Supporting Information

Prediction of soil heavy metal immobilization by biochar using machine learning

Kumuduni N. Palansooriya*,‡, Jie Li‡, Pavani D. Dissanayake,*,h, Manu Suvarna‡, Lanyu Li‡, Xiangzhou Yuan*, Binoy Sarkar*, Daniel C. W. Tsang‡, Jörg Rinklebe‡, Xiaonan Wang*,**, Yong Sik Ok*,*

*Korea Biochar Research Center, APRU Sustainable Waste Management Program & Division of Environmental Science and Ecological Engineering, Korea University, Seoul 02841, South Korea

‡Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117585, Singapore

*Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117585, Singapore

†Lancaster Environment Centre, Lancaster University, Lancaster LA1 4YQ, United Kingdom

‡Department of Chemical and Biomolecular Engineering, National University of Singapore, Singapore 117585, Singapore

§Department of Chemical Engineering, Tsinghua University, Beijing 100084, China

*Corresponding author: Email: yongsikok@korea.ac.kr (Y. S. Ok)

**Co-corresponding author: wangxiaonan@tsinghua.edu.cn (X. Wang)

††These authors contributed equally.

Summary:

Number of Pages: 15; Number of Tables: 3; Number of Figures: 8
Statistical data analysis

Prior to the development of the ML models, a thorough statistical analysis was conducted based on data collected from the literature. This analysis investigated the correlation between input and output variables. Fig. S1 presents the data analysis for HM immobilization efficiency alongside the N content in biochar, SA, and the biochar application rate.

The median HM immobilization efficiency was >90% when N content was >5%. This was mainly due to the availability of more N-containing functional groups (such as amines) on the biochar surface, which efficiently immobilize HMs in soils. Fig. S2a shows that C and O were the predominant elements in biochar, with median values of 72.3% and 11.04%, respectively; however, the N content was relatively low, with a median of 1.5%. Generally, the N content in biochar is very low, excluding biochars produced from feedstock with high N (e.g., biochars from poultry and swine manure). However, a positive correlation between HM immobilization and N content (Fig. S1a) highlights the significance of higher N content in biochar for better HM immobilization.

The SA of biochar is a critical physicochemical property for HM immobilization. The highest median value of HM immobilization was 81.10% when the biochar SA ranged from 300–950 m$^2$·g$^{-1}$ (Fig. S1b). This is indicative of a positive correlation between biochar SA and HM immobilization. A HM immobilization efficiency >50% was observed at an even lower SA (0.2 – 2.5 m$^2$·g$^{-1}$). This suggested that SA may not be the only governing factor for HM immobilization, and other parameters contribute to enhanced HM immobilization. For instance, H$_3$PO$_4$-modified biochars produced at higher temperatures exhibited a higher SA compared to lower temperature biochars. Interestingly, the high adsorption of Cu(II) and Cd(II) was observed in biochar with a lower SA. This suggests that the availability of surface oxygen-
containing functional groups is largely attributable to metal immobilization through surface complexation, despite the higher SA\textsuperscript{5}. The highest median of HM immobilization was achieved (88.08\%) when the biochar application rate was between 5\% and 10\% (Fig. S1c). The lowest median values were observed when the biochar application rate was <3\% (Fig. S1c). This suggests a positive correlation between the biochar application rate and HM immobilization.

A van Krevelen diagram (Fig. S2b), constructed for various biochars, clearly demonstrates very low H/C and O/C ratios for biochars produced at high temperatures compared to those produced at lower temperatures. The lower H/C and O/C ratios correspond to the high dehydration and deoxygenation of biomass during biochar production\textsuperscript{6}. Moreover, increasing pyrolysis temperature may cause the depolymerization of plant-based biomass into smaller, dissociated lignin and cellulose products\textsuperscript{6}. A decrease in the H/C and O/C ratios indicates an increase in aromaticity and a reduction in the surface polarity of biochar\textsuperscript{7, 8}.

Conversely, biochars produced at high pyrolysis temperatures were highly stable\textsuperscript{9}; the majority of biochar used in this study was produced at higher temperatures (Fig. S2b). The high pyrolysis temperatures increased the ash content in biochar, facilitating HM immobilization in soils\textsuperscript{7}. The observed median value for ash content was 16.89\% (Fig.S2a). The increased ash content in biochar was responsible for the higher biochar pH and the co-precipitation of HMs with cations and anions such as Mg\textsuperscript{2+}, Ca\textsuperscript{2+}, and PO\textsubscript{4}\textsuperscript{3−} derived from biochar ash\textsuperscript{1, 6}. 

