Journal of Environmental Management 293 (2021) 112808

Contents lists available at ScienceDirect

Journal of Environmental Management

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

Prediction of the sorption efficiency of heavy metal onto biochar using a robust combination of fuzzy C-means clustering and back-propagation neural network

Bo Ke [a,b], Hoang Nguyen [c,*], Xuan-Nam Bui [c,d], Hoang-Bac Bui [e,f], Trung Nguyen-Thoi [g,h]

[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 Vien St., Duc Thang Ward, Bac Tu Liem Dist., Hanoi, Viet Nam
[d] Center for Mining, Electro-Mechanical Research, Hanoi University of Mining and Geology, 18 Vien St., Duc Thang Ward, Bac Tu Liem Dist., Hanoi, 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] Division of Computational Mathematics and Engineering, Institute for Computational Science, Ton Duc Thang University, Ho Chi Minh City, Viet Nam
[h] Faculty of Civil Engineering, Ton Duc Thang University, Ho Chi Minh City, Viet Nam

ARTICLE INFO

Keywords:
Heavy metal
Biochar sorption
Wastewater
Hybrid machine learning model
Fuzzy C-Mean clustering
Neural networks

ABSTRACT

Heavy metal adsorption onto biochar is an effective method for the treatment of the heavy metal contamination of water and wastewater. This study aims to evaluate the heavy metals sorption efficiency of different biochar characteristics and propose a novel intelligence method for predicting the sorption efficiency of heavy metal onto biochar with high accuracy based on the back-propagation neural network (BPNN) and fuzzy C-means clustering algorithm (FCM), named as FCM-BPNN. Accordingly, the FCM algorithm was used to simulate the properties of metal adsorption data and divide them into clusters with similar features. The clustering results showed that the FCM algorithm simulated metal adsorption data's properties very well and classified them based on biochar characteristics and adsorption conditions. Afterward, BPNN models were well-developed based on these clusters, and their outcomes were then combined (i.e., FCM-BPNN). The results indicated that the FCM-BPNN model could predict heavy metal's sorption efficiency onto biochar with a promising result (i.e., RMSE of 0.036, R^2 of 0.987, RSE of 0.006, MAPE of 0.706, and VAF of 98.724). Whereas the BPNN model, without optimizing the FCM algorithm, was proved with lower performance (RMSE = 0.050, R^2 = 0.977, RSE = 0.011, MAPE = 0.802, and VAF = 97.662). These findings revealed that the FCM algorithm's presence impressively improved the BPNN model's accomplishment in predicting heavy metal's sorption efficiency onto biochar, and the proposed FCM-BPNN model can improve water/wastewater treatment plants' quality and provide a more efficient process for heavy metals with performance superiority.

1. Introduction

Environmental concerns related to heavy metal contamination in the water are of great interest to the community. Many countries have carried out efforts to protect water resources to eliminate heavy metal pollution and improve water quality. The metal ions in water are much higher than the maximum allowed (Hu et al., 2008). They are easily soluble in water and can be easily absorbed by plants and marine life as well. Therefore, they can easily get into the food chain and be dangerously noxious to the human body (Peralta-Videa et al., 2009) and become one of the harmful effects on human health and environmental ecosystem degradation (Haileslassie and Gebremedhin, 2015; Ojuederie and Babalola, 2017).

In recent years, the rapid development of industries, such as non-ferrous metallurgy, paper industry, mining activities, electroplating, mineral paints production, to name a few, has contributed a large number of heavy metals in wastewater (e.g., Cd (II), Cu (II), Fe (II), Zn (II), Ni, Pb, Hg) (Zhai et al., 2016; Ayangbenro and Babalola, 2017; Yang

* Corresponding author.
E-mail addresses: nguyenhoang@humg.edu.vn (H. Nguyen), buixuannam@humg.edu.vn (X.-N. Bui), buihoangbac@humg.edu.vn (H.-B. Bui), nguyenthoitrung@tdtu.edu.vn (T. Nguyen-Thoi).

https://doi.org/10.1016/j.jenvman.2021.112808
Received 26 July 2020; Received in revised form 14 May 2021; Accepted 15 May 2021
Available online 23 May 2021

et al., 2019; Thuy et al., 2020; Hung et al., 2020). Unlike organic pollutants, heavy metals are not biodegradable and are easily absorbed by living organisms (Atieh et al., 2017). Many heavy metals are known as good candidates for toxicity and carcinogenicity (Saleh et al., 2019).

In recent years, various technical solutions have been applied to remove heavy metals from water and wastewater, such as adsorption, flotation, membrane filtration, ion-exchange, chemical precipitation, flocculation, electrochemical, and coagulation (Muya et al., 2016). Sorption is considered a modern, simple, and effective technique in removing heavy metals from wastewater. Sorption techniques use various absorbing materials to remove dangerous pollutants from wastewater such as sawdust, rice husks (Duong et al., 2020), titanium phosphates, Lewatit FO 36 and Purolite Arsen Xnp (Kołodyńska et al., 2015), cysteine-grafted nonwoven geotextile (Vandenbossche et al., 2014), magnetized activated carbon (Rahmani-Sani et al., 2020), spent grain (Wierzba and Kłos, 2019), to name a few. However, the effectiveness of heavy metal sorption of materials is not the same. It depends not only on absorbing materials but also on environmental conditions and metal sources (Mejias Carpio et al., 2018; Wu et al., 2019). Therefore, predicting the effectiveness of heavy metal sorption is significant for wastewater treatment plants and the environment.

For estimating the sorption efficiency of heavy metals in water/wastewater, many techniques have been proposed based on soft computing methods. They are not only applied in wastewater treatment but also mining, civil, geological and environmental engineering (Shahabi et al., 2013; P. Asteris et al., 2016; Zhou et al., 2016; Nguyen et al., 2018; P. G. Asteris and Nikoo, 2019; Bui et al., 2019; Nguyen et al., 2019a; Wang et al., 2019; Zhou et al., 2019; Bo et al., 2018; Nhan et al., 2020; Hoang, 2020; Quyen et al., 2020) For instance, Al-Asheh et al. (2000) applied the theory of ideal adsorption solution based on the extended Langmuir, Freundlich, and Sips models to predict the sorption of Cu^{2+}, Cd^{2+}, and Ni^{2+}. Based on the advanced techniques, Turp et al. (2011) predicted the adsorption efficiency of Ni^{2+} by zeolite using an artificial neural network (ANN). Accordingly, a simple ANN was developed to estimate the removal of Ni^{2+} with highly precise (i.e., RMSE of 0.0222). Using a similar ANN model, Turan et al. (2011) predicted the Cu^{2+} adsorption from industrial influent with a great accuracy. It is argued that the ANN model was fit to predict the adsorption of Cu^{2+} from industrial leachate with R^2 of 0.999. In another firm, Parveen et al. (2017) developed two artificial intelligence (AI) systems for estimating the capacity of Cr^{4+} sorption, including support vector machine for regression (SVR) and ANN. The multiple linear regression (MLR) technique was also used to evaluate and compare with the SVR and ANN models in terms of the heavy metal sorption efficiency. Finally, a promising result was proven with the SVR model's performance reached an R^2 of 0.9986 and RMSE of 0.0159. Notably, the SVR was also introduced as the superior model for estimating the Cu^{2+}, As^{3+}, and Pb^{2+} sorption in wastewater (Parveen et al., 2016b, a). Fawzy et al. (2016) also successfully performed a fuzzy inference system of adaptive neural network (ANFIS) for estimating the efficiency of the removal of Ni^{2+} and Cd^{2+} in wastewater based on the role of pH. An encouraging result was given for the ANFIS model with $R^2 > 0.98$. Dolatabadi et al. (2018) also adopted a similar study to predict the sorption efficiency of Cu^{2+} using ANFIS and ANN. They concluded that these two models were the promising predicting techniques for simulation and prediction of heavy metals sorption. Zhu et al. (2019) also applied ANN and random forest (RF) algorithm to predict the sorption efficiency of heavy metals onto biochars with an acceptable result (i.e., RMSE of 0.106 and 0.059, R^2 of 0.948 and 0.973, for ANN and RF, respectively).

