]. The former, such as specific surface area, the pH at the point of zero charge, are largely determined by the original biomass and the synthesis conditions[11-13]. Optimizing synthesis routes can therefore markedly enhance adsorption performance. For instance, Janu et al[14]. established the relationship between the physicochemical characteristics of biochar and its FTIR spectral absorption properties, finding that pyrolysis temperature is more important than residence time for specificity of functional groups. Wang et al[13] investigated the effects of synthesis process on the physicochemical properties of biochar. Under the optimal conditions, a biochar with a specific surface area up to 3230.57 m^2/g was obtained, achieving a maximum methylene blue adsorption capacity of 661.7 mg/g. Zhang et al.[15] achieved a significant increase in the Langmuir maximum adsorption capacity for Pb^{2+} from 116 mg/g to 294 mg/g by grafting teak waste with acrylic acid. The latter, such as contact time, adsorption temperature, and pH, significantly influence the pollutant adsorption capacity. Zazycki et al.[16] showed pecan-shell biochar reached 80% of its capacity under optimized adsorption process conditions. Raut et al.[17] reported high Cr(VI) adsorption capacity (152.44 mg/g) and excellent cycling stability (71.8% efficiency after five cycles) for citrus-limetta-derived

biochar. These findings indicate that the adsorption capacity of biochar is strongly influenced by both the synthesis process and the operating conditions[18], creating a high-dimensional optimization problem that makes conventional trial-and-error design time and resource intensive[19].

In this context, machine learning is rapidly emerging as a data-driven strategy for the efficient adsorbent design and water treatment[20-22]. For instance, Bibi et al.[23] used XGB as the preferred model for optimizing congo red removal by biochar, yielding high predictive accuracy (R^2 = 0.99), and found the dosage of adsorbent as a key factor. Chen et al.[24] employed an integrated model to predict the specific surface area, pore size, and pore volume of biochar, identifying pyrolysis temperature as the common key parameter. With the further advancement of machine learning, its application is expanding from singular property prediction to comprehensive and interpretable optimization of the entire synthesis-to-application workflow. Zhang et al.[21] emphasize that regulating the entire process from synthesis to adsorption is a key direction for enhancing material design efficiency. Wang et al.[25] existing research predominantly relies on laboratory data. There is an urgent need to develop machine learning models that integrate experimental validation and provide reverse guidance for synthesis. To directly address these developmental trends and core challenges, this work proposes a sequentially coupled end-to-end machine learning framework. This framework establishes an interpretable synthesis-structure-performance holistic framework by sequentially coupling the process-property relationships at the biochar synthesis end with the property-performance relationships at the dye removal end. This provides implementable support for biochar screening and design in textile wastewater treatment.

To construct EEMLF, we compiled a multidimensional dataset encompassing biochar synthesis parameters, material physicochemical properties, and adsorption process conditions. Based on this dataset, we trained and evaluated six tree-based machine learning models. The best performing model can predict the adsorption

capacity of biochar for dyes, identification of optimal adsorption process conditions, and recommendation of synthesis parameters for preparing high-performance biochar. Ultimately, we established a sequential coupling framework based on the synthesis-structure-performance relationship. To enhance accessibility, we also developed a user-friendly interface to allow researchers to efficiently obtain these optimal parameters. These model-recommended synthesis conditions were further validated through targeted experiments.

2. Materials and methods

2.1 Data collection and preprocessing

2.1.1 Data acquisition

This work systematically investigates the structure-performance relationships between biochar and dye removal behavior by integrating ML techniques with experimental validation. Data were retrieved from the Web of Science database using the keywords "*Biochar Adsorption*" and "*Dye removal*", with additional subject term expansions (e.g., "methylene blue removal"). To ensure data quality and consistency, literature inclusion criteria were defined as follows, (1) published in JCR Q1&Q2 journals with an impact factor more than 3.0; (2) complete documentation of biochar synthesis parameters (e.g., raw biomass, pyrolysis conditions) and adsorption performance metrics (e.g., equilibrium adsorption capacity, Q_e); and (3) standardized adsorption conditions (Dye concentration range: 0-5000 mg/L; pH: 2-12). The inclusion criteria for literature were determined based on frontier research and the actual characteristics of dyeing and printing wastewater. Given that raw materials, chemicals, and operational conditions vary, significant differences in concentration and pH may exist across different dye wastewater streams[26, 27]. Hence, the aforementioned criteria were selected to ensure the practicality of the data. A total of 2300 data entries were curated. Two datasets were established based on the acquired data, Dataset #1 and Dataset #2. Dataset #1 was designed for predicting adsorption capacity and includes the following features, adsorption process conditions like molecular weight of dye (MW),

adsorbent dosage (AD, g/L), initial dye concentration (C_0, mg/L), pH, adsorption time (t_1, h), adsorption temperature (T_1, °C), and adsorption volume (AV, mL); and biochar characteristics like elemental composition (C, H, O, N, S; wt%), specific surface area (BET, m^2/g), pore diameter (D_p, nm), and zero-point charge (pH_{PZC}). Dataset #2 focuses on the biochar synthesis end, covering feedstock properties like raw biomass (RB, g), particle size distribution (PSD, mm); activation conditions like activator type (AT), activator to biomass ratio (ARm); and pyrolysis conditions like annealing atmosphere (Atm), annealing rate (RAnn, °C/min), pyrolysis temperature (T_2, °C), and residence time (t_2, h). For a detailed description of the dataset, shown in supplementary information **Text S1 and Fig. S1**. To ensure data integrity and prevent overfitting, we conducted systematic deduplication checks. Specifically, we compared entries based on combinations of 8 parameters from the biochar synthesis end and 15 parameters from the dye removal end. All entries with identical parameters shall be regarded as duplicate data, with only one instance retained. This process was completed prior to partitioning the data into training and test sets to avoid any data leakage.

2.1.2 Data preprocessing

After compiling the two initial datasets, a systematic preprocessing was applied to ensure data quality and model reliability. Data preprocessing involved three key steps: outlier removal, normalization, and data partitioning. Outliers were identified and removed based on violations of physicochemical principles, such as negative adsorption capacities or unrealistic parameter values. These non-physical entries, likely arising from data compilation errors, were excluded prior to further analysis. All numerical features were then normalized to the range of -1 to 1 to ensure consistent scaling for model training. Finally, the dataset was randomly split into a training set (80%) and a test set (20%) for model development and evaluation.

Before model development, a pearson correlation coefficient matrix analysis was performed to assess the linear relationships between variables[28]. Independent correlation features with correlation coefficients between 0 and 0.5 are fed directly into

the model to avoid covariance problems (Eqs. 1)[29, 30].

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

where x_i and y_i represent the observations of the two variables, $\overline{x}$ and $\overline{y}$ denote the means of the variables, and n represents the number of observations.

2.2 Development of machine learning models

This work evaluates six representative ML algorithms, including CB, DT, GBDT, HGB, RF, and XGB, to strike a balance between model accuracy and interpretability. The reasons for selecting these models are listed in **Text S2**.

To minimize overfitting and improve model robustness, the hyperparameters of the EEMLF were systematically optimized[31]. In this work, Bayesian Optimization was employed to fine-tune the hyperparameters of the selected model. Bayesian Optimization constructs a probabilistic model of the objective function, allowing for the efficient identification of the optimal hyperparameter combination[32, 33].

To ensure the accuracy and reliability of model evaluation, 10-fold cross-validation was implemented. Following completion of all data preprocessing, the dataset was randomly partitioned into ten equally sized subsets. During each validation cycle, one subset served as the test set, while the remaining nine subsets were aggregated as the training set. This process was iterated ten times, each subset being used as the test set once, thereby guaranteeing that every data point was evaluated. The training set was utilized for hyperparameter optimization and model construction, whereas the test set was employed to assess model performance. This approach makes optimal use of finite data, reducing potential evaluation bias arising from a single random partition through multiple training and testing iterations. Consequently, it provides a more robust estimate of the model's generalization capability.

2.3 Evaluation of machine learning models

To evaluate the performance of ML models, several metrics are employed,

including coefficient determination (R^2), root mean square error (RMSE), and mean absolute error (Eqs. 2-4)[34, 35]. These metrics provide a comprehensive understanding of the model's accuracy and reliability. Higher R^2 values indicate a better fit between the model's predictions and the empirical data. Lower values of RMSE and MAE indicate better model performance[36].

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(y_i - y)^2} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \quad (3)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}} \quad (4)$$

where y_i denotes the actual value, x_i denotes the predicted value, and y denotes the average of the actual values.

2.4 Model interpretation

In this work, we used SHAP (Shapley Additive exPlanations), an interpretability method based on game theory[37], to investigate the key biochar characteristics influencing its adsorption capacity. We calculated the Shapley values (Eqs. 5) for each feature to quantify their marginal contributions to the model's predictions[38]. Positive Shapley values indicate that a feature enhances adsorption capacity, whereas negative values suggest an inhibitory effect. The absolute magnitude of these values reflects the strength of each feature's influence on the model's output[39]. The color of the points changing from blue to red indicates an increase in feature values[40].

$$\phi_j(f, x) = \sum_{S \subseteq N/\{j\}} \frac{|S|!(|N| - |S| - 1)!}{|N|!}[f(\mathbf{X}_S \cup \{j\}) - f(\mathbf{X}_s)] \quad (5)$$

where, ϕ_j is the Shapley value of the j-th feature, *N* is the set of all features, S is any subset of features that does not include feature j, |S| is the number of features in set S, |N| is the total number of features, $f(x_S)$ is the predicted output of the model when considering feature set S, and $f(x_S \cup \{j\})$-$f(x_S)$ denotes the marginal contribution of

feature j for the set of S.

To further explore the nonlinear interactions between features, we used SHAP dependence plots. These plots, based on locally weighted regression, illustrate how variations in individual features affect the model's output. Additionally, they reveal synergistic or antagonistic relationships between key features through color gradients. This analytical approach integrates global feature importance with local effect, providing valuable insights into optimizing biochar synthesis processes and enhancing its application in dye removal systems.

2.5 Biochar synthesis and adsorption experiments

In order to verify the effectiveness of the EEMLF, reed straw from Dongting Lake was used in this work. First, reed straws were cut into 1-2 cm segments, washed with deionized water, and dried in 60°C. Subsequently, the reed straws were pulverized to 200 mesh. The reed powder was impregnated with an 85% H_3PO_4 solution at a mass ratio of 1:1 for 12 hours, followed by filtration and drying. It was then transferred to a tube furnace and heated under a nitrogen atmosphere to 800°C at a heating rate of 5°C/min, and held for 4 h. After pyrolysis, the sample was cooled to room temperature under nitrogen. The resulting biochar was washed with deionized water to remove ash, then dried, pulverized, and sieved to 200 mesh, yielding the final product RSCH-800. For comparison, the control biochar (RSC-800) was prepared following the identical protocol except omitting the H_3PO_4 activation step. The RSCOH-800 replaces H_3PO_4 with KOH, with all other steps remaining identical.

The adsorption performance of the prepared biochar was evaluated using methylene blue and the actual wastewater from a printing and dyeing factory in Changde, Hunan Province as the model pollutant. The adsorption capacity was calculated using the following formula[41]:

$$Q_e = \frac{(C_0 - C_e) \times V}{m} \tag{6}$$

Where: Q_e is the adsorption capacity (mg/g), C_0 is the initial concentration of methylene blue (mg/L), C_e is the equilibrium concentration of methylene blue (mg/L)

V is the volume of the solution (L), m is the mass of the biochar used (g). The biochar preparation process, adsorption procedures, and characterization methods are detailed in **Text S3**. The relevant parameters for the actual wastewater are shown in **Table S7**.

