

## ARTICLE OPEN



# Using machine learning to predict the efficiency of biochar in pesticide remediation

Amrita Nighojkar<sup>1,2,8</sup>, Shilpa Pandey<sup>3,8</sup>, Mino Naebe<sup>4</sup>, Balasubramanian Kandasubramanian<sup>1</sup>, Winston Wole Soboyejo<sup>5</sup>, Anand Plappally<sup>6</sup> and Xungai Wang<sup>7</sup>

Pesticides have remarkably contributed to protecting crop production and increase food production. Despite the improved food availability, the unavoidable ubiquity of pesticides in the aqueous media has significantly threatened human microbiomes and biodiversity. The use of biochar to remediate pesticides in soil water offers a sustainable waste management option for agriculture. The optimal conditions for efficient pesticide treatment via biochar are aqueous-matrix specific and differ amongst studies. Here, we use a literature database on biochar applications for aqueous environments contaminated with pesticides and employ ensemble machine learning models (i.e., CatBoost, LightGBM, and RF) to predict the adsorption behavior of pesticides. The results reveal that the textural properties of biochar, pesticide concentration, and dosage were the significant parameters affecting pesticide removal from water. The data-driven modeling intervention offers an empirical perspective toward the balanced design and optimized usage of biochar for capturing emerging micro-pollutants from water in agricultural systems.

*npj Sustainable Agriculture* (2023)1:1; <https://doi.org/10.1038/s44264-023-00001-1>

## INTRODUCTION

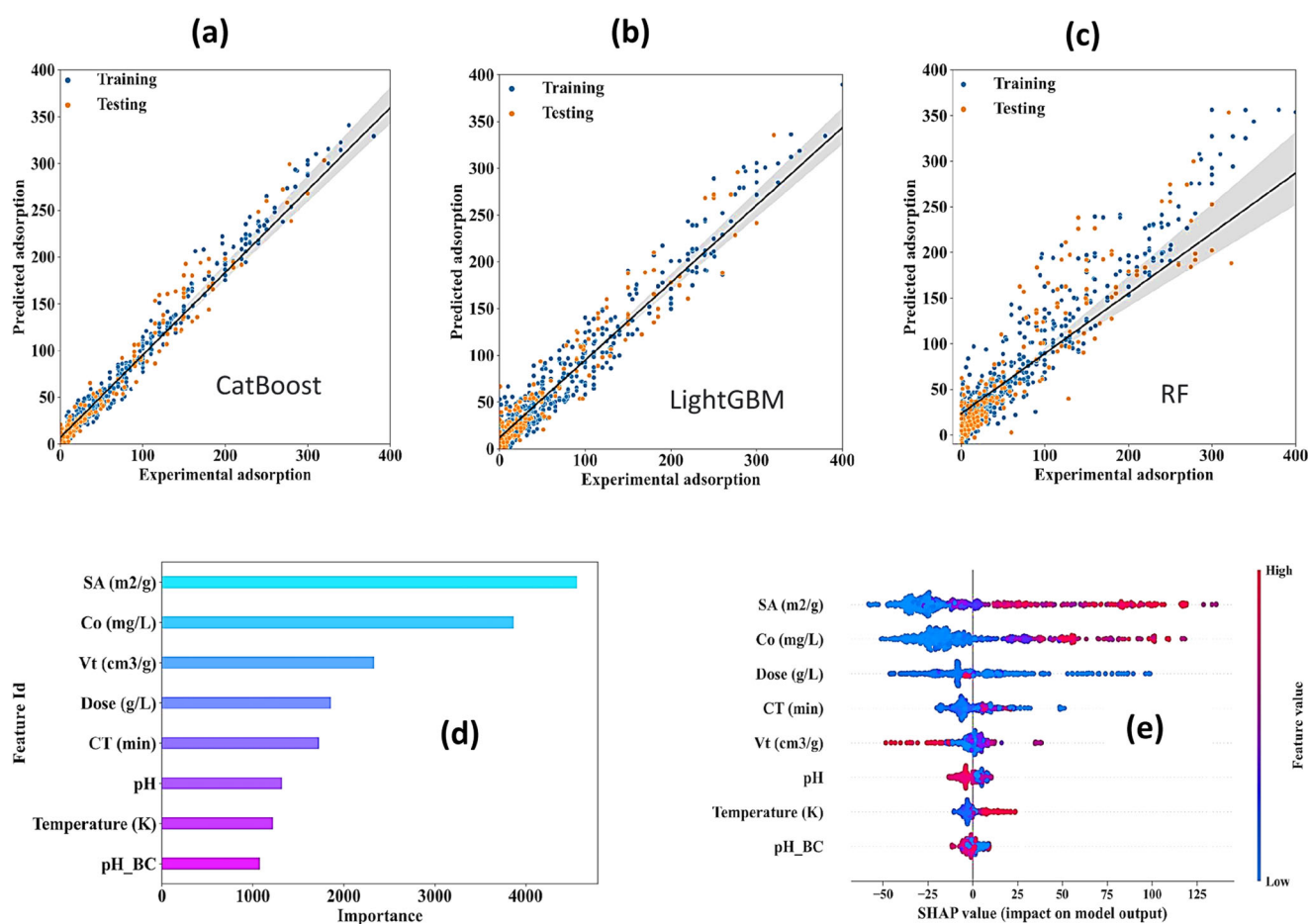
Water pollution caused by pesticides (PCs) is a major environmental concern that hinders the achievement of Sustainable Development Goals (UN), especially in emerging countries like India and China<sup>1</sup>. The rapid urbanization and industrialization that fuel agricultural modernization have accelerated the spread of pesticides in the ecosystem, posing serious threats to biota, water quality, and human health<sup>2</sup>. Anthropogenic activities, industrial wastewater, and agricultural runoff contribute to the contamination of surface and groundwater by pesticides<sup>1</sup>. These toxic and mobile contaminants eventually enter the food chain, leading to retinal degeneration, cardiovascular problems, muscle degeneration, cancer, and irreversible cell membrane damage<sup>3,4</sup>. Various ex-situ and in situ remediation processes have been implemented to treat water polluted with PCs. Among these, in-situ removal of PCs using biochar adsorption matrices has gained widespread recognition as a sustainable and cost-effective remediation procedure that is easy to design and efficient<sup>5,6</sup>.