Despite the previous studies proposed and evaluated the feasibility of AI models for predicting the heavy metal sorption with promising results; however, they were only proposed based on the simple AI techniques. Meanwhile, the complexity AI techniques have been recommended that can explain and get better accuracy of the simple models (Bo et al., 2021). Deep analyzing and learning help explain data and understand their relationship better (Zhang et al., 2020; Zhou et al., 2020a). For this reason, we developed a novel AI model with more in-depth data analyzed to provide insights in predicting the sorption efficiency of heavy metals onto biochar. The predictive models with high accuracy are useful for improving water/wastewater treatment plants' performance and removing heavy metals in the water or wastewater. Indeed, the fuzzy C-means clustering (FCM) method was made an effort to analyze and allocate the dataset in this study; then, the back-propagation neural network (BPNN) was developed based on classified clusters aiming to get better accuracies, called FCM-BPNN model. As an alternative, this study's primary objective is to propose the novel FCM-BPNN model for predicting the heavy metal sorption efficiency with high accuracy. It should be emphasized that the results of this study will be compared to previous studies based on the same dataset to demonstrate the superiority and improved accuracy of the proposed model.

Review of literature shows that the FCM-BPNN has not been applied in any fields of environment and engineering, especially for predicting the sorption efficiency of heavy metal onto biochar. Several previous studies proposed the FCM-ANN model, such as research on anomaly detection system (Pandeeswari and Kumar 2016), prediction of blast-induced ground vibration (Nguyen et al., 2020a), abnormal detection of chemical storage tank (Xiaojian et al., 2017), prediction of water demand (Papageorgiou et al., 2016). Nevertheless, these studies used the feedforward neural network instead of backpropagation neural network as per used in this study, and they did not apply the FCM-ANN model in terms of water/wastewater treatment for the removal of heavy metals. Therefore, the FCM-BPNN model was introduced as a novel model in terms of prediction of the sorption efficiency of heavy metal onto biochar in this work.

Furthermore, it is worth noting that the proposed FCM-BPNN model is developed based on the input variables, including pyrolysis temperature (T_P), pH of biochar in wastewater (pH_w), the ratio of hydrogen and carbon (H/C), percentage of carbon in biochar (C), ratio of oxygen and nitrate with carbon [(O + N)/C], the ratio of oxygen and carbon (O/C), percentage of ash (A), particle size of biochar (PSB), biochar surface area (BSA), cation exchange capacity (CEC), environmental temperature (T_{envi}), solution pH (pH_s), heavy metal concentration in wastewater (C_O). Therefore, by the tuning the input variables, the sorption efficiency of heavy metals will be changed. In other words, the outputs (i.e., the quality of water and wastewater) can be tuned to achieve specific end-use, including being safely returned to the environment.

2. Proposing the FCM-BPNN model for estimating the efficiency of heavy metal sorption

As mentioned above, this study's primary goal is to propose a novel soft computing model based on AI techniques for predicting the heavy metal sorption efficiency onto biochar with the accuracy improved. Indeed, the FCM-BPNN model's framework is proposed in this section based on combining the FCM and BPNN algorithms. Because of the FCM and BPNN algorithms' principle, they have been introduced in literature; therefore, they are not presented in this section, and they are presented in the supplementary materials. Accordingly, this state-of-the-art model includes five steps to generate the FCM-BPNN model, as follow:

- Step 1: Divide the dataset into two groups, 70% for developing the model and 30% for testing the model's accuracy.
- Step 2: Applying the FCM algorithm to classify the training dataset to distinct groups.
- Step 3: Developing the respectively BPNN models based on the classified groups.
- Step 4: Combination of the sorption efficiency predictions and evaluate the performance of the training phase.
- Step 5: Repeat steps 2 to 4 for the testing dataset.

The proposed framework of the FCM-BPNN model is shown in Fig. 1.

3. Sorption experiments of heavy metals and data pre-processing

To realize the experimentations on heavy metals' sorption study, feedstock materials, such as wood, spray-dried algae, plum stones, rice straw, sewage sludge, grape husks, wheat straw, nutshells, green waste, grape stalks, to name a few, were selected and pelletized to ensure the conditions pyrolysis reactor (300–750 °C). It is worth mentioning that 24 kinds of biomass were used in this study with diverse feedstock materials. The standards for preparing the above input materials, as well as the sorption experiments for each biochar can be referred to in the literature (Ronsse et al., 2013; 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). Subsequently, the slow pyrolysis reactions were set-up and carried out in a tubular reactor using stainless steel. The structure of the slow pyrolysis reactions is illustrated in Fig. S4 (supplementary materials).

Based on the previously published papers, as cited above, Zhu et al. (2019) collected a diverse dataset with different characteristics of biochar systems, feedstock materials, and environmental conditions. The details of the dataset used are presented in the supplementary materials. Herein, this dataset was used to develop a novel intelligent model for predicting heavy metals' sorption efficiency onto biochar. It was then compared with that of the developed models by Zhu et al. (2019) to demonstrate the outstanding achievement of the proposed FCM-BPNN model in this study. The database contains 353 experimental datasets with 14 variables, including pyrolysis temperature (T_P), pH of biochar in wastewater (pH_w), the ratio of hydrogen and carbon (H/C), percentage of carbon in biochar (C), ratio of oxygen and nitrate with carbon [(O + N)/C], the ratio of oxygen and carbon (O/C), percentage of ash (A), particle size of biochar (PSB), biochar surface area (BSA), cation exchange capacity (CEC), environmental temperature (T_{envi}), solution pH (pH_s), heavy metal concentration in wastewater (C_O), and sorption efficiency of heavy metal onto biochar (SB). Of those, SB is considered as the output variable; whereas, the remaining variables are the input variables. The data and distribution of the attributes are shown in Fig. 2.

Before developing the models, the pre-processing steps are needed to understand the dataset in-depth and ensure the models' accuracy. One of the interests is the correlation between the input variables. If they have a strong correlation, they should be removed to avoid overlap of information from the inputs (Zhou et al., 2020b). Therefore, a correlation matrix was analyzed in Table 1 to check the relationship between the input variables and the output variable.

The correlation matrix in Table 1 shows that most of them have a good correlation as the independent variables. However, it is worth mentioning that the O/C and (O + N)/C variables have an absolute correlation (e.g., equal to 1). This result shows that these two variables reflect the same information in the database, and one of these two variables must be removed. Herein, we decided to remove the O/C variable for the two following reasons: (i) the (O + N)/C variable reflects more information than the O/C variable because it includes oxygen, nitrate, and carbon. Whereas, the O/C variable only demonstrates the relationship between oxygen and carbon. (ii) The correlation between the (O + N)/C variable and the output variable is higher than the O/C variable. It is an excellent candidate to explain the features of the dataset used, as well as increasing the models' accuracy. The dataset features and their distribution are visualized in Fig. 3.