3. Results and discussion

3.1 Feature distribution analysis

An analysis of the established datasets (**Fig. 1a-f** and **Table S1**) first revealed the distribution patterns of key parameters governing biochar synthesis and adsorption, providing essential context for the subsequent machine learning modeling. For the biochar synthesis end, the activator to biomass ratio (ARm) and pyrolysis temperature (T_2) showed distinct concentrations (**Fig. 1a** and **1b**). Notably, the reported ARm values fell within the range of 0 to 10, with a 1:1 ratio being frequently employed. This common practice is supported by studies indicating that an appropriate amount of activator effectively enhances the specific surface area and structural stability of biochar[42, 43]. Similarly, T_2 used in reported studies were predominantly concentrated in the range of 600-800°C, and this temperature range is conducive to the full carbonization of biomass and the formation of well-developed pore structure[44]. To enhance transparency, Table S2 summarizes the statistical distribution of the top 20 key biochar types by data volume, categorized by biomass feedstock, pyrolysis temperature, and activation method. The table demonstrates that our dataset encompasses diverse conditions, with no single type being significantly overrepresented and a generally even distribution of data.

Consistent with established knowledge, the specific surface area of biochar is a critical determinant of its adsorption capacity, as a larger surface area provides more sites for pollutant attachment[45, 46]. In the current work, the specific surface area values were mainly centered around 200 m^2/g, with only a minority of biochars exceeding 2000 m^2/g (**Fig. 1c**). Regarding surface charge, the pH at the point of zero charge (pH_{PZC}) for most biochars in the dataset clustered within a near-neutral range(**Fig. 1d**). this may be related to the nearly neutral adsorption environment.

Notably, the broad distribution of both specific surface area (0-2500 m^2/g) and pH_{PZC} (0-12) demonstrates that precise modulation of biochar's properties can be achieved by adjusting the synthesis process conditions.

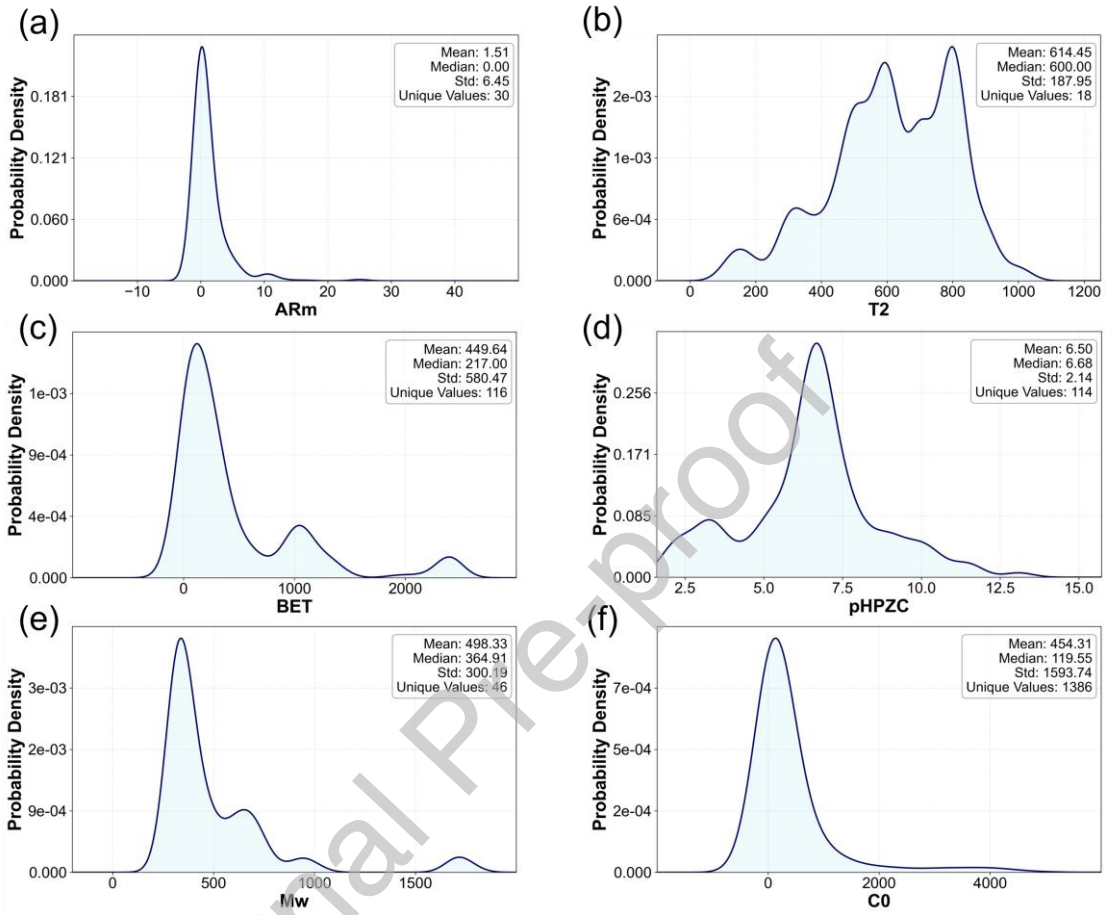


Fig. 1. Distribution of the activator to biomass ratio (a), annealing temperature (b), specific surface area (c), the pH at the point of zero charge (d), molecular weight of the dye (e), initial concentration of the dye (f) on the biochar synthesis and the dye removal with biochar adsorption.

The properties of the target dye molecules constitute critical variables in the adsorption system. Analysis of our dataset revealed that approximately 80% of the involved dyes have molecular weights (MW) between 250 and 500 Da (**Fig. 1e**), indicating a focus on low-to-medium molecular weight compounds. This range is mechanistically relevant, as such molecules are more likely to access the micropores and mesopores commonly found in biochar adsorbents. The initial concentration of the dyes (C_0) was an important driver of the adsorption process (**Fig. 1f**), spanning from 0

to 5000 mg/L. The density distribution of other features was shown in **Fig. S2-4.** Collectively, the broad yet physically interpretable distributions observed for biochar properties, process conditions, and dye characteristics establish a comprehensive and high-dimensional feature space. This diverse dataset is essential for training a robust machine learning model to decode the non-linear relationships within the complex biochar adsorption system.

3.2 Correlation-guided feature selection

To improve EEMLF's robustness and minimize feature redundancy, we performed a feature selection analysis prior to model training. The Pearson correlation coefficient was used to examine linear dependencies among key variables within each dataset. For features exhibiting strong collinearity ($|r| > 0.5$), a single representative feature was retained per correlated group based on its physicochemical interpretability and relevance. This feature selection strategy was applied independently to Dataset #1 (for predicting dye removal end) and Dataset #2 (for the biochar synthesis end).

For Dataset #1, the correlation matrix (**Fig. 2a)** revealed complex interactions. Most linear correlations were weak ($|r| < 0.50$). Notably, strong positive correlations were observed between the dye's molecular weight, molecular size, and ionic character ($r = 0.56$-0.93). For biochar elemental composition, carbon content (C, wt%) showed a strong negative correlation with the O/C and (O+N)/C ratios ($r = -0.59$ to -0.60). Additionally, specific surface area plays a positive correlation with the total pore volume (V_t) ($r = 0.66$). To minimize redundancy and multicollinearity, one representative feature was retained per group of highly correlated features[28]. For instance, molecular weight (MW) represented dye characteristics, and specific surface area represented porosity. The final input dataset #1 comprised 15 features, included dye properties(MW) adsorption conditions (AD, C_0, AV, pH, t_1, and T_1), biochar elemental composition (C, H, O, N, S, wt%), BET specific surface area, average pore width (D_p), and pH_{PZC}, totaling.

For Dataset #2, correlation analysis of biomass feedstock and synthesis conditions

(**Fig. 2b**) showed all pairwise coefficients below 0.42, indicating negligible linear collinearity. Therefore, all original 8 input features were retained for subsequent modeling.

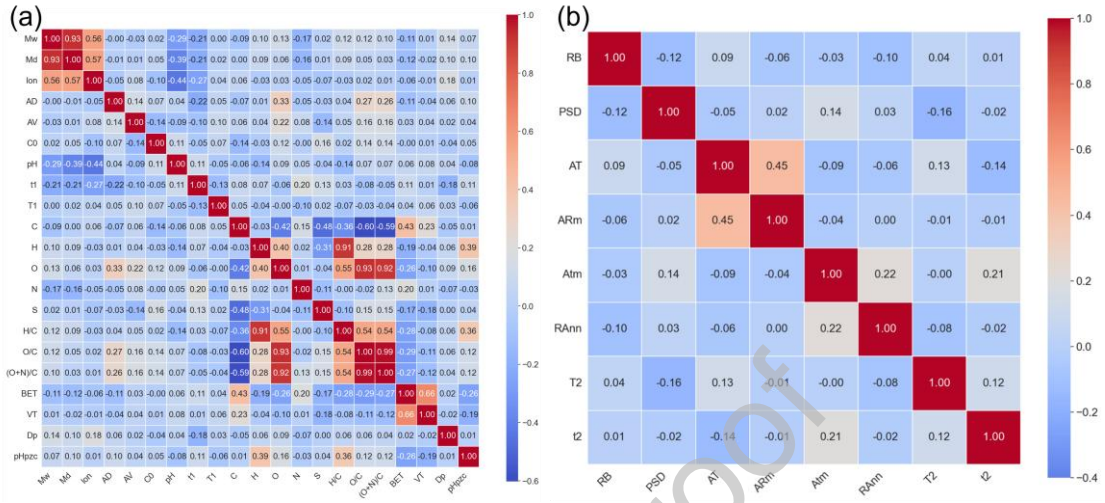


Fig. 2. Pearson plot of the predictive model for dye removal end (a), and biochar synthesis end (b).

3.3 EEMLF construction and performance

Building upon this refined feature set, the EEMLF was constructed using a suite of tree-based ML algorithms, including categorical boosting (CB), decision tree (DT), gradient boosting decision tree (GBDT), histogram gradient boosting (HGB), random forest (RF), and extreme gradient boosting (XGB). Model hyperparameters were optimized using Bayesian optimization with a Gaussian process surrogate[31]. And the performance was rigorously evaluated through 10-fold cross-validation to ensure robustness and prevent overfitting. The results showed that the XGB model performed optimally for predicting dye removal, while RF performed best for predicting biochar synthesis. The final, optimized models demonstrated strong predictive accuracy on the test set, with the XGB model achieving an R^2 of 0.99 and the RF achieving an R^2 of 0.97 (**Fig. 3, Fig. S5 and Table S4)**. Optimized via Bayesian, each ends reached its best configuration with minimal prediction error. In the dye removal end, XGB tuned four typical hyperparameters in an optimal configuration (*learning_rate = 0.99, n_estimators = 1311, max_depth = 7, min_child_weight = 0.59*), where RMSE reached

Journal Pre-proof

the minimum 0.01. Similarly, RF in biochar synthesis end tuned five typical hyperparameters in an optimal configuration (*max_depth = 92, max_samples = 0.99, min_samples_leaf = 1, min_samples_split = 2, n_estimators = 113*), where RMSE reached the minimum 0.04. The optimal hyperparameters for other models used in constructing the EEMLF are summarized in **Table S3**. These minimal RMSE values reflect the excellent generalization of EEMLF's two ends.

These results were further compared with previous works on data driven to biochar synthesis and application in **Table S5**. The prediction accuracy of the biochar specific surface area prediction model based on the EEMLF was improved by 21.25% (R^2 = 0.80)[47]. Meanwhile, the accuracy of the EEMLF in predicting the adsorption capacity of biochar was improved by 7.45% (R^2 = 0.94)[19]. These additional advantages demonstrate the superiority of the current work, in which a comprehensive model was developed for the biochar synthesis process.