Biochar has shown great potential for retrieving pesticides from contaminated ecosystems due to its superior physicochemical properties compared to primary feedstocks<sup>7</sup>. Biochar can be produced from various natural biological sources, such as agricultural residues, sludge, and animal manure, using different heating methods, such as fast pyrolysis, slow pyrolysis, hydrothermal carbonization, and gasification, to treat both organic and synthetic forms of pesticides, such as herbicides, insecticides, and fungicides in aqueous environments<sup>5</sup>. However, the PC adsorption efficiency of biochar-mediated aqueous systems depends on several factors, including feedstock choice, production conditions, water matrix (e.g., pH, concentration, temperature), experimental conditions (dose, treatment time), and pesticide properties (e.g., solubility, molecular weight)<sup>8</sup>. Numerous adsorption mechanisms,

such as hydrogen bonding, surface complexation, electrostatic interactions,  $\pi$ - $\pi$  interactions, Van der Waals' forces, and pore filling, have contributed to PC removal efficiency<sup>9-11</sup>. Despite extensive research, the optimal settings for maximizing PC adsorption in aqueous solutions via biochar remain highly variable. Concurrent adsorption studies examining the synergistic effect of all experimental parameters are challenging. Review studies, bibliometrics, and quantitative analyses have been used to assess the effectiveness of pesticide adsorption<sup>12-14</sup>. Still, these methods are time-consuming and complex in estimating the relative contribution of adsorption variables to removal efficiency. An empirical approach that uses a computational framework to optimize the factors related to biochar characteristics, experimental conditions, and aqueous matrix configuration and highlight their relative contribution to enhancing pesticide removal can lead to an overall understanding of complex adsorption behavior<sup>15</sup>. This approach can reduce the time, cost, and resources involved in biochar-based adsorption procedures and help determine the optimal conditions for maximum PC sequestration from water.

In the realm of chemical sciences, machine learning (ML) based data-driven models are gaining traction as powerful tools for promoting environmentally conscious research. Diverse machine learning techniques such as support vector machines (SVM), convolutional neural networks (CNN), random forests (RF), and artificial neural networks (ANN) have demonstrated effective utility in the identification of pesticides and their derivatives within real samples<sup>16-18</sup>. Additionally, ML models have been used to track the dissipation of pesticides in plants<sup>19</sup>, assess their effects on soil microbial communities<sup>20</sup>, and evaluate their potential genotoxic impact on humans<sup>21</sup>. However, the application of ensemble ML algorithms to predict the efficacy of biochar adsorption matrix in

<sup>1</sup>Nano Surface Texturing Laboratory, Defence Institute of Advanced Technology (DU), Pune, India. <sup>2</sup>Department of Technology Management, Defence Institute of Advanced Technology (DU), Pune, India. <sup>3</sup>Department of Computer Science and Engineering, Pandit Deendayal Energy University, Gujarat, India. <sup>4</sup>Institute for Frontier Materials, Deakin University, Waurn Ponds Campus, Geelong, Victoria, Australia. <sup>5</sup>Department of Mechanical Engineering, Worcester Polytechnic Institute, Massachusetts, USA. <sup>6</sup>Centre for Emerging Technologies for Sustainable Development, Indian Institute of Technology Jodhpur, Jodhpur, India. <sup>7</sup>School of Fashion and Textiles, The Hong Kong Polytechnic University, Hong Kong, Hongkong. <sup>8</sup>These authors contributed equally: Amrita Nighojkar, Shilpa Pandey. ✉email: minoo.naebe@deakin.edu.au; meetkbs@gmail.com



**Fig. 1 Performance of machine learning models and feature importance.** The figure presents the performance comparison of three machine learning models: **a** Categorical boost (Catboost), **b** Light Gradient Boosting Machine (LightGBM), and **c** Random Forest (RF) model in predicting pesticide adsorption capacity of biochars in aqueous system. In addition, **d** and **e** show feature importance insights using the CatBoost model and SHAP feature importance plots, respectively.

treating pesticides is not known. While various studies have been conducted to experimentally investigate pesticide remediation from water using different biochar systems and validate scientific hypotheses using classical sorption models, the exploration of machine-learning algorithms in this context remains unexplored<sup>22</sup>.