After processing, the whole dataset was considered and selected

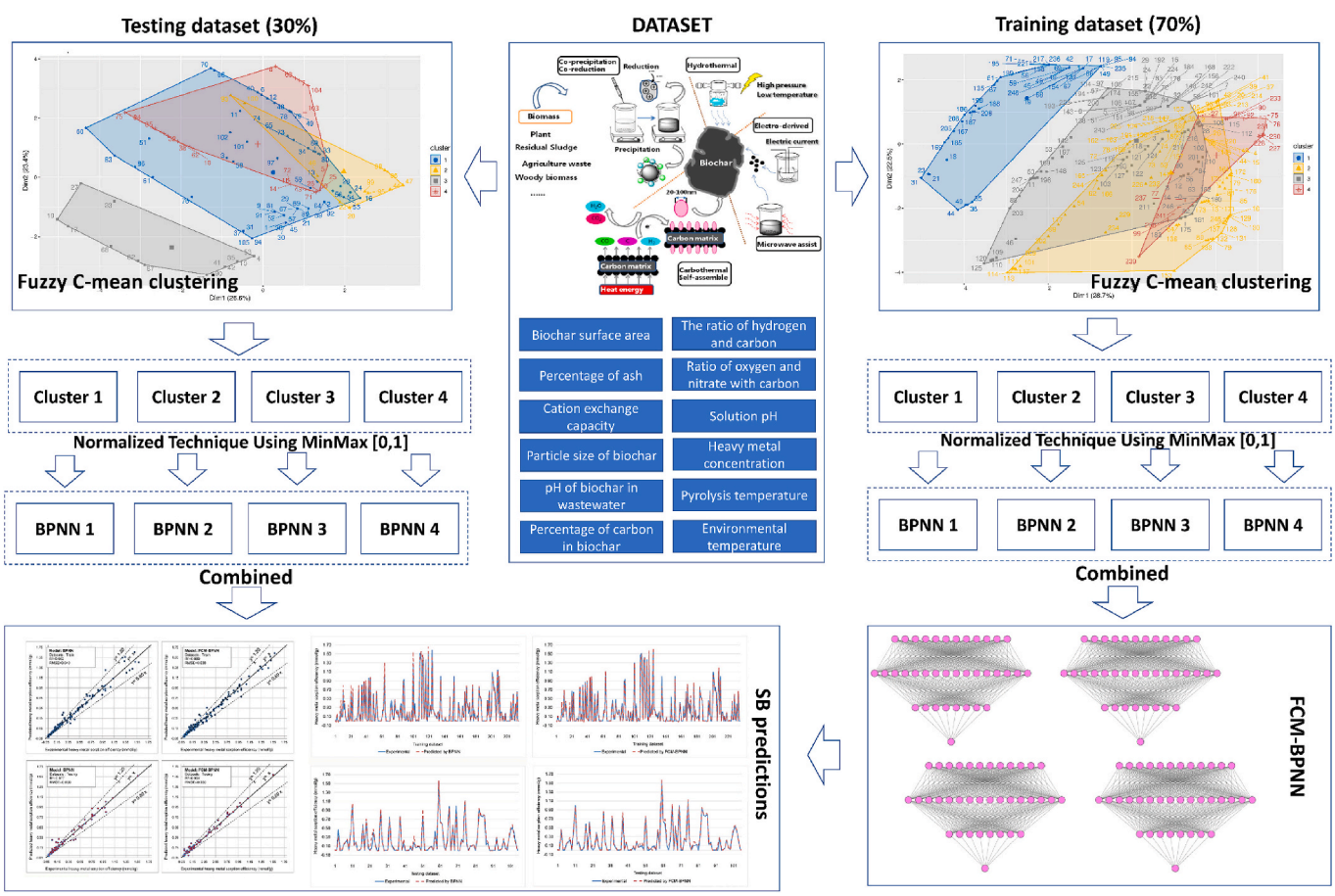

Fig. 1. Proposing the framework of the novel FCM-BPNN model for predicting the sorption efficiency of heavy metal onto biochar.

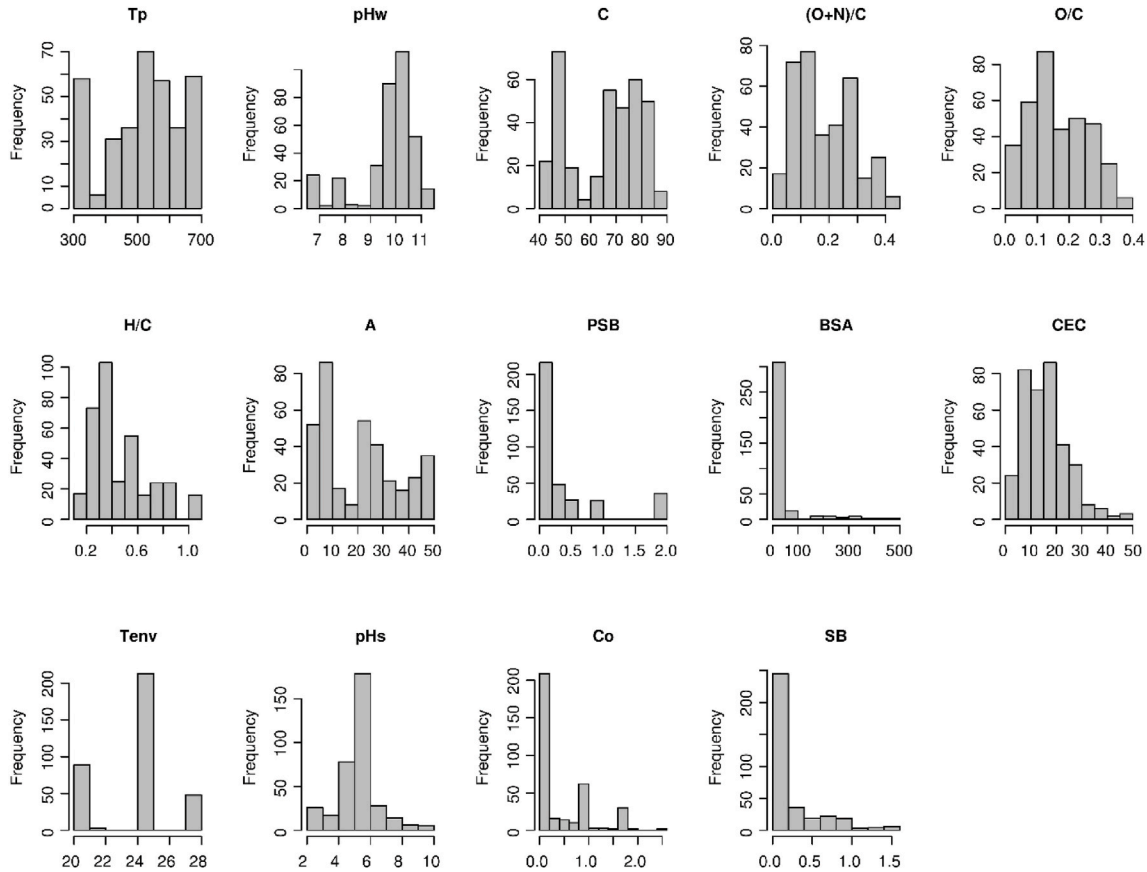

Fig. 2. The dataset before analyzing and processing.

Table 1
Correlation matrix of the dataset before processing..

	Tp	pHw	C	(O+N)/C	O/C	H/C	A	PSB	BSA	CEC	Tenv	pHs	Co	SB
Tp	1.00	0.06	0.17	-0.73	-0.72	-0.87	0.21	-0.04	0.30	-0.29	-0.36	-0.08	-0.04	0.00
pHw	0.06	1.00	-0.16	0.09	0.06	-0.32	0.23	-0.25	-0.13	0.40	0.28	-0.01	0.05	0.08
C	0.17	-0.16	1.00	-0.23	-0.19	-0.22	-0.81	-0.11	0.21	-0.05	-0.16	-0.02	-0.65	-0.59
(O+N)/C	-0.73	0.09	-0.23	1.00	1.00	0.83	-0.18	0.08	-0.32	0.54	0.68	0.04	0.05	0.12
O/C	-0.72	0.06	-0.19	1.00	1.00	0.83	-0.21	0.06	-0.30	0.53	0.69	0.04	0.01	0.10
H/C	-0.87	-0.32	-0.22	0.83	0.83	1.00	-0.22	0.07	-0.28	0.21	0.43	0.03	0.04	0.04
A	0.21	0.23	-0.81	-0.18	-0.21	-0.22	1.00	-0.03	-0.14	-0.15	-0.02	0.01	0.54	0.47
PSB	-0.04	-0.25	-0.11	0.08	0.06	0.07	-0.03	1.00	0.08	0.23	-0.01	-0.03	0.47	0.45
BSA	0.30	-0.13	0.21	-0.32	-0.30	-0.28	-0.14	0.08	1.00	-0.07	0.08	-0.13	-0.09	-0.13
CEC	-0.29	0.40	-0.05	0.54	0.53	0.21	-0.15	0.23	-0.07	1.00	0.49	-0.01	0.17	0.31
Tenv	-0.36	0.28	-0.16	0.68	0.69	0.43	-0.02	-0.01	0.08	0.49	1.00	-0.01	0.06	0.16
pHs	-0.08	-0.01	-0.02	0.04	0.04	0.03	0.01	-0.03	-0.13	-0.01	-0.01	1.00	0.05	0.27
Co	-0.04	0.05	-0.65	0.05	0.01	0.04	0.54	0.47	-0.09	0.17	0.06	0.05	1.00	0.82
SB	0.00	0.08	-0.59	0.12	0.10	0.04	0.47	0.45	-0.13	0.31	0.16	0.27	0.82	1.00

randomly as per the ratio of 70/30 (Nguyen et al., 2020b). Proportionately, 70% was used for training and developing the models; 30% was used for validating the performance of the developed sorption efficiency predictive models. Considering the distribution of the attributes in Fig. 3 shows that it is possible to see that they may be skewed Gaussian distributions and significantly affect the accuracy and the stable of the predictive models. Therefore, the normalized technique using MinMax [0,1] was applied to normalize the features of dataset with the [0,1] range before developing the models.