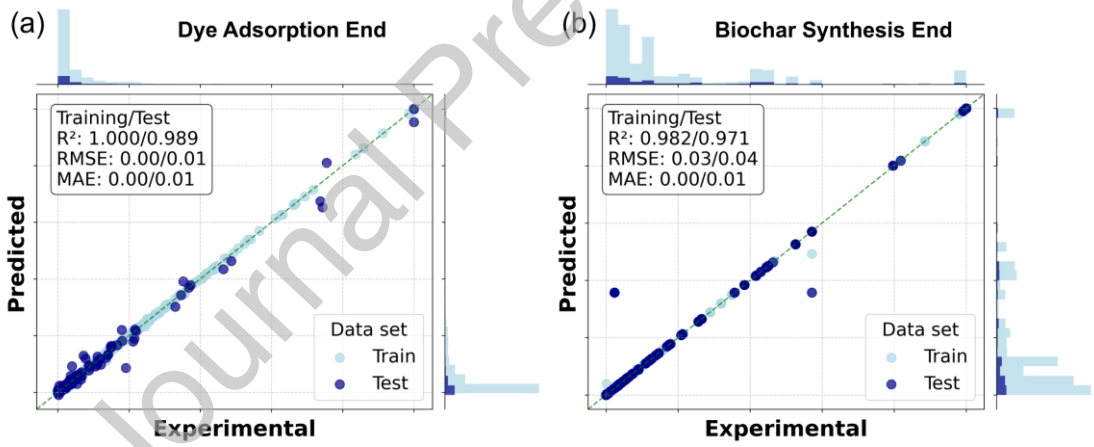


Fig. 3. Performance for EEMLF's dye removal end (a) and biochar synthesis end (b).

3.4 Interpretable optimization of parameters

3.4.1 SHAP Analysis of Key Features

To prioritize the most influential parameters in biochar synthesis and dye removal, we applied SHAP (SHapley Additive exPlanations) to interpret the model outputs and prioritize the most influential parameters governing biochar synthesis end and dye removal end[48]. For the dye removal end, **Fig. 4a-b** indicate that the C_0 is the most influential feature, with a strong positive correlation to the predicted adsorption

capacity. This is because at low concentrations, the biochar surface possesses more active sites to adsorb dye molecules; whereas at high concentrations, these active sites rapidly become saturated, with a limited number of adsorption sites available, rendering it incapable of adsorbing dye molecules[49, 50]. Furthermore, at high concentrations, the driving force increases to overcome the mass transfer resistance between dye molecules and the biochar[51]. H (wt%), BET, emerged as the most influential material properties. A higher H content (wt%) was associated with increased predicted adsorption capacity, which is consistent with its role as a proxy for surface polar groups (e.g., -OH, -COOH, etc.)that facilitate hydrogen bonding or electrostatic interactions with dye molecules[52, 53]. Concurrently, a larger specific surface area provides more sites for physical adsorption, especially in the porous structure of the biochar[48, 54].

It is worth emphasizing that, although hydrogen content is mechanistically crucial, we selected specific surface area as the key bridging variable connecting the biochar synthesis end and dye removal end primarily based on two considerations. Firstly, SHAP analysis confirms specific surface area as one of the most stable and significant structural features contributing to adsorption performance. Selecting it as the output at the biochar synthesis end and the input at the dye removal end effectively captures the core physical mechanisms determining adsorption capacity while ensuring model prediction robustness and practical operability. The SHAP importance scores for other features were lower, indicating their relatively limited contribution to predicting adsorption capacity. Furthermore, hydrogen content is frequently estimated indirectly via elemental difference calculations (100%-C%-N%-O%) within existing literature datasets. Significant variations exist in computational and reporting methodologies across different studies, rendering its use as a singular, robust bridging descriptor uncertain. In contrast, BET specific surface area represents the most standardized, widely reported, and relatively uniform physical characterization metric within materials science, offering superior accessibility and comparability within literature datasets. Therefore, this study establishes specific surface area as the core bridging

Journal Pre-proof

variable connecting material synthesis and performance evaluation within the framework.

In the biochar synthesis end, **Fig. 4c-d** indicate that the heating rate (RAnn) and the ARm had the most significant effects on specific surface area. Faster heating rates promoted activator diffusion and enhanced pore formation, facilitating the development of a well-structured porous architecture[55]. As the ARm increases, the pore properties of biochar can be enhanced, including surface area and porosity[56]. The number of functional groups on the biochar surface, such as phenolic, hydroxyl, and carboxyl groups, also rises[57]. Taking phosphoric acid as an example, its high dispersibility leads to a heterogeneous pore size distribution. The formation of phosphate bonds enhances crosslinking density, expelling volatile components and further expanding pores[58]. The etching interaction between carbon atoms and phosphorus elements creates pores within the solid structure[59]. Furthermore, with increased polarity, the adsorption rate of organic pollutants in wastewater becomes higher, but an excess of activators may also lead to the collapse of the carbon structure[60]. T_2 was also identified as a key parameter, with an optimal range that facilitates carbon skeleton rearrangement and pore development[61, 62], but avoids pore collapse at excessively high temperatures[63, 64]. The bidirectional distribution of residence time (t_2) suggests a more complex effect, which depends on the interaction between activation temperature and activator type.

Based on the importance analysis of the critical features, it can be concluded that at the dye removal end, the specific surface area and the initial dye concentration are the most influential features affecting adsorption capacity in biochar properties and adsorption performance, respectively. At the biochar synthesis end, the most critical factor is the mass ratio of activator to biomass. The feature importance analyses provide valuable insights for optimizing both adsorption capacity and biochar synthesis, thereby guiding experimental design. However, it should be noted that SHAP-based interpretations reflect model-dependent feature contributions rather than direct causal

relationships, and the identified effects may be influenced by feature correlations and data distribution characteristics. Consequently, there remains a need for deeper investigation into how key features influence biochar synthesis and adsorption capacity. Notably, these influencing factors are not purely univariate in nature; rather, they often involve complex interactions among multiple variables that must be considered to gain a comprehensive understanding.

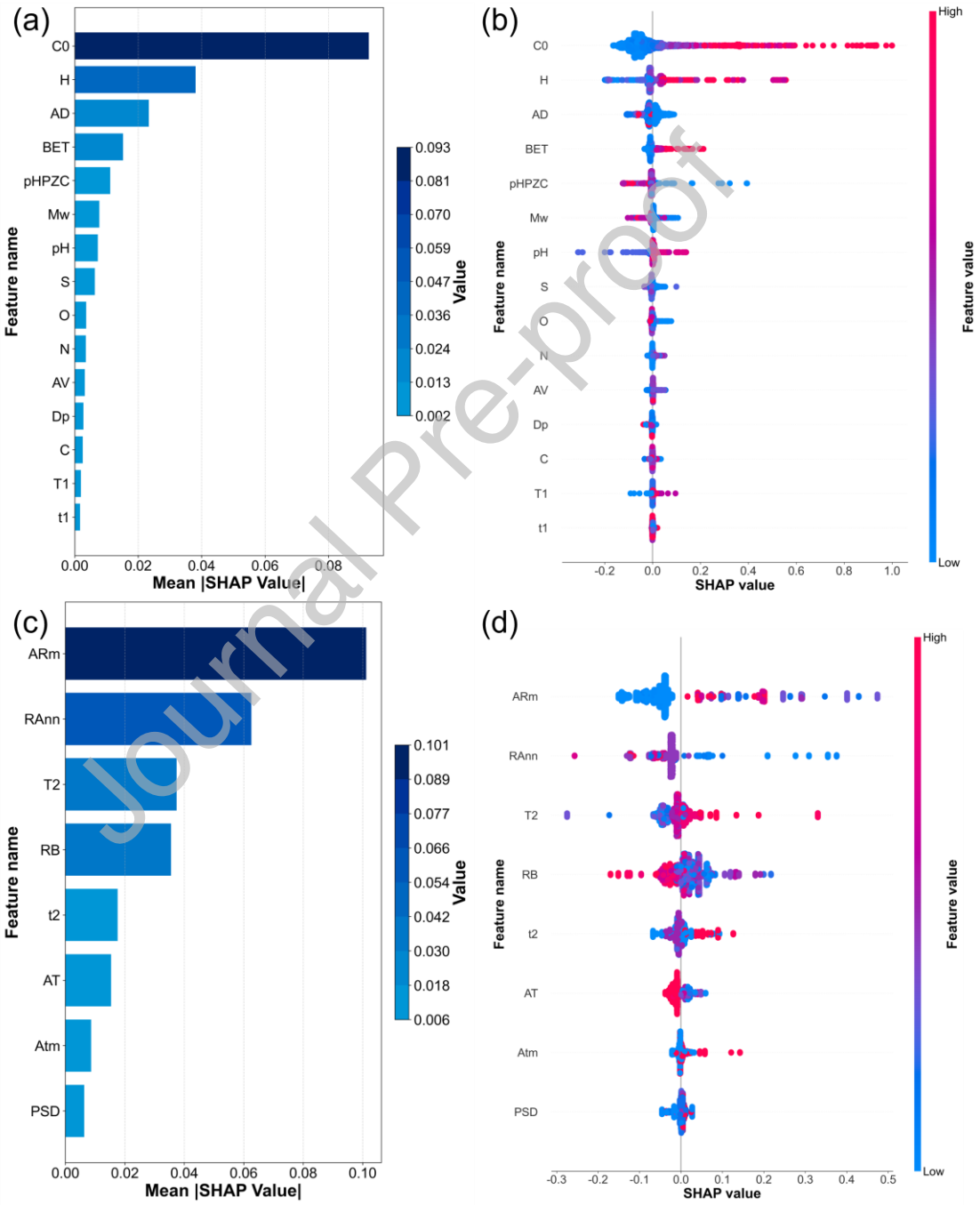


Fig. 4. SHAP-based interpretation of EEMLF, feature importance ranking (a), and

feature contribution analysis (b) on the biochar synthesis end, feature importance ranking (c), and feature contribution analysis (d) on the dye removal end.

3.4.2 SHAP Dependency Analysis

To gain deeper insights into EEMLF, we plotted SHAP dependency plots analysis to illustrate the non-linear relationships between features and model predictions, as well as the interactions among features. The results are presented in **Fig. 5** and **Fig. S6-S12.** **Fig. 5a** reveals the interaction between adsorption temperature (T_1) and adsorption time (t_1). Within the range of approximately 280 K to 320 K, the SHAP value exhibits minimal variation and no discernible monotonic trend. This indicates that temperature exerts a negligible influence on adsorption capacity within this temperature interval. This result correlates with the adsorption process likely being dominated by physical adsorption, as physical adsorption possesses lower activation energy and exhibits reduced sensitivity to temperature variations. Furthermore, this narrow temperature range may be insufficient to induce significant alterations in the adsorption mechanism. Regarding adsorption time, appropriately extending adsorption duration enhances adsorption capacity; however, once adsorption equilibrium is attained, the benefits of further prolonging adsorption time become markedly limited. Crucially, **Fig. 5b** reveals the partial dependence plot for the initial dye concentration (C_0), confirming its dominant positive influence on the predicted adsorption capacity, exhibiting near-linear growth. This non-linear trend at high concentrations may suggest a shift in adsorption mechanism, such as enhanced multi-layer formation or pore filling under a high concentration driving force. In contrast, the influence of temperature on adsorption capacity appeared less significant at similar initial concentrations, consistent with physisorption-dominated processes that are less sensitive to thermal energy within these limits.