S3
**Table S1.** Empirical categories and input features used to predict heavy metal immobilization efficiency in biochar added soils.

| Empirical categories | Input features | Unit | Abbreviation | Data range | No. of datapoints |
|----------------------|----------------|------|--------------|------------|-------------------|
| 1 Biochar production conditions | Pyrolysis temperature | °C | T °C | 200–850 | 162 |
| Biochar Properties | 2 pH | - | pH_BC | 5.26–12.39 | 162 |
| | 3 C content | % | C % | 11.55–93.7 | 162 |
| | 4 H content | % | H % | 0–6.9 | 162 |
| | 5 O content | % | O % | 1.14–51.16 | 162 |
| | 6 N content | % | N % | 0–43.66 | 162 |
| | 7 H/C | - | H/C | 0–1.57 | 162 |
| | 8 O/C | - | O/C | 0.01–0.91 | 162 |
| | 9 (O+N)/C | - | (O+N)/C | 0.03–1.00 | 162 |
| | 10 Ash content | % | Ash % | 1.24–84.3 | 162 |
| | 11 Surface area | m²·g⁻¹ | SA | 0.36–907.4 | 127 |
| 2 Experimental conditions | Biochar application rate | % | BC rate % | 0.5–10 | 162 |
| | 13 Experiment duration | months | Time | 0.03–5 | 162 |
| | 14 Available HM concentration (control treatment) | mg·kg⁻¹ | Avail. HM | 0.023–6,418 | 162 |
| 3 Soil properties after incubation | pH | - | pH_soil | 5.24–10.16 | 162 |
| | EC | dS·m⁻¹ | EC_soil | 0.067–12.3 | 152 |
| 4 Heavy metal properties | Molecular weight | G·mol⁻¹ | MW | 162 |
| | Electronegativity | Electronegativity | 1.55–2.33 | 162 |
| | Ion radius | nm | Ion radius | 0.07–0.119 | 162 |
| | Valency | - | Valency | 2 | 162 |
Table S2. The tuned hyper-parameters of RF, SVR, and NN model for the prediction of biochar surface area and HM immobilization efficiency.

| Prediction target               | ML models                | Hyper-parameters | max_depth | max_features |
|--------------------------------|--------------------------|------------------|-----------|--------------|
| Biochar surface area           | Random forest (RF)       | n_estimators     | 21        | 11           | auto        |
| HM immobilization efficiency   |                          | max_features     | 11        | auto         |
| Biochar surface area           | Supporting vector        | C                | 50        | 0.1          | -           |
| regression (SVR)               | epsilon                  | 0.01             |           |              |
| HM immobilization efficiency   |                          |                  |           |              |
|                                | Neural network (NN)      | Neurons in the   | 16        | 16           | -           |
| Biochar surface area           |                          | first hidden     |           |              |
| HM immobilization efficiency   |                          | lawyer           |           |              |
| Biochar surface area           |                          | Neurons in the   | 4         | 4            | -           |
| HM immobilization efficiency   |                          | second hidden    |           |              |
|                                |                          | lawyer           |           |              |
Table S3. Data collected from experimental work for further model validation\textsuperscript{10,11}.