4. Developing the models

4.1. BPNN model

For designing and building the BPNN model for predicting the sorption efficiency of heavy metal onto biochar, it isn't straightforward to determine how many hidden layers and hidden neurons are fit with the BPNN model. As reported by previous scholars, a simple BPNN model with only one hidden layer can rapidly solve most of the simple problems (Nguyen et al., 2019b). The BPNN models with more complexity structure (more than one hidden layer) are known as deep

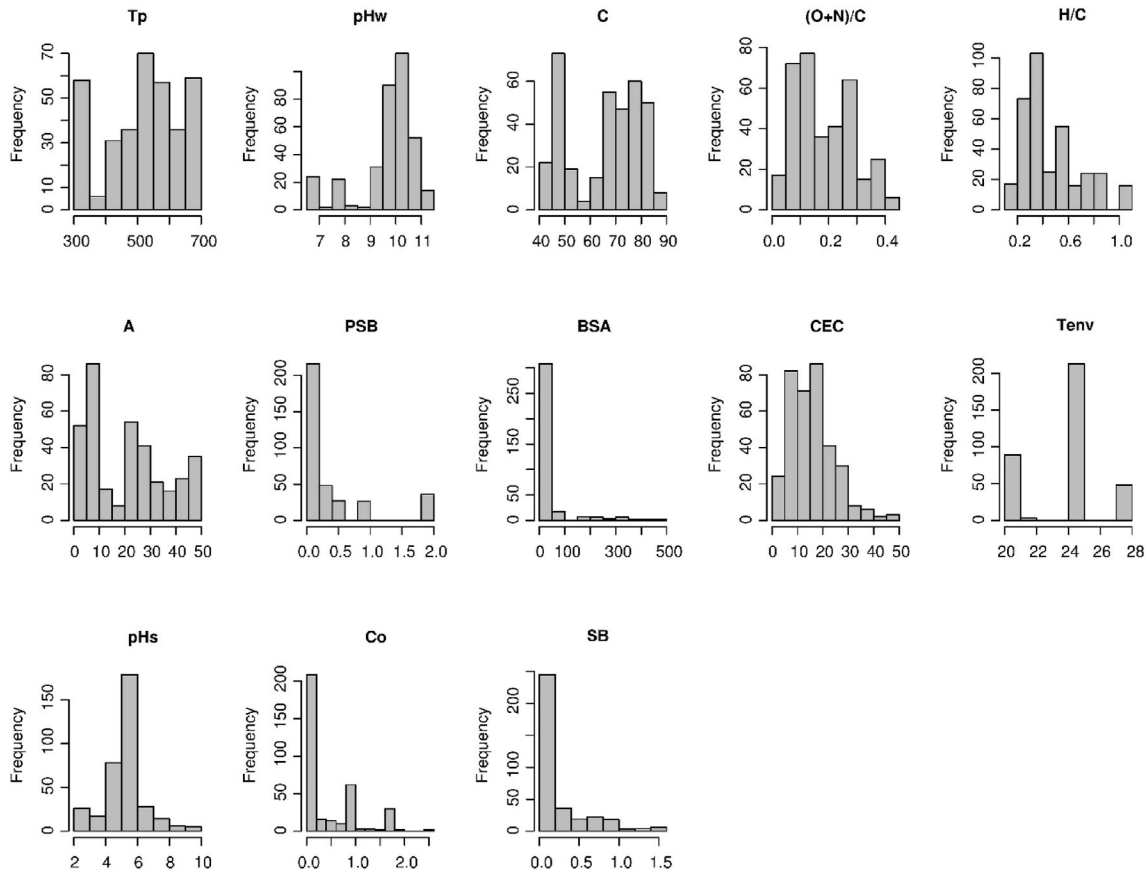

Fig. 3. The dataset after analyzing and processing.

neural networks for more complexity problems (Guo et al., 2019; Shang et al., 2019; Mamandipoor et al., 2020). Based on the analyzed correlations in Table 1, we can see the complicated relationship between the variables in this study. Therefore, a BPNN model with more than one hidden layer is the goal of this study for estimating the heavy metal sorption efficiency in this study. A preliminary procedure was used to design the network's structure (i.e., hidden layers and neurons). Eventually, an optimal structure of BPNN was determined, as shown in Fig. 4. Then, the weights (w) and biases (b) of the BPNN model were computed through the lines between the neurons. The model was then trained by the back-propagation algorithm and the w and b were adjusted to minimize the network's generalized error.

4.2. FCM-BPNN

To develop the FCM-BPNN model, the flowchart in Fig. 1 was applied. Accordingly, the sorption efficiency training dataset will be divided into clusters by the FCM model. The challenging is how many clusters are optimal in this case? To answer this question, the Elbow method (Kodinariya and Makwana, 2013) was applied to determine the optimal number of clusters for the training phase (Fig. 5). According to Elbow, the optimal cluster number is where the total within the sum of square does not change much from the previous place. Therefore, the optimal number of clusters for the dataset used herein should be equal to 4, as shown in Fig. 5. Accordingly, the FCM algorithm was divided the training dataset into 4 clusters (Fig. 6): (i) cluster 1: 84 observations; (ii) cluster 2: 49 observations; (iii) cluster 3: 98 observations; (iv) cluster 4: 17 observations.

In Fig. 6, "Dimension 1" is the variance of the data in the horizontal direction, and "Dimension 2 is the variance of the data in the vertical direction". Herein, the Dimension 1 is equal to 28.7% and Dimension 2 is equal to 22.5%. The total variance of the data in both directions is 51.2%. Cluster results show that the FCM algorithm split data based on biochar characteristics and adsorption conditions. T_P, C, H/C, PSB, and T_{env} are underlying variables used to divide the clusters by the FCM algorithm. Based on the cluster results, can be observed that the FCM algorithm correctly solved the similarity between the sorption efficiency observations of heavy metal onto biochar. Accordingly, heavy metal sorption onto biochar based on the biochar characteristics and adsorption conditions can be more accurately predicted by BPNN models. Indeed, four BPNN models have been developed on the four clusters of the training dataset. The details of these sub-BPNN models' development are similar to the BPNN model on the whole training dataset.

5. Results and discussion

Once the BPNN and FCM-BPNN models were well designed and optimized, their efficiency was evaluated through the performance indicators, such as RMSE, R^2, RSE, MAPE, and VAF. The equations for calculating these indicators are presented in the supplementary materials. In order to validate the modeling efficiency in practice, the testing dataset was used as the unseen data, and the performances on the models are comprehensive evaluated. It should be emphasized that the FCM algorithm was also applied with the same purpose for the testing dataset (Fig. 7). Four clusters were then also divided on the testing dataset, and they were used to evaluate four BPNN models (sub-BPNN models), respectively, as calculated in Table 2. Accordingly, the Dimension 1 in the testing dataset is equal to 26.6% and Dimension 2 in the testing dataset is equal to 23.4%. The total variance of the data in both directions is 50%. Based on the clusters plotted in two dimensions in both training and testing datasets, it can be seen that the variances and total variances are similar. In other words, the clustering is positive in the absorption dataset used.

The results in Table 2 show that the sub-BPNN models provided an

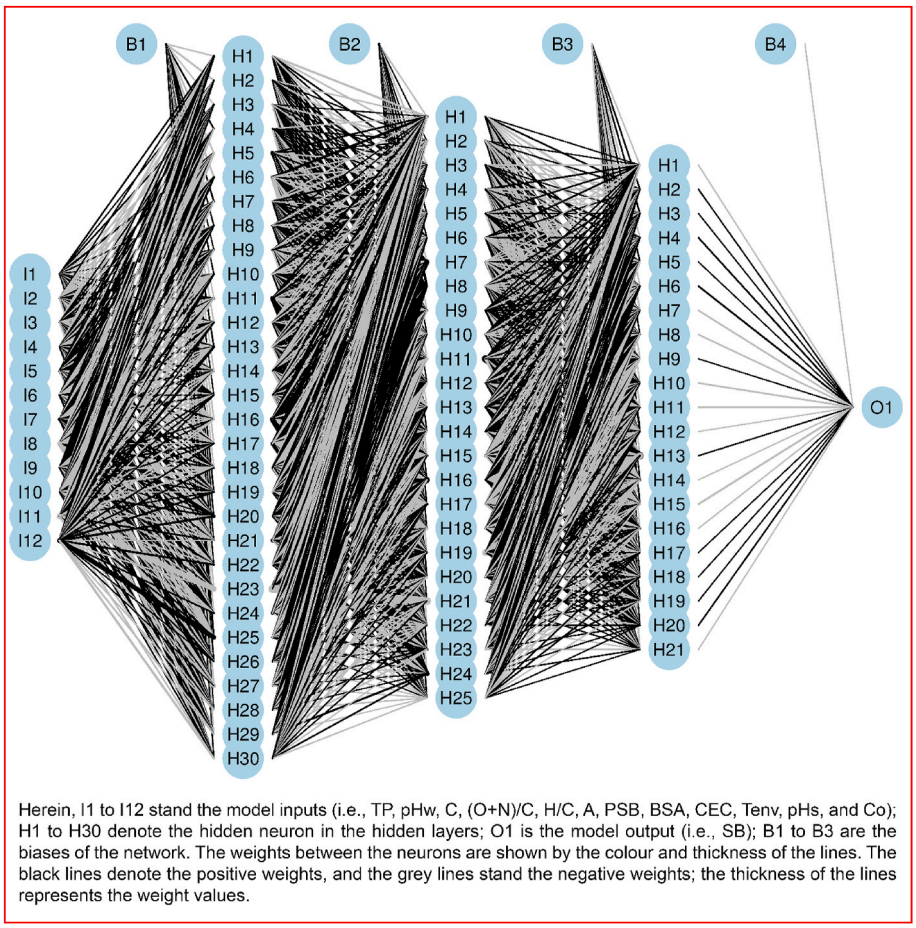

Fig. 4. The network of BPNN model for predicting the heavy metal sorption efficiency onto biochar in this study.

