Chemosphere 276 (2021) 130204

Contents lists available at ScienceDirect

Chemosphere

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

Predicting the sorption efficiency of heavy metal based on the biochar characteristics, metal sources, and environmental conditions using various novel hybrid machine learning models

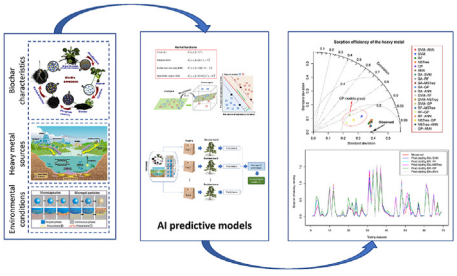
Bo Ke [a, b], Hoang Nguyen [c, *], Xuan-Nam Bui [c, d], Hoang-Bac Bui [e, f], Yosoon Choi [g], Jian Zhou [h], Hossein Moayedi [i], Romulus Costache [j], Thao Nguyen-Trang [k, l]

[a] School of Resources and Environmental Engineering, Wuhan University of Technology, Wuhan, Hubei, 430070, China
[b] School of Urban Construction, Wuchang University of Technology, Wuhan, 430223, China
[c] Department of Surface Mining, Mining Faculty, Hanoi University of Mining and Geology, 18 Pho Vien, Duc Thang Ward, Bac Tu Liem District, Hanoi, 100000, Viet Nam
[d] Center for Mining, Electro-Mechanical Research, Hanoi University of Mining and Geology, 18 Pho Vien, Duc Thang Ward, Bac Tu Liem District, Hanoi, 100000, Viet Nam
[e] Faculty of Geosciences and Geoengineering, Hanoi University of Mining and Geology, 18 Vien St., Duc Thang Ward, Bac Tu Liem Dist., Hanoi, 100000, Viet Nam
[f] Center for Excellence in Analysis and Experiment, Hanoi University of Mining and Geology, 18 Vien St., Duc Thang Ward, Bac Tu Liem Dist., Hanoi, 100000, Viet Nam
[g] Department of Energy Resources Engineering, Pukyong National University, Busan, 48513, South Korea
[h] School of Resources and Safety Engineering, Central South University, Changsha, Hunan, 410083, China
[i] Department of Energy Resources Engineering, Universiti Teknologi Malaysia, Johor Bahru, Malaysia
[j] Research Institute of the University of Bucharest, 90-92 Sos. Panduri, 5th District, Bucharest, Romania
[k] Division of Computational Mathematics and Engineering, Institute for Computational Science, Ton Duc Thang University, Ho Chi Minh City, 70000, Viet Nam
[l] Faculty of Mathematics and Statistics, Ton Duc Thang University, Ho Chi Minh City, 700000, Viet Nam

HIGHLIGHTS

- Biochar characteristics have significant effects on the SEoHM of biochar system.
- Metal sources, environmental conditions were investigated for the SEoHM.
- 15 novel hybrid models were proposed for predicting the SEoHM of biochar system.
- SVM-ANN was proposed as the best model for predicting the SEoHM of biochar system.
- Taylor diagram, boxplot and Q-Q plot were used to evaluate the models.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 12 December 2020
Received in revised form
17 February 2021

ABSTRACT

Heavy metals in water and wastewater are taken into account as one of the most hazardous environmental issues that significantly impact human health. The use of biochar systems with different materials helped significantly remove heavy metals in the water, especially wastewater treatment systems. Nevertheless, heavy metal's sorption efficiency on the biochar systems is highly dependent on the

* Corresponding author.
E-mail addresses: nguyenhoang@humg.edu.vn (H. Nguyen), buihoangbac@humg.edu.vn (H.-B. Bui), nguyenthaotrang@tdtu.edu.vn (T. Nguyen-Trang).

https://doi.org/10.1016/j.chemosphere.2021.130204

Abbreviations

(O + N)/C	Ratio of O and N with C
A	Percentage of ash
AI	Artificial intelligence
ANFIS	Adaptive neuro-fuzzy inference system
ANN	Artificial neural network
BA	Bagging
BSA	Biochar surface area
C	Percentage of carbon in biochar
CEC	Cation exchange capacity
C_O	Heavy metal concentration in wastewater
Cr	Chrome
Cu (II)	Copper (II)
GA	Genetic algorithm
GP	Gaussian process
GWI	Global Water Intelligence
H/C	Ratio of hydrogen and carbon
MAE	Mean absolute error
MLR	Multiple linear regression
Ni (II)	Nickel (II)
$n_{predictors}$	Number of randomly selected predictors
n_{tree}	Number of trees in the forest
O/C	Ratio of oxygen and carbon
pH_{sol}	Solution pH
pH_{ww}	pH of biochar in wastewater
PSB	Particle size of biochar
R^2	Determination coefficient
RF	Random forest
RMSE	Root-mean-squared error
SEoHM	Sorption efficiency of heavy metals
SVM	Support vector machine
T_{envi}	Environmental temperature
T_P	Pyrolysis temperature
Zn (II)	Zinc (II)

Accepted 4 March 2021
Available online 9 March 2021

Handling Editor: Derek Muir

Keywords:
Heavy metals
Biochar system
Sorption
Machine learning
Hybrid model
Ensemble model
Soft computing

biochar characteristics, metal sources, and environmental conditions. Therefore, this study implicates the feasibility of biochar systems in the heavy metal sorption in water/wastewater and the use of artificial intelligence (AI) models in investigating efficiency sorption of heavy metal on biochar. Accordingly, this work investigated and proposed 20 artificial intelligent models for forecasting the sorption efficiency of heavy metal onto biochar based on five machine learning algorithms and bagging technique (BA). Accordingly, support vector machine (SVM), random forest (RF), artificial neural network (ANN), M5Tree, and Gaussian process (GP) algorithms were used as the key algorithms for the aim of this study. Subsequently, the individual models were bagged with each other to generate new ensemble models. Finally, 20 intelligent models were developed and evaluated, including SVM, RF, M5Tree, GP, ANN, BA-SVM, BA-RF, BA-M5Tree, BA-GP, BA-ANN, SVM-RF, SVM-M5Tree, SVM-GP, SVM-ANN, RF-M5Tree, RF-GP, RF-ANN, M5Tree-GP, M5Tree-ANN, GP-ANN. Of those, the hybrid models (i.e., BA-SVM, BA-RF, BA-M5Tree, BA-GP, BA-ANN, SVM-RF, SVM-M5Tree, SVM-GP, SVM-ANN, RF-M5Tree, RF-GP, RF-ANN, M5Tree-GP, M5Tree-ANN, GP-ANN) are introduced as the novelty of this study for estimating the heavy metal's sorption efficiency on the biochar systems. Also, the biochar characteristics, metal sources, and environmental conditions were comprehensively assessed and used, and they are considered as a novelty of the study as well. For this aim, a dataset of sorption efficiency of heavy metal was collected and processed with 353 experimental tests. Various performance indexes were applied to evaluate the models, such as RMSE, R^2, MAE, color intensity, Taylor diagram, box and whiskers plots. This study's findings revealed that AI models could predict heavy metal's sorption efficiency onto biochar with high reliability, and the efficiency of the ensemble models is higher than those of individual models. The results also reported that the SVM-ANN ensemble model is the most superior model among 20 developed models. The predictive model proposed that heavy metal's efficiency sorption on biochar can be accurately forecasted and early warning for the water pollution by heavy metal.

1. Introduction

Water is considered an indispensable ingredient for human life. According to a Global Water Intelligence (GWI) survey, more than 1.2 billion people worldwide do not have clean water to use (Intelligence et al., 2011). The main reason is attributed to the significant influence of industrial development, mining process, thermal power plants, hydroelectricity, other environmental disasters, and fast population growth (Alrumman et al., 2016; Wang and Yang, 2016; Cheng et al., 2019; Tran et al., 2020; Nguyen et al., 2020). According to the statistics of Programme and UN-Water (2009), an average of more than 2 million tons of industrial, agricultural, and domestic wastewater discharges into the water environment worldwide every day. Water waste seriously affects human health, and an estimated 14,000 people die from wastewater every day (Bolisetty et al., 2019). Besides, pollution of water from waste is also the primary cause of drastic changes in the biosphere. According to scientists, arsenic, cadmium, chromium, lead, mercury, pesticides, and radiation are the major pollutants in wastewater (Fig. 1). Water scarcity has also pushed the concentration of polluted components, and heavy metals in the water increased significantly. Therefore, wastewater treatment is an urgent issue that significantly impacts human health, requiring the whole community's cooperation.

According to scientists, the heavy metal in wastewater is the most dangerous and difficult problem to handle (Hua et al., 2012). Mining activities, industrial zones, commercial centers, and industrial wastewater are the primary agents that increase heavy metals concentration in wastewater. Even natural water can also be contaminated with cadmium, arsenic, and iron (Nguyen and Trinh, 2020). They are the most critical pollutants contributing to human health risks and other cancers (e.g., liver, lungs, stomach)

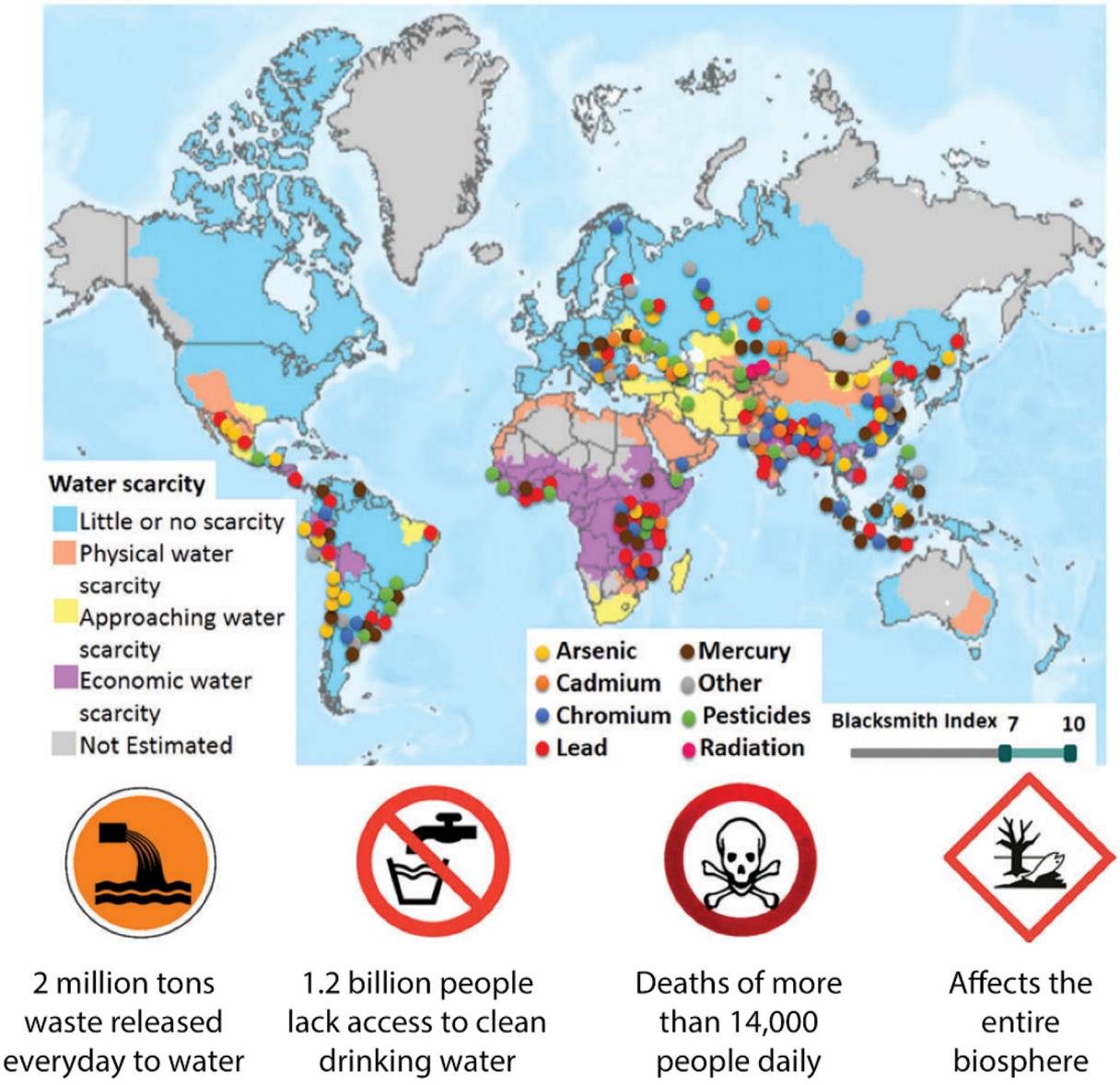


Fig. 1. Summary of water pollution by heavy metals, radionuclide, and pesticides, as well as the water scarcity worldwide (Bolisetty et al., 2019).

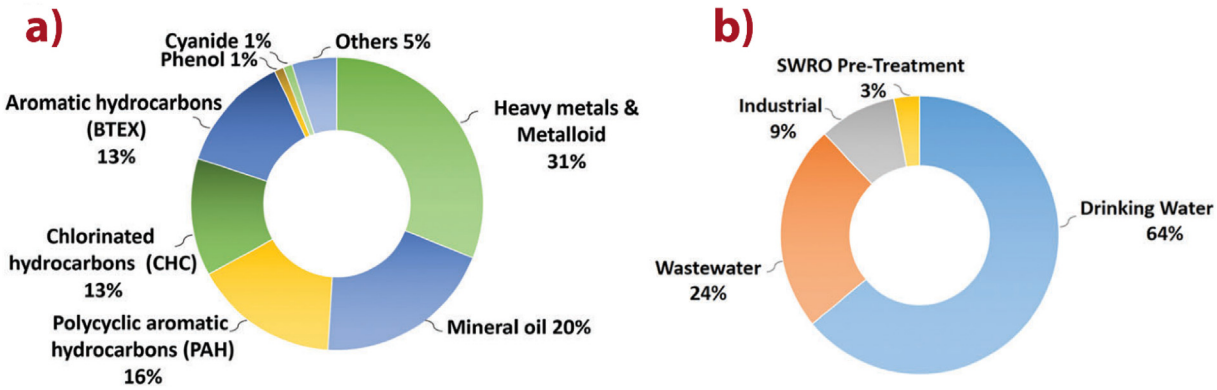


Fig. 2. A review of composition and volume of treated water (Bolisetty et al., 2019). (a) Components of water pollutants in industrial and commercial areas; (b) The volume of treated water according to demands in use.

(Muchuweti et al., 2006; Zhao et al., 2012, 2014). The primary pollutant components in European industrial and commercial areas are illustrated in Fig. 2a. Besides, the percentage of water treated by each field is shown in Fig. 2b. As can be seen, heavy metals in the water are dangerous materials, and they should be removed. In recent years, there are several methods to remove heavy metals, such as physicochemical methods (e.g., mechanical screening, hydrodynamic classification, attrition scrubbing, magnetic separation, gravity concentration, flotation, and electrostatic separation, chemical precipitation, ion exchange, and electrochemical deposition), chemical precipitation, membrane filtration, coagulation and flocculation, electrodialysis, electrochemical treatments, and biological methods (Gunatilake, 2015; Yu et al., 2017; Gan et al., 2018; Hernández-Cocoletzi et al., 2020; Sankaran et al., 2020). Of those, biological methods (i.e., biochar systems) are recommended as effective solutions for absorbing heavy metals due to the following reasons:

(i) Biochar systems use various low-cost adsorbents (agricultural waste, industrial waste, mining products, forest waste, algae, to name a few) to remove heavy metals (Inyang et al., 2012, 2016; Tran et al., 2020; Zhang et al., 2020). These adsorbents are also known as the recycled materials aiming to reduce environmental pollution.
(ii) Biochar systems have been confirmed with the maximum removal of heavy metals in water/wastewater (Gomez-Serrano et al., 1998; Gupta et al., 2001; Idris et al., 2010; Mohan et al., 2014; Inyang et al., 2016).

As per research by previous researchers, the biochar system's heavy metal sorption efficiency depends significantly on its characteristics, metal sources, and environmental conditions (Tytłak et al., 2015; Li et al., 2017; Hass and Lima, 2018). Therefore, the prediction of heavy metal sorption efficiency (SEoHM) of biochar systems based on the biochar characteristics, metal sources, and environmental conditions is necessary and is very important for improving water quality.