| Data points          | 1  | 2  | 3  | 4  | 5  | 6  | 7  | 8  |
|----------------------|----|----|----|----|----|----|----|----|
| Heavy metal types    | Pb | Cu | Pb | Cu | Cu | Pb | Cd | Pb |
| T(°C)                | 500| 750| 750| 750| 750| 500| 500| 500|
| C (%)                | 86.28| 86 | 86 | 86 | 86 | 50.8 | 50.8 | 50.8 |
| H (%)                | 3.12| 1.49| 1.49| 1.49| 1.49| 1.72 | 1.72 | 1.72 |
| O (%)                | 7.35| 12.06| 12.06| 12.06| 12.06| 45.82 | 45.82 | 45.82 |
| N (%)                | 3.25| 0.45| 0.45| 0.45| 0.45| 1.66 | 1.66 | 1.66 |
| Ash (%)              | 52.37| 11.9| 11.9| 11.9| 11.9| 42.7 | 42.7 | 42.7 |
| pH\textsubscript{BC} | 10.5| 9.5 | 9.5 | 9.5 | 9.5 | 10 | 10 | 10 |
| SA (m\textsuperscript{2}·g\textsuperscript{-1}) | 13.7| 907.4| 907.4| 907.4| 907.4| 36.7 | 36.7 | 36.7 |
| BC rate (%)          | 2.22| 1 | 5 | 1 | 5 | 1 | 5 | 1 |
| Time (Months)        | 15.8| 12 | 12 | 12 | 12 | 12 | 12 | 12 |
| pH\textsubscript{soil} | 6.09| 4.83| 4.80| 4.83| 4.98| 5.32 | 5.43 | 5.13 |
| EC\textsubscript{soil} (dS/m) | 0.57| 0.29| 0.29| 0.28| 0.19| 0.63 | 0.47 | 0.31 |
| Avail. HM (mg·kg\textsuperscript{-1}) | 0.155| 34.4| 23.1| 34.4| 34.4| 23.1 | 0.5 | 23.1 |
| Electronegativity    | 2.33| 1.9 | 2.33| 1.9 | 1.9 | 2.33 | 1.69 | 2.33 |
| Immobilization (%)   | 36.8| 31.9| 28.1| 26.8| 47.7| 48.1 | 25.8 | 29 |
| Predicted Immobilization (%) | 28.70| 23.17| 26.39| 23.17| 31.24| 46.90 | 27.46 | 40.58 |
| Prediction error (%) | 22.01| 27.36| 6.07| 13.54| 34.51| 2.50 | -6.42 | -39.95 |
Fig. S1. Statistical data visualization of the impact of: (a) N content (n=162); (b) surface area (n=127); and (c) the biochar application rate (n= 162) on heavy metal immobilization in biochar-amended soils.
Fig. S2. Box plot showing the: (a) range of biochar composition (n=162); and (b) Van Krevelen diagram (n=162) of biochar (low temperature biochar: 200–250 °C; and high temperature biochar: 250–850 °C).
Fig. S3. Results on hyper-parameter tuning for: (a) supporting vector regression; (b) neural network models, and prediction performance of the optimal models with respect to: (c) supporting vector regression; and (d) neural network models for the surface area prediction of biochar.
Fig. S4. Results on hyper-parameter tuning for: (a) random forest; (b) supporting vector regression; and (c) neural network models to predict heavy metal immobilization in biochar-amended soils.
|                  | T °C  | pH_BC | C % | H %    | O %    | N %    | (O+N)/C | O/C  | H/C   | Ash % | SA   | BC % | Time | Avail. HM | pH_soil | EC_soil | WM | Electronegativity | Ion radius | Valency |
|------------------|-------|-------|-----|--------|--------|--------|----------|-------|--------|-------|------|------|------|-----------|----------|---------|----|------------------|------------|---------|
|                  | 1.00  | 0.65  | 0.31 | -0.84 | -0.43 | 0.31   | -0.14   | -0.35 | -0.84 | 0.11  | 0.71 | 0.18 | -0.19 | 0.03      | 0.12     | -0.01   | 0.08 | -0.14           | 0.07       | -0.06   |
|                  | 1.00  | 1.00  | 0.14 | -0.76 | -0.29 | 0.03   | -0.17   | -0.21 | 0.73  | 0.34  | 0.36 | -0.22 | 0.14  | 0.08      | 0.05     | 0.13    | 0.18 | -0.22           | 0.14       | -0.14   |
|                  | 0.31  | 0.14  | 1.00 | -0.03 | -0.47 | 0.06   | -0.71   | -0.64 | 0.44  | 0.81  | 0.02 | 0.30 | -0.53 | 0.15      | 0.51     | 0.32    | 0.15 | 0.27            | 0.07       | 0.20    |