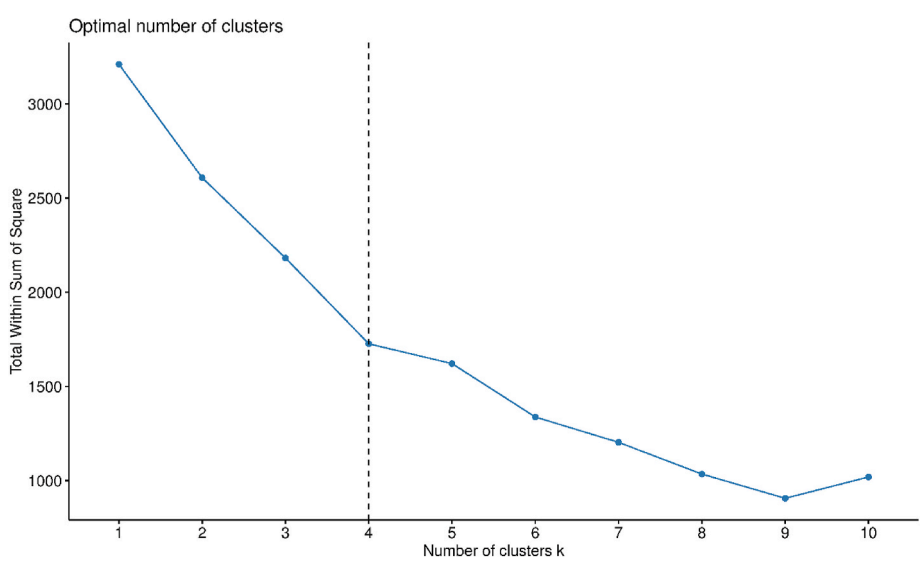

Fig. 5. The optimal number of clusters.

outstanding performance on each data cluster. They indicated that the FCM algorithm adequately explained the data's relationship and grouped them into clusters with similar characteristics. The data's similar characteristics are good candidates for the BPNN model, and they contributed a crucial role in understanding and improving the models' accuracies. Analyzing Table 2 carefully shows that despite the RMSE of the BPNN model on cluster 4 is lowest on the training dataset (i.e., RMSE = 0.025); however, it is the highest accuracy on the testing dataset (i.e., RMSE = 0.051). The reason for this problem is a small dataset was trained on cluster 4 during developing the BPNN model (i.e., 17 observations). Besides, neither training nor testing phases provided the lowest MAPE for the BPNN model. It is evident that overfitting did

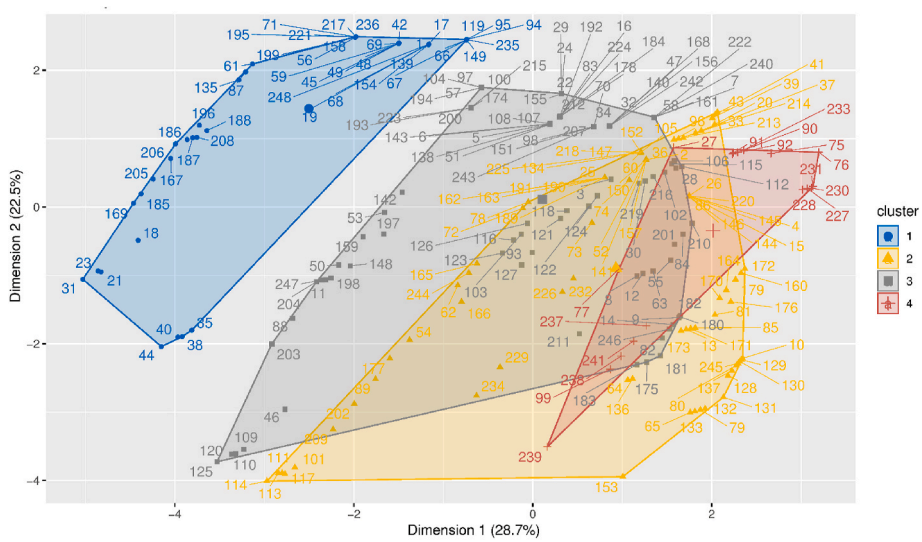

Fig. 6. The clusters of the training dataset used.

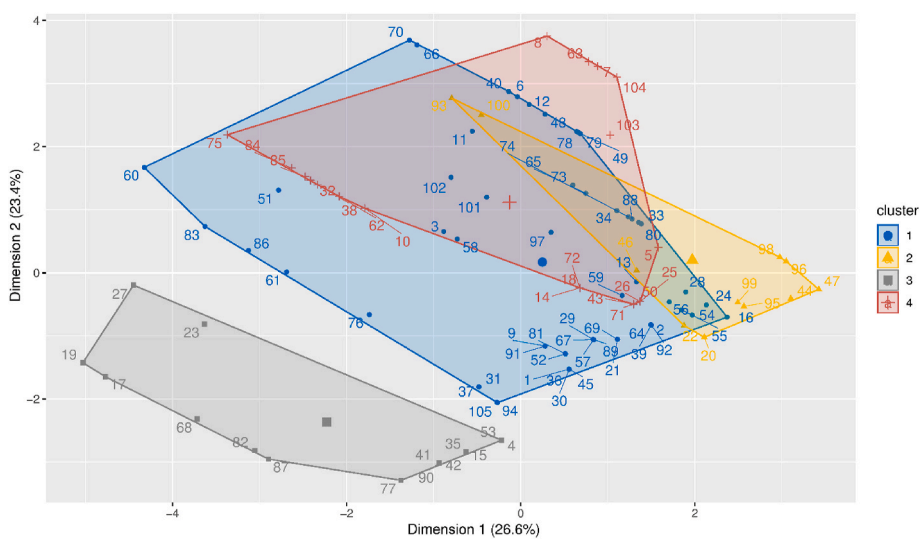

Fig. 7. Clusters divided based on the testing dataset.

Table 2
Performance of the sub-BPNN models on the groups divided by the FCM algorithm.

sub-BPNN models	Training phase					Testing phase				
	RMSE	R^2	RSE	MAPE	VAF	RMSE	R^2	RSE	MAPE	VAF
Group 1	0.048	0.985	0.007	0.525	98.488	0.048	0.987	0.011	0.571	98.165
Group 2	0.028	0.993	0.066	0.405	99.270	0.013	0.986	0.074	0.646	98.170
Group 3	0.034	0.991	0.488	0.685	99.024	0.038	0.991	0.553	0.820	99.063
Group 4	0.025	0.974	0.084	0.286	97.143	0.051	0.987	0.096	0.301	98.308

Table 3
Overall performance of the BPNN and FCM-BPNN models.

Models	Training phase					Testing phase				
	RMSE	R^2	RSE	MAPE	VAF	RMSE	R^2	RSE	MAPE	VAF
BPNN	0.048	0.982	0.005	1.143	98.174	0.050	0.977	0.011	0.802	97.662
FCM-BPNN	0.038	0.989	0.004	0.702	98.866	0.036	0.987	0.006	0.706	98.724

not occur for the BPNN model on cluster 4 due to data similarity. Furthermore, the techniques used for developing the BPNN model in cluster 4 are the same as those used to develop the BPNN model on the whole dataset. Once the sub-BPNN models' outcome predictions on the four groups obtained, they were combined as the outcome predictions of the FCM-BPNN model, and their results are compared in Table 3.

Based on the performances in Table 3, it can be asserted that the FCM-BPNN model yielded promising results. The predictive efficiency of heavy metal adsorption onto the proposed FCM-BPNN model's biochar is much superior to the BPNN model. The MAPE percentage of the FCM-BPNN model is only 0.706%, while the BPNN model is 0.802%. Similar results on the RSE also show that the proposed FCM-BPNN model's error rate is shallow. In addition, the correlation between actual sorption efficiency and predicted sorption efficiency of both BPNN and FCM-BPNN models is very high ($R^2 > 0.97$). Of those, the proposed FCM-BPNN model is fitter than the BPNN model with an R^2 of 0.987. These results are also visualized in Fig. 8 for a more accurate assessment.