2. Mechanism of biochar systems for heavy metal sorption

As one of the main objectives of this work, mechanisms of biochar systems play an essential role in the sorption of heavy metals in water/wastewater, including pH, surface charge, surface area, porosity, functional groups, and mineral contents (Appel and Ma, 2002; Uchimiya et al., 2012; Suliman et al., 2016).

In this regard, the characteristics of surface area and porosity are significant physical properties of biochar systems. Under the high temperature of biochar systems (pyrolyze), micropores are established in biochar systems due to the dehydration process (Plaza et al., 2014). The size of biochar pore has a significantly influent on the metal sorption capacity of biochar systems. It is worth mentioning that the surface area and porosity are highly dependent on biochar systems' temperature (Leng et al., 2020). Besides, the biochar feedstock compositions are also essential in evaluating biochar systems' efficiency (Li et al., 2019). Similarity characteristics are also claimed for pH with temperature and feedstock (Al-Wabel et al., 2013). Next is the surface charge properties of biochar systems. It is highly dependent on the pH solution when using biochar systems for treating water/wastewater. In addition to the surface area, porosity, pH solution, and surface charge, functional groups play essential roles in biochar systems' heavy metal sorption. They are controlled by pyrolysis temperature and feedstock based on the carbonization degree. Functional groups are diversity in biochar systems and they may be lost under increasing temperature (Li and Chen, 2018). The last factor is the mineral components in biochar systems, such as calcium (Ca), potassium (K), phosphorus (P), and magnesium (Mg). They are responsible for metal sorption from water/wastewater. Their primary mission is to exchange or precipitate with heavy metals to reduce or remove their availability (Uddin, 2017).

Regarding the heavy metal sorption mechanisms, five mechanisms have been proposed to exercise control over heavy metals sorption by biochar systems, including complexation, cation exchange, precipitation, electrostatic interactions, and chemical reduction (Li et al., 2017), as illustrated in Fig. 3. Depending on the heavy metals, the role of each mechanism is different. Further assessments can be referred to the literature (Sajidu et al., 2008; Deze et al., 2012; Lu et al., 2012; Trakal et al., 2016).

3. Related works

Concerning the heavy metal sorption efficiency of the biochar system, various materials and methods have been applied. Indeed, Šćiban et al. (2007) used the wood sawdust to absorb heavy metals from electroplating wastewater. Their findings demonstrated that the wood sawdust could better absorb copper ions in wastewater than metal solutions. However, other ions have been thought to reduce cadmium ions' sorption efficiency - one of the most dangerous ions in wastewater. In another study, Sarkar and Majumdar (2011) applied the response surface method to optimize the SEoHM in wastewater (i.e., Cu (II), Ni (II), and Zn (II)) of the biochar system by surfactant modified chitosan bead. A quadratic equation has also been used to predict the SEoHM with a coefficient determined of 0.8326. Mehta and Gaur (2005) also used algae as a special material to absorb heavy metals in wastewater. The findings showed that algae could effectively remove heavy metals from wastewater. Furthermore, the authors' recommendations on the development of the SEoHM predictive models of biochar systems have been given as a problem to be addressed in future studies.

In recent years, artificial intelligence (AI) is considered a useful tools in real-life problems (Pal and Deswal, 2009; Qu et al., 2014; Rodriguez-Galiano et al., 2014; Bui et al., 2019a, 2019b, 2019c, 2020; Nguyen, 2019; Nguyen and Bui, 2019; Nguyen et al., 2019a; Shariati

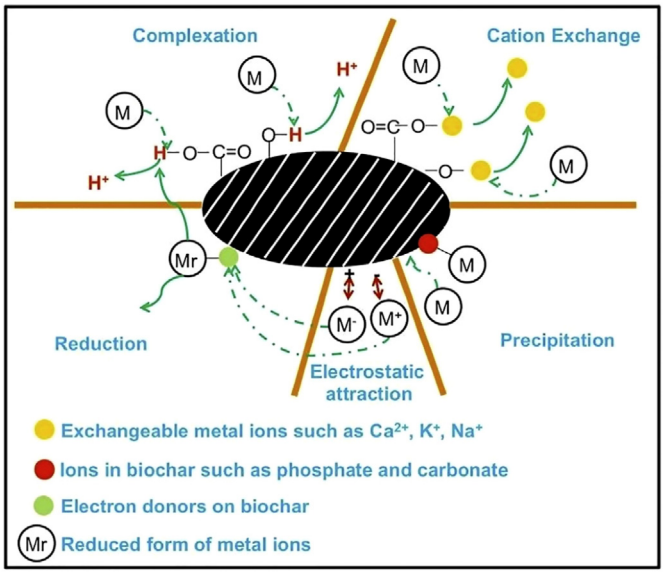


Fig. 3. Illustrating the mechanism of biochar systems to absorb heavy metals (Li et al., 2017).

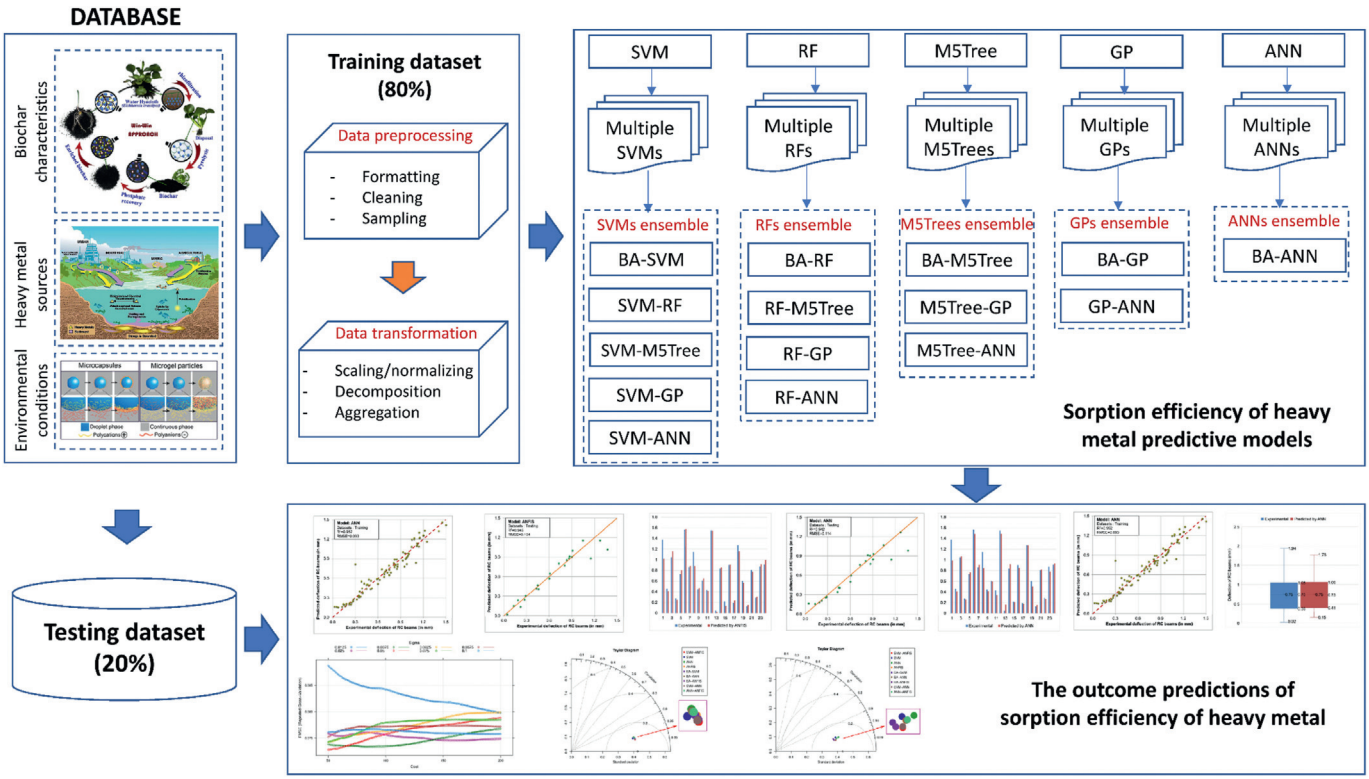


Fig. 4. Proposing the framework of novel ensemble models for predicting sorption efficiency.

et al., 2019, 2020a, 2020b; Trung et al., 2019; Le and Dang, 2020; Vu et al., 2020), as well as in terms of sorption efficiency prediction of heavy metals in wastewater. Prakash et al. (2008) developed an artificial neural network (ANN) model for predicting the sorption efficiency of Cu (II) ions in wastewater using sawdust. Their findings showed that ANN could predict the sorption efficiency of Cu (II) ions impressive. A similar ANN technique has also been applied to remove the nickel (II) ions from an aqueous solution by zeolite (Turp et al., 2011). In another study, Parveen et al. (2017) developed an intelligence model for predicting the sorption efficiency of the Cr (IV) ions using the support vector machine algorithm (SVM). MLR and ANN were also developed to compare and assess the developed SVM model's feasibility in predicting the sorption efficiency of heavy metal. The results showed that the SVM model could predict more precisely than those of the MLR and ANN models. However, only certain environmental conditions were considered as input variables in their study. Dolatabadi et al. (2018) also developed ANN and ANFIS models to forecast heavy metal removal (i.e., Basic Red46 (BR46) and Cu) from wastewater of a biochar system using sawdust based on the characteristics of the metal source (e.g., contact time, initial Cu (II), initial dye, pH, and adsorbent dosage). To improve the accuracy of the baseline predictive models, Sutherland et al. (2018) used the genetic algorithm aiming to optimize an initial ANN model (i.e., ANN-GA) in predicting the SEoHM. Metal sources and environmental conditions were taken into account to assess the Cu (II) ions removal from aqueous solutions. The ANN-GA model predicted its sorption efficiency with a promising result. Besides, the biochar characteristics were also claimed that have a significant influence on the heavy metal sorption efficiency of the biochar system (Zhu et al., 2019). A random forest (RF) and an ANN model were then developed to predict the SEoHM in wastewater while considering the biochar characteristics. The results interpreted the high convergence of the dataset on the RF and ANN models. In other words, the RF and ANN models can predict the SEoHM with high reliability, and the influence of the biochar characteristics on the sorption efficiency has also been properly addressed.

An overview of the literature shows that AI is a powerful tool to predict the SEoHM in biochar systems' wastewater. However, previous studies just only passed simple problems and models. Besides, summarizing the above reviews' conclusions showed that the biochar characteristics, metal sources, and environmental conditions have a significant impact on the SEoHM of the biochar system. Nevertheless, they have not been adequately considered and explained. Therefore, this study aims to develop some novel hybrid AI models for predicting the SEoHM based on biochar characteristics, metal sources, and environmental conditions. Accordingly, SVM, RF, ANN, M5Tree, and Gaussian process (GP) algorithms are used as the key algorithms for predicting the sorption efficiency of heavy metal. Subsequently, the bagging techniques (BA) is applied

Table 1
Properties of the input and output variables used herein.

Categories	T_P	pH_{ww}	C	(O + N)/C	O/C	H/C	A
Min.	300	6.78	43.29	0.048	0.033	0.16	1.57
1st Qu.	450	9.69	48.69	0.095	0.08	0.299	7.52
Median	550	10.03	69.04	0.168	0.142	0.392	21.25
Mean	540.9	9.71	65.8	0.1848	0.1676	0.4607	21.3
3rd Qu.	650	10.4	77	0.26	0.243	0.569	33.1
Max.	700	11.1	88.2	0.421	0.359	1.084	47.93
Catergories	**PSB**	**BSA**	**CEC**	**T_{envi}**	**pH_{sol}**	**C_0**	**SEoHM**
Min.	0.15	2.46	3.17	20	2	0.0029	0.0003
1st Qu.	0.2	12.78	7.41	20	5	0.008	0.004
Median	0.2	20.1	14.6	25	5.5	0.1429	0.0823
Mean	0.4554	43.8	15.68	24.12	5.504	0.4511	0.2324
3rd Qu.	0.5	41.12	21.8	25	6	0.8897	0.2829
Max.	2	465	45.7	28	10	2.4131	1.5835

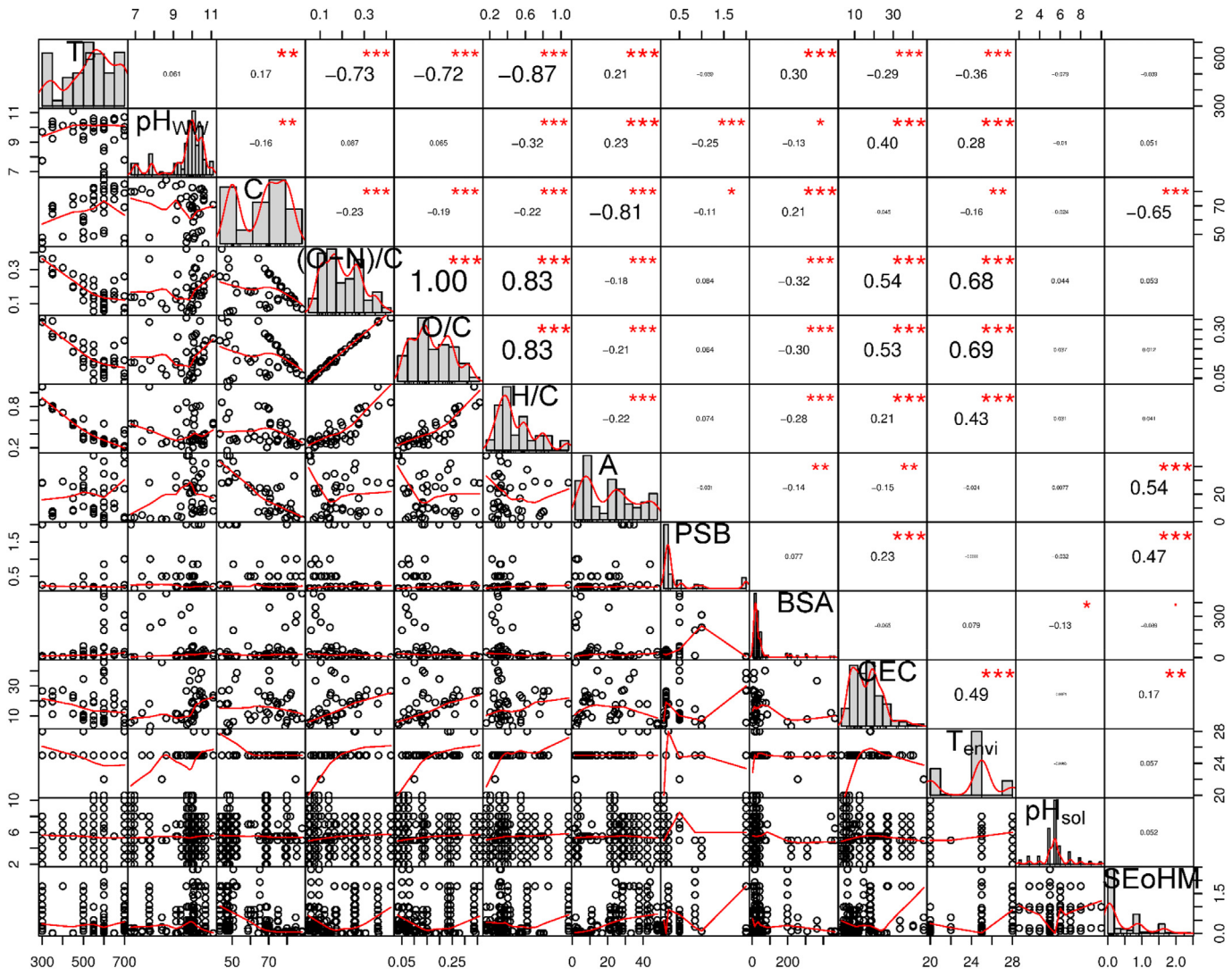


Fig. 5. The chart of the correlation matrix of the original database.

Table 2
Properties of the input and output variables in the training dataset.