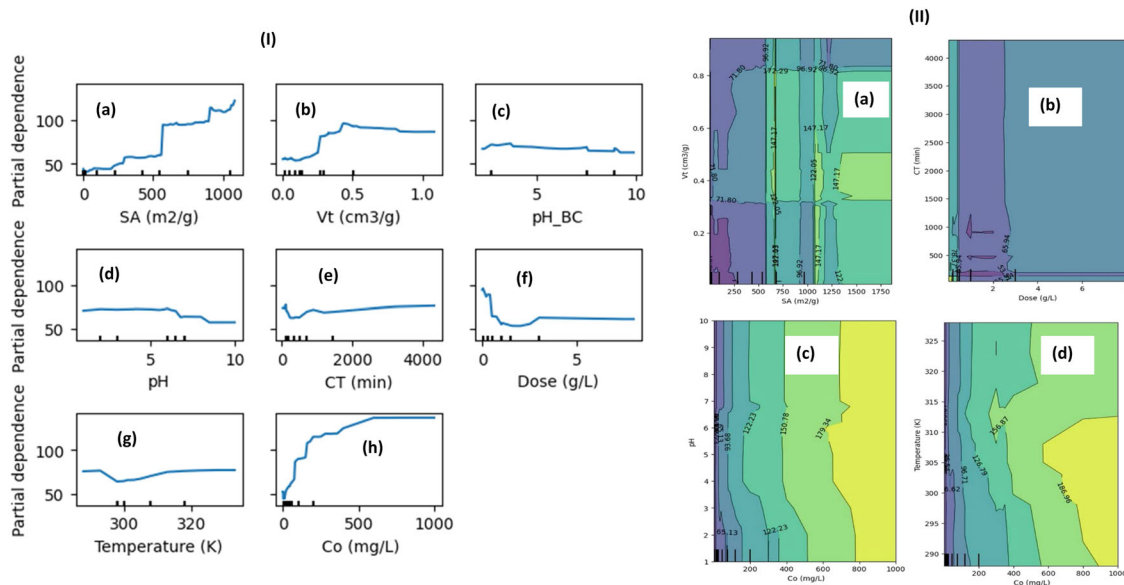
Furthermore, previous studies have utilized molecular simulations to forecast biochar adsorption<sup>23,24</sup>. However, these investigations have focused on capturing minuscule interactions rather than evaluating the impact of physicochemical parameters on the remediation of pesticides using bulk biochar. To fill the existing research gap, we utilize three ensemble machine learning (ML) models, namely Categorical boost (CatBoost), Light gradient boost machine (LightGBM), and Random forest (RF), to predict the adsorption efficiency of biochar in removing pesticides from aqueous environments. While traditional statistical methods can establish linear or quadratic relationships between an individual independent variable and the target variable, ML models can simultaneously consider all correlated features and establish a more complex relationship with the output variable. Moreover, the developed models can assist researchers in designing experiments for biochar-based pesticide remediation. In this study, we considered ten adsorption attributes linked to three different aspects of the adsorbate-adsorbent system, namely (i) biochar properties, (ii) aqueous matrix configuration, and (iii) experimental conditions, and assessed their impact on pesticide adsorption in biochar-treated aqueous media.

## RESULTS

### Modeling pesticide adsorption on biochars

Decision-tree-based ensemble machine learning (ML) models, including CatBoost, LightGBM, and Random forest (RF), were utilized to predict the efficacy of adsorptive sequestration of pesticides from an aqueous medium using biochar. These models integrate multiple individual models to enhance the accuracy of predictions. The decision tree-based model flow utilized eight input attributes to split the data into smaller subsets based on specific features. This process was recursively repeated to achieve a decision or prediction. Grid search optimization was employed to determine the best possible combination of hyperparameters, minimizing the root-mean-squared error (RMSE) on both the train and test sets. More information regarding the hyperparameters of the fine-tuned models can be found in Supplementary Table 1 in the supplementary information (SI).

Figure 1a–c shows scatter plots that juxtapose the experimentally determined pesticide adsorption efficacy values with those predicted by the model. The CatBoost framework yielded the highest regularization coefficient ( $R^2$ ) values of 0.968 and 0.956 for the training and test subsets, correspondingly surpassing LightGBM ( $R^2_{\text{train}}=0.931$ ,  $R^2_{\text{test}}=0.862$ ) and Random forest ( $R^2_{\text{train}}=0.820$ ,  $R^2_{\text{test}}=0.796$ ) models. The CatBoost was the best-performing model due to less variation between the train and test losses compared to LightGBM, while the RF predictive error was very high than the loss functions of boosting models (refer SI, Supplementary Fig. 1). The feature importance graph



**Fig. 2 Attribute analysis for pesticide adsorption on biochar.** (i) 1D-PDP of Model Attributes: One-dimensional Partial Dependence Plots (1D-PDP) showcasing the relationship between individual model attributes and pesticide adsorption capacity. (ii) 2D-PDP on Pesticide Adsorption Capacity: Two-dimensional Partial Dependence Plots (2D-PDP) exploring the combined effects of model attributes on pesticide adsorption capacity of biochars.

(Fig. 1d) indicated that surface area (SA) had the most significant effect on the model, followed by pesticide concentration ( $C_o$ ) and pore volume, respectively. Figure 1e demonstrates the SHAP values, which visually represent each feature's relative contributions to the final predictions generated by the CatBoost model.

These results suggested the superiority of CatBoost in predicting pesticide adsorption on biochars. The influence of each input feature on the output or PC adsorption performance in the CatBoost model, which delivered the best performance, was assessed utilizing both the CatBoost-built-in feature importance criteria and Shapley Additive exPlanations. Figure 3d exhibits the feature significance of the trained model based on the total percentage gain with the feature. The findings of this study highlight surface area (SA) as the most significant physicochemical characteristic affecting pesticide adsorption on biochar, followed by pesticide concentration ( $C_o$ ), pore volume (Vt), and biochar dose (Fig. 1d, e).