|                  | -0.84 | -0.76 | 0.03 | 1.60  | 0.29  | -0.15 | 0.01    | 0.13  | 0.88  | -0.37 | -0.59 | -0.04 | -0.16 | 0.01      | 0.04     | -0.21   | 0.20 | 0.30            | 0.14       | 0.20    |
|                  | -0.43 | -0.29 | -0.47 | 0.29  | 1.00  | 0.26  | 0.71    | 0.04  | 0.42  | 0.09  | -0.37 | 0.19  | 0.39  | 0.00      | 0.25     | 0.49    | 0.04 | 0.04            | 0.11       | 0.06    |
|                  | 0.31  | -0.03 | -0.06 | -0.15 | -0.26 | 1.00  | 0.39    | -0.21 | -0.17 | -0.12 | 0.02  | 0.46  | -0.31 | 0.04      | 0.13     | -0.12   | -0.05 | -0.13          | 0.06       | 0.12    |
|                  | -0.14 | -0.17 | -0.71 | -0.01 | 0.71  | 0.39  | 1.00    | 0.80  | 0.24  | 0.30  | -0.38 | 0.03  | 0.30  | -0.07     | 0.30      | 0.45    | -0.05 | 0.21           | 0.03       | 0.20    |
|                  | -0.35 | -0.21 | -0.64 | 0.13  | 0.94  | 0.21  | 0.86    | 1.00  | 0.36  | 0.30  | -0.38 | -0.17 | 0.45  | 0.04      | 0.30      | 0.58    | -0.10 | 0.10           | 0.10       | 0.10    |
|                  | -0.84 | -0.73 | -0.44 | 0.88  | 0.42  | 0.17  | 0.24    | 0.36  | 1.00  | 0.01  | 0.69  | -0.14 | 0.07  | 0.06      | 0.20     | 0.15    | 0.20 | 0.12           | 0.13       | 0.13    |
|                  | 0.11  | 0.34  | -0.69 | -0.37 | 0.09  | -0.12 | 0.36    | 0.30  | 0.01  | 1.00  | -0.18 | -0.19 | 0.52  | -0.06     | -0.48     | 0.11    | -0.09 | -0.15          | -0.05      | -0.12   |
|                  | 0.71  | 0.30  | 0.52  | 0.30  | -0.37 | 0.02  | -0.38   | -0.38 | 0.60  | -0.18 | 1.00  | 0.24  | -0.19 | 0.10      | 0.13     | -0.12   | -0.08 | -0.09          | 0.01       | 0.01    |
|                  | 0.18  | -0.22 | -0.30 | -0.04 | -0.19 | 0.46  | 0.03    | -0.17 | -1.14 | -0.19 | 0.24  | 1.00  | -0.27 | 0.32      | 0.29     | -0.17   | 0.18  | 0.23           | 0.13       | 0.10    |
|                  | -0.19 | 0.11  | -0.53 | -0.16 | 0.33  | -0.31 | 0.35    | 0.45  | 0.67  | 0.52  | 0.18  | 0.27  | 1.00  | 0.14      | 0.07     | -0.19   | -0.17 | -0.24          | -0.07      | -0.28   |
|                  | 0.03  | -0.06 | 0.15  | 0.01  | 0.00  | 0.04  | -0.07   | -0.04 | -0.06 | 0.10  | 0.32  | 0.14  | 1.00  | 0.20      | 0.08     | 0.19    | 0.20 | 0.16           | 0.15       | 0.15    |
|                  | 0.12  | 0.06  | 0.51  | 0.04  | 0.25  | 0.13  | -0.30   | -0.30 | 0.26  | -0.48 | 0.13  | 0.29  | -0.67 | 0.20      | 1.00     | 0.09    | 0.05 | -0.12          | -0.02      | 0.16    |
|                  | 0.01  | 0.13  | 0.32  | 0.21  | 0.49  | -0.12 | 0.45    | 0.58  | 0.12  | 0.11  | -0.12 | -0.17 | 0.19  | -0.08     | 0.09     | 1.00    | 0.09 | 0.24           | -0.02      | 0.08    |
|                  | -0.08 | -0.18 | 0.15  | 0.20  | 0.04  | -0.05 | -0.05   | 0.05  | 0.02  | 0.15  | -0.09 | -0.08 | 0.18  | 0.17      | 0.19     | 0.05    | -0.09 | 1.00           | 0.85       | 0.96    |
|                  | -0.14 | -0.22 | 0.27  | 0.30  | 0.04  | -0.13 | -0.21   | -0.10 | 0.20  | -0.15 | -0.02 | 0.23  | -0.24 | 0.20      | 0.13     | -0.24   | 0.85 | 1.00           | 0.72       | 0.73    |
|                  | -0.07 | -0.14 | 0.07  | 0.14  | 0.11  | 0.06  | 0.03    | 0.10  | 0.12  | -0.05 | -0.09 | 0.13  | -0.07 | 0.16      | -0.02    | -0.02   | 0.96 | 0.72           | 1.00       | 0.57    |
|                  | -0.06 | -0.14 | -0.20 | 0.20  | 0.06  | -0.12 | 0.20    | -0.10 | 0.13  | -0.12 | 0.91  | 0.10  | -0.20 | 0.15      | 0.16     | -0.08   | 0.66 | 0.73           | 0.57       | 1.00    |