A significant antagonistic interaction was observed between the heating rate (RAnn) and the annealing atmosphere (Atm), as shown in **Fig. 5c**. When RAnn was below 5 °C/min, positive SHAP values indicated that slower heating can enhance the

specific surface area in the N_2 atmosphere. In contrast, when RAnn exceeded 5 °C/min, negative SHAP values were observed, suggesting that rapid heating may suppress the development of pore structures[62]. This aligns with prior research indicating that heating rate significantly influences its porosity and specific surface area. A low heating rate favours the preservation of structural complexity, whereas high heating rates induce localised melting of biomass, phase transitions, and expansion of cellular structures[65]. Zeng et al. observed that reducing the heating rate while increasing temperature promotes the ordering of carbonized tissue structures[66]. Researchers further noted that as the heating rate increases, the pore volume and specific surface area initially rise before subsequently decreasing[67]. When the particle size distribution (PSD) of the raw materials is lower than 1mm, the ARm was also found to be a critical determinant of specific surface area (**Fig. 5d**). In the absence of activators, a weak negative correlation was observed. With the introduction of activation agents, SHAP values increased, indicating improved porosity. However, excessive ARm values (more than 2:1) led to a decline in SHAP values, likely due to pore structure collapse caused by over-etching[68].

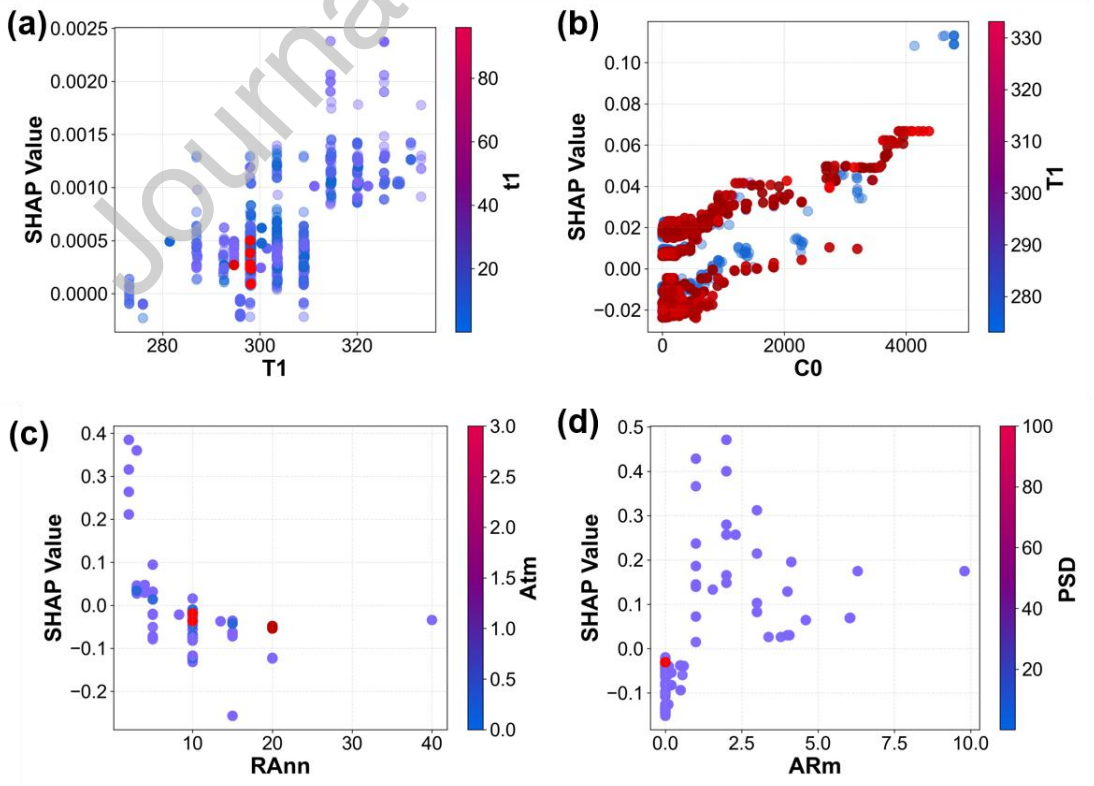


Fig. 5. Marginal effect analysis of key features on Q_e for dye removal end and specific surface area for biochar synthesis end in the EEMLF, adsorption temperature and adsorption time (a), dye initial concentration and adsorption temperature (b), annealing rate and annealing atmosphere, and the number of the Atm, 0 is N_2; 1 is Air; 2 is NH_3; 3 is Ar (c), activator to biomass ratio and particle size distribution (d).

Based on the above analysis combined with **Figs. S6-S12**, it can be concluded that, during the biochar synthesis end, pyrolysis at 800 °C under a nitrogen atmosphere with a heating rate of 5 °C/min and an activator-to-biomass ratio of 2:1 is favorable for biochar formation. In addition, at the dye removal end, adsorption experiments conducted at 45 °C benefit from the use of biochar adsorbents with higher carbon (C) and hydrogen (H) content and pH_{pzc} close to 7.

3.5 User-oriented online prediction platform

In order to enhance the practical utility of the EEMLF, we developed a user-oriented online prediction platform (https://huggingface.co/spaces/Biochar3.0). This platform provides an intuitive interface that visualizes the complex processes of model training and prediction, improving model transparency and accessibility. Simultaneously, it reduces the need of extensive and costly laboratory experiments, enabling users without coding expertise to obtain reliable predictions efficiently (**Fig. 6**). The platform functions as a standardized web application, accessible via mainstream browsers (such as Chrome, Firefox) across different operating systems (Windows, macOS, Linux) without requiring local installation. For typical single prediction tasks, server response times are generally within 1 minutes. For ease of use, **Fig. 6** annotates the platform's main interface, input areas, functional zones, and result display screen. Users may switch between the decolorization and biochar synthesis modules within the model selection area according to their research task. Users may upload .csv files via the data input module or manually enter relevant parameters. Within the feature input panel, input parameters are categorized into three main groups: biochar properties, dye characteristics, and adsorption conditions. In the feature contribution visualisation area,

Journal Pre-proof

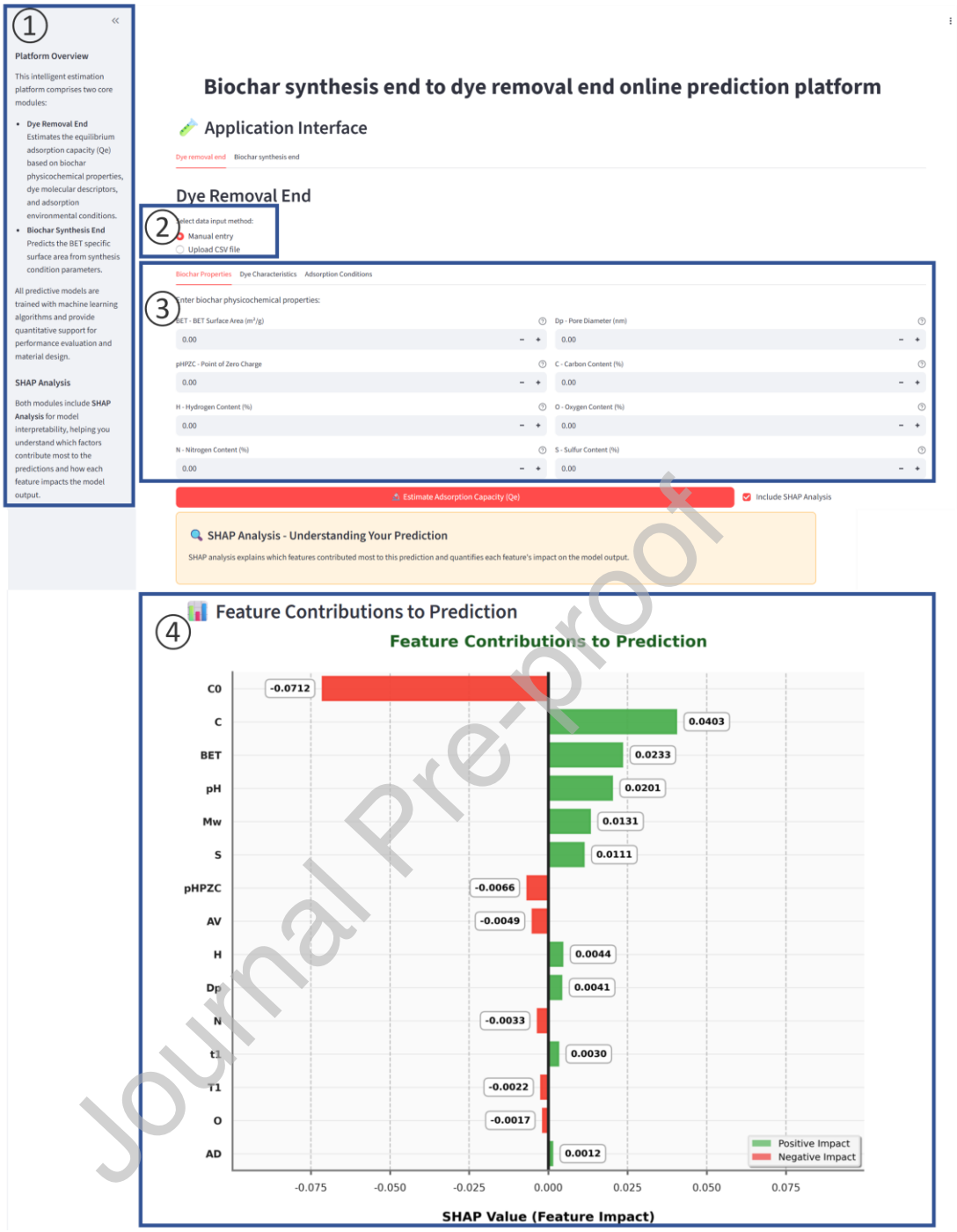


Fig 6. Schematic illustration of the EEMLF online prediction platform and its functional modules, ①: Module switch; ②:Data input mode; ③: Feature input panels; ④: Feature contribution visualization.

users can assess each feature's contribution to predicting Qe values via SHAP bar charts. Taking the biochar synthesis module as an example, inputting a .csv file containing biomass properties, activation conditions, and pyrolysis parameters enables

users to view key biochar characteristics on the results visualization interface, such as specific surface area and zero-point charge. A key advancement in this platform is the integration of SHAP analysis. This feature provides immediate visual interpretation of the model predictions, illustrating the magnitude and direction of influence that each input variable (e.g., pyrolysis temperature, adsorbent dosage) has on the prediction targets (BET surface area and adsorption capacity). This greatly enhances model interpretability and offers valuable, data-driven insights for optimizing experimental design and process parameters.

Input ranges for several core parameters are recommended, including activator ratio (ARm, 0-5), pyrolysis temperature (T_2, 200-1200 °C), pyrolysis residence time (t_2, 0-8h), average pore size (D_p, 0.5-200 nm), molecular weight of the dye (100-1500 g/mol), initial dye concentration (10-5000 mg/L), adsorption temperature (5-60°C), and adsorbent dosage (0.1-20 g/L). Each component of the platform interface has been carefully designed to ensure that users can easily upload and process data, train models, and fine-tune parameters.