#### Model feature insights and their implication on sustainability in agriculture

The partial dependence plots (PDP), depicted in Fig. 2i, II, offer valuable insights into the impact of biochar attributes (SA, Vt, pH\_BC), experimental factors (CT, dose), and aqueous conditions (pH,  $C_o$ , and T) on the biochar capacity to treat pesticide-contaminated water in agricultural fields. Among the biochar attributes, the textural characteristics of biochar, especially surface area, have the most significant influence on the model prediction results. In contrast, the biochar pH contributes the least among all the input parameters. It can be observed that the increment in the surface area (SA) and pore volume (Vt) demonstrate a direct correlation with the increase in the pesticide adsorption capacity of biochar within a specific range (0.25–1000  $m^2/g$ , 0.004–0.5  $cm^3/g$ ) (Fig. 2I-a, Fig. 2II-a). The expansion in the surface area leads to an increased number of active adsorption sites<sup>25–28</sup>. At the same time, the interconnected pore structure of well-structured biochar facilitates efficient diffusion and transfer of adsorbate molecules through its pores, thereby augmenting adsorption performance<sup>29</sup>. This information can help researchers, agricultural scientists, and biochar producers optimize biochar production processes for targeted remediation of specific pesticides, thereby promoting the

responsible and efficient use of biomass feedstocks to produce biochar with effective textural properties.

Among the experimental factors, biochar dose significantly influenced the model predictions (Fig. 2I-f). The pesticide adsorption capacity of biochar decreased with the increasing biochar dose and was found to be highest for a biochar dose of less than 1 g/L. As the quantity of biochar increases, the available surface area per unit mass decreases, or pore blockage occurs due to particles' aggregation, which reduces the accessibility of active adsorption sites, resulting in reduced adsorption capacity<sup>30</sup>. The biochar dose of less than 1 g/L with a treatment duration of fewer than 500 min provided the biochar's maximum pesticide adsorption capacity (Fig. 2II-b). These findings can aid in optimizing the biochar dose with the treatment time, which maximizes the removal of specific pesticides from agricultural water. By optimizing the biochar dose and contact time during pesticide treatment, the residual biochar in the treated water can be utilized for subsequent soil amendment<sup>31,32</sup>.

Among the water matrix parameters (i.e.  $C_o$ , pH, and T), pesticide concentration significantly impacts the model predictions, followed by pH and water temperature. The biochar adsorption capacity showed a linear increment with the increasing concentration of pesticide in water (0.2–2000 mg/L) (Fig. 2I-h). Higher pesticide concentrations result in steeper concentration gradients between the solution and the biochar surface. This concentration gradient promotes more efficient mass transfer of pesticides from the solution to the biochar, allowing for increased adsorption capacity<sup>33</sup>. The model results can help in optimizing the water matrix parameters for maximum adsorption of pesticides on Biochar (Fig. 2II-c, Fig. 2II-d). The model can improve water quality near agricultural fields by reducing pesticide runoff and subsequent contamination.

With the acquired insights into the capacity of biochar to adsorb pesticides by manipulating model attributes across a wide range of parameters, machine learning-mediated modeling can be pivotal in advancing sustainable agricultural practices. This can be achieved through optimizing resource utilization to achieve a well-balanced biochar design, enhancing soil health via applying biochar for soil amendments, and mitigating pesticide runoff and environmental contamination. The present data utilized to train

the model primarily comprises experiments conducted on laboratory-scale solutions. Nonetheless, to enhance the predictive capability of the machine learning model, it is essential to incorporate real-world data obtained from diverse agricultural scenarios and environments, thus ensuring its applicability and generalizability.

## DISCUSSION

### Benefits of the ML intervention and its pathway to impact agriculture

Utilizing a machine learning framework to predict pesticide removal from agricultural systems using biochar holds significant advantages for various stakeholders in the agricultural sector. Firstly, agrarian practitioners can significantly benefit from adopting machine learning models to assess pesticide removal using biochar. By doing so, farmers can gain valuable insights into the performance and effectiveness of biochar-based remediation methods. With this knowledge, individuals can make well-informed decisions when implementing these techniques to address irrigation water concerns. This approach not only minimizes pesticide exposure but also fosters sustainable agricultural practices.

Moreover, researchers investigating sustainable agricultural practices and water quality can leverage this technology to their advantage. By utilizing predictions generated by the machine learning model, they can explore the impact of biochar on pesticide adsorption and its subsequent effects on soil and water systems. Through extensive data analysis, researchers can develop robust models capable of predicting the efficiency of biochar in removing specific pesticides from water. Consequently, this promotes subsequent investigations aimed at optimizing the utilization of biochar as an adsorbent and gaining insights into the kinetics of pesticide elimination. Ultimately, this enhances our comprehension of the intricate interactions among biochar characteristics, soil water chemistry, and pesticide behavior.

Furthermore, biochar producers can harness this technology to tailor their products for adsorbing pesticides in agricultural runoffs. Producers can optimize their production processes by understanding the relationship between biochar characteristics and pesticide adsorption efficiency and offer highly effective biochar products for pesticide treatment. This contributes to biochar's commercialization and drives its adoption as a sustainable solution for treating agricultural water and reducing pesticide contamination.

Lastly, environmental agencies can rely on the machine learning model to assess the potential impacts of pesticides from agricultural runoffs on water bodies. By incorporating the predicted adsorption efficiency of biochar, these agencies can evaluate the effectiveness of different biochar types in reducing pesticide concentrations in soil and water. The resulting knowledge enables the development of more targeted policies and guidelines to mitigate pesticide contamination, thereby safeguarding the environment and public health.