Looking at Fig. 8, we can see that the BPNN and FCM-BPNN models in this study are well suited to predict heavy metal's sorption efficiency onto biochar. The level of regression on the models is very high for both training and testing phases. Looking closely at Fig. 8 shows that the regression rate on the FCM-BPNN model is higher than the BPNN model (without clustering). The models' 80% confidence level showed that the number of predicted values outside this confidence level was very low on the FCM-BPNN model. Meanwhile, indicated points outside the 80% confidence level of the BPNN model appear more. The models' accuracies are interpreted in Fig. 9.

In addition to the performance metrics, a Taylor diagram was also analyzed to compare and evaluate the developed sorption efficiency of heavy metal prediction models through standard deviation and correlation coefficient (Fig. 10). Accordingly, it is easy to see that the FCM-BPNN model is closer to the observed model than the BPNN model. The Taylor diagram also indicates that the standard deviation of the FCM-BPNN model is lower than the BPNN model, and the correlation coefficient is the same as discussed in Table 3.

Compared with the previous study (Zhu et al., 2019) based on the same database, it can be concluded that both advanced models in this study (i.e., BPNN and FCM-BPNN) are superior to those of the models that were developed in (Zhu et al., 2019) (i.e., ANN and RF), as shown in Table 4. Remarkable, the ANN model in the previous study also used the back-propagation algorithm for training the ANN; however, the obtained performance in the previous study was lower than the BPNN model in this study. This comparison indicated that the structure of the BPNN model in this study (with three hidden layers) is more optimized than the ANN model's structure in the previous study (with two hidden layers). Furthermore, the proposed FCM-BPNN model provided the most dominant accuracy among this study's models and the previous study.

6. Conclusions

Management of water resources and treatment of waste/wastewater containing heavy metals is an urgent issue that needs to be implemented in all countries. The biological absorption systems aiming to remove heavy metals from waste/wastewater brought certain economic and environmental benefits. However, under different environmental conditions and various biochar characteristics, the adsorption of heavy metals efficiency is not the same. This research developed and proposed two artificial intelligence models capable of predicting heavy metal

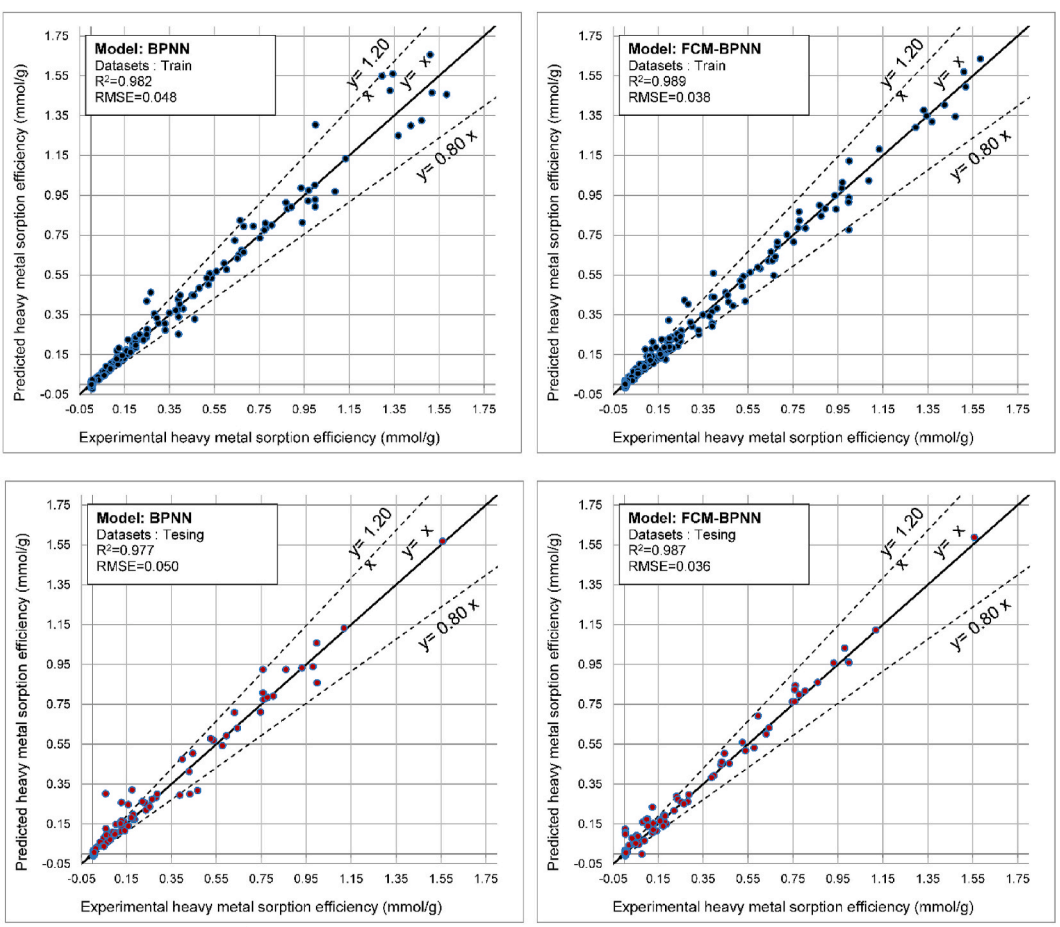

Fig. 8. Correlation between the actual sorption efficiency and predicted sorption efficiency by the BPNN and FCM-BPNN models and their 80% confident level.

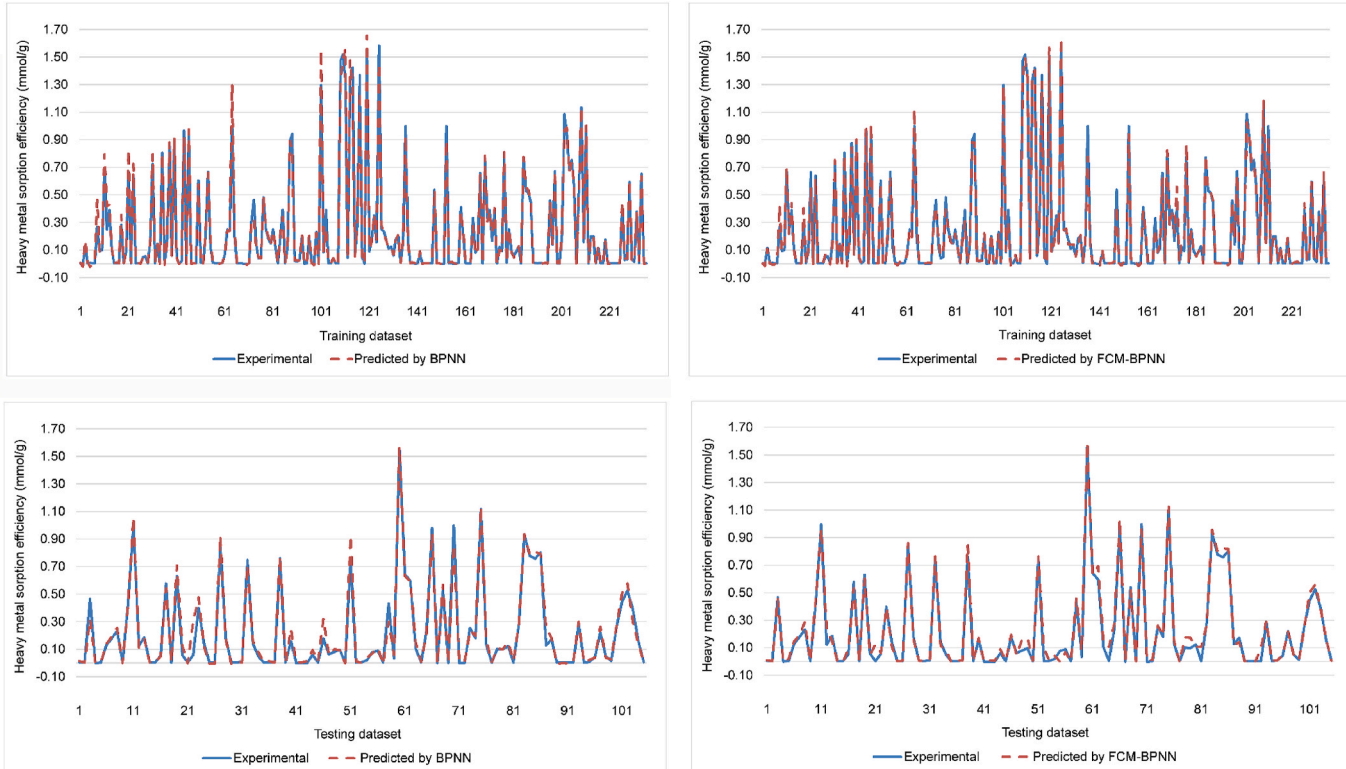

Fig. 9. The accuracy of the BPNN and FCM-BPNN models on the datasets.