3.6 EEMLF-guided experimental validation

3.6.1 Characterization of EEMLF-Guided Biochars

Following the guidance of the EEMLF online prediction platform, we conducted experimental verification using reed stalks. The model outputs are as follows: Atm: N_2, t_2: 4-5 h, PSD: 0-5 mm, ARm: 0-5, AT: 85% H_3PO_4, RAnn: 5 °C/min, T_2: 500-1000 °C. The final selection involved soaking reed fragments passing a 200-mesh sieve in 85% H_3PO_4 at a 1:1 ratio for 12 hours. Following filtration and drying, the material was pyrolyzed for 4 hours in N_2 at a heating rate of 5 °C/min up to 800 °C, as shown in **Fig. 7a**. The primary objective of this experimental study was to validate the feasibility and effectiveness of the proposed framework, rather than to determine the optimal biochar material. Finally, we synthesized RSC-800 and RSCH-800, with their morphological and structural evolution detailed in **Fig. 7b-d** . Scanning electron microscopy (SEM) revealed a clear progression: the RS showed an irregular, dense, and rough surface;

pyrolysis produced RSC-800 with a smooth, lamellar structure; and phosphoric acid activation transformed RSCH-800 into a material with a folded, multilayered lamellar morphology suggestive of a high surface area .Elemental mapping provided chemical insight into this transformation. While carbon formed the primary matrix for all samples, EDS analysis confirmed the successful incorporation of the activator in RSCH-800, evidenced by a distinct phosphorus signal (1.62 at.%). Concurrently, a markedly enhanced and uniform oxygen distribution was observed in RSCH-800 compared to RSC-800, indicating the introduction of oxygen-containing functional groups critical for surface reactivity.

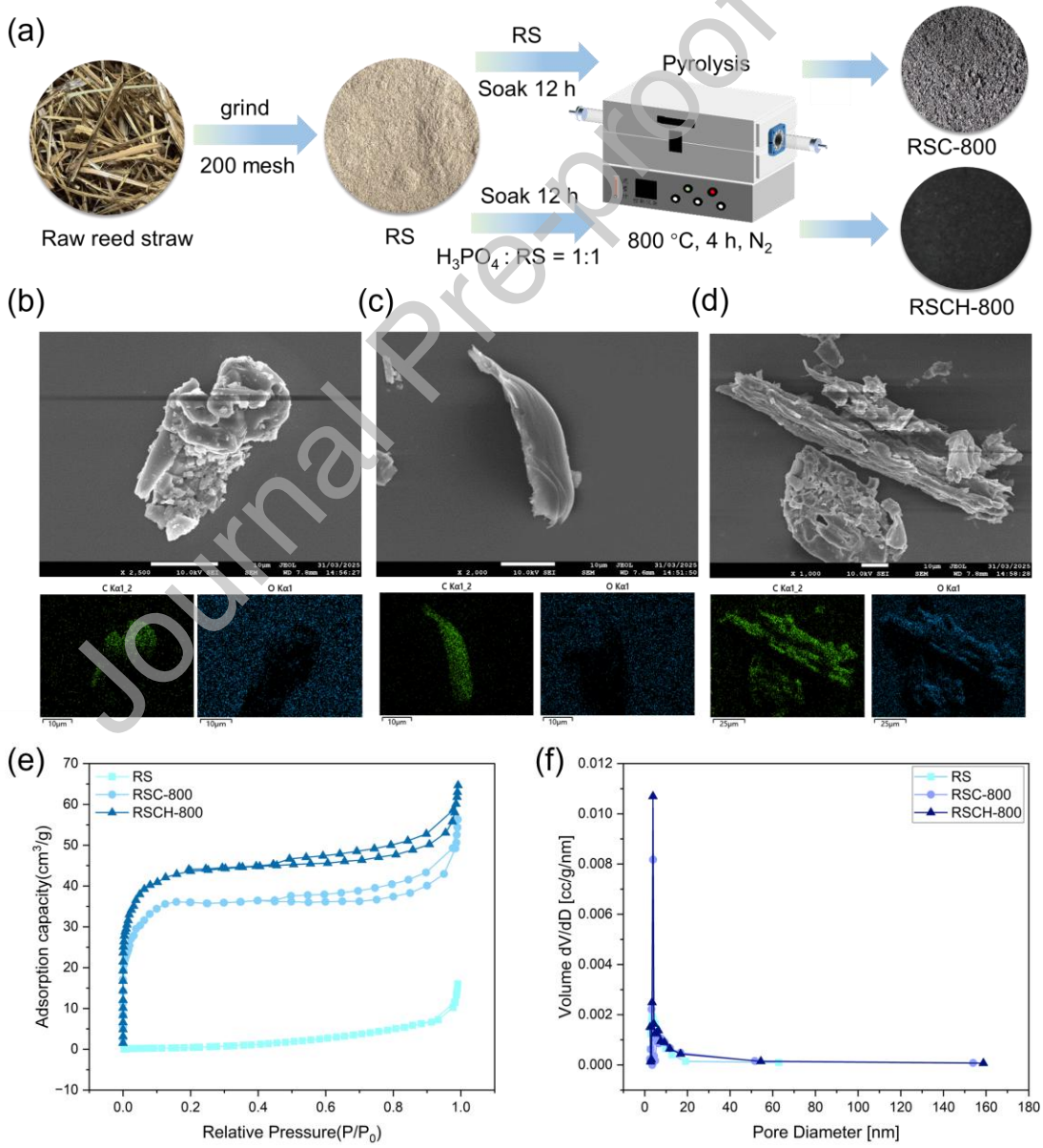


Fig. 7. Biochar RSCH-800 and RSC-800 synthesis guided by EEMLF (a); SEM and

EDS spectra of RS(b),RSC-800(c), RSCH-800(d); BET(e) and D_P(f) diagrams for different biochar.

As can be seen from **Fig. 7e, f and Table 1**, the adsorption isotherm of three materials exhibit a Type I characteristic[69], reed straw (RS) exhibited extremely low specific surface area (1.14 m^2/g) and pore volume (0.02 cm^3/g), indicating a closed surface and lack of effective pore structure. After carbonization treatment at 800°C, the specific surface areas of RSC-800 and RSCH-800 were significantly enhanced to 133.67 m^2/g and 159.62 m^2/g, where phosphoric acid activation increased the specific surface area of RSCH-800 by 19.4%, confirming its effective optimization of pore development. The results of pore size distribution showed that the average pore size of RS was 43.11 nm belonging to mesoporous structure, but its pore volume was very low (0.02 cm^3/g), which indicated that the number of pores was very small; whereas, the average pore sizes of RSC-800 and RSCH-800 were centered at 1.25-1.30 nm (in line with the criterion of microporous), and the pore volumes were increased to 0.08 cm^3/g and 0.1 cm^3/g, respectively, in which the microporous capacity of RSCH-800 was increased by 25% compared with that of the control group, highlighting the potentiating effect of activation on functional micropores.

In summary, the EEMLF-guided activation process directionally engineered RSCH-800 with an optimized hierarchical pore structure and modified surface chemistry . These quantitative structural data align with the framework's predictions , providing direct validation of its efficacy in guiding the rational design of high-performance adsorbents.

Table 1 Specific surface area, pore size, pore volume and elemental composition of different materials

The sample	BET (m^2/g)	D_p	V_T (cm^3/g)	EDS,wt%				
				C	O	N	S	P
RS	1.14	43.11	0.02	24.14	58.11	15.39	0.59	1.77

RSC-800	133.67	1.30	0.08	43.38	55.19	0	0	1.44
RSCH-800	159.62	1.25	0.10	40.31	57.97	0	0.10	1.62

3.6.2 Experimental validation of adsorption capacity

The adsorption performance was evaluated using methylene blue (MB) as a model pollutant. Isotherm data were fitted with Langmuir, Freundlich, and Sips models (**Fig. 8a, Table S6).** with the Sips model providing the best fit (R^2 = 0.99 for RSC-800; 0.98 for RSCH-800) compared to the Langmuir and Freundlich models. The Sips model, effectively describes the adsorption behavior transitioning from heterogeneous site occupation at low concentrations towards monolayer saturation at higher concentrations[70]. The substantial increase in the Freundlich constant (K_F from 19.56 to 41.52) confirms the enhanced overall adsorption affinity due to phosphoric acid activation. The low heterogeneity index (1/n = 0.19; less than 0.5) suggests the presence of high-energy sites with strong selectivity at low MB concentrations[71, 72]. The heterogeneity parameter from the Sips model (1-m ≈ 0.7 for both biochar) indicates a transition toward a more homogeneous, monolayer-saturation type of adsorption at higher concentrations. This combined adsorption behavior can be attributed to the hierarchical porous structure of the biochar. Specifically, MB molecules are likely adsorbed through multiple mechanisms, including π–π stacking interactions, electrostatic attraction, and hydrogen bonding within the multi-scale pore network[73]. Comparing the two materials, the maximum adsorption capacity (Q_m) increased significantly from 66.08 mg/g (RSC-800) to 125.25 mg/g (RSCH-800), representing an 89.6% enhancement. It is worth mentioning that the adsorption capacity of RSCH-800 with a specific surface area of 159.62 m^2/g under the guidance of the EEMLF increased 18.71% at least for MB as compared to other studies (**Table 2**).

To further validate the EEMLF framework's potential for guiding different activation pathways, KOH-activated biochar (RSCOH-800) was prepared under identical conditions and subjected to adsorption testing. Within the identical concentration range (10-200 mg/L), RSCOH-800 exhibited methylene blue removal

rates approaching 100% at all tested concentrations(**Fig. S13**), with complete adsorption of the dye from the solution. This phenomenon demonstrates RSCOH-800's superior adsorption potential across this concentration range. KOH, as a strong alkaline activator, typically generates a more developed microporous structure, thereby providing an adsorption site density far exceeding that of unactivated biochar. Although RSCOH-800 did not reach saturation adsorption within this concentration range, its demonstrated dye removal capability in practical testing strongly validates the EEMLF framework's efficacy and guiding value in synthesizing biochar with ultra-high adsorption performance.

Reusability is a critical parameter for adsorbents in practical applications. Research indicates that ethanol solvent demonstrates high efficiency in the desorption process of MB dye[74]. Consequently, ethanol was employed to desorb MB from RSCH-800. The removal rate decreased with increasing cycle numbers. However, after five cycles, the removal efficiency stabilized at 55.62%, as shown in **Fig. 8b**. The decline in removal rate was attributed to the irreversibility of some adsorption sites[75]. It is also possible that portions of MB adsorbed within mesopores or micropores were not completely desorbed during the elution process[76]. The results indicate that RSCH-800 possesses favorable adsorption potential for MB removal due to its excellent regenerative properties and ease of separation.