The pathway to impact in agriculture involves empowering stakeholders with knowledge and tools to make informed decisions about agricultural water treatment, pesticide management, and water resource protection. The technology helps optimize the use of biochar with a balanced design as an effective adsorbent for pesticide removal, promoting sustainable agricultural practices, minimizing environmental contamination, and supporting soil and water conservation efforts.

The application of biochar to remediate pesticides from agricultural runoffs offers a sustainable waste management and water treatment solution. The permeation of ensemble machine learning tools can help in the balanced design and use of biochar-based pesticide removal processes. This study demonstrates the

potential of the CatBoost framework to derive insights from the biochar adsorption data and help forecast the relevant attributes impacting the retrieval of pesticides from aqueous solutions. The CatBoost reveals the dominance of textural characteristics, pesticide concentration, and biochar dose on the adsorptive capture of pesticides. As more pertinent data related to pesticide properties and selectivity parameters of biochar are available, the CatBoost technique, with the acquired prediction accuracy of ~96%, can be extended into real-world treatment systems to formulate a more holistic model framework that will help researchers and agricultural scientists to predict the potential of biochar to treat specific pesticides in water and soil systems. Moreover, the current theme of the research focuses on the role of ML in analyzing pesticide adsorption behavior, but in the future, it will be intriguing to explore the role of ML-assisted management models to address the post-adsorption environmental challenges related to desorbed pesticides and the management of pesticide saturated biochars.

## METHODS

The sections below briefly discuss the pre-processing steps in the pesticide treatment data preparation using biochars, model development, and feature importance analysis.

### Pesticide remediation dataset based on biochars

**Compilation of data.** A comprehensive collection of 96 academic research articles was compiled to investigate the adsorption of pesticides in biochar-mediated aqueous systems. The articles were gathered using reputable online search engines such as Google Scholar, Scopus Index, and Web of Science. Pesticides treated using biochar adsorption matrix methods are depicted using a tag cloud (Fig. 3). Supplementary Table 2 provides an elaborate tabular representation of feedstock biomass data, while Supplementary Fig. 2 offers a visual overview of the biochar dataset. A comprehensive and detailed description of the gathered data is presented in an Excel file, which can be accessed through a downloadable link provided in Supplementary Table 3. The file contains extensive information about the specific biochar utilized for pesticide treatment, research locations, publication years, and the dataset.

The information about the PC adsorption capacity from the collected research articles was extracted from graphs using Plot Digitizer Software<sup>34</sup> or was acquired from tables or calculated using Eq. (1).

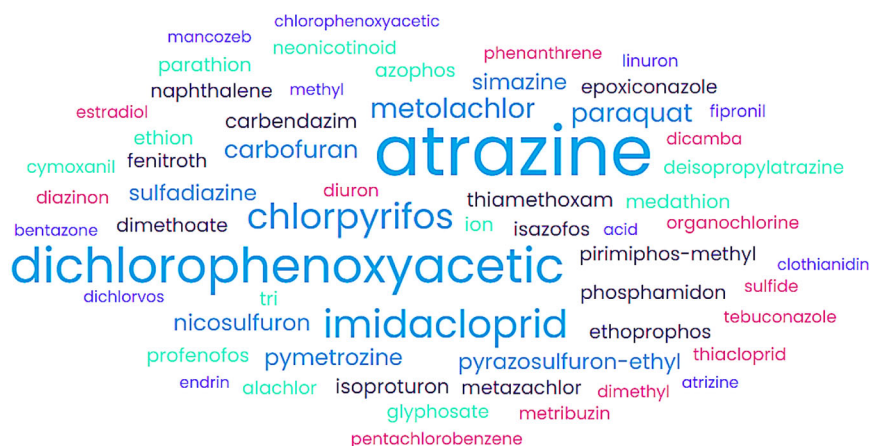
$$PC \text{ adsorption capacity} = \frac{(C_{PO} - C_{PB}) * V}{C_{PO} * m} \quad (1)$$

where  $C_{PO}$  refers to the initial PC concentration in the solution and  $C_{PB}$  is the PC concentration post-biochar treatment at time  $t$  of the pesticide solution.  $V$  designates the volume of the aqueous solution, and  $m$  denotes the mass of the biochar. The pesticide adsorption capacity was expressed in  $\text{mg g}^{-1}$ . The biochar-mediated adsorption process in PC-contaminated solutions describes a decrease in their concentration, expressed in  $\text{mg L}^{-1}$ . The domain knowledge was used to group the ten input attributes, i.e., the surface area of Biochar (SA), pH of biochar (pH\_BC), total pore volume (Vt), biochar dose (dose), cation exchange capacity (CEC), pH of the solution (pH), initial pesticide concentration ( $C_0$ ), treatment or contact time (CT) and temperature (T) into empirical categories of the adsorbate-adsorbent system mentioned in the Introduction.

### Imputation of missing data and assessment of correlations

The research studies did not report all the values associated with the ten sets of attributes. The missing value imputation was used for the feature for which most data was available. In this work, a threshold of





**Fig. 3** Visualization of a cloud representing various pesticides adsorbed on biochar, with further details available in Supplementary Table 2.