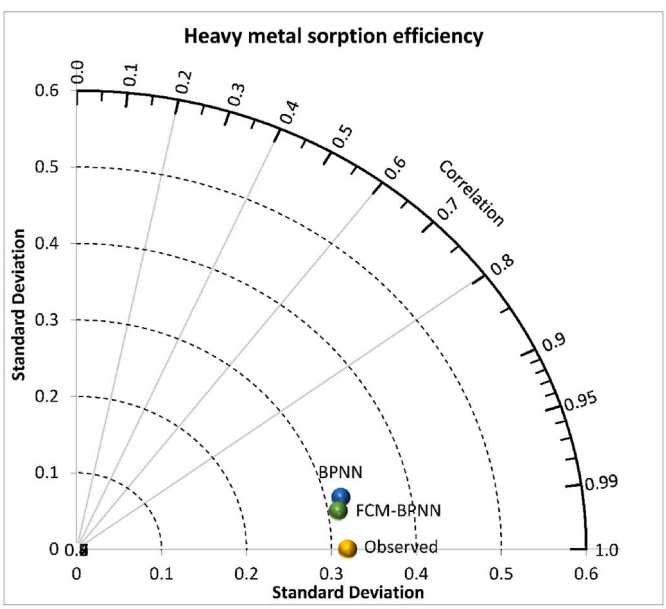

Fig. 10. Taylor diagram of the BPNN and FCM-BPNN models.

Table 4
The obtained results of the present work versus the previous work.

Models	Performance	
	RMSE	R^2
BPNN	0.048	0.982
FCM-BPNN	0.038	0.989
ANN (Zhu et al., 2019)	0.079	0.948
RF (Zhu et al., 2019)	0.057	0.973

sorption efficiency onto biochar with high accuracy based on neural networks and clustering algorithms (i.e., BPNN and FCM-BPNN models). Of those, the FCM-BPNN model was proposed as the most intelligence model with the highest accuracy in predicting the sorption efficiency of heavy metal onto biochar. Based on the obtained results, the proposed FCM-BPNN model can be used as an alternative method to predict the sorption efficiency of heavy metal onto biochar with different characteristics of the biochar systems and the different biochar conditions. This model is a useful tool for wastewater treatment plants, maximum support in water resources management, and pollution of countries' wastewater.

Credit author statement

Hoang Nguyen and Xuan-Nam Bui: Conceptualization, Methodology, Software, Data curation, Visualization, Writing- Original draft preparation and revise the manuscript. Hoang Bac Bui and Trung Nguyen-Thoi: Data curation, Visualization, Original draft preparation. Bo Ke: Conceptualization, Methodology, Software, Data curation, Visualization, and revise the manuscript.

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.

Acknowledgements

The authors would like to thank the Center for Mining, ElectroMechanical 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.jenvman.2021.112808.

References

Al-Asheh, S., Banat, F., Al-Omari, R., Duvnjak, Z., 2000. Predictions of binary sorption isotherms for the sorption of heavy metals by pine bark using single isotherm data. Chemosphere 41 (5), 659–665.

Asteris, P., Kolovos, K., Douvika, M., Roinos, K., 2016. Prediction of self-compacting concrete strength using artificial neural networks. Eur. J. Environ. Civil Eng. 20 (Suppl. 1), s102–s122.

Asteris, P.G., Nikoo, M., 2019. Artificial bee colony-based neural network for the prediction of the fundamental period of infilled frame structures. Neural Comput. Appl. https://doi.org/10.1007/s00521-018-03965-1 [journal article].

Atieh, M.A., Ji, Y., Kochkodan, V., 2017. Metals in the environment: toxic metals removal. Bioinorgan. Chem. Appl. 2017.

Ayangbenro, A.S., Babalola, O.O., 2017. A new strategy for heavy metal polluted environments: a review of microbial biosorbents. Int. J. Environ. Res. Publ. Health 14 (1), 94.

Bo, Ke, Hoang, Nguyen, Xuan-Nam, Bui, et al., 2021. Predicting the sorption efficiency of heavy metal based on the biochar characteristics, metal sources, and environmental conditions using various novel hybrid machine learning models. Chemosphere 276, 130204. https://doi.org/10.1016/j.chemosphere.2021.130204.

Bo, Ke, Keping, Zhou, Chaoshui, Xu, et al., 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. In this issue.

Bui, X.-N., Lee, C.W., Nguyen, H., Bui, H.-B., Long, N.Q., Le, Q.-T., et al., 2019. Estimating PM10 concentration from drilling operations in open-pit mines using an assembly of SVR and PSO. Appl. Sci. 9 (14), 2806.

Cui, X., Fang, S., Yao, Y., Li, T., Ni, Q., Yang, X., et al., 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.

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.

Duong Thanh, Nguyen, Nu Thi, Nguyen, Ha Thi Ngoc, Pham, et al., 2020. Rice husk ash and its utilization in soil improvement: An overview. J. Mining Earth Sci. 61 (03), 1–11. https://doi.org/10.46326/JMES.2020.61(3).01.

Fawzy, M., Nasr, M., Adel, S., Nagy, H., Helmi, S., 2016. Environmental approach and artificial intelligence for Ni (II) and Cd (II) biosorption from aqueous solution using Typha domingensis biomass. Ecol. Eng. 95, 743–752.

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

Guo, H., Zhou, J., Koopialipoor, M., Jahed Armaghani, D., Tahir, M.M., 2019. Deep neural network and whale optimization algorithm to assess flyrock induced by blasting. Eng. Comput. https://doi.org/10.1007/s00366-019-00816-y.

Haileslassie, T., Gebremedhin, K., 2015. Hazards of heavy metal contamination in ground water. Int. J. Technol. Enhanc. Emerg. Eng. Res. 3 (2), 1–6.

Hoang, Nguyen, 2020. Application of the k - nearest neighbors algorithm for predicting blast - induced ground vibration in open - pit coal mines: a case study. J. Mining Earth Sci. 61 (06), 22–29. https://doi.org/10.46326/JMES.2020.61(6).03.

Hu, J.S., Zhong, L.S., Song, W.G., Wan, L.J., 2008. Synthesis of hierarchically structured metal oxides and their application in heavy metal ion removal. Adv. Mater. 20 (15), 2977–2982.

Hung Van, Nguyen, Son Truong, Bui, Hai Huu, Phung, Ha Thi Ngoc, Pham, 2020. The initial research on the compressive strength of mortar when using bottom ash from thermal power plants to replace natural sand in construction. J. Mining Earth Sci. 61 (03), 12–18. https://doi.org/10.46326/JMES.2020.61(3).02.

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

Kodinariya, T.M., Makwana, P.R., 2013. Review on determining number of cluster in K-means clustering. Int. J. Adv. Res. Computer Sci. Manag. Stud. 1 (6), 90–95.

Kołodyńska, D., Kowalczyk, M., Hubicki, Z., Shvets, V., Golub, V., 2015. Effect of accompanying ions and ethylenediaminedisuccinic acid on heavy metals sorption using hybrid materials Lewatit FO 36 and Purolite Arsen Xnp. Chem. Eng. J. 276, 376–387.

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.

Mamandipoor, B., Majd, M., Sheikhalishahi, S., Modena, C., Osmani, V., 2020. Monitoring and detecting faults in wastewater treatment plants using deep learning. Environ. Monit. Assess. 192 (2), 148. https://doi.org/10.1007/s10661-020-8064-1.

Mejias Carpio, I.E., Ansari, A., Rodrigues, D.F., 2018. Relationship of biodiversity with heavy metal tolerance and sorption capacity: a meta-analysis approach. Environ. Sci. Technol. 52 (1), 184–194.

Muya, F.N., Sunday, C.E., Baker, P., Iwuoha, E., 2016. Environmental remediation of heavy metal ions from aqueous solution through hydrogel adsorption: a critical review. Water Sci. Technol. 73 (5), 983–992.

Nguyen, H., Bui, X.-N., Bui, H.-B., Mai, N.-L., 2018. A comparative study of artificial neural networks in predicting blast-induced air-blast overpressure at Deo Nai open-pit coal mine. Neural Comput. Appl. 1–17. https://doi.org/10.1007/s00521-018-3717-5. Vietnam. [journal article].

Nguyen, H., Bui, X.-N., Tran, Q.-H., Mai, N.-L., 2019a. A new soft computing model for estimating and controlling blast-produced ground vibration based on hierarchical K-means clustering and cubist algorithms. Appl. Soft Comput. 77, 376–386. https://doi.org/10.1016/j.asoc.2019.01.042.