Building on its excellent performance with model pollutants, the practical application potential of RSCH-800 was evaluated using actual dyeing and printing wastewater. Given that actual dyeing and printing wastewater constitutes a complex mixture of components, making it difficult to determine specific molecular structures, we employed ultraviolet-visible spectrophotometry (UV-Vis) and three-dimensional excitation-emission matrix (3D-EEM) fluorescence spectroscopy for holistic characterization of its organic constituents. Spectral analysis revealed (**Fig. 8c and Fig. S14**) that the pre-treated wastewater exhibited significant ultraviolet absorption near 240 nm. Furthermore, the 3D-EEM spectrum displayed intense fluorescence peaks

Journal Pre-proof

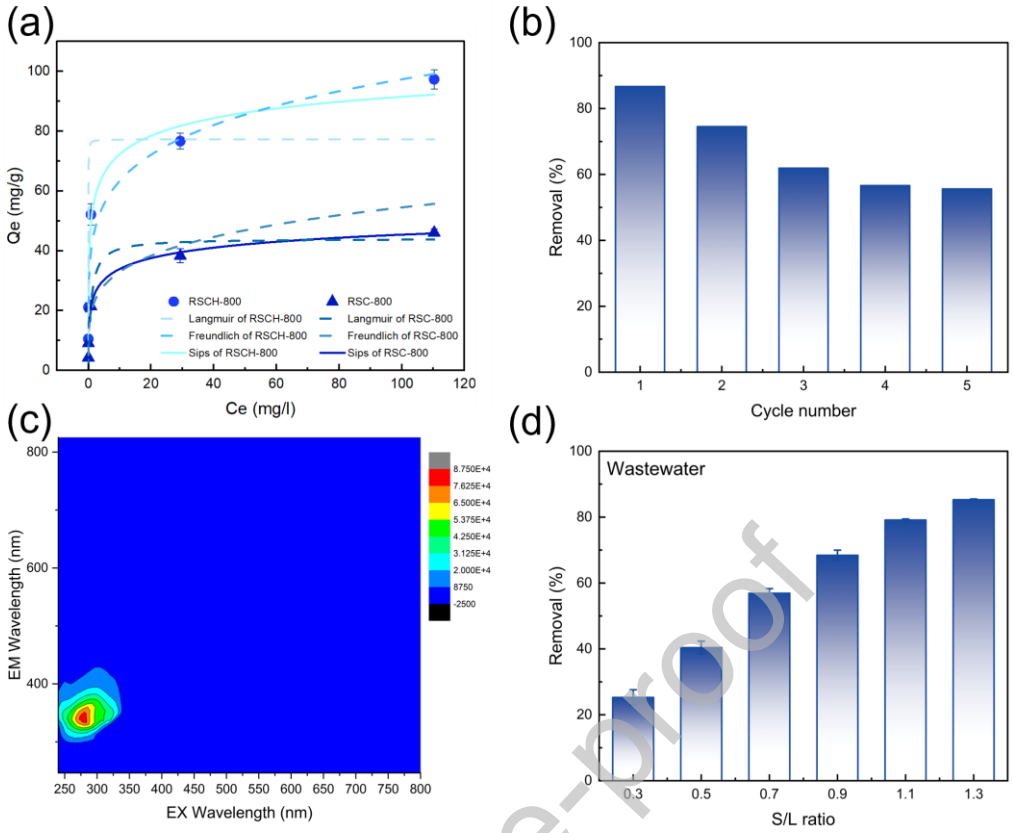


Fig. 8. adsorption isotherms for MB on RSCH-800 and RSC-800(a), Cyclic Stability of RSCH-800(b), 3D-EEM spectrum of the actual wastewater(c), and TOC removal efficiency of the RSCH-800 for actual wastewater(d).

within the excitation/emission wavelength range of 250-300 nm / 300-400 nm, a region typically associated with characteristic aromatic compounds and certain dye intermediates in wastewater. The actual wastewater was characterized by a high total organic carbon content of 672.6 mg/L and a low BOD_5/COD ratio of 0.31, indicating a high proportion of recalcitrant organic compounds resistant to conventional degradation. As shown in **Fig. 8d**, RSCH-800 achieved a high TOC removal efficiency of approximately 85% at 1.3g/L, demonstrating its effective broad-spectrum adsorption capability against a complex pollutant mixture. However, the material's performance in complex systems differs significantly from that observed in simulated wastewater containing a single pollutant. Under identical adsorption conditions, RSCH-800 achieved a TOC removal rate of 93.68% for simulated MB wastewater at 100 mg/L. In

Journal Pre-proof

contrast, when treating complex actual wastewater at a solid-liquid ratio of 0.9 g/L, the TOC removal rate was 68.43%. This discrepancy clearly demonstrates the competitive adsorption effects arising from the coexistence of multiple pollutants within complex matrices. Although adsorption efficiency is reduced by competition, the 68.43% removal rate still demonstrates the practical efficacy and application potential of the EEMLF-guided biochar synthesis for treating actual recalcitrant textile wastewater.

Table 2 The adsorption capacity of MB on different materials

Biomass	activation	BET (m^2/g)	Qm (mg/g)	Reference
Algae	H_3PO_4	239.8	108.2	[77]
coconut shell	H_3PO_4	259.77	63.7	[77]
coconut coat	H_3PO_4	190.74	0.6	[77]
Rice straw	Chitosan and CH_3COOH	5.61	27.96	[41]
Rice straw	KOH	86.54	19.62	[41]
Rice straw	KOH and Chitosan	31.03	62.04	[41]
cotton seed	NaOH	0.73	24.87	[78]
Tea	NaOH	178	105.44	[79]
Bagasse	NaOH	8.2175	114.94	[80]
Sawdust	Coating okra	-	78.13	[81]
Reed	HNO_3	37.5	33.33	[82]
Reed straw	-	133.67	66.08	This work
	H_3PO_4	159.62	125.25	This work

4. Conclusion

This work develops a rapid EEMLF, which was demonstrated strong predictive performance across both biochar synthesis and dye removal tasks for dyeing wastewater treatment, yielding R_1^2/R_2^2 of 0.99/0.97 and $RMSE_1/RMSE_2$ =0.01/0.04, respectively. The EEMLF effectively identified specific surface area as the most influential

characteristic governing biochar adsorption capacity, while pyrolysis temperature and activator ratio were recognized as key synthesis parameters controlling specific surface area. To enhance the framework's practical utility, an online prediction platform was developed to support rapid parameter selection and reduce experimental workload and cost. Guided by these insights, the RSCH-800 biochar was successfully synthesized under optimized conditions (800°C and 1:1 activator ratio), yielding a specific surface area of 159.62 m^2/g and an adsorption capacity of 125.25 mg/g for methylene blue, representing an 18.71% improvement compared with previous studies.

Despite its contributions, this work retains limitations that warrant urgent exploration, which also point to future research directions. Firstly, the predictive reliability of the current framework is highly dependent upon the feature space covered by the training data. Its outputs are intended to provide recommendations for parameter ranges with superior performance, rather than single precise values; thus, decision-making in practical applications must incorporate engineering expertise. Secondly, the current EEMLF constitutes a sequentially coupled open-loop prediction framework. Future research may evolve this into an integrated closed-loop autonomous optimization system encompassing "design-synthesis-testing-learning" by incorporating real-time feedback mechanisms. For instance, dynamically feeding rapid adsorption screening results from newly synthesized biochar back into the synthesis parameter recommendation algorithm could enable iterative, adaptive optimization. This paradigm, akin to active learning or autonomous laboratories, holds promise for significantly accelerating the discovery of high-performance materials by continuously narrowing the gap between prediction and experimentation.

Author Contributions

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used Kimi to assist in correcting

grammatical errors in the text. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

Declaration of competing interest

The authors declare no competing financial interest.

Acknowledgement

The authors acknowledge financial was supported by the National Key Research and Development Program of China (2023YFC3207000). And we are grateful for resources from the High-Performance Computing Center of Central South University.

Data availability

Data will be made available on request.

Reference

[1] Y. Li, P. Cao, S. Wang, X. Xu. Research on the treatment mechanism of anthraquinone dye wastewater by algal-bacterial symbiotic system. Bioresource Technology, 347 (2022) 126691. https://doi.org/10.1016/j.biortech.2022.126691.

[2] K.C. Bedin, I.P.A.F. Souza, A.L. Cazetta, L. Spessato, A. Ronix, V.C. Almeida. CO_2-spherical activated carbon as a new adsorbent for Methylene Blue removal: Kinetic, equilibrium and thermodynamic studies. Journal of Molecular Liquids, 269 (2018) 132-139. https://doi.org/10.1016/j.molliq.2018.08.020.

[3] M. Bagheri, E. Wetterlund. Introducing hydrothermal carbonization to sewage sludge treatment systems—a way of improvin

g energy recovery and economic performance? Waste Management, 170 (2023) 131-143. https://doi.org/10.1016/j.wasman.2023.08.006.

[4] S. Sağlam, F.N. Türk, H. Arslanoğlu. Use and applications of metal-organic frameworks (MOF) in dye adsorption: Review. Journal of Environmental Chemical Engineering, 11(5) (2023)

110568. https://doi.org/10.1016/j.jece.2023.110568.

[5] C. Yao, H. Fan, A. Adogwa, H. Xiong, M. Yang, F. Liu, Z. Chen, Y. Lou. Recent advances in carbon dioxide selective hydrogenation and biomass valorization via single-atom catalysts. Resources Chemicals and Materials, 2(3) (2023) 189-207. https://doi.org/10.1016/j.recm.2023.05.003.

[6] S. Sutar, P. Patil, J. Jadhav. Recent advances in biochar technology for textile dyes wastewater remediation: A review. Environmental Research, 209 (2022) 112841. https://doi.org/10.1016/j.envres.2022.112841.

[7] D. Laishram, S.-B. Kim, S.-Y. Lee, S.-J. Park. Advancements in Biochar as a Sustainable Adsorbent for Water Pollution Mitigation. Advanced Science, 12(19) (2025) 2410383. https://doi.org/10.1002/advs.202410383.

[8] D.C.C.d.S. Medeiros, C. Nzediegwu, C. Benally, S.A. Messele, J.-H. Kwak, M.A. Naeth, Y.S. Ok, S.X. Chang, M. Gamal El-Din. Pristine and engineered biochar for the removal of contaminants co-existing in several types of industrial wastewaters: A critical review. Science of The Total Environment, 809 (2022) 151120. https://doi.org/10.1016/j.scitotenv.2021.151120.

[9] N. Cao, J. Ji, C. Li, M. Yuan, X. Guo, X. Zong, L. Li, Y. Ma, C. Wang, S. Pang. Rapid and efficient removal of multiple aqueous pesticides by one-step construction boric acid modified biochar. RSC Advances, 13(13) (2023) 8765-8778. https://doi.org/10.1039/D2RA07684E.

[10] J. Cui, F. Wu, W. Zhang, L. Yang, J. Hu, Y. Fang, P. Ye, Q. Zhang, X. Suo, Y. Mo, X. Cui, H. Chen, H. Xing. Direct prediction of gas adsorption via spatial atom interaction learning. Nature Communications, 14(1) (2023) 7043. https://doi.org/10.1038/s41467-023-42863-6.

[11] C. Zhang, W.-H. Chen, S.-H. Ho, A.T. Hoang, Y. Zhang. Tetracycline-adsorbed biochar for solid biofuel usage to achieve waste utilization for environmental sustainability. Environmental Research, 237 (2023) 116959. https://doi.org/10.1016/j.envres.2023.116959.

[12] Z. Dai, L. Zhao, H. Tian, S. Peng, Z. Yue. Study on removal of oxytetracycline by KOH-modified chestnut shell biochar: Based on two-dimensional spectroscopy and molecular density functional theory. Journal of Water Process Engineering, 56 (2023) 104356. https://doi.org/10.1016/j.jwpe.2023.104356.

Journal Pre-proof

[13] K. Wang, Y. Yang, C. Liu, Y. Li, C.-Q. Cao, P.-G. Duan, L. Pei, J. Duo. Preparation of high-performance adsorption carbon materials via hydrothermal carbonization of agricultural solid waste-grapevines. Journal of Environmental Management, 383 (2025) 125353. https://doi.org/10.1016/j.jenvman.2025.125353.

[14] R. Janu, V. Mrlik, D. Ribitsch, J. Hofman, P. Sedláček, L. Bielská, G. Soja. Biochar surface functional groups as affected by biomass feedstock, biochar composition and pyrolysis temperature. Carbon Resources Conversion, 4 (2021) 36-46. https://doi.org/10.1016/j.crcon.2021.01.003.

[15] J.-W. Zhang, N.D. Hai, M.A. Kholif, H.-P. Chao. Adsorption of heavy metals from water using teak-based carbon material through graft copolymerization. Carbon Resources Conversion, 8(4) (2025) 100332. https://doi.org/10.1016/j.crcon.2025.100332.

[16] M.A. Zazycki, M. Godinho, D. Perondi, E.L. Foletto, G.C. Collazzo, G.L. Dotto. New biochar from pecan nutshells as an alternative adsorbent for removing reactive red 141 from aqueous solutions. Journal of Cleaner Production, 171 (2018) 57-65. https://doi.org/10.1016/j.jclepro.2017.10.007.