**Table 1.** Representation of data variability using the quartile range (QR).

QR	Input parameters								Output parameters
	SA (m <sup>2</sup> /g)	Vt (cm <sup>3</sup> /g)	pH_BC	pH	CT (min)	Dose (g/L)	T (K)	C <sub>o</sub> (mg/L)	Pesticide adsorption (mg/g)
Minimum	0.25	0.0024	2.29	2	40	0.001	288	0.2	0.1
First Quartile	15.08	0.0751	2.73	5	180	0.4	298	10	6
Median	275	0.134	6.24	6.5	360	1	303	25	25
Third Quartile	660	0.296	8.91	7	720	1	308	100	95.83
Maximum	2192	1.085	10.12	10	4320	8	333	1000	1120

SA surface area, Vt total pore volume, pH\_BC biochar pH, pH pH of water, CT contact time between biochar and pesticide-contaminated water, dose quantity of biochar used, T temperature condition, Co pesticide concentration in water.

50% was employed, meaning attributes with missing values greater than the threshold were discarded from the dataset. Thus the cation exchange capacity (CEC) attribute was excluded from the dataset, and the median value was employed to impute missing values for the remaining features. The median was selected as it demonstrated the highest representativeness for the comprehensive data and did not introduce bias. Pearson's correlation matrix, consisting of nine sets of features, is presented in Supplementary Fig. 3. Subsequently, the dataset matrix with dimensions of 878 × 9, encapsulating data variability as shown in Table 1, was employed to generate ensemble machine learning (ML) models.

### Development and evaluation of models

**Data division, scaling, and hyperparameter tuning.** Predictive models were developed using three machine-learning techniques: CatBoost, LightGBM, and Random Forest (RF). Figure 4 showcases a flowchart that provides an overview of the technical aspects of the machine learning pipeline. This flowchart encompasses the steps involved in data preparation, model training, validation, and, ultimately, the prediction of pesticide adsorption. A sample space was established to perform ML modeling, comprising 878 data records encompassing nine attributes: SA, Vt, pH\_BC, pH, CT, Dose, T, Co, and the desired output attribute, Pesticide Adsorption. By using a stratified random sampling algorithm, 703 samples were chosen as the training set, and 175 samples were designated as the test set from the previously mentioned sample space. The grid search algorithm was utilized to identify the optimal hyperparameters in the training process for each specific model (CatBoost, LightGBM, and RF). The fundamentals of each ML model are described in Supplementary Methods in SI. To mitigate the risk of overfitting in our study, we implemented a 5-fold cross-validation

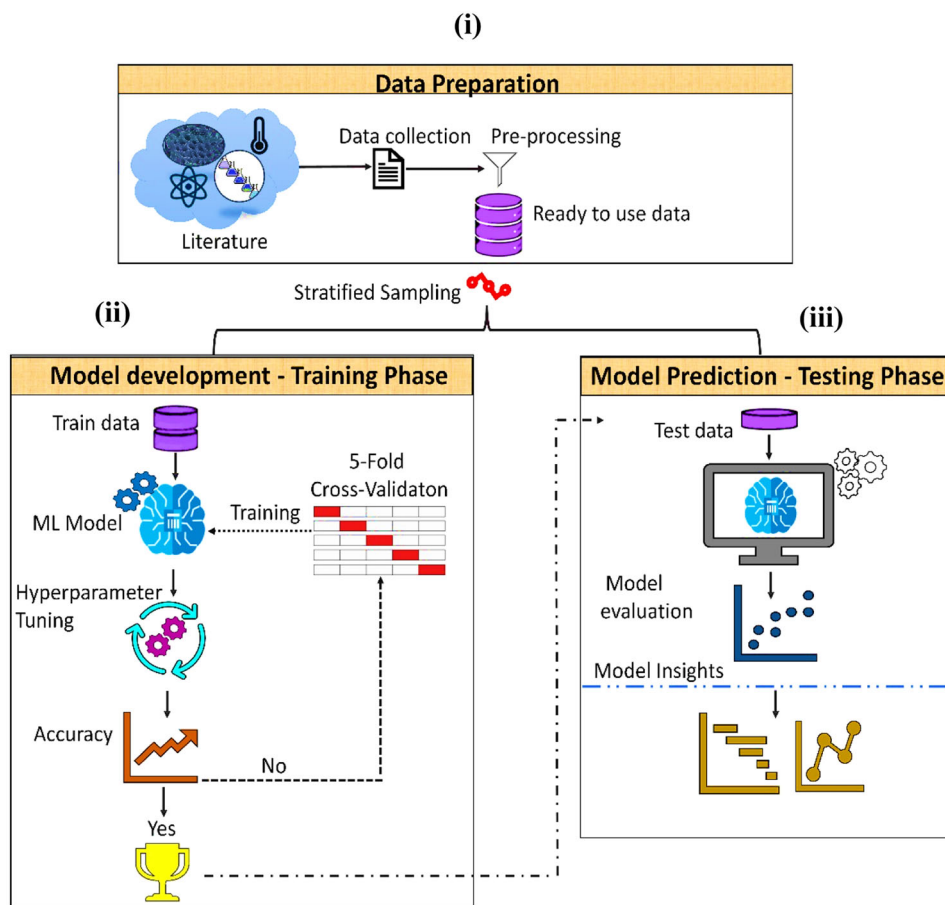
approach. This technique was employed to assess the predictive performance of our models on unseen data samples, thereby enhancing the reliability and generalizability of our results. Finally, the precision of the predictive models was measured using the coefficient of regularization ( $R^2$ ) and the root-mean-squared error (RMSE).  $R^2$  and RMSE values were computed using Eqs. (2) and (3).