Categories	T_P	pH_{ww}	C	(O + N)/C	H/C	A	PSB
Min.	300	6.78	43.29	0.048	0.16	1.57	0.15
1st Qu.	450	9.69	48.69	0.095	0.306	7.425	0.2
Median	550	10	69.04	0.17	0.392	21.25	0.2
Mean	537.1	9.68	65.81	0.1854	0.4675	21.331	0.4528
3rd Qu.	650	10.4	79	0.26	0.569	33.1	0.5
Max.	700	11.1	88.2	0.421	1.084	47.93	2
Catergories	BSA	CEC	T_{envi}	pH_{sol}	C_0	SEoHM	–
Min.	2.46	3.17	20	2	0.0029	0.0003	–
1st Qu.	12.78	7.2	22	5	0.008	0.003975	–
Median	20.1	14.6	25	5.5	0.131	0.08285	–
Mean	47	15.53	24.16	5.523	0.4376	0.22813	–
3rd Qu.	41.12	21.8	25	5.625	0.8897	0.28375	–
Max.	465	45.7	28	10	2.4131	1.5835	–

Table 3
Properties of the input and output variables in the testing dataset.

Categories	T_P	pH_{ww}	C	(O + N)/C	H/C	A	PSB
Min.	300	6.78	43.29	0.048	0.16	2.2	0.15
1st Qu.	500	9.81	51.69	0.12	0.291	7.52	0.2
Median	550	10.08	69.04	0.168	0.376	21.25	0.2
Mean	556.5	9.83	65.79	0.1824	0.4326	21.15	0.4659
3rd Qu.	650	10.5	76.3	0.26	0.549	32.27	0.25
Max.	700	11.1	85.3	0.365	1.084	47.93	2
Catergories	BSA	CEC	T_{envi}	pH_{sol}	C_0	SEoHM	–
Min.	2.46	3.17	20	2	0.0029	0.0003	–
1st Qu.	14.03	7.85	20	5	0.0089	0.004	–
Median	20.1	17.4	25	5.5	0.2	0.0823	–
Mean	30.64	16.26	23.97	5.428	0.5067	0.2502	–
3rd Qu.	41.12	21.8	25	6	1	0.2818	–
Max.	309.29	40.2	28	8	1.9305	1.5568	–

to generate novel hybrid models (ensemble models) for similar purposes with improved performance, namely BA-SVM, BA-RF, BA-M5Tree, BA-GP, BA-ANN, SVM-RF, SVM-M5Tree, SVM-GP, SVM-ANN, RF-M5Tree, RF-GP, RF-ANN, M5Tree-GP, M5Tree-ANN, and GP-ANN. The details of the models' development, as well as their properties, are presented step-by-step in the next sections.

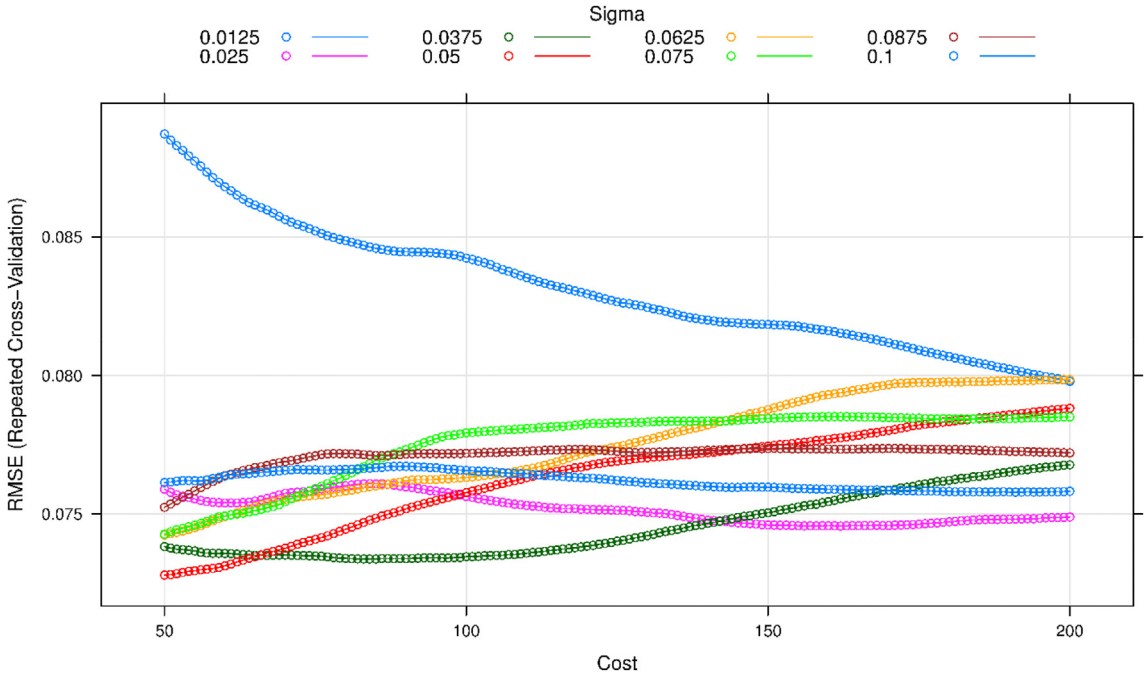


Fig. 6. Grid search for the development of the SVM model in predicting sorption efficiency.

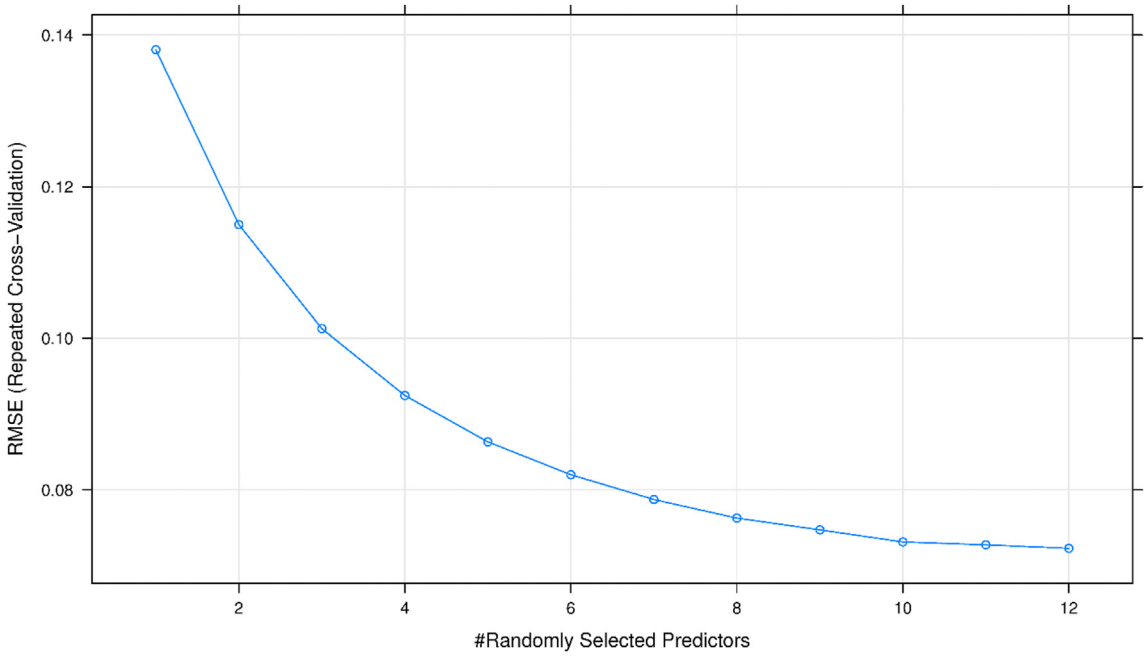


Fig. 7. RMSE of the RF model on the training phase with different randomly selected predictors.

4. Methodology

As mentioned above, this study applies the BA techniques to develop various novel hybrid models (ensemble models), including BA-SVM, BA-RF, BA-M5Tree, BA-GP, BA-ANN, SVM-RF, SVM-M5Tree, SVM-GP, SVM-ANN, RF-M5Tree, RF-GP, RF-ANN, M5Tree-GP, M5Tree-ANN, and GP-ANN. Therefore, the methodology of the hybrid models will be presented in detail in this section. Besides, the sub-methods background, such as SVM, RF, M5Tree, GP, and ANN, are presented in the Supplementary material.

Ensemble models are known as an approach to improving forecasting models (Nguyen et al., 2020; Ke et al., 2018; Zhou et al., 2021; Zhou et al., 2021). They rely on bagging to combine many weak predictors to create a new model with better predictability and higher reliability (Breiman, 1996; Quinlan, 1996; Dudoit and Fridlyand, 2003). Based on bagging theory, the efforts of this study are to combine single models (i.e., SVM, RF, M5Tree, GP, and ANN) to generate ensemble models (hybrid models) with better performance, including BA-SVM, BA-RF, BA-M5Tree, BA-GP, BA-ANN, SVM-RF, SVM-M5Tree, SVM-GP, SVM-ANN, RF-M5Tree, RF-

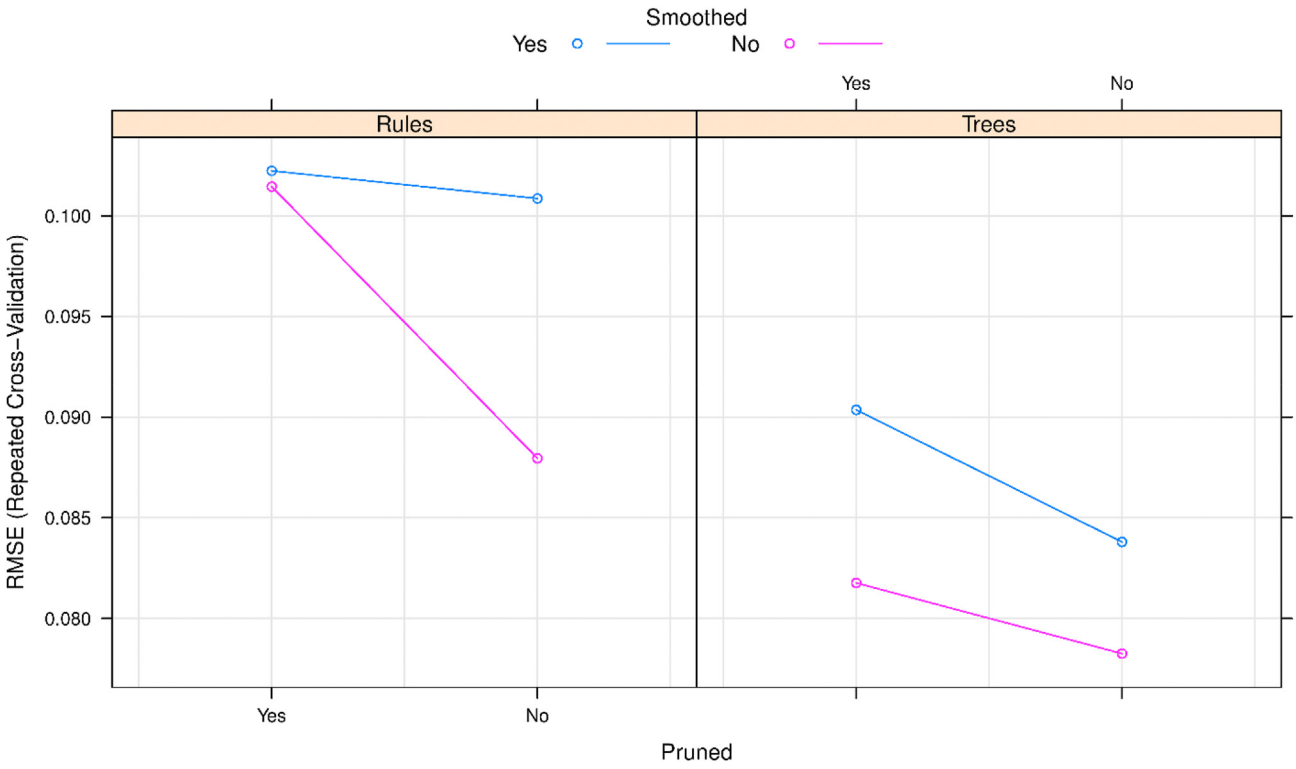


Fig. 8. RMSE of the M5Tree model on the training phase.

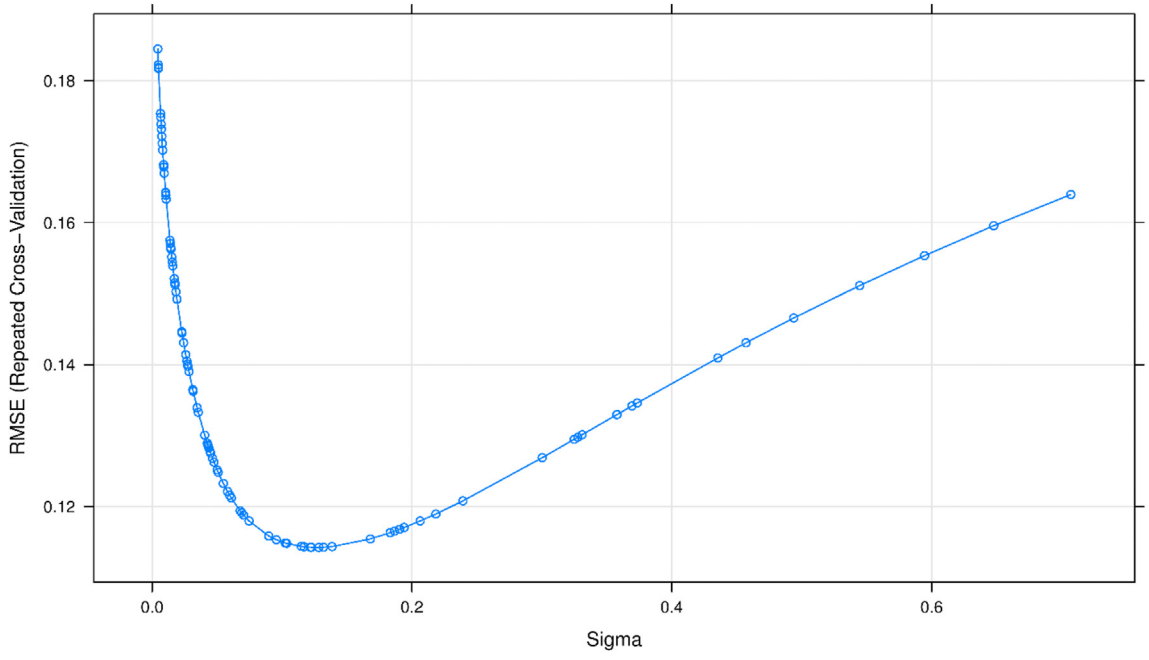


Fig. 9. RMSE of the GP model on the training phase with various values of σ

GP, RF-ANN, M5Tree-GP, M5Tree-ANN, and GP-ANN. The framework of novel ensemble models for predicting the SEoHM of the biochar system in this study is presented in Fig. 4.

The original dataset is split into two parts to developing the ensemble models: 80% for training and 20% for testing (Nguyen, 2020). The single models are then developed based on the training dataset, including SVM, RF, M5Tree, GP, and ANN. Note that these single models are different from the optimal developed single models before. They should be weak models, and they are considered as the sub-models for the ensemble models. In the next step, the sub-models are combined based on the outcome predictions. Ultimately, the testing dataset is used to test and evaluate the performance, as well as the accuracy of the ensemble models in practice. The process of developing ensemble models to predict the

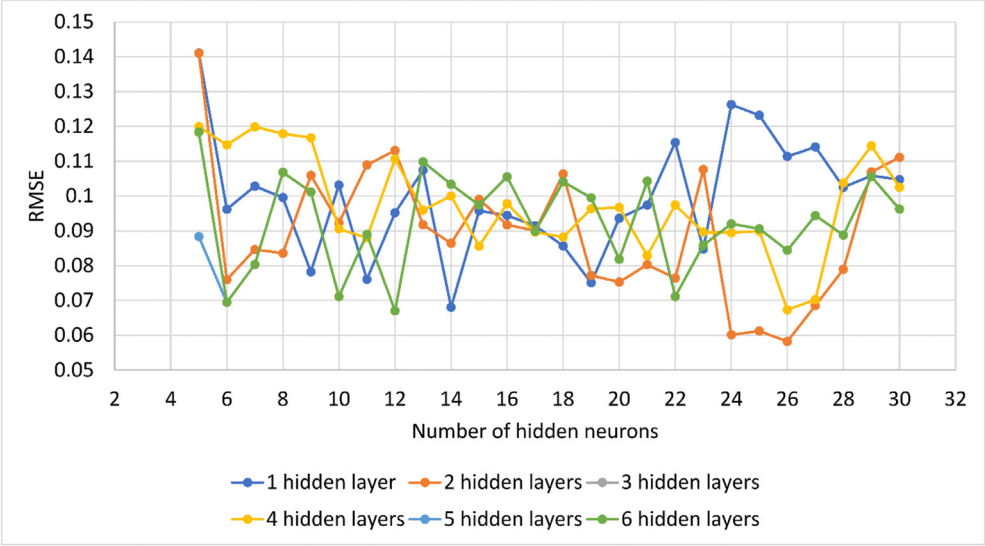


Fig. 10. Performance of the ANN model with different hidden layers and hidden neurons.