[17] S.S. Raut, D. Kumari, S. Singh, D. Pandey, A. Daverey, K. Dutta. Fabrication of different magnetic biochar from Citrus limetta peels for Cr (VI) adsorption: Characterisation, and performance. Resources Chemicals and Materials 4(3) (2025). https://doi.org/10.1016/j.recm.2025.100118.

[18] Z. Han, J. Yue, X. Zeng, J. Yu, F. Wang, Y. Guan, X. Liu, F. Ding, L. Fu, X. Jia, X. Song, C. Wang, Y. Cui, L. Shi, K. Wang, V. Zivkovic, D. Bai, G. Xu. Major challenges and recent advances in characterizing biomass thermochemical reactions. Resources Chemicals and Materials, 3(2) (2024) 146-158. https://doi.org/10.1016/j.recm.2023.10.001.

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

[20] X. Zhu, D.C.W. Tsang, L. Wang, Z. Su, D. Hou, L. Li, J. Shang. Machine learning exploration

Journal Pre-proof

of the critical factors for CO_2 adsorption capacity on porous carbon materials at different pressures. Journal of Cleaner Production, 273 (2020) 122915. https://doi.org/10.1016/j.jclepro.2020.122915.

[21] W. Zhang, R. Chen, J. Li, T. Huang, B. Wu, J. Ma, Q. Wen, J. Tan, W. Huang. Synthesis optimization and adsorption modeling of biochar for pollutant removal via machine learning. Biochar, 5(1) (2023) 25. https://doi.org/10.1007/s42773-023-00225-x.

[22] P. Rajput, S. Yadav, C. Liu, P. Balasubramanian. Predicting biochar adsorption capacity for methylene blue removal using machine learning. Journal of Water Process Engineering, 69 (2025) 106749. https://doi.org/10.1016/j.jwpe.2024.106749.

[23] A. Bibi, H. Khan, S. Hussain, M. Arshad, F. Wahab, M. Usama, K. Khan, F. Akbal. Sustainable wastewater purification with crab shell-derived biochar: Advanced machine learning modeling & experimental analysis. Bioresource Technology, 390 (2023) 129900. https://doi.org/10.1016/j.biortech.2023.129900.

[24] C. Chen, Y. Hu, Y. Ge, J. Tao, B. Yan, Z. Cheng, X. Lv, X. Cui, G. Chen. Integrated learning framework for enhanced specific surface area, pore size, and pore volume prediction of biochar. Bioresource Technology, 424 (2025) 132279. https://doi.org/10.1016/j.biortech.2025.132279.

[25] W. Wang, J.-S. Chang, D.-J. Lee. Machine learning applications for biochar studies: A mini-review. Bioresource Technology, 394 (2024) 130291. https://doi.org/10.1016/j.biortech.2023.130291.

[26] A. Kausar, M. Iqbal, A. Javed, K. Aftab, Z.-i.-H. Nazli, H.N. Bhatti, S. Nouren. Dyes adsorption using clay and modified clay: A review. Journal of Molecular Liquids, 256 (2018) 395-407. https://doi.org/10.1016/j.molliq.2018.02.034.

[27] S.A. Ishak, M.F. Murshed, H. Md Akil, N. Ismail, S.Z. Md Rasib, A.A. Al-Gheethi, The Application of Modified Natural Polymers in Toxicant Dye Compounds Wastewater: A Review. Water, 2020, p. 2032.

[28] X. Yuan, M. Suvarna, S. Low, P.D. Dissanayake, K.B. Lee, J. Li, X. Wang, Y.S. Ok. Applied Machine Learning for Prediction of CO2 Adsorption on Biomass Waste-Derived Porous Carbons. Environmental Science & Technology, 55(17) (2021) 11925-11936.

https://doi.org/10.1021/acs.est.1c01849.

[29] J. Li, X. Zhu, Y. Li, Y.W. Tong, Y.S. Ok, X. Wang. Multi-task prediction and optimization of hydrochar properties from high-moisture municipal solid waste: Application of machine learning on waste-to-resource. Journal of Cleaner Production, 278 (2021) 123928. https://doi.org/10.1016/j.jclepro.2020.123928.

[30] X. Zhu, X. Wang, Y.S. Ok. The application of machine learning methods for prediction of metal sorption onto biochars. Journal of Hazardous Materials, 378 (2019) 120727. https://doi.org/10.1016/j.jhazmat.2019.06.004.

[31] Z. Lin, S. Pan, H. Li, J. Yao. Unveiling similarities and differences in oxidation processes of oxidants and derived reactive oxygen species through machine learning interpretation of oxidation rate constants. Separation and Purification Technology, 340 (2024) 126649. https://doi.org/10.1016/j.seppur.2024.126649.

[32] J. Löfgren, D. Tarasov, T. Koitto, P. Rinke, M. Balakshin, M. Todorović. Machine Learning Optimization of Lignin Properties in Green Biorefineries. ACS Sustainable Chemistry & Engineering, 10(29) (2022) 9469-9479. https://doi.org/10.1021/acssuschemeng.2c01895.

[33] A.H. Victoria, G. Maragatham. Automatic tuning of hyperparameters using Bayesian optimization. Evolving Systems, 12(1) (2021) 217-223. https://doi.org/10.1007/s12530-020-09345-2.

[34] G. Gorreta, C. Maesani, S. Alini, M. Blangetti, C. Prandi. Towards fiber-to-fiber recycling of synthetic textile wastes: the problem of decolorization. Waste Management, 205 (2025) 114979. https://doi.org/10.1016/j.wasman.2025.114979.

[35] X. Yuan, J. Li, J.Y. Lim, A. Zolfaghari, D.S. Alessi, Y. Wang, X. Wang, Y.S. Ok. Machine Learning for Heavy Metal Removal from Water: Recent Advances and Challenges. ACS ES&T Water, 4(3) (2024) 820-836. https://doi.org/10.1021/acsestwater.3c00215.

[36] Z. Lin, J. Chen, Y. Fang, S.-h. Deng, H. Li, Y. Yang, J. Yao. Rapidly tailor metal–organic frameworks for arsenate removal using graph convolutional neural networks. Separation and Purification Technology, 354 (2025) 129334. https://doi.org/10.1016/j.seppur.2024.129334.

[37] S.M. Lundberg, G. Erion, H. Chen, A. DeGrave, J.M. Prutkin, B. Nair, R. Katz, J. Himmelfarb,

N. Bansal, S.-I. Lee. From local explanations to global understanding with explainable AI for trees. Nature Machine Intelligence, 2(1) (2020) 56-67. https://doi.org/10.1038/s42256-019-0138-9.

[38] Q. Dong, S. Zhang, B. Wu, M. Pi, Y. Xiong, H. Zhang. Co-pyrolysis of Sewage Sludge and Rice Straw: Thermal Behavior and Char Characteristic Evaluations. Energy & Fuels, 34(1) (2020) 607-615. https://doi.org/10.1021/acs.energyfuels.9b03800.

[39] J. Yao, H. Li, H.Y. Yang. Predicting adsorption capacity of pharmaceuticals and personal care products on long-term aged microplastics using machine learning. Journal of Hazardous Materials, 458 (2023) 131963. https://doi.org/10.1016/j.jhazmat.2023.131963.

[40] J. Li, L. Pan, M. Suvarna, Y.W. Tong, X. Wang. Fuel properties of hydrochar and pyrochar: Prediction and exploration with machine learning. Applied Energy, 269 (2020) 115166. https://doi.org/10.1016/j.apenergy.2020.115166.

[41] X. Su, X. Wang, Z. Ge, Z. Bao, L. Lin, Y. Chen, W. Dai, Y. Sun, H. Yuan, W. Yang, J. Meng, H. Wang, S.C. Pillai. KOH-activated biochar and chitosan composites for efficient adsorption of industrial dye pollutants. Chemical Engineering Journal, 486 (2024) 150387. https://doi.org/10.1016/j.cej.2024.150387.

[42] C. Wang, X. Lin, X. Zhang, P.L. Show. Research advances on production and application of algal biochar in environmental remediation. Environmental Pollution, 348 (2024) 123860. https://doi.org/10.1016/j.envpol.2024.123860.

[43] C. Li, X. Zhu, H. He, Y. Fang, H. Dong, J. Lü, J. Li, Y. Li. Adsorption of two antibiotics on biochar prepared in air-containing atmosphere: Influence of biochar porosity and molecular size of antibiotics. Journal of Molecular Liquids, 274 (2019) 353-361. https://doi.org/10.1016/j.molliq.2018.10.142.

[44] E. Sørmo, L. Silvani, N. Bjerkli, N. Hagemann, A.R. Zimmerman, S.E. Hale, C.B. Hansen, T. Hartnik, G. Cornelissen. Stabilization of PFAS-contaminated soil with activated biochar. Science of The Total Environment, 763 (2021) 144034. https://doi.org/10.1016/j.scitotenv.2020.144034.

[45] L. Ren, G. Zhao, L. Pan, B. Chen, Y. Chen, Q. Zhang, X. Xiao, W. Xu. Efficient Removal of

Journal Pre-proof

Dye from Wastewater without Selectivity Using Activated Carbon-Juncus effusus Porous Fibril Composites. ACS Applied Materials & Interfaces, 13(16) (2021) 19176-19186. https://doi.org/10.1021/acsami.0c22104.

[46] T. Ai, X. Jiang, Q. Liu, L. Lv, H. Wu. Daptomycin adsorption on magnetic ultra-fine wood-based biochars from water: Kinetics, isotherms, and mechanism studies. Bioresource Technology, 273 (2019) 8-15. https://doi.org/10.1016/j.biortech.2018.10.039.

[47] A. Hai, G. Bharath, M.F.A. Patah, W.M.A.W. Daud, R. K, P. Show, F. Banat. Machine learning models for the prediction of total yield and specific surface area of biochar derived from agricultural biomass by pyrolysis. Environmental Technology & Innovation, 30 (2023) 103071. https://doi.org/10.1016/j.eti.2023.103071.

[48] J.O. Eniola, B. Sizirici, A. Khaleel, I. Yildiz. Fabrication of engineered biochar-iron oxide from date palm frond for the effective removal of cationic dye from wastewater. Journal of Water Process Engineering, 54 (2023) 104046. https://doi.org/10.1016/j.jwpe.2023.104046.

[49] J. Ma, J. Li, L. Weng, X. Ouyang, Y. Chen, Y. Li. Phosphorus-Enhanced and Calcium-Retarded Transport of Ferrihydrite Colloid: Mechanism of Electrostatic Potential Changes Regulated via Adsorption Speciation. Environmental Science & Technology, 57(10) (2023) 4219-4230. https://doi.org/10.1021/acs.est.2c09670.

[50] K.S. Obayomi, S.Y. Lau, M.K. Danquah, J. Zhang, T. Chiong, O.V. Obayomi, L. Meunier, M.M. Rahman. A response surface methodology approach for the removal of methylene blue dye from wastewater using sustainable and cost-effective adsorbent. Process Safety and Environmental Protection, 184 (2024) 129-150. https://doi.org/10.1016/j.psep.2024.01.106.

[51] Z. Jia, L. Wu, C. Han, D. Zhang, M. Li, R. Wei. Preparation of auto-suspending hollow silica microfiber derived from biotemplate and its highly selective adsorption for organic dyes. Materials Today Communications, 34 (2023) 105437. https://doi.org/10.1016/j.mtcomm.2023.105437.