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_{m,i} - y_{e,i})^2}{\sum_{i=1}^N (y_{e,i} - y_{m,av})^2} \quad (2)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_{m,i} - y_{e,i})^2}{n}} \quad (3)$$

where  $y_{m,i}$  associates with the value of output predicted by the model and  $y_{e,i}$  corresponds to the output value acquired using experiments,  $y_{m,av}$  is the average of all the outcomes predicted by the model, and  $n$  is the sample size in the training or testing subsets. We utilized the Python programming language and relevant libraries for modeling implementation. The Scikit-Learn library<sup>31</sup> was employed for performance analysis, while the conda packages of CatBoost, LightGBM, and RF libraries were utilized for implementing the model algorithms. The pyplot libraries were also used for exploratory data analysis and graph generation.

**Feature importance.** The significance and impact of each attribute (feature) on the output attribute (pesticide adsorption capacity) were computed in two ways in Python. First, the built-in feature importance function provided by the CatBoost library was



**Fig. 4** illustrates the machine learning workflow employed to assess the potential of a biochar-mediated aqueous system for pesticide treatment. The first block (i) represents the compilation of pesticide adsorption data, i.e., biochar's textural properties, water matrix parameters, experimental factors, and adsorption capacity, followed by data pre-processing for model development. The second block (ii) outlines the steps involved in the model training process, where a five-fold cross-validation methodology was implemented to address overfitting concerns. The third block (iii) pertains to model performance evaluation using the test data and the utilization of the best predictive model for feature exploration.

used to plot features in the order of their impact on model fit. The systematic correlation between the input adsorption attributes and output was illustrated by integrating partial dependence plots generated using CatBoost. Another feature importance criterion, known as SHAP (Shapley's Addition Description Method)<sup>35,36</sup>, was also employed to anticipate the impact of various attributes on the model predictions.

#### DATA AVAILABILITY

The data used in this study can be downloaded from the following GitHub link: [GitHub [https://github.com/shilpapedey/Pesticide\\_study](https://github.com/shilpapedey/Pesticide_study)].

#### CODE AVAILABILITY

The source code is currently employed to conduct an in-depth analysis of comprehensive data and will be provided upon request from the corresponding author.

Received: 5 May 2023; Accepted: 6 July 2023;

Published online: 04 October 2023

#### REFERENCES

- Tang, F. H. M., Lenzen, M., McBratney, A. & Maggi, F. Risk of pesticide pollution at the global scale. *Nat. Geosci.* **14**, 206–210 (2021).
- Madhav, S. et al. in *Sensors in Water Pollutants Monitoring: Role of Material* (eds Pooja, D., Praveen Kumar, P., Singh, P. & Patil, S.) 43–62 (Springer, 2020).
- de Barros Rodrigues, M. et al. Association between exposure to pesticides and allergic diseases in children and adolescents: a systematic review with meta-analysis. *J. Pediatrics* **98**, 551–564 (2022).
- Sandoval-Insauti, H. et al. Intake of fruits and vegetables according to pesticide residue status in relation to all-cause and disease-specific mortality: results from three prospective cohort studies. *Environ. Int.* **159**, 107024 (2022).
- Varjani, S., Kumar, G. & Rene, E. R. Developments in biochar application for pesticide remediation: current knowledge and future research directions. *J. Environ. Manag.* **232**, 505–513 (2019).
- Dai, Y., Zhang, N., Xing, C., Cui, Q. & Sun, Q. The adsorption, regeneration and engineering applications of biochar for removal organic pollutants: a review. *Chemosphere* **223**, 12–27 (2019).
- Blanco-Canqui, H. Biochar and water quality. *J. Environ. Quality* **48**, 2–15 (2019).
- Yavari, S., Malakhammad, A. & Sapari, N. B. Biochar efficiency in pesticides sorption as a function of production variables—a review. *Environ. Sci. Pollut. Res.* **22**, 13824–13841 (2015).
- Zhang, P., Sun, H., Yu, L. & Sun, T. Adsorption and catalytic hydrolysis of carbaryl and atrazine on pig manure-derived biochars: Impact of structural properties of biochars. *J. Hazardous Mater.* **244–245**, 217–224 (2013).
- Suo, F., You, X., Ma, Y. & Li, Y. Rapid removal of triazine pesticides by P doped biochar and the adsorption mechanism. *Chemosphere* **235**, 918–925 (2019).
- Yoon, J.-Y. et al. Assessment of adsorptive behaviors and properties of grape pomace-derived biochar as adsorbent for removal of cymoxanil pesticide. *Environ. Technol. Innovation* **21**, 101242 (2021).
- Anupam, K. et al. In *Pesticides Remediation Technologies from Water and Wastewater* (eds Dehghani, M. H., Karri, R. R. & Anastopoulos, I.) 77–95 (Elsevier, 2022).