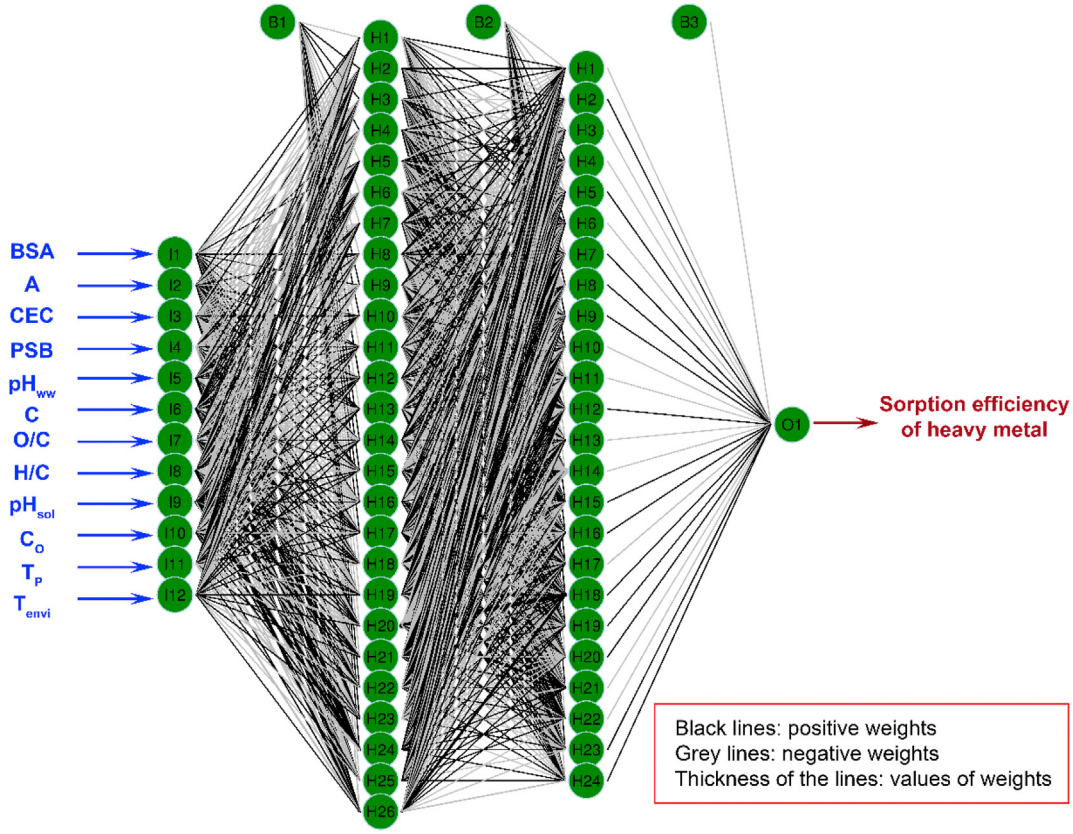


Fig. 11. The ANN structure for predicting the sorption efficiency of heavy metal on biochar.

SEoHM in this study is proposed in Fig. 4.

To evaluate the performance of the machine learning models, statistical indicators are used to evaluate the accuracy of the model and model compatibility with the dataset used, including root-mean-squared error (RMSE), determination coefficient (R^2), and mean absolute error (MAE). In addition to these statistical indicators, additional model evaluation methods based on the color spectrum, data distribution, and standard deviation were also performed, including the boxplot and Taylor diagram. The details of these methods are described in Supplementary material.

Table 4
Performance of the single and novel ensemble models in predicting the SEoHM.

Model	Training			Testing		
	RMSE	R^2	MAE	RMSE	R^2	MAE
SVM	0.073	0.953	0.044	0.089	0.951	0.050
RF	0.064	0.972	0.035	0.066	0.973	0.033
M5Tree	0.072	0.962	0.041	0.079	0.959	0.031
GP	0.114	0.897	0.066	0.136	0.885	0.082
ANN	0.058	0.972	0.030	0.075	0.965	0.036
BA-SVM	0.080	0.975	0.038	0.071	0.972	0.043
BA-RF	0.057	0.980	0.029	0.060	0.977	0.032
BA-M5Tree	0.070	0.967	0.030	0.073	0.965	0.034
BA-GP	0.121	0.889	0.070	0.133	0.892	0.080
BA-ANN	0.054	0.973	0.031	0.071	0.970	0.038
SVM-RF	0.055	0.978	0.034	0.063	0.974	0.036
SVM-M5Tree	0.078	0.955	0.047	0.090	0.952	0.051
SVM-GP	0.168	0.813	0.080	0.194	0.793	0.086
SVM-ANN	0.036	0.995	0.018	0.046	0.987	0.026
RF-M5Tree	0.066	0.969	0.031	0.072	0.967	0.033
RF-GP	0.176	0.891	0.148	0.184	0.916	0.152
RF-ANN	0.079	0.952	0.043	0.096	0.948	0.050
M5Tree-GP	0.139	0.905	0.078	0.152	0.903	0.082
M5Tree-ANN	0.069	0.976	0.030	0.074	0.967	0.035
GP-ANN	0.057	0.977	0.029	0.064	0.974	0.036

and density of this pair of variables in Fig. 5 shows that their similarity is almost absolute. Therefore, one of the two variables needs to be removed to ensure the objectivity of the model. Looking at the composition of the two variables (i.e., (O + N)/C and O/C), it can be seen that the (O + N)/C variable includes the ratio of oxygen and nitrogen with carbon. Meanwhile, the O/C variable does not reflect the influence of nitrogen on carbon. Therefore, the O/C variable was removed in this research.

Before developing the machine learning models mentioned above, the dataset is split according to the ratio of 80/20 with 80% of the whole sorption efficiency dataset was randomly selected to develop the models through the learning of the algorithms. Subsequently, the remaining 20% is used to evaluate/validate the developed models' performance. The characteristics of the training and testing datasets are listed in Tables 2 and 3, respectively.

6. Development of the models

In this section, the detail of the models' development is described step-by-step. To develop the SVM model, the dataset was normalized using the BoxCox, center, and scaled methods to avoid

Fig. 12. Color spectrum for model assessment. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

5. Data acquisition and analyses

To carry out this study, a dataset of 353 observations of heavy metal adsorption was collected from the previous 12 studies (Sun et al., 2014; Trakal et al., 2014; Shen et al., 2015, 2017a, 2017b; Cui et al., 2016a, 2016b; Ding et al., 2016; Jiang et al., 2016; Zama et al., 2017; Li et al., 2018; Gao et al., 2019). The dataset characteristics include biochar characteristics, adsorption conditions, the initial concentration ratio of heavy metals to biochar, and heavy metal properties. However, the heavy metal properties, such as ion radius, charge number, and electronegativity, were indicated that not significant with the sorption efficiency of heavy metal (Zhu et al., 2019). Therefore, they were not used to predict the SEoHM in the present study. Instead, biochar characteristics, metal sources, and environmental conditions were taken into account to predict ESoHM in this study with 13 input variables, such as biochar surface area (BSA), percentage of ash (A), cation exchange capacity (CEC), particle size of biochar (PSB), pH of biochar in wastewater (pH_{ww}), percentage of carbon in biochar (C), the ratio of oxygen and carbon (O/C), the ratio of hydrogen and carbon (H/C), ratio of O and N with C [(O + N)/C], solution pH (pH_{sol}), heavy metal concentration in wastewater (C_O), pyrolysis temperature (T_P), and environmental temperature (T_{envi}). The characteristics of the dataset are presented in Table 1.

In machine learning, data needs to be well understood before developing machine learning models in order to increase the reliability of predictive models (Nguyen et al., 2020). In other words, the data needs to be analyzed to select the models as well as the appropriate model training methods. As can be seen, the number of input variables in the dataset is high, so an analysis of the correlation as well as the distribution of them is needed to minimize the number of input variables if the level is similarities high. A correlation and distribution matrix, as well as the density of the input variables, are analyzed in Fig. 5. Accordingly, it can be seen that the correlation between pairs of variables (O + N)/C and O/C is absolute (i.e., correlation value equal to 1). A closer look at the distribution overfitting the model (Grace and Durham, 2001; Schnitzer et al., 2003). Besides, the 10-fold cross-validation technique was also used to improve the SVM model's accuracy. For the regression problems (i.e., SEoHM), kernel functions are applied to support the model's mapping data, as presented above. A review of the literature shows that the radial basis function (RBF) often provides better results for regression problems (Nguyen, 2019; Nguyen et al., 2019b). Therefore, RBF was selected as the primary kernel function to develop the SVM model for predicting SEoHM in this study with σ and *Cost* (τ) are the main parameters used to control the SVM model's accuracy. The grid search with σ in the range of [0.0125, 0.1] and τ in the range of [50, 200] was established to define the optimal SVM model for the present problem. Finally, 1207 SVM models were taken into account and computed for performance with different parameters, as illustrated in Fig. 6. The results indicated that the best SVM model for estimating the SEoHM in this study reached at $\sigma = 0.05$ and $\tau = 50$.

For the RF model, the same techniques were applied to develop the RF model as those used for the SVM model. The number of randomly selected predictors ($n_{predictors}$) and the number of trees in the forest (n_{tree}) are the critical parameters in developing the RF model. According to previous researchers, n_{tree} should be diverse to ensure that the predictive results represent high confidence. Therefore, the n_{tree} was selected equal to 2000 to meet the above requirements. In this study, 12 predictors were used to predict the SEoHM as introduced above; therefore, the $n_{predictors}$ should be selected in the range of [1,12]. The results of calculating the RF model's performance on the training dataset with different parameters show that the best RF model at $n_{tree} = 2000$ and $n_{predictors} = 12$ (Fig. 7).

For the M5Tree model, the regression tree development was performed to predict the SEoHM. To develop regression trees, pruning may be necessary to minimize errors of the model. Subsequently, the smooth task can be applied to smooth out the pruned locations, minimize errors and losses. The rules were also considered whether they could be applied or not to improve the

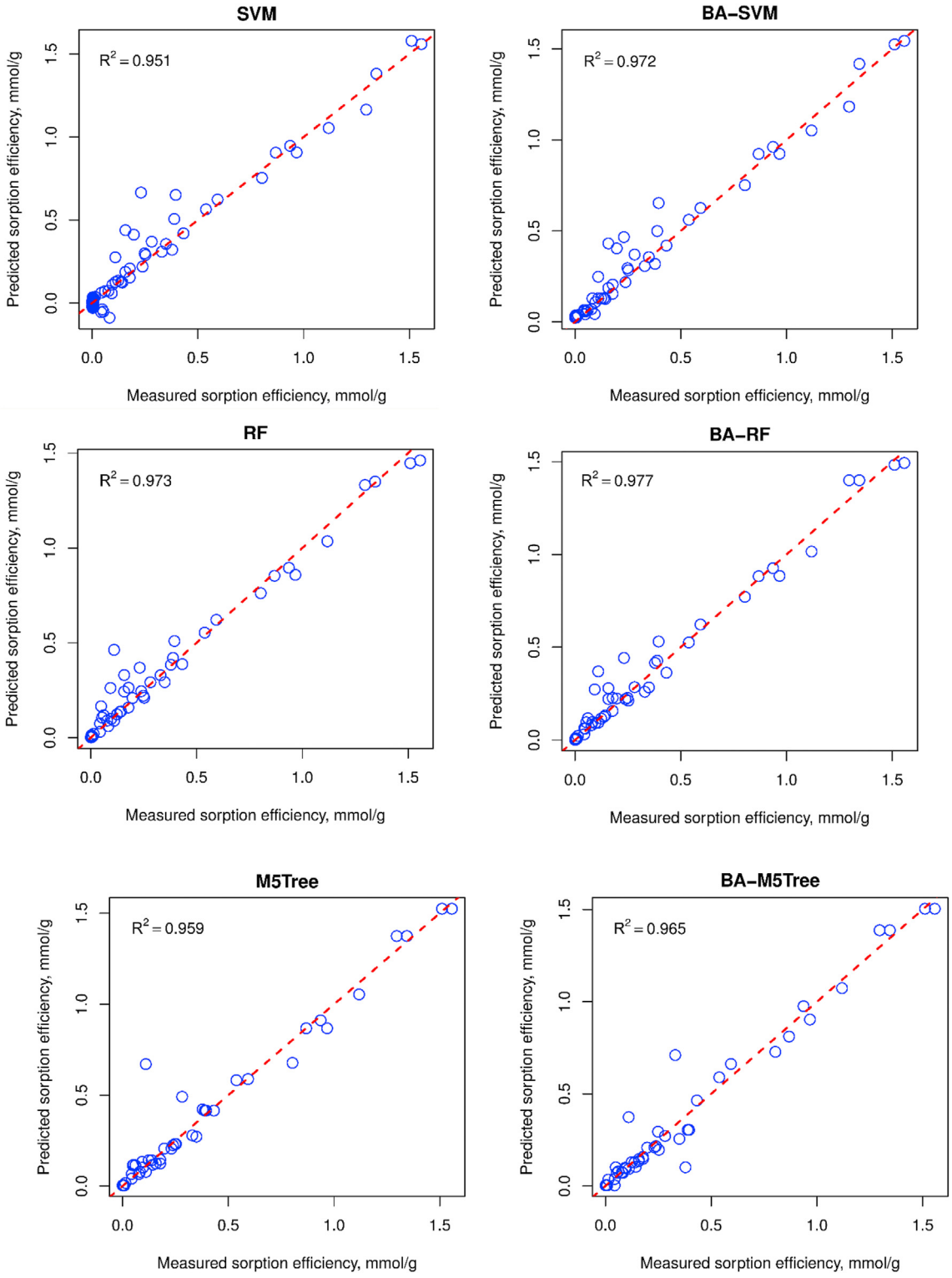


Fig. 13. Correlation of measured and predicted by different intelligence models.

model's accuracy. Finally, eight M5Tree models were developed based on three parameters: pruned, smoothed, and rules. The optimal M5Tree model was then defined, as shown in Fig. 8.

For the GP model, the development process is similar to the SVM model. The RBF was also applied to support the GP model with the Gaussian distribution. However, there is only one parameter involved in adjusting the GP model's performance, i.e., σ. A grid search with σ in the range of [0, 0.7] was established to determine

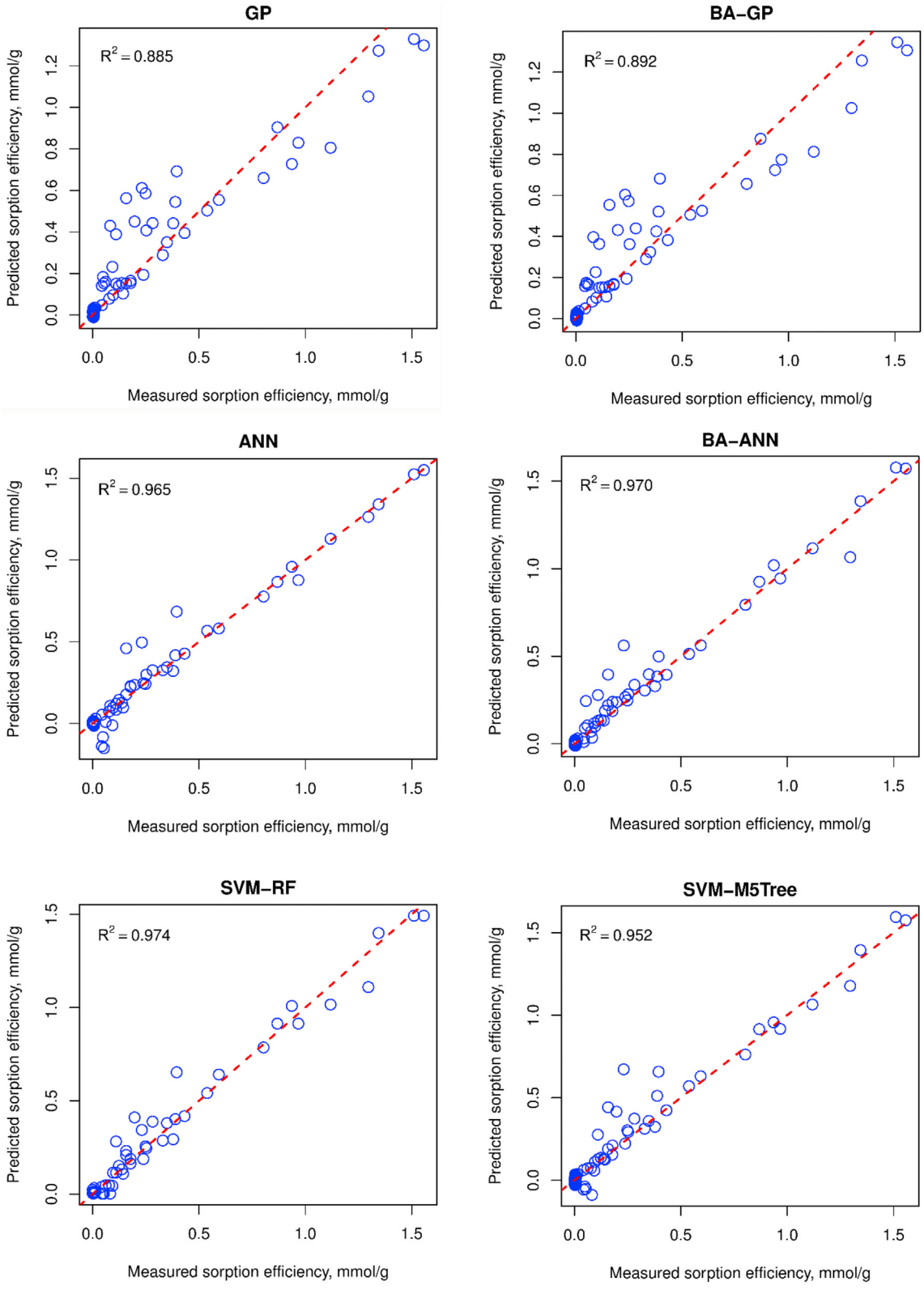


Fig. 13. (*continued*).

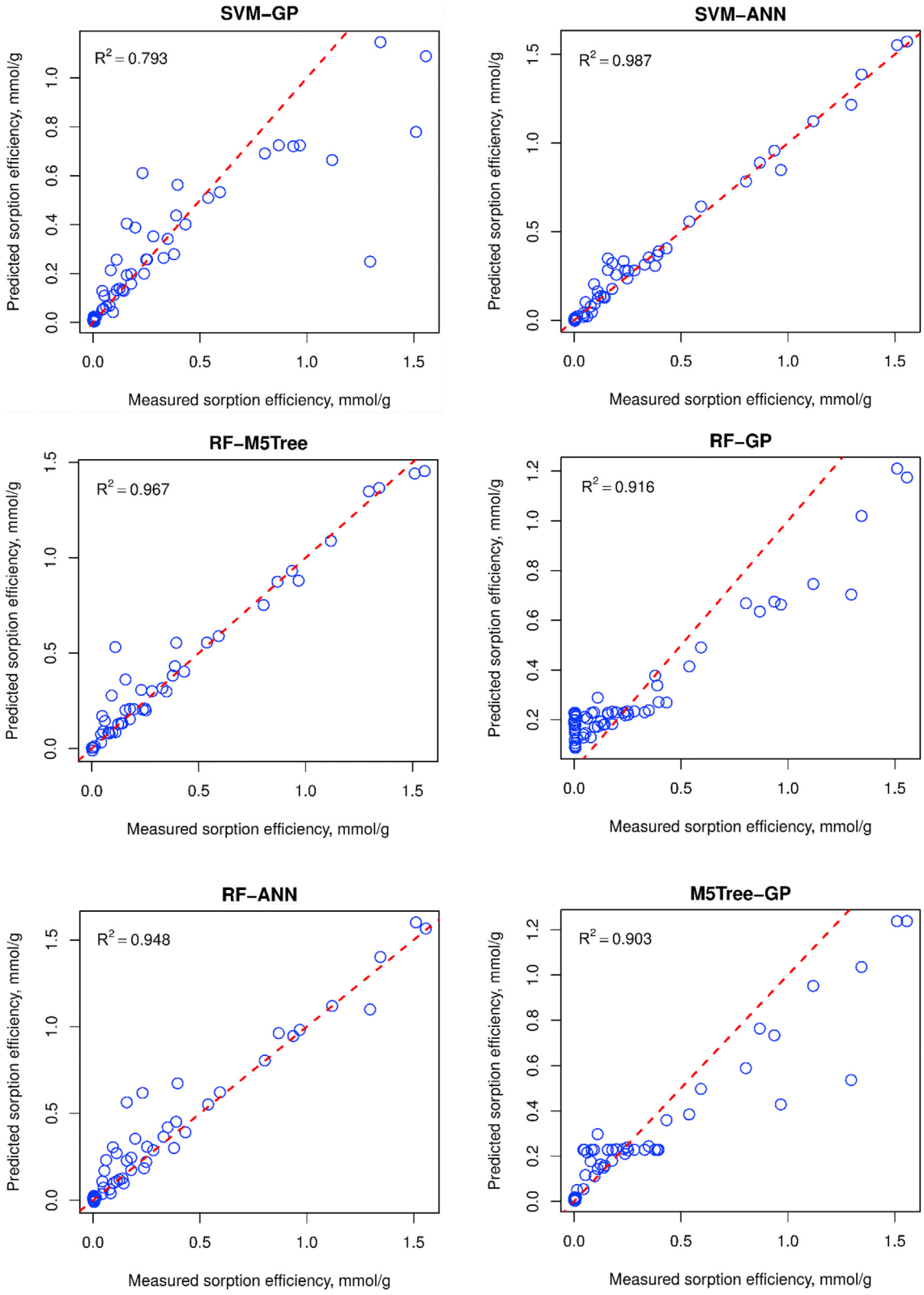


Fig. 13. (*continued*).

the optimal GP model. Finally, the optimal GP model was defined with $\sigma = 0.128$, as shown in Fig. 9.

For the ANN model, the normalized technique of MinMax in the range [0,1] was applied to prevent overfitting occurrence. Then, the BP algorithm was selected as the potential algorithm for training the ANN model. Different hidden layers and hidden neurons was applied to select the ANN model's best structure, as shown in Fig. 10. Accordingly, the optimal structure of the ANN should include two hidden layers with 26 and 24 hidden neurons in the first and second hidden layers, respectively (Fig. 10). Eventually, an optimal ANN was proposed to predict the SEoHM of biochar based on the BP algorithm, as introduced in Fig. 11.

To develop the ensemble models, the framework in Fig. 4 was applied. Accordingly, the single models with low performance (i.e., weak regressors) were used as the sub0-models and combined to generate a new ensemble model with higher accuracy. Herein, we selected five sub-models for each ensemble model and combined them with each other. Subsequently, their performance on the training and testing datasets was computed and compared through statistical metrics, color intensity, boxplot, and Taylor diagram. The development of the novel ensemble models for predicting the SEoHM in this study was implemented as follows:

- BA-SVM model: generate five weak SVM models, then combine them by the SVM.
- BA-RF model: generate five weak RF models, then combine them by the RF.
- BA-M5Tree model: generate five weak M5Tree models, then combine them by the M5Tree.
- BA-GP model: generate five weak GP models, then combine them by the GP.
- BA-ANN model: generate five weak ANN models, then combine them by the ANN.
- SVM-RF model: generate five weak SVM models, then combine them by the RF.
- SVM-M5Tree model: generate five weak SVM models, then combine them by the M5Tree.
- SVM-GP model: generate five weak SVM models, then combine them by the GP.
- SVM-ANN model: generate five weak SVM models, then combine them by the ANN.
- RF-M5Tree model: generate five weak RF models, then combine them by the M5Tree.
- RF-GP model: generate five weak RF models, then combine them by the GP.
- RF-ANN model: generate five weak RF models, then combine them by the ANN.
- M5Tree-GP model: generate five weak M5Tree models, then combine them by the GP.
- M5Tree-ANN model: generate five weak M5Tree models, then combine them by the ANN.
- GP-ANN model: generate five weak GP models, then combine them by the ANN.

7. Results and discussion

Prior develop the SEoHM predictive models, three statistical indices, including RMSE, R^2, and MAE were pointed out as the standards for evaluating the models. The evaluations are performed on both training and testing datasets in order to evaluate the performance of modeling in theory and practice. The SEoHM predictive models' results are computed and listed in Table 4. Also, the color spectrum method (Fig. 12) was applied to highlight the models' quality.

Table 4 indicated that the machine learning models predicted the SEoHM very well. Among the single models, the RF model provided the most superior performance (i.e., RMSE of 0.064 and 0.066; R^2 of 0.972 and 0.973; and MAE of 0.035 and 0.033) on the training and testing datasets. In contrast to the RF model, the GP model provided the lowest performance on the datasets (e.g., RMSE of 0.114 and 0.136; R^2 of 0.897 and 0.885; MAE of 0.066 and 0.082). These results revealed that the Gaussian distribution does not appear strong enough to explain clearly the relationship between the SEoHM and the input variables. Further experimental tests were conducted on the GP-based models (ensemble models based on the GP model) to confirm this statement.

Looking at the new ensemble models' results, we can see that most ensemble models predicted the SEoHM better than individual models. Indeed, the values in Table 4 revealed that combining multiple SVM models (weak SVM models) to create new hybrid models, such as BA-SVM, SVM-RF, SVM-M5Tree, and SVM-ANN, is better than the single SVM model. In particular, the color spectrum showed that the SVM-ANN model is the best of the 20 models developed to predict SEoHM in this study with a promising result (i.e., RMSE = 0.046, R^2 = 0.987, and MAE = 0.026 on the testing

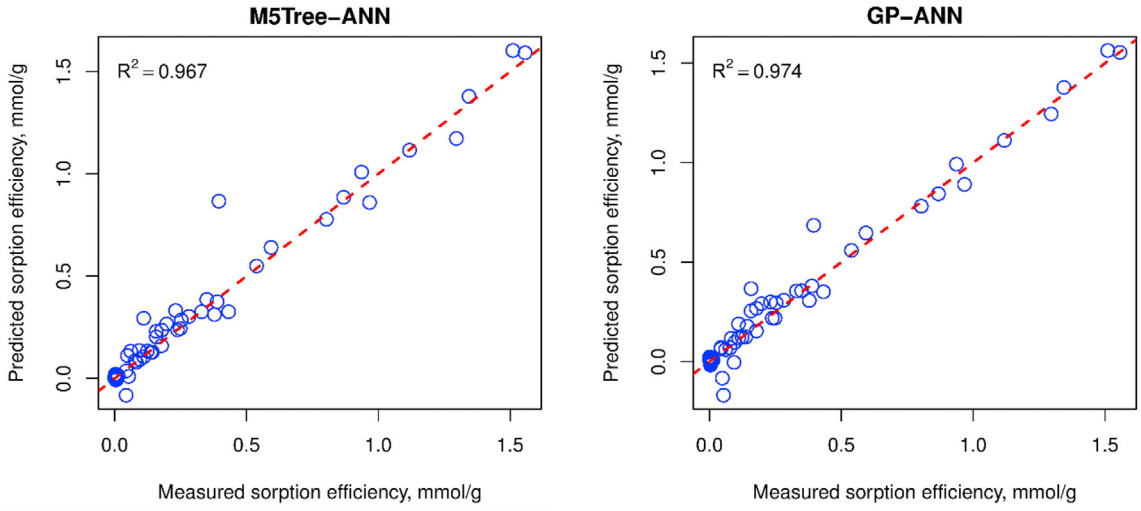


Fig. 13. (*continued*).

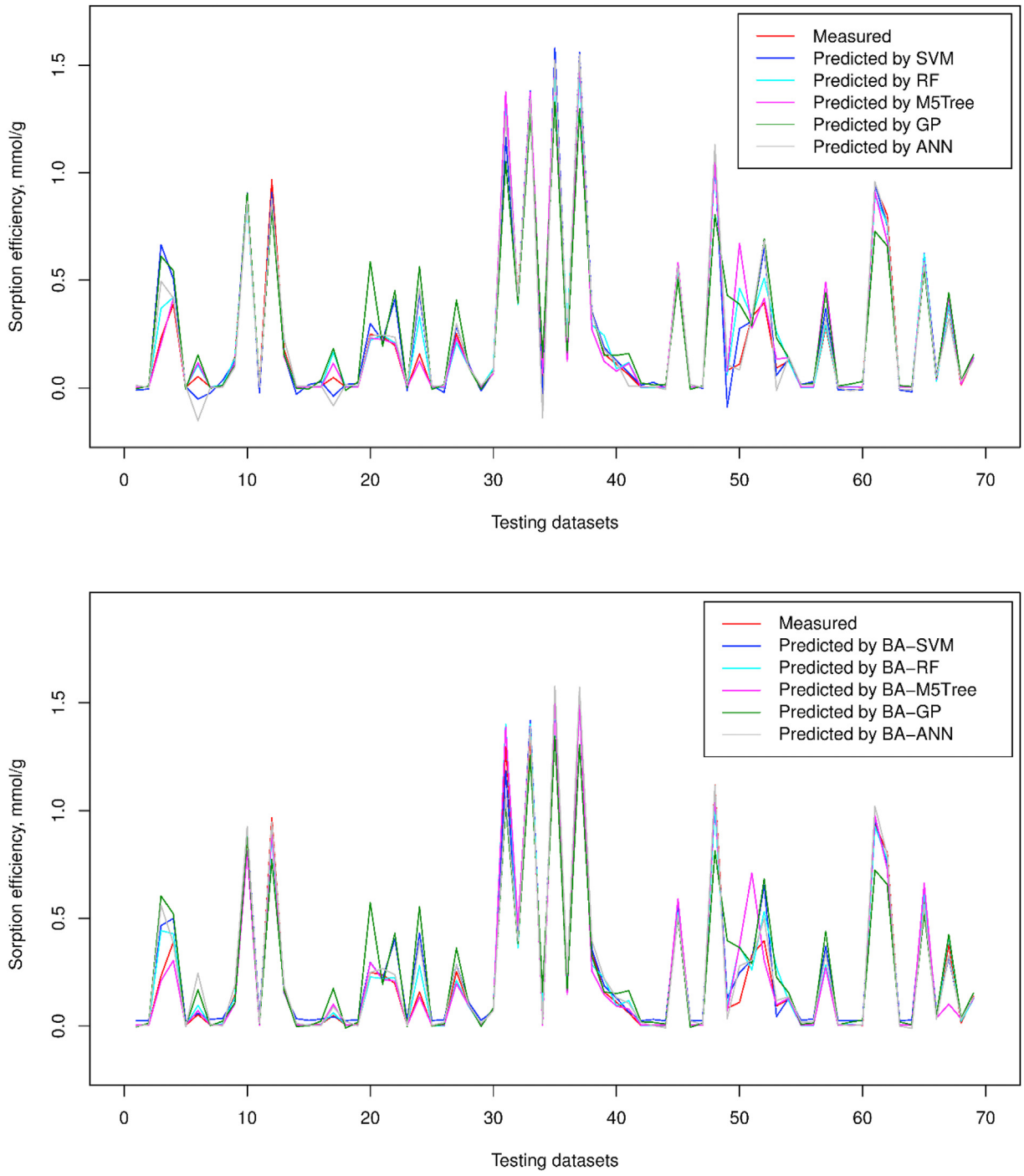


Fig. 14. Predictions resulting from the individual models.

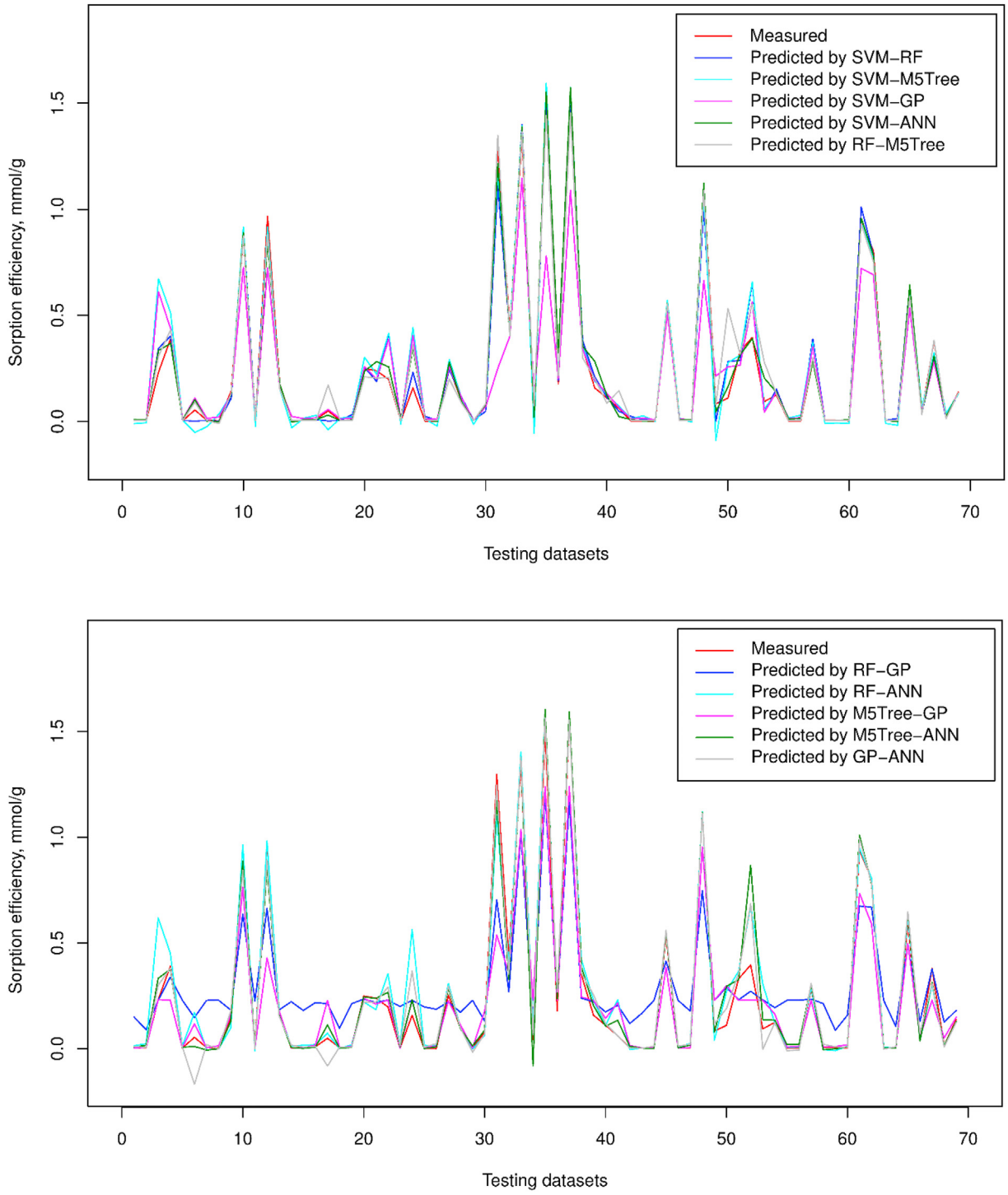


Fig. 14. (*continued*).

dataset). However, not all ensemble models are superior to single models. Indeed, comparing the SVM model's performance with the SVM-M5Tree model, the performance between them is not too different. More significantly, the combination of many sub-SVM models by GP has led to the SVM-GP model's performance being even lower than the single GP model. Similar problems have also occurred for the RF-GP and M5Tree-GP models. This finding explains the Gaussian distribution mismatch for the SEoHM data in this study, as stated above. However, the combination of multiple Gaussian distributions (i.e., weak GP models) by ANN (GP-ANN) seems to yield higher performance than the single ANN model in the SEoHM prediction. This finding shows that ANN can explain the Gaussian relationship of weak GP models in this study well. To better assess the feasibility of the developed models in predicting the SEoHM of the biochar system, let's look at the correlation of predictive models and actual in Fig. 13.

It is a fact that ensemble models based on sub-models of SVM, RF, and M5Tree seem to be better suited to the actual dataset. The ensemble models' predicted values based on sub-models of SVM, RF, and M5Tree are closer to the regression line than the remaining

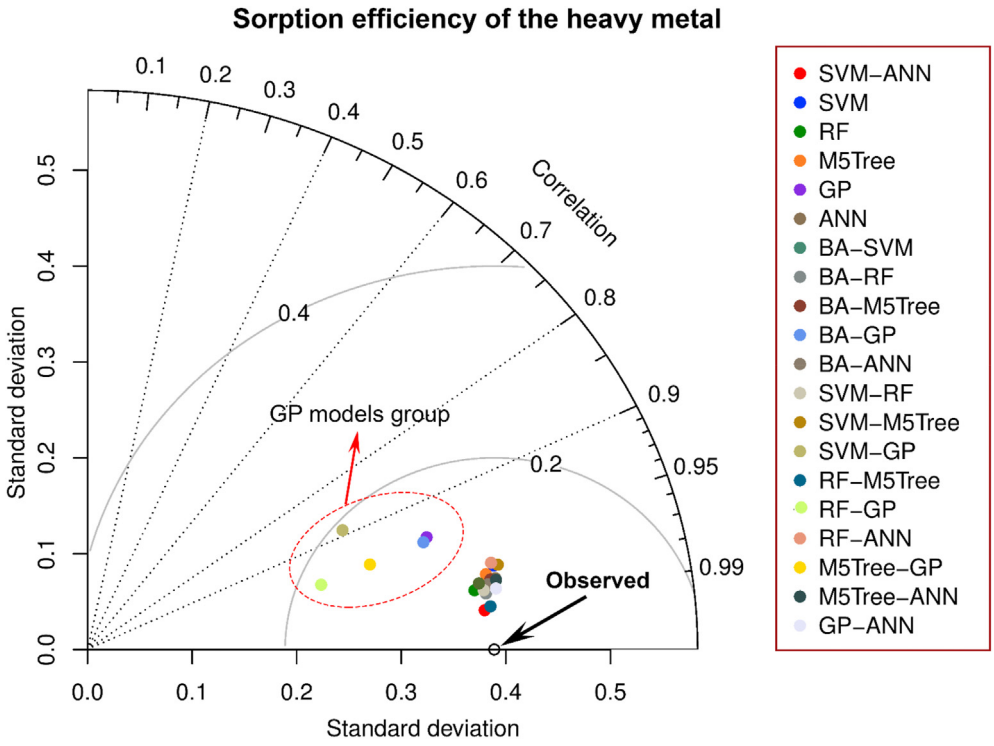


Fig. 15. Taylor diagram for models' assessment through standard deviation and correlation metrics.

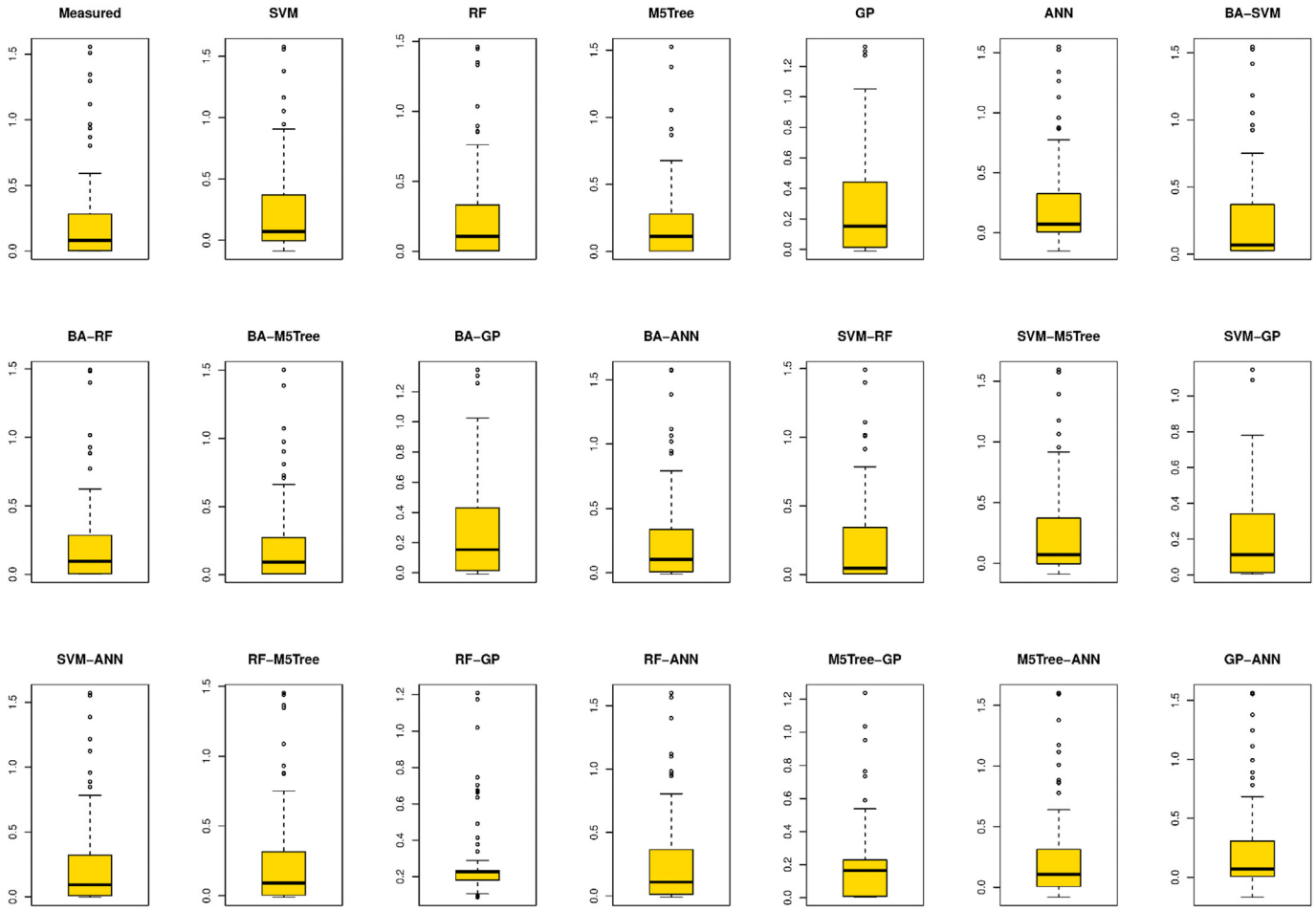


Fig. 16. Boxplot of the sorption efficiency predictive models.

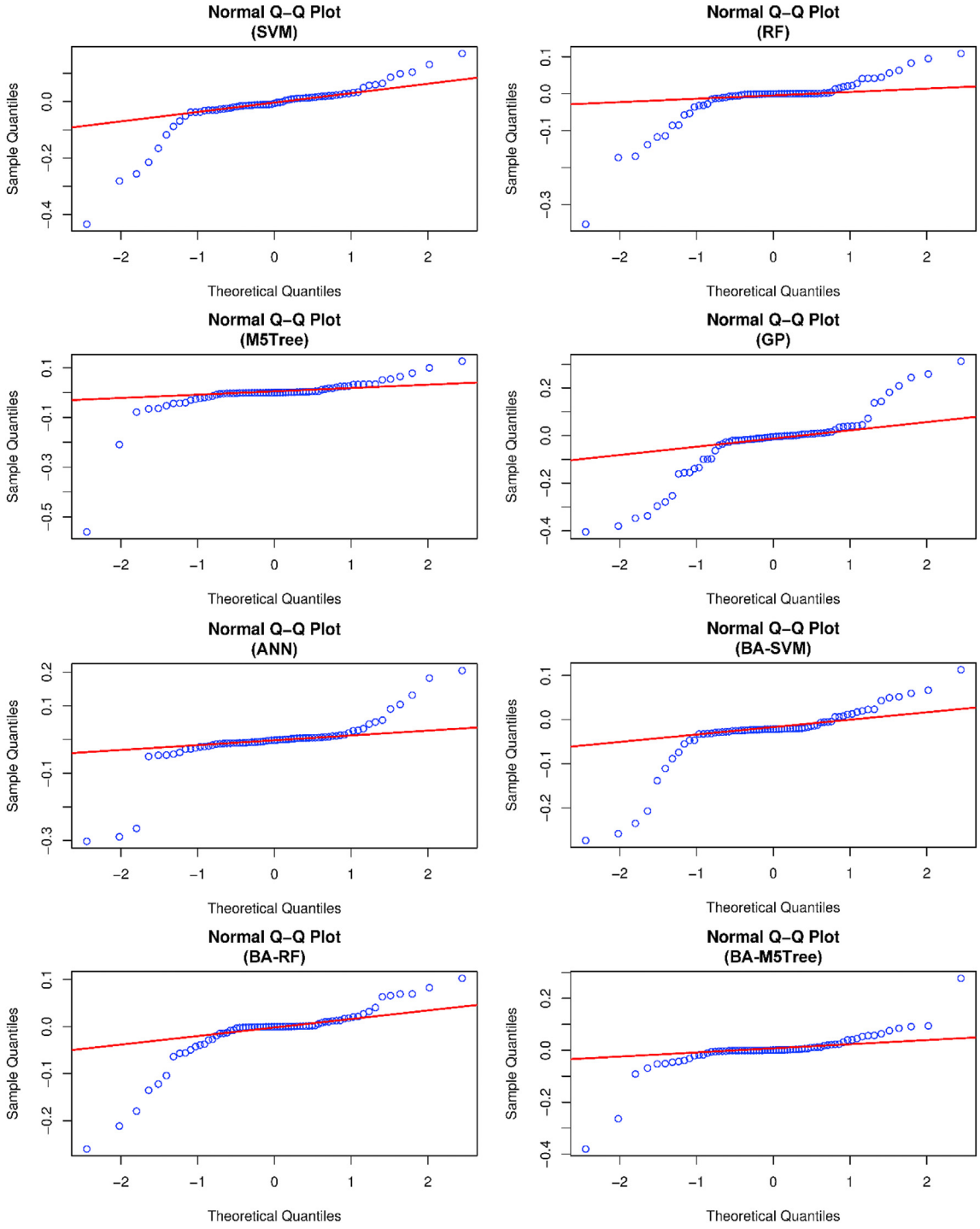


Fig. 17. Q-Q plot of the sorption efficiency predictive models for evaluating the fitness of the models.

models. Meanwhile, the ensemble models based on the GP's sub-models have provided predicted values far from the regression line. A closer look at the ensemble models for predicting the SEoHM by GP showed that the greater the sorption efficiency, the higher the error of the model. In other words, GP seems to be not suitable to predict the SEoHM with high adsorption efficiency (>0.2 mmol/g in this study). To compare the absolute accuracy in practical, the predicted values and the actual values are illustrated in Fig. 14. The prediction resulting of the models on the testing dataset is presented in Supplementary materials. Furthermore, a Taylor diagram and a boxplot were also analyzed to assess further the models developed, as shown in Figs. 15 and 16.

As seen in the Taylor graph (Fig. 15), it is easy to recognize that the SVM-ANN model is the best model for predicting the SEoHM in

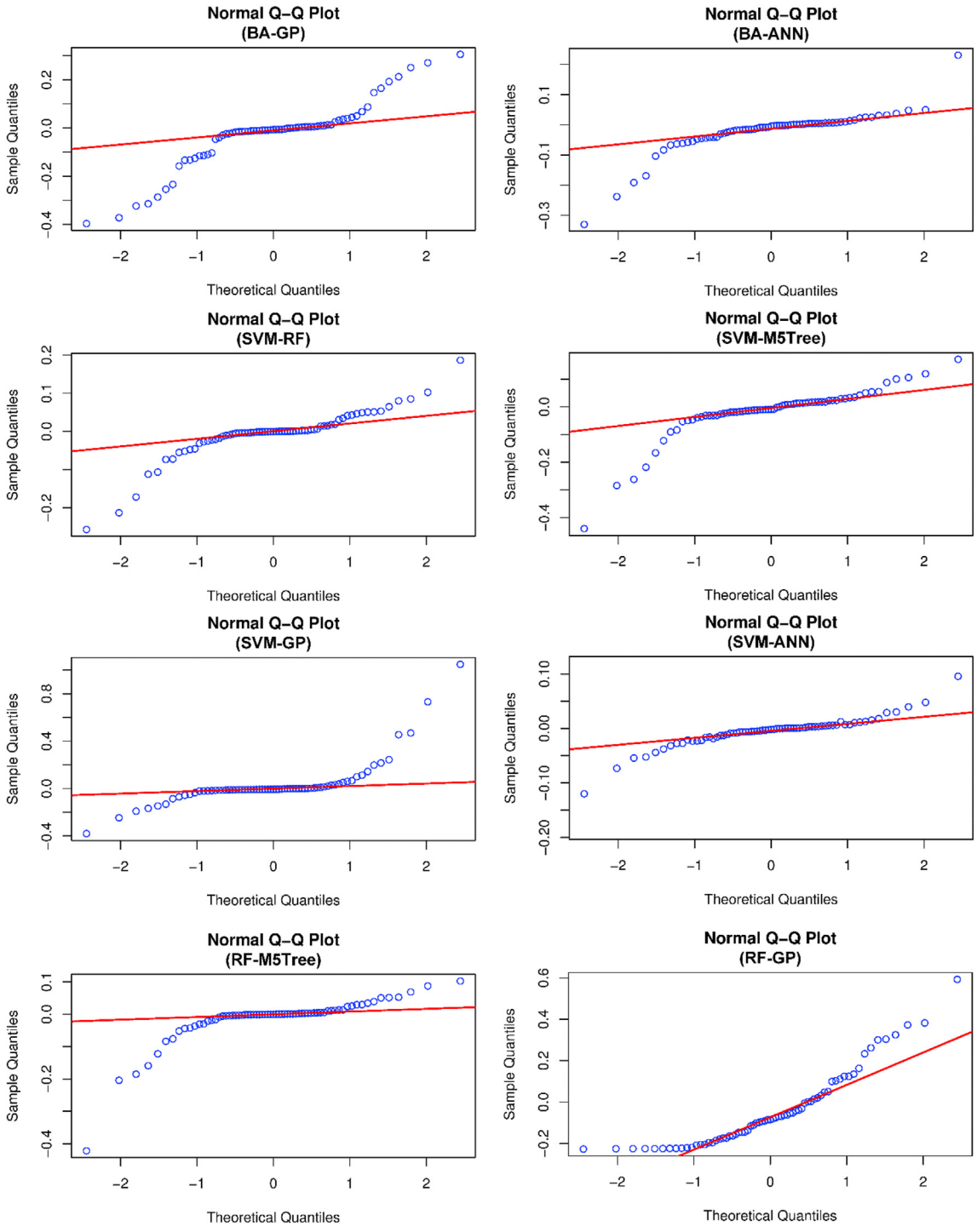


Fig. 17. (*continued*).

this study, with the standard error is lower than the observed and the correlation is highest. The RF-M5Tree model seems to provide a slightly lower correlation; however, its standard error is higher than the SVM-ANN model. Observing the distribution of the models' output (predicted SEoHM) on the boxplot shows that all models appear outliers. However, the SVM-ANN model's predicted SEoHM distribution has the highest similarity to the actual model (measured SEoHM). They show the high stability of the SVM-ANN model in predicting the SEoHM in this study. Notably, the RF-GP model has the smallest IQR, much lower than the actual model. The outliers also appeared more than the actual model. They show significant differences between the RF-GP model and the remaining models in predicting the SEoHM.

Fig. 16 shows that the dataset's distribution through the boxes

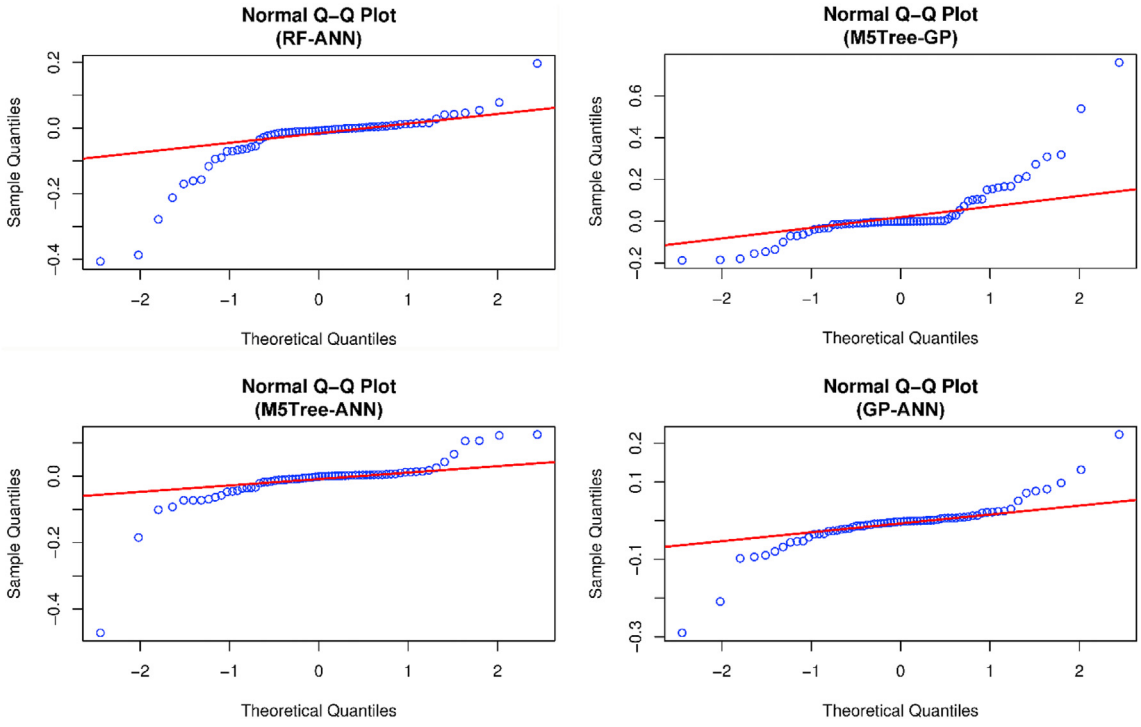


Fig. 17. (*continued*).

and whiskers showed that the SVM-ANN model is very close to the actual model. Some other models, such as GP, BA-GP, SVM-RF, SVM-GP, RF-M5Tree, provided fewer outliers; however, their distribution is very different from the actual model. Remarkably, the RF-GP model's data distribution showed that the average of the predicted SEoHM is much higher than the actual model, and the data tends to lean heavily above the mean. These findings show that the predicted SEoHM values of the RF-GP model have large fluctuations, instability, and low quality. The GP algorithm's machine learning models also provided large deviations, indicating the models' instability.

Although the models' accuracy, performance, and stability have been evaluated in detail as presented above. However, the models' inherent random errors need to be checked whether they are extracted from a normal distribution or not. A quantile-quantile plot (Q-Q plot) was analyzed in Fig. 17, aiming to evaluate the normal distribution and the models' random errors. Accordingly, the normality of the residuals of the models was investigated through the Q-Q plots. The findings show that the ensemble models have the normality of the residuals higher than those of the single models. Of those, the SVM-ANN model is the best fit for the SEoHM database with the highest normal distribution and the lowest random errors inherent. Based on all the obtained results and analysis, it is possible to see that SVM-ANN is the best model in this study for predicting the SEoHM with outstanding accuracy and stability.

8. Conclusions and remarks

Heavy metals in water and wastewater are dangerous components that seriously affect human health. Applying biochar technology to adsorb heavy metals in water and wastewater has important implications in life and brings co-benefits to the economy and the environment. However, the heavy metal adsorption system's effectiveness depends on many factors, such as biochar characteristics, metal sources, and environmental conditions. This study has successfully developed a series of artificial intelligence models to predict the sorption efficiency of the heavy metal in the biochar system with high reliability, including SVM, RF, M5Tree, GP, ANN, BA-SVM, BA-RF, BA-M5Tree, BA-GP, BA-ANN, SVM-RF, SVM-M5Tree, SVM-GP, SVM-ANN, RF-M5Tree, RF-GP, RF-ANN, M5Tree-GP, M5Tree-ANN, GP-ANN. Of those, the SVM-ANN model was introduced as the most accurate predictive model with the highest reliability. The results have proved that the ensemble models have the potential to improve the accuracy of individual prediction models. However, the selection of training algorithms should also be considered under the dataset used. Based on the models proposed in this study, water/wastewater treatment plants can effectively control the adsorption of heavy metals in water/wastewater to ensure the quality of the outlet water. In addition, the water/wastewater treatment capacity of the plants can also be adjusted accordingly based on the SEoHM prediction models proposed in this study.

Although this work's findings are outstanding, and the obtained results have important significance in the wastewater treatment industry and mitigate environmental water pollution. Notably, this study used a dataset collected from 12 previous studies with various biochar systems and adsorbents. The results showed that the developed models, especially the SVM-ANN model, have good generalizability for various biochar systems and adsorbents. The developed models' accuracy can be improved if they are applied for only a stand-alone biochar system with an adsorbent. However, this study's limitation is the size of a dataset for a stand-alone biochar system with an adsorbent. It is too few samples in each previous study to develop and validate the AI models. Therefore, increasing the samples (datasets) of each biochar system with an adsorbent and developing AI models for estimating the SEoHM of each biochar system with high accuracy should be considered and employed in future studies.

Declaration of competing interest

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

Acknowledgments

The authors would like to thank the Center for Mining, Electro-Mechanical research of Hanoi University of Mining and Geology (HUMG), Hanoi, Vietnam, and the research team of Innovations for Sustainable and Responsible Mining (ISRM) of HUMG.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemosphere.2021.130204.

Credit author statement

Hoang Nguyen and Xuan-Nam Bui: Conceptualization, Methodology, Software, Data curation, Visualization, Writing – original draft and revise the manuscript. Hoang Bac Bui, Yosoon Choi, Jian Zhou, Hossein Moayedi, Romulus Costache, and Thao Nguyen-Trang: Software, Data curation, Visualization, Writing – original draft. Bo Ke: Conceptualization, Methodology, Software, Data curation, Visualization, and revise the manuscript.

References

Al-Wabel, M.I., Al-Omran, A., El-Naggar, A.H., Nadeem, M., Usman, A.R., 2013. Pyrolysis temperature induced changes in characteristics and chemical composition of biochar produced from conocarpus wastes. Bioresour. Technol. 131, 374–379.

Alrumman, S., Keshk, S., El Kott, A., 2016. Water pollution: source & treatment. Am. J. Environ. Eng. 88–98.

Appel, C., Ma, L., 2002. Concentration, pH, and surface charge effects on cadmium and lead sorption in three tropical soils. J. Environ. Qual. 31, 581–589.

Bolisetty, S., Peydayesh, M., Mezzenga, R., 2019. Sustainable technologies for water purification from heavy metals: review and analysis. Chem. Soc. Rev. 48, 463–487.

Breiman, L., 1996. Bagging predictors. Mach. Learn. 24, 123–140.

Bui, H.-B., Nguyen, H., Choi, Y., Bui, X.-N., Nguyen-Thoi, T., Zandi, Y., 2019a. A novel artificial intelligence technique to estimate the gross calorific value of coal based on meta-heuristic and support vector regression algorithms. Appl. Sci. 9, 4868.

Bui, X.-N., Lee, C.W., Nguyen, H., Bui, H.-B., Long, N.Q., Le, Q.-T., Nguyen, V.-D., Nguyen, N.-B., Moayedi, H., 2019b. Estimating PM10 concentration from drilling operations in open-pit mines using an assembly of SVR and PSO. Appl. Sci. 9, 2806.

Bui, X.-N., Muazu, M.A., Nguyen, H., 2020. Optimizing Levenberg–Marquardt backpropagation technique in predicting factor of safety of slopes after two-dimensional OptumG2 analysis. Eng. Comput. 36, 941–952.

Bui, X.-N., Nguyen, H., Le, H.-A., Bui, H.-B., Do, N.-H., 2019c. Prediction of blast-induced air over-pressure in open-pit mine: assessment of different artificial intelligence techniques. Nat. Resour. Res. 29, 571–591.

Cheng, D., Ngo, H.H., Guo, W., Chang, S.W., Nguyen, D.D., Liu, Y., Wei, Q., Wei, D., 2019. A critical review on antibiotics and hormones in swine wastewater: water pollution problems and control approaches. J. Hazard Mater. 121682.

Cui, X., Fang, S., Yao, Y., Li, T., Ni, Q., Yang, X., He, Z., 2016a. Potential mechanisms of cadmium removal from aqueous solution by Canna indica derived biochar. Sci. Total Environ. 562, 517–525.

Cui, X., Hao, H., Zhang, C., He, Z., Yang, X., 2016b. Capacity and mechanisms of ammonium and cadmium sorption on different wetland-plant derived biochars. Sci. Total Environ. 539, 566–575.

Deze, E.G., Papageorgiou, S.K., Favvas, E.P., Katsaros, F.K., 2012. Porous alginate aerogel beads for effective and rapid heavy metal sorption from aqueous solutions: effect of porosity in Cu2+ and Cd2+ ion sorption. Chem. Eng. J. 209, 537–546.

Ding, Z., Wan, Y., Hu, X., Wang, S., Zimmerman, A.R., Gao, B., 2016. Sorption of lead and methylene blue onto hickory biochars from different pyrolysis temperatures: importance of physicochemical properties. J. Ind. Eng. Chem. 37, 261–267.

Dolatabadi, M., Mehrabpour, M., Esfandyari, M., Alidadi, H., Davoudi, M., 2018. Modeling of simultaneous adsorption of dye and metal ion by sawdust from aqueous solution using of ANN and ANFIS. Chemometr. Intell. Lab. Syst. 181, 72–78.

Dudoit, S., Fridlyand, J., 2003. Bagging to improve the accuracy of a clustering procedure. Bioinformatics 19, 1090–1099.

Gan, Y.Y., Ong, H.C., Show, P.L., Ling, T.C., Chen, W.-H., Yu, K.L., Abdullah, R., 2018. Torrefaction of microalgal biochar as potential coal fuel and application as bio-adsorbent. Energy Convers. Manag. 165, 152–162.

Gao, L.-Y., Deng, J.-H., Huang, G.-F., Li, K., Cai, K.-Z., Liu, Y., Huang, F., 2019. Relative distribution of Cd2+ adsorption mechanisms on biochars derived from rice straw and sewage sludge. Bioresour. Technol. 272, 114–122.

Gomez-Serrano, V., Macias-Garcia, A., Espinosa-Mansilla, A., Valenzuela-Calahorro, C., 1998. Adsorption of mercury, cadmium and lead from aqueous solution on heat-treated and sulphurized activated carbon. Water Res. 32, 1–4.

Grace, D.R., Durham, J.T., 2001. Methods for Normalization of Experimental Data. Google Patents.

Gunatilake, S., 2015. Methods of removing heavy metals from industrial wastewater. Methods 1, 14.

Gupta, V.K., Gupta, M., Sharma, S., 2001. Process development for the removal of lead and chromium from aqueous solutions using red mud—an aluminium industry waste. Water Res. 35, 1125–1134.

Hass, A., Lima, I.M., 2018. Effect of feed source and pyrolysis conditions on properties and metal sorption by sugarcane biochar. Environ. Technol. Innov. 10, 16–26.

Hernández-Cocoletzi, H., Salinas, R.A., Águila-Almanza, E., Rubio-Rosas, E., Chai, W.S., Chew, K.W., Mariscal-Hernández, C., Show, P.L., 2020. Natural hydroxyapatite from fishbone waste for the rapid adsorption of heavy metals of aqueous effluent. Environ. Technol. Innov. 20, 101109.

Hua, M., Zhang, S., Pan, B., Zhang, W., Lv, L., Zhang, Q., 2012. Heavy metal removal from water/wastewater by nanosized metal oxides: a review. J. Hazard Mater. 211, 317–331.

Idris, A., Hassan, N., Ismail, N.S.M., Misran, E., Yusof, N.M., Ngomsik, A.-F., Bee, A., 2010. Photocatalytic magnetic separable beads for chromium (VI) reduction. Water Res. 44, 1683–1688.

Intelligence, G.W., Yearbook, I.D., Summit, G.W., Card, R., 2011. Global water intelligence. Global Water Intelligence 12, 1–72.

Inyang, M., Gao, B., Yao, Y., Xue, Y., Zimmerman, A.R., Pullammanappallil, P., Cao, X., 2012. Removal of heavy metals from aqueous solution by biochars derived from anaerobically digested biomass. Bioresour. Technol. 110, 50–56.

Inyang, M.I., Gao, B., Yao, Y., Xue, Y., Zimmerman, A., Mosa, A., Pullammanappallil, P., Ok, Y.S., Cao, X., 2016. A review of biochar as a low-cost adsorbent for aqueous heavy metal removal. Crit Rev Environ Sci Technol 46, 406–433.

Jiang, S., Huang, L., Nguyen, T.A., Ok, Y.S., Rudolph, V., Yang, H., Zhang, D., 2016. Copper and zinc adsorption by softwood and hardwood biochars under elevated sulphate-induced salinity and acidic pH conditions. Chemosphere 142, 64–71.

Ke, B., Zhou, K., Xu, C., Deng, H., Li, J., Bin, F., 2018. Dynamic mechanical property deterioration model of sandstone caused by freeze–thaw weathering. Rock Mech. Rock Eng. 51, 2791–2804. https://doi.org/10.1007/s00603-018-1495-0.

Le, D.N., Dang, C.V., 2020. Application of fuzzy-logic to design fuzzy compensation controller for speed control system to reduce vibration of CBIII-250T drilling machine in mining industry. J. Min. Earth Sci. 61 (6), 90–96. https://doi.org/10.46326/JMES.2020.61(6).10.

Leng, L., Xiong, Q., Yang, L., Li, H., Zhou, Y., Zhang, W., Jiang, S., Li, H., Huang, H., 2020. An overview on engineering the surface area and porosity of biochar. Sci. Total Environ. 144204.

Li, H., Dong, X., da Silva, E.B., de Oliveira, L.M., Chen, Y., Ma, L.Q., 2017. Mechanisms of metal sorption by biochars: biochar characteristics and modifications. Chemosphere 178, 466–478.

Li, S., Chen, G., 2018. Thermogravimetric, thermochemical, and infrared spectral characterization of feedstocks and biochar derived at different pyrolysis temperatures. Waste Manag. 78, 198–207.

Li, S., Harris, S., Anandhi, A., Chen, G., 2019. Predicting biochar properties and functions based on feedstock and pyrolysis temperature: a review and data syntheses. J. Clean. Prod. 215, 890–902.

Li, Y., Liu, X., Zhang, P., Wang, X., Cao, Y., Han, L., 2018. Qualitative and quantitative correlation of physicochemical characteristics and lead sorption behaviors of crop residue-derived chars. Bioresour. Technol. 270, 545–553.

Lu, H., Zhang, W., Yang, Y., Huang, X., Wang, S., Qiu, R., 2012. Relative distribution of Pb2+ sorption mechanisms by sludge-derived biochar. Water Res. 46, 854–862.

Mehta, S., Gaur, J., 2005. Use of algae for removing heavy metal ions from wastewater: progress and prospects. Crit. Rev. Biotechnol. 25, 113–152.

Mohan, D., Sarswat, A., Ok, Y.S., Pittman Jr., C.U., 2014. Organic and inorganic contaminants removal from water with biochar, a renewable, low cost and sustainable adsorbent–a critical review. Bioresour. Technol. 160, 191–202.

Muchuweti, M., Birkett, J., Chinyanga, E., Zvauya, R., Scrimshaw, M.D., Lester, J., 2006. Heavy metal content of vegetables irrigated with mixtures of wastewater and sewage sludge in Zimbabwe: implications for human health. Agric. Ecosyst. Environ. 112, 41–48.

Nguyen, H., 2019. Support vector regression approach with different kernel functions for predicting blast-induced ground vibration: a case study in an open-pit coal mine of Vietnam. SN Appl. Sci. 1, 283.

Nguyen, H., 2020. Application of the k - nearest neighbors algorithm for predicting blast - induced ground vibration in open - pit coal mines: a case study. J. Min. Earth Sci. 61 (6), 22–29. https://doi.org/10.46326/JMES.2020.61(6).03. In this issue.

Nguyen, H., Bui, X.-N., 2019. Predicting blast-induced air overpressure: a robust artificial intelligence system based on artificial neural networks and random forest. Nat. Resour. Res. 28, 893–907.
Nguyen, H., Bui, X.-N., Tran, Q.-H., Moayedi, H., 2019a. Predicting blast-induced peak particle velocity using BGAMs, ANN and SVM: a case study at the Nui Beo open-pit coal mine in Vietnam. Environ. Earth Sci. 78, 479.
Nguyen, D.T, Nguyen, N.T, Pham, H.N.T, Phung, H.H., Van Nguyen, H., 2020. Rice husk ash and its utilization in soil improvement: an overview. J. Min. Earth Sci 61 (3), 1–11. https://doi.org/10.46326/JMES.2020.61(3).01.
Nguyen, H., Bui, X.-N., Tran, Q.-H., Van Hoa, P., Nguyen, D.-A., Hoa, L.T.T., Le, Q.-T., Do, N.-H., Bao, T.D., Bui, H.-B., Moayedi, H., 2020. A comparative study of empirical and ensemble machine learning algorithms in predicting air overpressure in open-pit coal mine. Acta Geophys. 68, 325–336.
Nguyen, H., Choi, Y., Bui, X.-N., Nguyen-Thoi, T., 2019b. Predicting blast-induced ground vibration in open-pit mines using vibration sensors and support vector regression-based optimization algorithms. Sensors 20, 132.
Nguyen, A.D., Tran, H.Q., Tran, B.D., Soukhanouvong, P., 2020. Prediction of the peak velocity of blasting vibration based on various models at Ninh Dan quarry, Thanh Ba district, Phu Tho province. J. Min. Earth Sci. 61 (4), 102–109. https://doi.org/10.46326/JMES.2020.61(4).11.
Nguyen, N.V., Trinh, H.L., 2020. Determination of water quality parameters in the Tan Rai exploiting area (Lam Dong province) using Sentinel-2 MSI and Landsat 8 data. J. Min. Earth Sci. 61 (2), 126–134. https://doi.org/10.46326/JMES.2020.61(2).14.
Pal, M., Deswal, S., 2009. M5 model tree based modelling of reference evapotranspiration. Hydrol. Process. 23, 1437–1443.
Parveen, N., Zaidi, S., Danish, M., 2017. Development of SVR-based model and comparative analysis with MLR and ANN models for predicting the sorption capacity of Cr (VI). Process Saf. Environ. Protect. 107, 428–437.
Plaza, M., González, A., Pis, J., Rubiera, F., Pevida, C., 2014. Production of microporous biochars by single-step oxidation: effect of activation conditions on CO2 capture. Appl. Energy 114, 551–562.
Prakash, N., Manikandan, S., Govindarajan, L., Vijayagopal, V., 2008. Prediction of biosorption efficiency for the removal of copper (II) using artificial neural networks. J. Hazard Mater. 152, 1268–1275.
Programme, W.W.A., UN-Water, 2009. Water in a Changing World. Earthscan.
Qu, M., Li, W., Zhang, C., 2014. Spatial distribution and uncertainty assessment of potential ecological risks of heavy metals in soil using sequential Gaussian simulation. Hum. Ecol. Risk Assess. 20, 764–778.
Quinlan, J.R., 1996. Bagging, Boosting, and C4. 5, vol. 1. AAAI/IAAI, pp. 725–730.
Rodriguez-Galiano, V., Mendes, M.P., Garcia-Soldado, M.J., Chica-Olmo, M., Ribeiro, L., 2014. Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: a case study in an agricultural setting (Southern Spain). Sci. Total Environ. 476, 189–206.
Sajidu, S., Persson, I., Masamba, W., Henry, E., 2008. Mechanisms of heavy metal sorption on alkaline clays from Tundulu in Malawi as determined by EXAFS. J. Hazard Mater. 158, 401–409.
Sankaran, R., Show, P.L., Ooi, C.-W., Ling, T.C., Shu-Jen, C., Chen, S.-Y., Chang, Y.-K., 2020. Feasibility assessment of removal of heavy metals and soluble microbial products from aqueous solutions using eggshell wastes. Clean Technol. Environ. Policy 22, 773–786.
Sarkar, M., Majumdar, P., 2011. Application of response surface methodology for optimization of heavy metal biosorption using surfactant modified chitosan bead. Chem. Eng. J. 175, 376–387.
Schnitzer, J., Rice, D., Cruickshank, R., Zajkowski, A., 2003. Data Normalization. Google Patents.
Šćiban, M., Radetić, B., Kevrešan, Ž., Klašnja, M., 2007. Adsorption of heavy metals from electroplating wastewater by wood sawdust. Bioresour. Technol. 98, 402–409.
Shariati, M., Mafipour, M.S., Haido, J.H., Yousif, S.T., Toghroli, A., Trung, N.T., Shariati, A., 2020a. Identification of the most influencing parameters on the properties of corroded concrete beams using an Adaptive Neuro-Fuzzy Inference System (ANFIS). Steel Compos. Struct. 34, 155–170.
Shariati, M., Mafipour, M.S., Mehrabi, P., Ahmadi, M., Wakil, K., Trung, N.T., Toghroli, A., 2020b. Prediction of concrete strength in presence of furnace slag and fly ash using Hybrid ANN-GA (Artificial Neural Network-Genetic Algorithm). Smart Struct. Syst. 25, 183–195.
Shariati, M., Mafipour, M.S., Mehrabi, P., Zandi, Y., Dehghani, D., Bahadori, A., Shariati, A., Trung, N.T., Salih, M.N., Poi-Ngian, S., 2019. Application of Extreme Learning Machine (ELM) and Genetic Programming (GP) to design steel-concrete composite floor systems at elevated temperatures. Steel Compos. Struct. 33, 319–332.
Shen, Z., Jin, F., Wang, F., McMillan, O., Al-Tabbaa, A., 2015. Sorption of lead by Salisbury biochar produced from British broadleaf hardwood. Bioresour. Technol. 193, 553–556.
Shen, Z., Zhang, Y., Jin, F., McMillan, O., Al-Tabbaa, A., 2017a. Qualitative and quantitative characterisation of adsorption mechanisms of lead on four biochars. Sci. Total Environ. 609, 1401–1410.
Shen, Z., Zhang, Y., McMillan, O., Jin, F., Al-Tabbaa, A., 2017b. Characteristics and mechanisms of nickel adsorption on biochars produced from wheat straw pellets and rice husk. Environ. Sci. Pol. 24, 12809–12819.
Suliman, W., Harsh, J.B., Abu-Lail, N.I., Fortuna, A.-M., Dallmeyer, I., Garcia-Perez, M., 2016. Influence of feedstock source and pyrolysis temperature on biochar bulk and surface properties. Biomass Bioenergy 84, 37–48.
Sun, Y., Gao, B., Yao, Y., Fang, J., Zhang, M., Zhou, Y., Chen, H., Yang, L., 2014. Effects of feedstock type, production method, and pyrolysis temperature on biochar and hydrochar properties. Chem. Eng. J. 240, 574–578.
Sutherland, C., Marcano, A., Chittoo, B., 2018. Artificial Neural Network-Genetic Algorithm Prediction of Heavy Metal Removal Using a Novel Plant-Based Biosorbent Banana Floret: Kinetic, Equilibrium, Thermodynamics and Desorption studiesDesalin. Water Trea, pp. 385–411.
Trakal, L., Bingöl, D., Pohořelý, M., Hruška, M., Komárek, M., 2014. Geochemical and spectroscopic investigations of Cd and Pb sorption mechanisms on contrasting biochars: engineering implications. Bioresour. Technol. 171, 442–451.
Trakal, L., Veselská, V., Šafařík, I., Vítková, M., Číhalová, S., Komárek, M., 2016. Lead and cadmium sorption mechanisms on magnetically modified biochars. Bioresour. Technol. 203, 318–324.
Tran, T.T.T., Pham, H.K., Nguyen, H.M., 2020. Assessing the current status of rural domestic solid waste management in Nam Dinh province. J. Min. Earth Sci. 61 (6), 82–89. https://doi.org/10.46326/JMES.2020.61(6).09.
Tran, H.N., Tomul, F., Nguyen, H.T.H., Nguyen, D.T., Lima, E.C., Le, G.T., Chang, C.-T., Masindi, V., Woo, S.H., 2020. Innovative spherical biochar for pharmaceutical removal from water: insight into adsorption mechanism. J. Hazard Mater. 122255.
Trung, N.T., Shahgoli, A.F., Zandi, Y., Shariati, M., Wakil, K., Safa, M., Khorami, M., 2019. Moment-rotation prediction of precast beam-to-column connections using extreme learning machine. Struct. Eng. Mech. 70, 639–647.
Turp, S.M., Eren, B., Ates, A., 2011. Prediction of adsorption efficiency for the removal of nickel(ii) ions by zeolite using artificial neural network(ann) approach. Fresenius Environ. Bull. 20, 3158–3165.
Tytłak, A., Oleszczuk, P., Dobrowolski, R., 2015. Sorption and desorption of Cr (VI) ions from water by biochars in different environmental conditions. Environ. Sci. Pol. 22, 5985–5994.
Uchimiya, M., Bannon, D.I., Wartelle, L.H., 2012. Retention of heavy metals by carboxyl functional groups of biochars in small arms range soil. J. Agric. Food Chem. 60, 1798–1809.
Uddin, M.K., 2017. A review on the adsorption of heavy metals by clay minerals, with special focus on the past decade. Chem. Eng. J. 308, 438–462.
Vu, Q.T., Pham, M.N., Dang, C.M., Vuong, H.H., Duong, H.D., Bui, T.T., Dang, T.T., Vu, M.T., Vu, D.M., 2020. Fuzzy logic in controlling the forest fire - level forecast warning signage. J. Min. Earth Sci. 61 (4), 126–136. https://doi.org/10.46326/JMES.2020.61(4).14.
Wang, Q., Yang, Z., 2016. Industrial water pollution, water environment treatment, and health risks in China. Environ. Pollut. 218, 358–365.
Yu, K.L., Show, P.L., Ong, H.C., Ling, T.C., Lan, J.C.-W., Chen, W.-H., Chang, J.-S., 2017. Microalgae from wastewater treatment to biochar–feedstock preparation and conversion technologies. Energy Convers. Manag. 150, 1–13.
Zama, E.F., Zhu, Y.-G., Reid, B.J., Sun, G.-X., 2017. The role of biochar properties in influencing the sorption and desorption of Pb (II), Cd (II) and as (III) in aqueous solution. J. Clean. Prod. 148, 127–136.
Zhang, P., O'Connor, D., Wang, Y., Jiang, L., Xia, T., Wang, L., Tsang, D.C., Ok, Y.S., Hou, D., 2020. A green biochar/iron oxide composite for methylene blue removal. J. Hazard Mater. 384, 121286.
Zhao, H., Xia, B., Fan, C., Zhao, P., Shen, S., 2012. Human health risk from soil heavy metal contamination under different land uses near Dabaoshan Mine, Southern China. Sci. Total Environ. 417, 45–54.
Zhao, Q., Wang, Y., Cao, Y., Chen, A., Ren, M., Ge, Y., Yu, Z., Wan, S., Hu, A., Bo, Q., 2014. Potential health risks of heavy metals in cultivated topsoil and grain, including correlations with human primary liver, lung and gastric cancer, in Anhui province, Eastern China. Sci. Total Environ. 470, 340–347.
Zhou, J., Qiu, Y., Armaghani, D.J., Zhang, W., Li, C., Zhu, S., Tarinejad, R., 2021. Predicting TBM penetration rate in hard rock condition: A comparative study among six XGB-based metaheuristic techniques. Geosci. Front. 12 (03) https://doi.org/10.1016/j.gsf.2020.09.020.
Zhu, X., Wang, X., Ok, Y.S., 2019. The application of machine learning methods for prediction of metal sorption onto biochars. J. Hazard Mater. 378, 120727.
Zhou, J., Qiu, Y., Yao, Y., Zhu, S., Armaghani, D.J., Li, C., Nguyen, H., Yagiz, S., 2021. Optimization of support vector machine through the use of metaheuristic algorithms in forecasting TBM advance rate. Eng. Appl. Artif. Intell. 97 https://doi.org/10.1016/j.engappai.2020.104015.