**Fig. S5.** Results on the Pearson correlation coefficient (PCC) in terms of input features.
Fig. S6. Graphic user interface (GUI) development in the webpage through the integration of Python (version 3.7) and Flask (version 1.1.2). This was based on the well-trained random forest model to predict heavy metal immobilization efficiency in biochar-amended soils (online prediction software: https://ljplj.pythonanywhere.com/biochar_for_soil_HM_immobilization).
Fig. S7 Further validation results of the developed ML model through the graphical user interface (GUI) with new experimental data.
Fig. S8. Partial dependence plot of each input feature from the updated random forest model. Note: T (°C): pyrolysis temperature; pH_BC: biochar pH; ash %: ash content of biochar; biochar pH; C, O, N content (%): C, O, N content of biochar as a percentage, respectively; ash %: ash content of biochar; SA: surface area of biochar; BC rate %: biochar application rate in soil; time: experimental duration; Avail. HM: available heavy metal content in soil; pH_soil: soil pH; EC_soil: soil electrical conductivity; electronegativity: electronegativity of heavy metals.
References

1. Park, J. H.; Choppala, G. K.; Bolan, N. S.; Chung, J. W.; Chuasavathi, T., Biochar reduces the bioavailability and phytotoxicity of heavy metals. *Plant and soil* 2011, *348*, (1-2), 439.

2. Lei, S.; Shi, Y.; Qiu, Y.; Che, L.; Xue, C., Performance and mechanisms of emerging animal-derived biochars for immobilization of heavy metals. *Science of The Total Environment* 2019, *646*, 1281-1289.

3. Shen, Z.; Tian, D.; Zhang, X.; Tang, L.; Su, M.; Zhang, L.; Li, Z.; Hu, S.; Hou, D., Mechanisms of biochar assisted immobilization of Pb2+ by bioapatite in aqueous solution. *Chemosphere* 2018, *190*, 260-266.

4. Palansooriya, K. N.; Shaheen, S. M.; Chen, S. S.; Tsang, D. C. W.; Hashimoto, Y.; Hou, D.; Bolan, N. S.; Rinklebe, J.; Ok, Y. S., Soil amendments for immobilization of potentially toxic elements in contaminated soils: A critical review. *Environment International* 2020, *134*, 105046.

5. Peng, H.; Gao, P.; Chu, G.; Pan, B.; Peng, J.; Xing, B., Enhanced adsorption of Cu(II) and Cd(II) by phosphoric acid-modified biochars. *Environmental Pollution* 2017, *229*, 846-853.

6. Ahmad, M.; Rajapaksha, A. U.; Lim, J. E.; Zhang, M.; Bolan, N.; Mohan, D.; Vithanage, M.; Lee, S. S.; Ok, Y. S., Biochar as a sorbent for contaminant management in soil and water: A review. *Chemosphere* 2014, *99*, 19-33.

7. Igalavithana, A. D.; Kwon, E. E.; Vithanage, M.; Rinklebe, J.; Moon, D. H.; Meers, E.; Tsang, D. C. W.; Ok, Y. S., Soil lead immobilization by biochars in short-term laboratory incubation studies. *Environment International* 2019, *127*, 190-198.

8. Hassan, M.; Liu, Y.; Naidu, R.; Parikh, S. J.; Du, J.; Qi, F.; Willett, I. R., Influences of feedstock sources and pyrolysis temperature on the properties of biochar and functionality as adsorbents: A meta-analysis. *Science of The Total Environment* 2020, *744*, 140714.

9. Ahmad, M.; Lee, S. S.; Rajapaksha, A. U.; Vithanage, M.; Zhang, M.; Cho, J. S.; Lee, S.-E.; Ok, Y. S., Trichloroethylene adsorption by pine needle biochars produced at various pyrolysis temperatures. *Bioresource Technology* 2013, *143*, 615-622.

10. El-Naggar, A.; Lee, M.-H.; Hur, J.; Lee, Y. H.; Igalavithana, A. D.; Shaheen, S. M.; Ryu, C.; Rinklebe, J.; Tsang, D. C.; Ok, Y. S., Biochar-induced metal immobilization and soil biogeochemical process: an integrated mechanistic approach. *Science of the Total Environment* 2020, *698*, 134112.

11. Yang, X.; Liu, J.; McGrouther, K.; Huang, H.; Lu, K.; Guo, X.; He, L.; Lin, X.; Che, L.; Ye, Z., Effect of biochar on the extractability of heavy metals (Cd, Cu, Pb, and Zn) and enzyme activity in soil. *Environmental Science and Pollution Research* 2016, *23*, (2), 974-984.