13. Mehmood, K. et al. Biochar research activities and their relation to development and environmental quality. A meta-analysis. *Agronomy Sustainable Dev.* **37**, 1–15 (2017).
14. Mandal, A. & Singh, N. Optimization of atrazine and imidacloprid removal from water using biochars: Designing single or multi-staged batch adsorption systems. *Int. J. Hygiene Environ. Health* **220**, 637–645 (2017).
15. Nighojkar, A. et al. Application of neural network in metal adsorption using biomaterials (BMs): a review. *Environ. Sci.: Adv.* **2**, 11–38 (2023).
16. Feng, C. et al. Evaluation and application of machine learning-based retention time prediction for suspect screening of pesticides and pesticide transformation products in LC-HRMS. *Chemosphere*. **271**, 129447 (2021).
17. Richardson, A. K. et al. Rapid direct analysis of river water and machine learning assisted suspect screening of emerging contaminants in passive sampler extracts. *Analytical Methods*. **13**, 595–606 (2021).
18. Tun, W. S. T. et al. A machine learning colorimetric biosensor based on acetylcholinesterase and silver nanoparticles for the detection of dichlorvos pesticides. *Mater. Chem. Front.* **6**, 1487–1498 (2022).
19. Shen, Y., Zhao, E., Zhang, W., Baccarelli, A. A. & Gao, F. Predicting pesticide dissipation half-life intervals in plants with machine learning models. *J. Hazardous Mater.* **436**, 129177 (2022).
20. Ke, M. et al. Development of a machine-learning model to identify the impacts of pesticides characteristics on soil microbial communities from high-throughput sequencing data. *Environ. Microbiol.* **24**, 5561–5573 (2022).
21. Wang, L. et al. Ensemble machine learning to evaluate the in vivo acute oral toxicity and in vitro human acetylcholinesterase inhibitory activity of organophosphates. *Arch. Toxicol.* **95**, 2443–2457 (2021).
22. Liang, L. et al. Review of organic and inorganic pollutants removal by Biochar and biochar-based composites. *Biochar* **3**, 255–281 (2021).
23. Mandal, A., Kumar, A. & Singh, N. Sorption mechanisms of pesticides removal from effluent matrix using Biochar: conclusions from molecular modelling studies validated by single-, binary and ternary solute experiments. *J. Environ. Manag.* **295**, 113104 (2021).
24. Mroziak, W. et al. Valorisation of agricultural waste derived biochars in aquaculture to remove organic micropollutants from water—experimental study and molecular dynamics simulations. *J. Environ. Manag.* **300**, 113717 (2021).
25. Taheri, E., Fatehizadeh, A., Lima, E. C. & Rezakazemi, M. High surface area acid-treated biochar from pomegranate husk for 2, 4-dichlorophenol adsorption from aqueous solution. *Chemosphere*. **295**, 133850 (2022).
26. Ahmed, M. J., Okoye, P. U., Hummadi, E. H. & Hameed, B. H. High-performance porous biochar from the pyrolysis of natural and renewable seaweed (*Gelidium acerosa*) and its application for the adsorption of methylene blue. *Bioresource Technol.* **278**, 159–164 (2019).
27. Qu, J. et al. KOH-activated porous biochar with high specific surface area for adsorptive removal of chromium (VI) and naphthalene from water: Affecting factors, mechanisms and reusability exploration. *J. Hazardous Materials* **401**, 123292 (2021).
28. Chen, S. et al. Study on the adsorption of dyestuffs with different properties by sludge-rice husk biochar: adsorption capacity, isotherm, kinetic, thermodynamics and mechanism. *J. Mol. Liquids* **285**, 62–74 (2019).
29. Piai, L. et al. Diffusion of hydrophilic organic micropollutants in granular activated carbon with different pore sizes. *Water Res.* **162**, 518–527 (2019).
30. Saarela, T., Lafdani, E. K., Laurén, A., Pumpanen, J. & Palviainen, M. Biochar as adsorbent in purification of clear-cut forest runoff water: Adsorption rate and adsorption capacity. *Biochar* **2**, 227–237 (2020).
31. Beckinghausen, A. et al. Post-pyrolysis treatments of biochars from sewage sludge and A. meamsii for ammonia (NH<sub>4</sub>-n) recovery. *Appl. Energy* **271**, 115212 (2020).
32. Yavari, S. et al. Bio-efficacy of imidazolinones in weed control in a tropical paddy soil amended with optimized agrowaste-derived biochars. *Chemosphere*. **303**, 134957 (2022).
33. Bisaria, K., Singh, R., Gupta, M., Mathur, A. & Dixit, A. Novel acoustic-activated alkali-functionalized *Trapa bispinosa* peel biochar for green immobilization of chlorpyrifos from wastewater: artificial intelligence modelling and experimental validation. *Bio-mass Conv. Bioref* <https://doi.org/10.1007/s13399-022-02898-z> (2022).
34. WebPlotDigitizer - Copyright 2010-2022 Ankit Rohatgi, (n.d.). <https://apps.automeris.io/wpd/> (accessed 22 February 2023).
35. Shahbeik, H. et al. Characterizing sludge pyrolysis by machine learning: towards sustainable bioenergy production from wastes. *Renew. Energy* **199**, 1078–1092 (2022).
36. Yang, H. et al. Predicting heavy metal adsorption on soil with machine learning and mapping global distribution of soil adsorption capacities. *Environ. Sci. Technol.* **55**, 14316–14328 (2021).

## ACKNOWLEDGEMENTS

The authors would like to express sincere gratitude to the Vice Chancellor, Dr. C.P. Ramanarayanan, and the esteemed Nano Surface Texturing Lab members at the Defence Institute of Advanced Technology for their invaluable support and guidance throughout the research process. Their unwavering commitment to academic excellence and dedication to fostering a vibrant research environment has significantly contributed to the successful completion of this study.

## AUTHOR CONTRIBUTIONS

The manuscript was written with the contributions of all authors. All authors have approved the final version of the manuscript. Amrita Nighojkar and Shilpa Pandey designed and executed the model simulations and wrote the original draft. Minoo Naebe and Balasubramanian Kandasubramanian conceptualized the research work and performed the formal analysis. Winston Soboyejo, Anand Plappally, and Xungai Wang prepared the data and reviewed and edited the manuscript.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s44264-023-00001-1>.

**Correspondence** and requests for materials should be addressed to Minoo Naebe or Balasubramanian Kandasubramanian.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



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

© The Author(s) 2023