Nguyen, H., Bui, X.-N., Tran, Q.-H., Moayedi, H., 2019b. Predicting blast-induced peak particle velocity using BGAMs, ANN and SVM: a case study at the Nui Beo open-pit coal mine in Vietnam. [journal article]. Environ. Earth Sci. 78 (15), 479. https://doi.org/10.1007/s12665-019-8491-x.

Nguyen, H., Drebenstedt, C., Bui, X.-N., Bui, D.T., 2020a. Prediction of blast-induced ground vibration in an open-pit mine by a novel hybrid model based on clustering and artificial neural network. Nat. Resour. Res. 29 (2), 691–709.

Nguyen, H., Moayedi, H., Foong, L.K., Al Najjar, H.A.H., Jusoh, W.A.W., Rashid, A.S.A., et al., 2020b. Optimizing ANN models with PSO for predicting short building seismic response. Eng. Comput. 36 (3), 823–837. https://doi.org/10.1007/s00366-019-00733-0.

Nhan Thi, Pham, Pan, Bing, Nghia Viet, Nguyen, 2020. Study on the effect of some parameters of soil nails on the stability of vertical slopes. J. Mining Earth Sci. 61 (06), 30–37. https://doi.org/10.46326/JMES.2020.61(6).04.

Ojuederie, O.B., Babalola, O.O., 2017. Microbial and plant-assisted bioremediation of heavy metal polluted environments: a review. Int. J. Environ. Res. Publ. Health 14 (12), 1504.

Pandeeswari, N., Kumar, G., 2016. Anomaly detection system in cloud environment using fuzzy clustering based ANN. Mobile Network. Appl. 21 (3), 494–505.

Papageorgiou, E. I., Poczęta, K., & Laspidou, C. Hybrid model for water demand prediction based on fuzzy cognitive maps and artificial neural networks. In 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), vol. 2016 (pp. 1523-1530): IEEE.

Parveen, N., Zaidi, S., Danish, M., 2016a. Artificial Intelligence (AI) based tools for predicting the removal efficiency of heavy metals by adsorption. In: Pharmaceutical and Food Industry Application of Adsorption and Ion Exchange Chromatography in Chemical. Materials Research Forum (LLC), USA, pp. 194–238.

Parveen, N., Zaidi, S., Danish, M., 2016b. Support vector regression model for predicting the sorption capacity of lead (II). Perspect. Sci. 8, 629–631.

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.

Peralta-Videa, J.R., Lopez, M.L., Narayan, M., Saupe, G., Gardea-Torresdey, J., 2009. The biochemistry of environmental heavy metal uptake by plants: implications for the food chain. Int. J. Biochem. Cell Biol. 41 (8–9), 1665–1677.

Quyen Thi, Vu, Minh Ngoc, Pham, Chinh Manh, Dang, et al., 2020. Fuzzy logic in controlling the forest fire - level forecast warning signage. J. Mining Earth Sci. 61 (04), 126–136. https://doi.org/10.46326/JMES.2020.61(4).14.

Rahmani-Sani, A., Singh, P., Raizada, P., Lima, E.C., Anastopoulos, I., Giannakoudakis, D.A., et al., 2020. Use of chicken feather and eggshell to synthesize a novel magnetized activated carbon for sorption of heavy metal ions. Bioresour. Technol. 297, 122452.

Ronsse, F., Van Hecke, S., Dickinson, D., Prins, W., 2013. Production and characterization of slow pyrolysis biochar: influence of feedstock type and pyrolysis conditions. Gcb Bioenergy 5 (2), 104–115.

Saleh, H.N., Panahande, M., Yousefi, M., Asghari, F.B., Conti, G.O., Talaee, E., et al., 2019. Carcinogenic and non-carcinogenic risk assessment of heavy metals in groundwater wells in Neyshabur Plain, Iran. Biol. Trace Elem. Res. 190 (1), 251–261.

Shahabi, H., Ahmad, B.B., Khezri, S., 2013. Evaluation and comparison of bivariate and multivariate statistical methods for landslide susceptibility mapping (case study: zab basin). Arab. J. Geosci. 6 (10), 3885–3907. https://doi.org/10.1007/s12517-012-0650-2.

Shang, Y., Nguyen, H., Bui, X.-N., Tran, Q.-H., Moayedi, H., 2019. A novel artificial intelligence approach to predict blast-induced ground vibration in open-pit mines based on the firefly algorithm and artificial neural network. Nat. Resour. Res. 29 (2), 723–737. https://doi.org/10.1007/s11053-019-09503-7.

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. Pollut. Control Ser. 24 (14), 12809–12819.

Sun, Y., Gao, B., Yao, Y., Fang, J., Zhang, M., Zhou, Y., et al., 2014. Effects of feedstock type, production method, and pyrolysis temperature on biochar and hydrochar properties. Chem. Eng. J. 240, 574–578.

Thuy Thi Thanh, Tran, Huy Khanh, Pham, Hoa Mai, Nguyen, 2020. Assessing the current status of rural domestic solid waste management in Nam Dinh province. J. Mining Earth Sci. 61 (06), 82–89. https://doi.org/10.46326/JMES.2020.61(6).09.

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.

Turan, N.G., Mesci, B., Ozgonenel, O., 2011. The use of artificial neural networks (ANN) for modeling of adsorption of Cu (II) from industrial leachate by pumice. Chem. Eng. J. 171 (3), 1091–1097.

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 (12), 3158–3165.

Vandenbossche, M., Casetta, M., Jimenez, M., Bellayer, S., Traisnel, M., 2014. Cysteine-grafted nonwoven geotextile: a new and efficient material for heavy metals sorption–Part A. J. Environ. Manag. 132, 107–112.

Wang, Y., Hong, H., Chen, W., Li, S., Panahi, M., Khosravi, K., et al., 2019. Flood susceptibility mapping in Dingnan County (China) using adaptive neuro-fuzzy inference system with biogeography based optimization and imperialistic competitive algorithm. J. Environ. Manag. 247, 712–729. https://doi.org/10.1016/j.jenvman.2019.06.102.

Wierzba, S., Kłos, A., 2019. Heavy metal sorption in biosorbents–Using spent grain from the brewing industry. J. Clean. Prod. 225, 112–120.

Wu, Y., Pang, H., Liu, Y., Wang, X., Yu, S., Fu, D., et al., 2019. Environmental remediation of heavy metal ions by novel-nanomaterials: a review. Environ. Pollut. 246, 608–620.

Xiaojian, Y., Yue, Z., Jinghui, Q., 2017. Research on abnormal detection of chemical storage tank based on FCM-ANN. Computer Appl. Software 2 (38).

Yang, X., Wan, Y., Zheng, Y., He, F., Yu, Z., Huang, J., et al., 2019. Surface functional groups of carbon-based adsorbents and their roles in the removal of heavy metals from aqueous solutions: a critical review. Chem. Eng. J. 366, 608–621. https://doi.org/10.1016/j.cej.2019.02.119. In this issue.

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.

Zhai, Y., Liu, X., Zhu, Y., Peng, C., Wang, T., Zhu, L., et al., 2016. Hydrothermal carbonization of sewage sludge: the effect of feed-water pH on fate and risk of heavy metals in hydrochars. Bioresour. Technol. 218, 183–188.

Zhang, H., Nguyen, H., Bui, X.-N., Nguyen-Thoi, T., Bui, T.-T., Nguyen, N., et al., 2020. Developing a novel artificial intelligence model to estimate the capital cost of mining projects using deep neural network-based ant colony optimization algorithm. Resour. Pol. 66, 101604.

Zhou, J., Li, E., Yang, S., Wang, M., Shi, X., Yao, S., et al., 2019. Slope stability prediction for circular mode failure using gradient boosting machine approach based on an updated database of case histories. Saf. Sci. 118, 505–518.

Zhou, J., Li, X., Mitri, H.S., 2016. Classification of rockburst in underground projects: comparison of ten supervised learning methods. J. Comput. Civ. Eng. 30 (5), 04016003.

Zhou, J., Qiu, Y., Armaghani, D.J., Zhang, W., Li, C., Zhu, S., et al., 2020a. Predicting TBM penetration rate in hard rock condition: A comparative study among six XGB-based metaheuristic techniques. Geosci. Front. 12 (3), 101091 https://doi.org/10.1016/j.gsf.2020.09.020.

Zhou, J., Qiu, Y., Zhu, S., Armaghani, D.J., Khandelwal, M., Mohamad, E.T., 2020b. Estimation of the TBM advance rate under hard rock conditions using XGBoost and Bayesian optimization. Undergr. Space. https://doi.org/10.1016/j.undsp.2020.05.008 in press.

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.