[52] K.S.D. Premarathna, A.U. Rajapaksha, B. Sarkar, E.E. Kwon, A. Bhatnagar, Y.S. Ok, M. Vithanage. Biochar-based engineered composites for sorptive decontamination of water: A review. Chemical Engineering Journal, 372 (2019) 536-550.

https://doi.org/10.1016/j.cej.2019.04.097.

[53] J. Guo, L. Zheng, Z. Li, X. Zhou, S. Cheng, L. Zhang, Q. Zhang. Effects of various pyrolysis conditions and feedstock compositions on the physicochemical characteristics of cow manure-derived biochar. Journal of Cleaner Production, 311 (2021) 127458. https://doi.org/10.1016/j.jclepro.2021.127458.

[54] G. Kwon, A. Bhatnagar, H. Wang, E.E. Kwon, H. Song. A review of recent advancements in utilization of biomass and industrial wastes into engineered biochar. Journal of Hazardous Materials, 400 (2020) 123242. https://doi.org/10.1016/j.jhazmat.2020.123242.

[55] C.E.R. Barquilha, M.C.B. Braga. Adsorption of organic and inorganic pollutants onto biochars: Challenges, operating conditions, and mechanisms. Bioresource Technology Reports, 15 (2021) 100728. https://doi.org/10.1016/j.biteb.2021.100728.

[56] X.F. Tan, S.B. Liu, Y.G. Liu, Y.L. Gu, G.M. Zeng, X.J. Hua, X. Wang, S.H. Liu, L.H. Jiang. Biochar as potential sustainable precursors for activated carbon production: Multiple applications in environmental protection and energy storage. Bioresource Technology, 227 (2017) 359-372. https://doi.org/10.1016/j.biortech.2016.12.083.

[57] N.L. Panwar, A. Pawar. Influence of activation conditions on the physicochemical properties of activated biochar: a review. Biomass Conversion and Biorefinery, 12(3) (2022) 925-947. https://doi.org/10.1007/s13399-020-00870-3.

[58] F. Guo, K. Peng, S. Liang, X. Jia, X. Jiang, L. Qian. Evaluation of the catalytic performance of different activated biochar catalysts for removal of tar from biomass pyrolysis. Fuel, 258 (2019) 116204. https://doi.org/10.1016/j.fuel.2019.116204.

[59] B. Sajjadi, W.-Y. Chen, N.O. Egiebor. A comprehensive review on physical activation of biochar for energy and environmental applications. Reviews in Chemical Engineering, 35(6) (2019) 735-776. https://doi.org/10.1515/revce-2017-0113.

[60] D. Akhil, D. Lakshmi, A. Kartik, D.-V.N. Vo, J. Arun, K.P. Gopinath. Production, characterization, activation and environmental applications of engineered biochar: a review. Environmental Chemistry Letters, 19(3) (2021) 2261-2297. https://doi.org/10.1007/s10311-020-01167-7.

[61] A. Raj, A. Yadav, S. Arya, R. Sirohi, S. Kumar, A.P. Rawat, R.S. Thakur, D.K. Patel, L. Bahadur, A. Pandey. Preparation, characterization and agri applications of biochar produced by pyrolysis of sewage sludge at different temperatures. Science of The Total Environment, 795 (2021) 148722. https://doi.org/10.1016/j.scitotenv.2021.148722.

[62] B. Qiu, Q. Shao, J. Shi, C. Yang, H. Chu. Application of biochar for the adsorption of organic pollutants from wastewater: Modification strategies, mechanisms and challenges. Separation and Purification Technology, 300 (2022) 121925. https://doi.org/10.1016/j.seppur.2022.121925.

[63] S.A. Dhar, T.U. Sakib, L.N. Hilary. Effects of pyrolysis temperature on production and physicochemical characterization of biochar derived from coconut fiber biomass through slow pyrolysis process. Biomass Conversion and Biorefinery, 12(7) (2022) 2631-2647. https://doi.org/10.1007/s13399-020-01116-y.

[64] M.Z. Yameen, S.R. Naqvi, D. Juchelková, M.N.A. Khan. Harnessing the power of functionalized biochar: progress, challenges, and future perspectives in energy, water treatment, and environmental sustainability. Biochar, 6(1) (2024) 25. https://doi.org/10.1007/s42773-024-00316-3.

[65] L. Leng, H. Huang. An overview of the effect of pyrolysis process parameters on biochar stability. Bioresource Technology, 270 (2018) 627-642. https://doi.org/10.1016/j.biortech.2018.09.030.

[66] K. Zeng, D.P. Minh, D. Gauthier, E. Weiss-Hortala, A. Nzihou, G. Flamant. The effect of temperature and heating rate on char properties obtained from solar pyrolysis of beech wood. Bioresource Technology, 182 (2015) 114-119. https://doi.org/10.1016/j.biortech.2015.01.112.

[67] Y. Li, B. Xing, Y. Ding, X. Han, S. Wang. A critical review of the production and advanced utilization of biochar via selective pyrolysis of lignocellulosic biomass. Bioresource Technology, 312 (2020) 123614. https://doi.org/10.1016/j.biortech.2020.123614.

[68] X. Pan, Z. Gu, W. Chen, Q. Li. Preparation of biochar and biochar composites and their application in a Fenton-like process for wastewater decontamination: A review. Science of The Total Environment, 754 (2021) 142104. https://doi.org/10.1016/j.scitotenv.2020.142104.

Journal Pre-proof

[69] Z. Su, M. Duan, Y. Wang, H. Fan, Z. Tian, Z. Chen, C. Yang, J. Shangguan. In-situ preparation of Cu-BTC modified with organic amines for H2S removal under ambient conditions. Resources Chemicals and Materials, 2(2) (2023) 143-150. https://doi.org/10.1016/j.recm.2023.03.001.

[70] Y. Nie, C. Zhao, Z. Zhou, Y. Kong, J. Ma. Hydrochloric acid-modified fungi-microalgae biochar for adsorption of tetracycline hydrochloride: Performance and mechanism. Bioresource Technology, 383 (2023) 129224. https://doi.org/10.1016/j.biortech.2023.129224.

[71] A. Ahmad, N. Khan, B.S. Giri, P. Chowdhary, P. Chaturvedi. Removal of methylene blue dye using rice husk, cow dung and sludge biochar: Characterization, application, and kinetic studies. Bioresource Technology, 306 (2020) 123202. https://doi.org/10.1016/j.biortech.2020.123202.

[72] A. Herath, M. Salehi, S. Jansone-Popova. Production of polyacrylonitrile/ionic covalent organic framework hybrid nanofibers for effective removal of chromium(VI) from water. Journal of Hazardous Materials, 427 (2022) 128167. https://doi.org/10.1016/j.jhazmat.2021.128167.

[73] F. Wu, L. Chen, P. Hu, X. Zhou, H. Zhou, D. Wang, X. Lu, B. Mi. Comparison of properties, adsorption performance and mechanisms to Cd(II) on lignin-derived biochars under different pyrolysis temperatures by microwave heating. Environmental Technology & Innovation, 25 (2022) 102196. https://doi.org/10.1016/j.eti.2021.102196.

[74] X.-J. Liu, M.-F. Li, S.K. Singh. Manganese-modified lignin biochar as adsorbent for removal of methylene blue. Journal of Materials Research and Technology, 12 (2021) 1434-1445. https://doi.org/10.1016/j.jmrt.2021.03.076.

[75] Y. Sun, T. Wang, C. Han, X. Lv, L. Bai, X. Sun, P. Zhang. Facile synthesis of Fe-modified lignin-based biochar for ultra-fast adsorption of methylene blue: Selective adsorption and mechanism studies. Bioresource Technology, 344 (2022) 126186. https://doi.org/10.1016/j.biortech.2021.126186.

[76] X. Yao, L. Ji, J. Guo, S. Ge, W. Lu, L. Cai, Y. Wang, W. Song, H. Zhang. Magnetic activated biochar nanocomposites derived from wakame and its application in methylene blue adsorption. Bioresource Technology, 302 (2020) 122842. https://doi.org/10.1016/j.biortech.2020.122842.

[77] J. Xu, M. Fu, Q. Ma, X. Zhang, C. You, Z. Shi, Q. Lin, X. Wang, W. Feng. Modification of biochar by phosphoric acid via wet pyrolysis and using it for adsorption of methylene blue. RSC Advances, 13(22) (2023) 15327-15333. https://doi.org/10.1039/d3ra00680h.

[78] C.T. Primaz, A. Ribes-Greus, R.A. Jacques. Valorization of cotton residues for production of bio-oil and engineered biochar. Energy, 235 (2021) 121363. https://doi.org/10.1016/j.energy.2021.121363.

[79] Y. Mu, H. Ma. NaOH-modified mesoporous biochar derived from tea residue for methylene Blue and Orange II removal. Chemical Engineering Research and Design, 167 (2021) 129-140. https://doi.org/10.1016/j.cherd.2021.01.008.

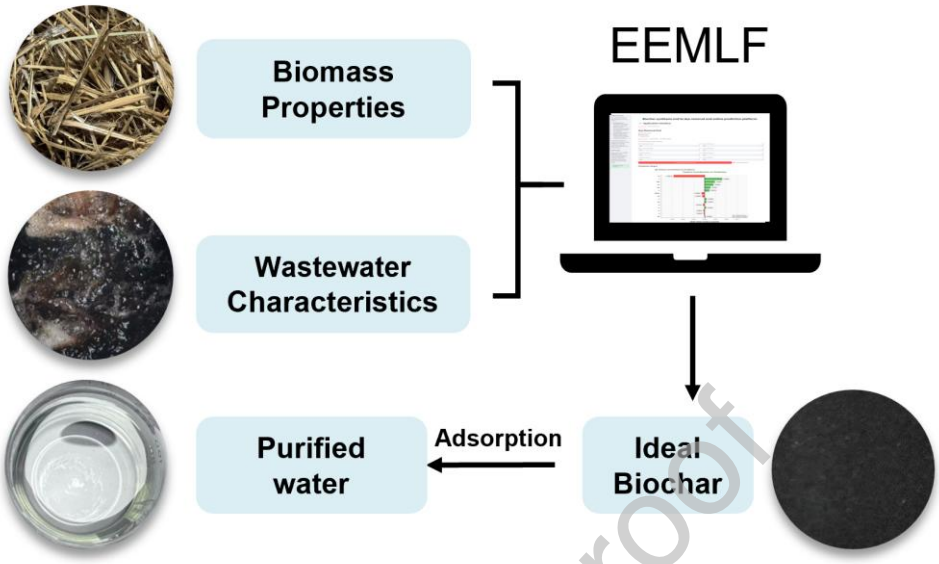
[80] A.E. Moharm, G.A. El Naeem, H.M.A. Soliman, A.I. Abd-Elhamid, A.A. El-Bardan, T.S. Kassem, A.A. Nayl, S. Bräse. Fabrication and Characterization of Effective Biochar Biosorbent Derived from Agricultural Waste to Remove Cationic Dyes from Wastewater. Polymers, 14(13) (2022) 2587. https://doi.org/10.3390/polym14132587.

[81] H. Nath, A. Saikia, P.J. Goutam, B.K. Saikia, N. Saikia. Removal of methylene blue from water using okra (Abelmoschus esculentus L.) mucilage modified biochar. Bioresource Technology Reports, 14 (2021) 100689. https://doi.org/10.1016/j.biteb.2021.100689.

[82] Y. Wang, Y. Zhang, S. Li, W. Zhong, W. Wei. Enhanced methylene blue adsorption onto activated reed-derived biochar by tannic acid. Journal of Molecular Liquids, 268 (2018) 658-666. https://doi.org/10.1016/j.molliq.2018.07.085.

Journal Pre-proof

Graphical abstract



Declaration of 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.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof