Gonbad Kavous University

Journal of New Approaches in Water Engineering and Environment

Online ISSN: 2981-0280

Homepage: https://nawee.gonbad.ac.ir

Nitrate adsorption modeling using SVM and LSSVM models

Masumeh Farasati[1*], Seyed Morteza Seyedian[2], Seyed Javad sajadi[3]

[1] Department of water engineering, faculty of agriculture and natural resource, Gonbad kavous University, Gonbad kavous, Iran.

[2]Department of water engineering, faculty of agriculture and natural resource, Gonbad kavous University, Gonbad kavous, Iran.

[3]Department of plant production, faculty of agriculture and natural resource, Gonbad kavous University, Gonbad kavous, Iran.

Article Info

Article type:

Research Article

Article history:

Received 05 March 2025

Revised form 29 February 2025

Accepted 20 June 2025

Published online 26 June 2025

Keywords:

Nitrate

Adsorption

SVM

Lssvm

ABSTRACT

Objective: Nitrate compounds are among the pollutants of groundwater resources, and in recent years, in terms of agricultural development and human activities, their average rate has been increasing. The purpose of this study is modeling of nitrate adsorption by using SVM, LSSVM and Random Forest model.

Material and Methods: The nitrate adsorption data used in this study were first randomized and standardized and then divided into two groups of training and testing. 70% of the data were in the training group and the remaining 30% in the experimental group. Validation of model training was performed using k-fold cross-validation method with a value of k = 5 in order to prevent over-fitting of models. The parameters of Random Forest, SVM and LS-SVM models were determined using a Bayesian optimization algorithm. The objective function of the optimization algorithm was to minimize the MSE error value of the model.

Results and Discussion: Based on the results, the Random Forest model was used with the Bagging algorithm and the parameters of minimum node size, number of trees and number of variables used were equal to 2, 10 and 3, respectively. The SVM model was trained with the RBF kernel function and the parameters of Box Constraint and Epsilon equal to 2.2156 and 0.0891, respectively, along with standardization of input and output data of the model. The LS-SVM model was also trained with RBF kernel function and setting parameters and kernel function equal to 3160/3160 and 19.7891/19, respectively.

Conclusions: Taylor diagram results showed that the stochastic forest model and SVM had a higher correlation between observational and estimated data. Therefore, based on the results, the stochastic forest model is more consistent with the observation data and predicts nitrate changes well.

*Corresponding author, Email: z.khani1060@yahoo.com.

Cite this article: Farasati, Masumeh., seyedian, morteza., sajadi, javad. Nitrate adsorption modeling using SVM and LSSVM models. Journal of New Approaches in Water Engineering and Environment, 4(2),160-179. https://doi.org/10.22034/nawee.2025.510813.1145.

Publisher: Gonbad Kavous University.

DOI: https://doi.org/10.22034/nawee.2025.510813.1145

1.Introduction

The average amount of nitrate compounds, which are among the contaminant factors of groundwater sources, has recently increased due to the expansion of agriculture and human activities (Yousefi and Qomian, 2013). This ion may enter water wells when water passes through soils or may enter underground water sources as a result of water pollution through organic substances, accumulation of urban and industrial waste, accumulation of manure and chemical fertilizers or the leakage of urban sewage facilities. However, in the last few decades, the increase in the use of nitrogenous chemical fertilizers has led to the addition of nitrate levels in surface and underground waters (McLay, 2001).

Water with a high nitrate concentration is potentially harmful to infants and children. According to the guidelines of the World Health Organization (WHO) and the latest national standard of Iran, the maximum permissible nitrate ion in drinking water is 50 mg/l (Iran Institute of Standards and Industrial Research, 2018).

In recent decades, extensive research has been conducted on the use of inexpensive adsorbents as nitrate removal agents. Although these adsorbents do not have the same ability of activated carbon to adsorb pollutants, their low cost has attracted the attention of scientists to use these materials (Sayadi et al., 2020). The theory of turning agricultural residues into anion exchangers came to power in 1970. Due to abundance, low cost, and biodegradability, use of these residues has attracted great interest, especially in recent years (Manju et al., 1998; Simkovic et al., 1997).

Organic waste is continuously or intermittently decomposed, metabolized, synthesized and finally stabilized under the synergistic action of various microorganisms. There are some primary metabolites as precursors of humus generation, and the thorough and complete mineralization processes do not always occur (Guo et al., 2019).

Biochar is produced from the pyrolysis of various biomasses under anaerobic conditions (or low oxygen conditions) (Lin et al., 2011). The adsorption capacity of biochar depends on its physico-chemical characteristics, which are influenced by various factors such as raw material type, particle size, pyrolysis temperature, rate of temperature changes and temperature holding time (Acikalin and Bulat, 2012; Chan and Xu, 2009). Researchers have reported that biochar can be employed as a modifier to improve the fertility of agricultural land upon removing nitrate from polluted water without any particular risk to the environment (Yao et al., 2013).

SVM and LSSVM models are one of the supervised learning methods used for classification and regression nitrate adsorption.

In a research conducted on the removal of nitrates from contaminated water sources using modified Phragmites Australis Nanoparticles, Farasati et al (2011) concluded that the removal efficiency increases from 60 to 80% by increasing the pH of the solution from 2 to 10 and

reaches its maximum at a pH of 6. An equilibrium time of 2 h was also obtained. By increasing the initial nitrate concentration from 5 to 120 mg/l, the removal efficiency has decreased from 90 to 67%. Moreover, they found that with an increase in the adsorbent mass, the adsorption efficiency increases.

Sayadi et al. (2020) investigated nitrate removal from water using modified biochar from Conocarpus and Palnoia plants. The results of their study showed that nitrate adsorption increases with time and reaches its maximum after 60 min. Moreover, maximum nitrate adsorption was found at pH=2. As the initial concentration of nitrate increased, the removal efficiency increased so that the highest removal efficiency was observed in Palnoia modified biochar (94.65%) at a concentration of 5 mg/l.

Examining the removal and recovery of cadmium from aqueous solutions using Conocarpus nanostructure, Pourmohamad et al (2017) arrived at the fact that the optimum pH for cadmium adsorption is 6 and estimated the maximum removal efficiency and cadmium adsorption capacity as 80.9% and 0.86 mg/g, respectively. The comparison of the Langmuir adsorption model and kinetic models showed that the Langmuir isotherm and Howe's kinetic models provided a better fit and better description compared to other counterparts.

The evaporation phenomenon from the free surface of water was simulated using support vector machine (SVM) and least squares support vector machine (LS-SVM) models (Farasati et al., 2020). The results showed that among the input patterns to SVM and LSSVM models, pattern 16 with the input parameters of minimum relative humidity, maximum relative humidity, wind speed and sundial meets the highest R^2 and the least root mean square error (RMSE) and multistate bar examination (MBE) values. The LS-SVM model at Bandar Turkmen station was also found to have the best prediction compared to the other two stations. Moreover, in all examined stations, the LS-SVM model has yielded a higher R^2 and less RMSE and MBE than the SVM counterpart. The aim of this research is the application of SVM and LSSM model for nitrate adsorption.

2.Materials and methods

2.1. Preparation of the adsorbents

In order to select the appropriate adsorbent for removing nitrate, Conocarpus and Palnoia plants were modified with different methods, as described below:

First method: Conocarpus and Palnoia plants were washed with distilled water and dried in an oven at 60 °C for 12 h. Then, they were pyrolyzed in an electric furnace at 300 and 600 °C for 4 h (Sayadi et al., 2017).

Second method: the biochars obtained at temperatures of 300 and 600 °C were mixed with

potassium hydroxide at a ratio of 1 to 3 and left for 24 h. They were filtered and then washed with distilled water and placed in an oven to dry. Finally, they were pyrolyzed again in an electric furnace at a temperature of 600 °C for 2 h (Sayadi et al., 2017).

Third method: the raw plants of Conocarpus and Palnoia were combined with 1 Molar iron chloride (pH=10) and rested for 2 h. After the specified time, they were filtered, washed with distilled water and placed in an oven at a temperature of 60 °C for 12 h to dry. Then, they were

pyrolyzed in an electric furnace at a temperature of 600 °C for 4 h (Sayadi et al., 2017).

Table 1 shows the physical properties of modified Conocarpus and Palnoia adsorbents. As can be observed, the modified Palnoia adsorbent has a more specific surface than the modified Conocarpus one. Furthermore, the modified Conocarpus adsorbent has a lower ash percentage and weighted moisture than that of the modified Palnoia. It is saying that specific surface is one

of the influential factors on adsorption and, the higher the specific surface, the higher the adsorption rate.

Table 1- Physical properties of modified Conocarpus and Palnoia adsorbents

Adsorbent	Specific surface (m^2/g)	Ash (%)	Organic materials (%)	Density (kg/m^3)	Weighted moisture (%)
$FeCL_2$ Conocarpus	85.1	88.6	11.4	0.01	1.14
$FeCL_2$ Palnoia	99.8	97.6	2.4	0.002	2.4

Finally, five different biochars were prepared for each of the examined plants, which include Conocarpus 300, Conocarpus 600, potassium hydroxide Conocarpus 300, potassium hydroxide Conocarpus 600, $FeCL_2$ Conocarpus, Palnoia 300, Palnoia 600, potassium hydroxide Palnoia 300, potassium hydroxide Palnoia 600 and $FeCL_2$ Palnoia.

Each of the prepared biochars was used to conduct nitrate experiments. Nitrate concentration was measured with a spectrophotometry device with a wavelength of 410 nm (Amininejad et al., 2017; Sayadi, 2020). Moreover, the removal efficiency and nitrate adsorption capacity have been

calculated using Eqs. (1) and (2), respectively:

$$\%R = \frac{C_i - C_f}{C_i} \times 100 \quad (1)$$

$$q = \frac{C_i - C_f}{m} \times V \quad (2)$$

Where q defines the adsorption rate of the dissolved matter per unit mass of the adsorbent (mg/g), C_i is the initial concentration of the solute (mg/g), C_f stands for the remaining concentration of the solute (mg/g) after the equilibration time has passed, m is the adsorbent amount (g) and V denotes the solution volume (l).

Conducting experiments using SVM, LSSVM and random forest models, nitrate removal has been predicted and compared with that via laboratory data.

2.2.SVM model

SVM is one of the supervised learning methods used for classification and regression purposes. SVM was introduced by Vapnik (1998) and was built based on the statistical learning theory. This method is employed for performing binary classification in an arbitrary feature space, and therefore, it is considered as an appropriate approach for solving prediction problems. In most real-world problems, the linear approximation function has limited practical application.

Using the concept of inner product, Vapnik (1998) showed that the input vector x can be initially transferred to a high-dimensional space through a nonlinear transformation and perform the inner product. Therefore, it can be proven that if the conditions of Mercer's theorem hold for a symmetric kernel, applying this kernel to the input space with low dimension can highly reduce the product. If the input vector x_i is mapped into the eigenspace using a nonlinear function $\boldsymbol{\Phi}$ (x), the decision function will be as follows:

SVM can be used to solve regression problems. The main relationship for the statistical learning process is given below:

$$y=f(X)=\sum_{i=1}^{M} W_i\varphi_i(X)=W\varphi(X) \quad 3$$

Where the output of the model is the linear part of M and its converter represents the nonlinear model denoted by (X)φ. This equation is transformed into equation 2 for the support vector machine.

$$y=f(X)=\left\{\sum_{i=1}^{N} W_iK(X_i,X)\right\} -b \quad 4$$

The model parameters have been estimated by maximizing the objective function. SVM is used in a number of special kernel functions that transform the input vector into the input data of the nonlinear function. Choosing a suitable kernel function is complex and often the standard kernel function is used. The linear regression model of Eq. (5) aims to find the linear function that provides the best fit for the training points.

$$y= f(X)=\langle W.X\rangle + b \quad 5$$

Parameters W and b have been determined using the least squares method as follows:

$$\sum_{i=f1}^{l}(y_i - \langle W.X\rangle - b)^2 \quad 6$$

In order to consider the difference between the observed data and the estimated values, the error value ε is entered into the above model as in Eq. (5).

$$y_i - W.X - b < \varepsilon$$

$$y_i - W.X + b \leq \varepsilon \quad 7$$

Using Eq. (5), a range is considered around function f(X), and some penalty is considered for the points outside this range, except in cases where a covariate $\in_i$ is employed. This covariate is zero for points within the range, while it increases exponentially for data outside it. This regression approach is called ε SV, which is the most common modeling method. In this model,

the value function is defined according to the following equation:

$$|\epsilon|_{\varepsilon}=|y-f(X)|_{\varepsilon}=\begin{cases}0 & \text{if}|y-f(X)|\le\varepsilon\\ |y-f(X)-\varepsilon| & \text{O.W.}\end{cases} \quad 8$$

The two layers around the function f(X) should be estimated by maximizing the boundary of the regions that have an inverse relationship with the Euclidean and vector $\|W\|^2$ planes. Then, the Euclidean and vector plane should be minimized as follows:

$$\text{Min}\frac{1}{2}\|W\|^2+C\left(\sum_i^l\epsilon_i^*+\sum_i^l\epsilon_i\right)$$

$$\text{S.t}\quad y_i-\langle W.X\rangle-b\le\varepsilon+\epsilon_i \quad 9$$

$$\langle W.X\rangle+b-y_i\le\varepsilon+\epsilon_i^* \qquad \epsilon_i,\epsilon_i^*\ge 0$$

To classify the data with high complexity, the data are taken to a higher-dimensional space using a kernel function (φ). The kernel function maps the data from the input space to another one with higher dimensions in such a way that it is possible to linearly separate them in that space. In fact, without any change in the nature of the optimization problem, the data are mapped to a space with higher dimensions using a mapping function φ, so that the decision level is determined linearly in this space. In this study, the radial basis function (RBF) kernel has been employed in SVM classification, according to the following relationship:

$$K(X,X_i)=\exp(-\|X-X_i\|^2/2\sigma^2) \quad 10$$

2.3.LS-SVM model

RBF kernel has been used here to implement LS-SVM. Using this kernel, LS-SVM determines parameter ɣ and kernel function parameter (σ^2) in such a way to minimize the training error and lead to a less complex parameter estimation function. There are three optimization algorithms in LS-SVM approach. To start the determination process of the parameters, first, the proper starting point is determined for each method and each kernel via a coupled simulated annealing (CSA) algorithm. This algorithm searches within the range exp(-10) to exp(10). Then, the starting points determined in the previous step are introduced to one of the three described optimization algorithms that are selected by the user. CSA has been proven to be more effective than the multi-start gradient descent optimization algorithm. Also, this method reduces the sensitivity of the algorithm to the initial parameters and improves the optimization process if there is an issue with the parameter optimization. By default, five multiple origins are used in CSA, and the optimized parameters include ɣ and the square of the kernel function (or σ^2 in the case of the RBF kernel). This technique originates from SVM and it is a powerful method for function estimation. According to the LS-SVM method, nonlinear functions can be represented as:

$$f(x)=wT.\boldsymbol{\varphi}(x)+b \quad (11)$$

The regression problem can be solved using minimization Eq. (10) along with the constraint of Eq. (11).

$$\text{Min}w,e,b\ j(w,e)=\frac{1}{2}w^{T}w+\frac{\gamma}{2}\sum_{i=1}^{m}e_i^2 \quad (12)$$

$$yi=wT.\varphi(xi)+b+ei \qquad (i=1,2,\ldots,m) \quad (13)$$

To solve the optimization problem of Eq. (10), the objective function can be solved by transforming the resulting constraint into an infinite problem and by introducing the Lagrange coefficients according to the following equation:

$$L(w,e,b,\boldsymbol{\alpha})=J(w,e)-\sum_{i=1}^{m}\alpha_i\{wT\varphi(xi)+b+ei-y_i\} \quad (14)$$

According to the Karush-Kuhn-Tucker conditions, optimal conditions can be obtained by considering the partial derivative of the above equation with respect to W, b, e and α as follows:

$$W^{T}\varphi(X_i)+b+e_i-\gamma_i=0\ ،\alpha_i=\gamma e_i\ ،\sum_{i=1}^{m}\alpha_i=0\ ,W=\sum_{i=1}^{m}\alpha_i\varphi(X_i) \quad (15)$$

Then, the optimized linear equations and kernel function can be defined as follows:

$$K(x,xj)=\boldsymbol{\varphi}(\mathbf{X})^{T}\boldsymbol{\varphi}(\mathbf{X_i}) \qquad I=1,\ldots,m \quad (16)$$

Also, LSSVM regression is given as:

$$f(X)=\sum_{i=1}^{m}\alpha_iK(X,X_i)+b \quad (17)$$

RBF kernel is commonly used for regression problems. In this research, this kernel function has been used, which can be calculated using Eq. (8).

2.4. RF model

RF is a set of unpruned trees, each coming with a recursive partitioning algorithm (Beriman, 2001). The algorithm for building a RF with T trees from a data set with n observations and p variables is as follows: (i) Using the bootstrap method, a random sample is extracted by sampling in a volume of n observations. (ii) For the selected bootstrap sample, a classification tree is grown using a recursive partitioning algorithm. In each node, partitioning is done based on a m tuple random sample of p predictor variables. (iii) The recursive partitioning algorithm continues until the tree reaches its largest size (that is, one final node for each observation), without pruning the tree. (iv) Steps (i) to (iii) are repeated T times to build an RF (Beriman, 1984). The common choices are 1000 trees for T, $\sqrt{p}$ and log(p) for m. All analyses have been mostly conducted using the Development Core Team R software. Moreover, RF analysis has been performed using the random forest function.

2.5. Data determination for model training and testing

The data used in this research were first randomized and standardized and then classified into two groups of training and testing. 70% of the data were included in the training group, while the remaining 30% were dedicated to the testing one. Validation of training models was performed using k-fold cross-validation approach with k = 5 in order to avoid overfitting the models. Also, the parameters of RF, SVM and LS-SVM models were determined using Bayesian optimization algorithm in which the objective function was considered to be MSE value of the model.

In order to evaluate the accuracy of the models, the statistical indices of coefficient of determination (R^2), RMSE and MBE were used as follows:

$$R^2 = \left(\frac{\sum(E_i - \bar{E}_i)(E_O - \bar{E}_O)}{\sqrt{\sum(E_i - \bar{E}_i)^2 \sum(E_O - \bar{E}_O)^2}}\right)^2 \quad (18)$$

$$RMSE = \sqrt{\frac{\sum(E_O - \bar{E}_i)^2}{N}} \quad (19)$$

$$MBE = \frac{\sum_{i=1}^{n}(E_o - E_i)}{N}$$

In these relationships, E_i and E_o stand for the observed and estimated adsorptions, respectively and N indicates the number of data.

3.Results and discussion

3.1.SEM test results

The results of measuring the surface morphology of the modified Conocarpus and Palnoia adsorbents are exhibited in Figures 1 and 2, respectively.

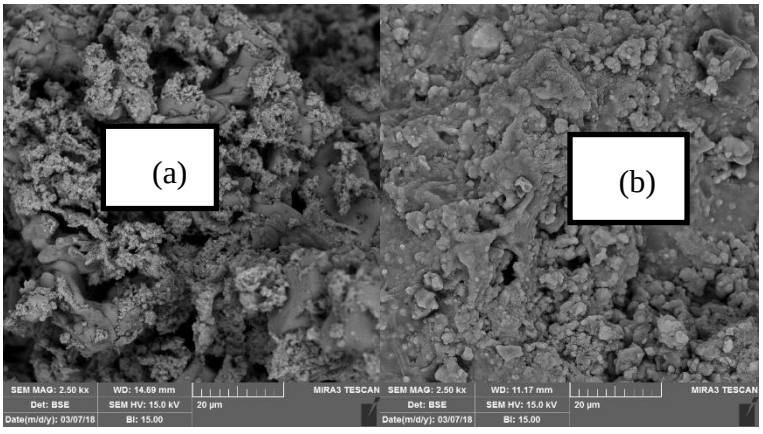

Figure 1-SEM image of modified adsorbent(a)$Fecl_2$ conocarpus(b) Conocarpus potassium hydroxide 600

According to Figure 1(a), it can be observed that many small and large dimples on the surface of Conocarpus adsorbent modified with $FeCL_2$, which suggests that the adsorbent has become rougher. This further indicates an increase in the specific surface area of the adsorbent and the non-uniform distribution of energy on its surface. As would be observed in Figure 1(b), the surface of Conocarpus adsorbent modified with potassium hydroxide 600 is softer and cottony.

According to Figure 2(a), the surface of Palnoia adsorbent modified with potassium hydroxide 300 has become rougher with many empty pores, which allows the adsorbent to accommodate more contaminants on the surface and in empty spaces. In figures (2(b and c)), the surface of Palnoia absorbents modified with iron chloride and 600 has become softer and smoother, increasing the specific surface area of the absorbents. As can be seen from Figures 2(b)-(c), the surface of Palnoia adsorbents modified with $FeCL_2$ and 600 has become softer and smoother, which has led to an increment in their specific surface area.

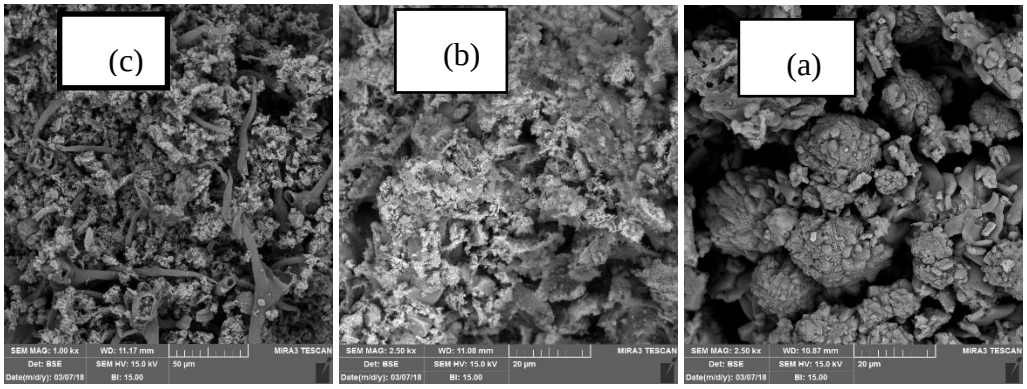

Figure 2- SEM images of modified attractants (a) palonia potassium hydroxide 300, (b) palonia iron chloride and (c) palonia 600

3.2. Investigation of the chemical bonds of modified Conocarpus and Palnoia adsorbents

An FTIR device was employed to determine the functional groups present in each adsorbent. Table 2 presents the relationship between the wavelength and the type of FTIR chemical bonds.

Table 2-Relationship between the wavelength and the type of chemical bonds in the FTIR test

type of chemical bonds	wavelength (cm^{-1})
C-H	۳۰۰۰-۲۸۵۰
O-H	۳۵۰۰-۳۲۰۰
C-O ،C-N ،C-C	۱۳۵۰-۱۰۰۰

C=O	۱۷۴۰-۱۷۲۰

3.3. Investigation of the infrared (IR) spectrum of modified Conocarpus adsorbents

Figure 3 depicts the IR spectrum of modified Conocarpus adsorbents.

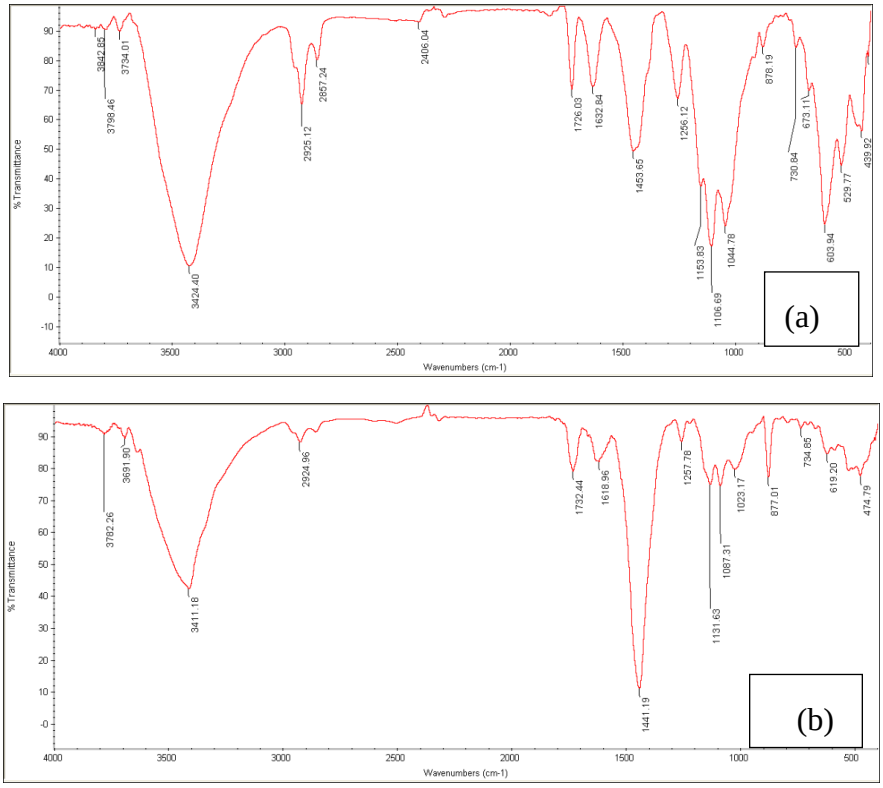

Figure3- IR spectrum associated with the modified adsorbents of (a) Conocarpus iron chloride and (b) Conocarpus potassium hydroxide 300

Figure 3(a) presents the IR spectrum of modified Conocarpus with $FeCl_2$. As would be observed, the bond vibrational frequencies are obtained as 603.94, 1044.76, 1106.169, 1453.65 and 3234.40 cm^{-1} for the vibrations of C-H bonds (alkenes), C-O, C-O, stretching aromatic vibrations of C=C, O-H and N-H, respectively.

Figure 3(b) depicts the IR spectrum of Conocarpus adsorbent modified with potassium hydroxide 600. Here, the bond vibrational frequencies have been obtained as: 1441.119 cm^{-1} corresponding to the stretching aromatic vibrations of C=C and 3418.118 cm^{-1} for those of O-H and N-H bonds.

3.4. Investigation of the infrared spectrum of modified Palnoia adsorbents

The IR spectrum of modified Palnoia adsorbents have been shown in Figure 4.

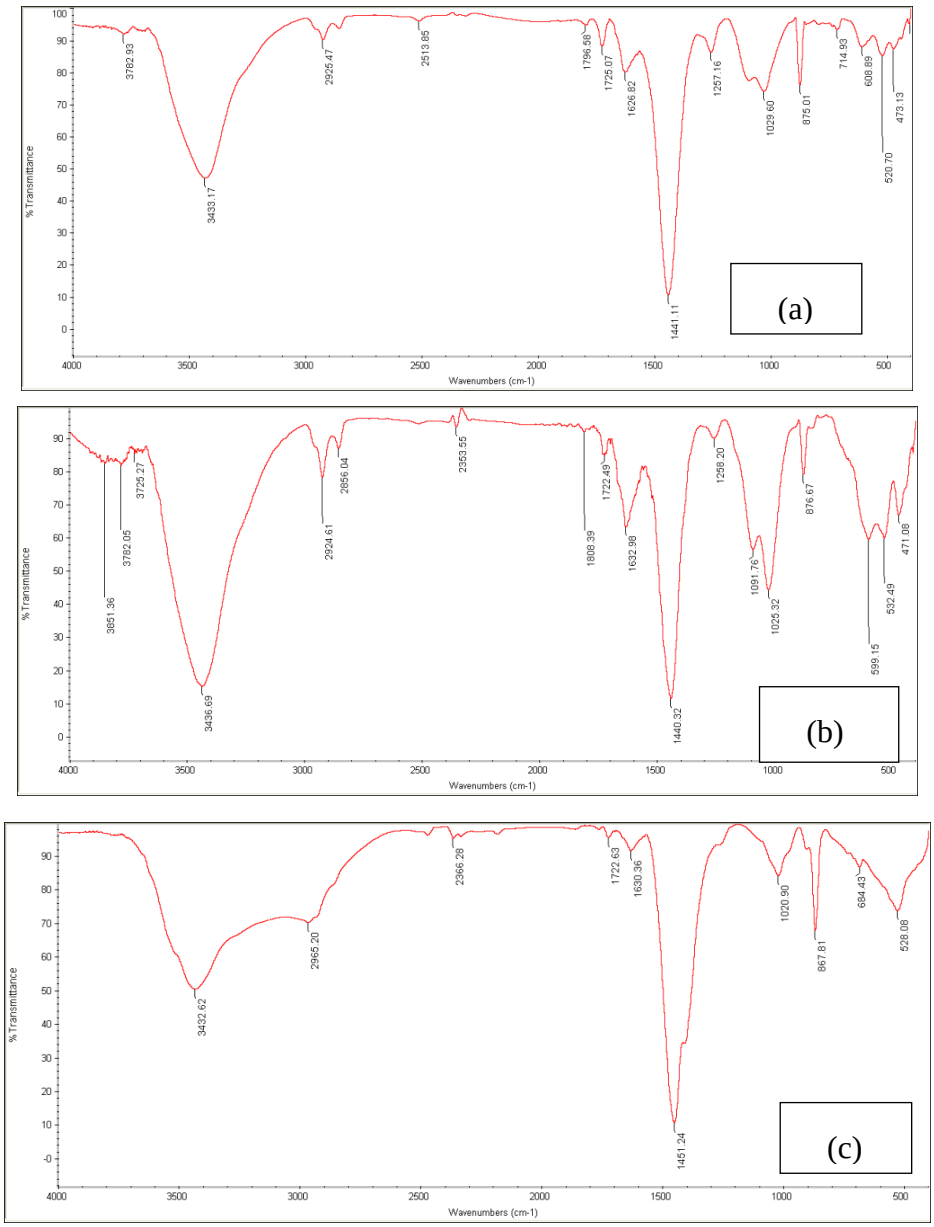

Figure4- IR spectrum corresponding to the modified absorbents of (a) Palnoia 600, (b) Palnoia iron chloride and (c) Palnoia potassium hydroxide 300

Figure 4(a) presents the IR spectrum of modified Palnoia 600. As would be observed, the bond vibrational frequencies are obtained as 875.01, 1044.76, 1029.60, 1441.11 and 3433.17 cm^{-1} for the vibrations of C-H bonds (alkenes), C-O, C-H, stretching aromatic vibrations of O-H and C-H, respectively.

Figure 4(b) depicts the IR spectrum of the modified $FeCL_2$ Palnoia adsorbent. Here, the bond vibrational frequencies have been achieved as: 1025.32 cm^{-1} corresponding to the vibrations of C-

O bonds, 1632.98 cm^{-1} for the stretching aromatic vibrations of C=C and C=N, 2924.61 cm^{-1} for the vibrations of O-H and C-H bonds and 3436.69 cm^{-1} corresponding to the stretching vibrations of O-H bonds.

The IR spectrum of Palnoia adsorbent modified with potassium hydroxide 300 has been plotted in Figure 4(c). The wave frequencies of 603.94, 1044.76 and 1106.69 cm^{-1} were found to be associated with the vibrations of C-H (alkenes), C-O and C-O bonds, respectively, while those of 1453.65 and 3424.40 cm^{-1} were related to the stretching vibrations of C=C, O-H and N-H bonds, respectively.

3.5. The results of determining the specific surface area of modified Conocarpus and Palnoia adsorbents

The specific surface area of the adsorbent examined in this study has been calculated using the methylene blue adsorption method and presented in Table 3.

Table 3- specific surface area of study adsorbent

Adsorbent	specific surface area(m^2/g)
$Fecl_2$ conocarpus	85.1
Potasium hydroxid 600 conocarpus	102
palonia 600	93.4
Potassium hydroxid 300 palonia	77.3
$Fecl_2$ palonia	99.8

3.6. Data determination for model training and testing

The results corresponding to the optimization of parameter values are listed in Table 4. Based on the obtained results, RF model with Bagging algorithm was employed with the parameters of the minimum size of nodes, number of trees and number of variables used equal to 2, 10 and 3, respectively. SVM model was trained with RBF kernel function as well as Box Constraint and ε parameters equal to 2.2156 and 0.08691, respectively, along with the standardization of input and output data of the model. Furthermore, LS-SVM model was trained with RBF kernel function and adjustment parameters of 31.8160 and 19.7891, respectively.

Table 4- Optimized parameters of Random Forest ،SVM and LS-SVM models

model	Parameter	Range of search		Optimize parameter
		minimum	maximum	
Random Forest	Minimum Leaf Size	1	43	2
	Number of Learners	10	500	10
	Number of Predictors to Sample	1	3	3
SVM	Box Constrait	0.001	1000	2.2156
	Epsilon	0.025861	2586.1351	0.08691
	Standardize Data	true	false	true
LS-SVM	Regularization Parameter	1	40	31.8160
	Parameter Kernel Function	1	40	19.7891

The results of evaluating the effectiveness of RF, SVM and LSSVM models in the training and testing phases using RMSE, R^2, MSE and MAE statistical parameters are presented in Table 5. According to the obtained results, based on the error criteria of RMSE, MSE and MAE, RF and SVM models have less error in the training and test phases compared to the LS-SVM counterpart. The values of these parameters for the above models have a smaller difference than the LS-SVM model. Based on the criterion of the coefficient of determination, the examined models in the training phase have been found to have R^2 values close to each other. However, in the testing phase, a significant difference has been observed between the R^2 values of RF and SVM models with LS-SVM one. Among the investigated models, RF model has shown a better performance in estimating the adsorption values compared to other ones in the training and test stages, which is mainly due to its lower MAE value.

Table 5- Efficiency of Random Forest, SVM and LSSVM models in training and testing stages

Model	Training phase				Test phase			
	RMSE	R^2	MSE	MAE	RMSE	R^2	MSE	MAE
RF_Optimized_Bysian_BaggedTrees	7.69	0.93	59.21	4.91	4.05	0.97	16.41	2.30
SVM_Cubic_Optimized_Bysian	8.66	0.91	75.067	5.65	3.94	0.97	15.52	2.65
LSSVM_Optimized_Bysian_RBF	17.09	0.89	292.35	11.05	10.37	0.81	107.64	6.65

The linear regression between the measured and the estimated adsorption values due to RF, SVM and LS-SVM models in the testing phase is shown in Figures 5, 6 and 7, respectively. In these figures, the regression equation between the estimated and the observed adsorption values as well as the obtained R^2 statistic, is presented. According to the present achievements, RF and SVM models have shown a similar performance in terms of higher correlation between the observed and the estimated adsorption values. The LS-SVM model has been found to have less correlation between the observations and the estimated values.

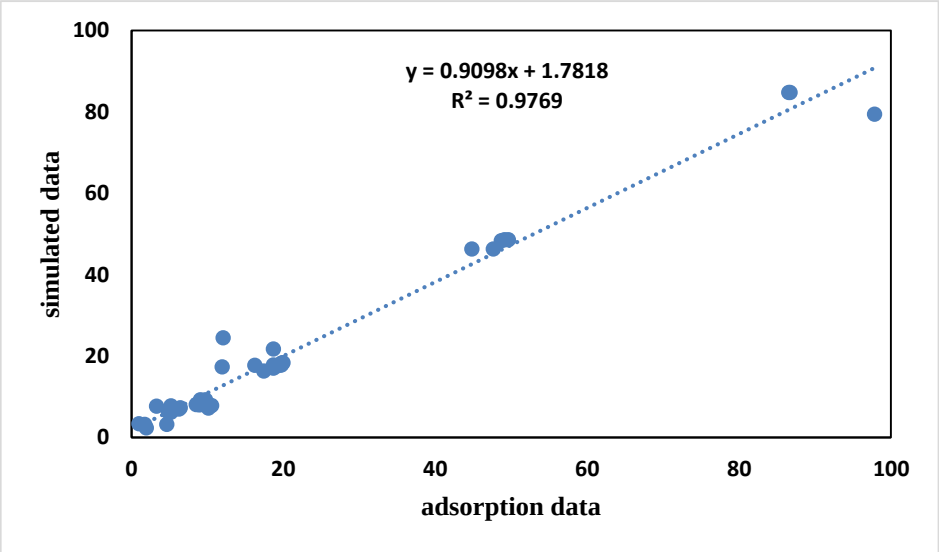

Figure 5- Linear regression between the measured and the estimated adsorption values due to RF

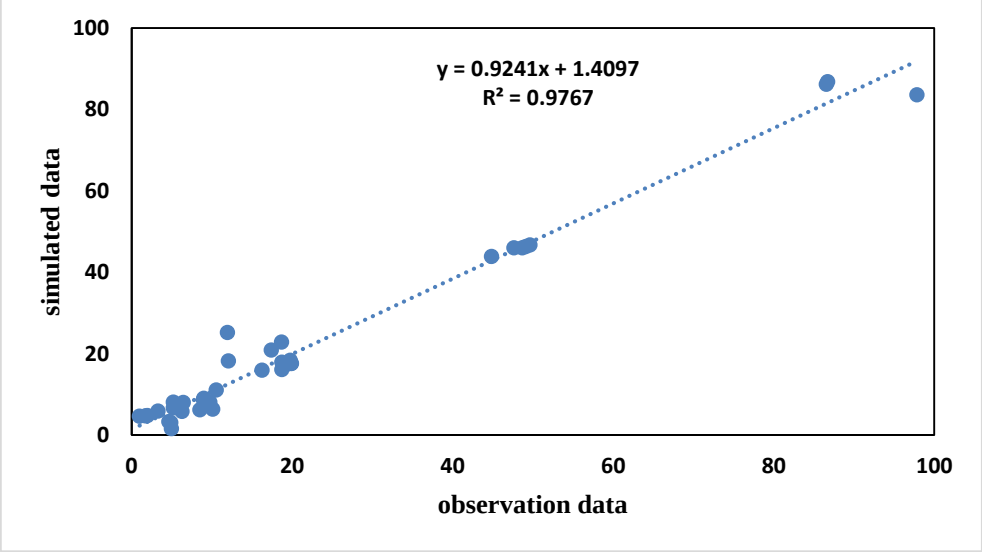

Figure 6- Linear regression between the measured and the estimated adsorption values due to SVM

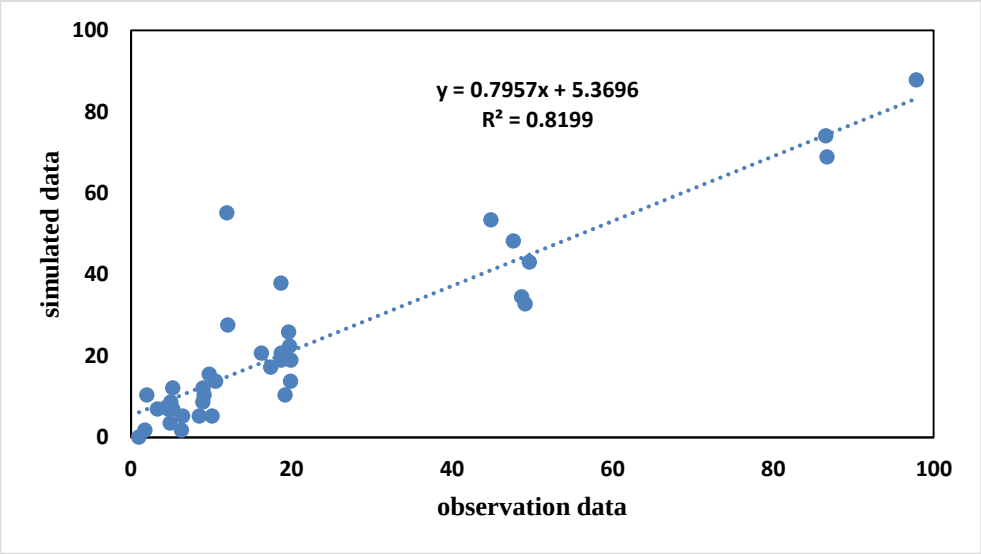

Figure 7-Linear regression between the measured and the estimated adsorption values due to LS-SVM

The measured and estimated adsorption values in the testing phase by means of RF, SVM and LS-SVM models are compared with each other in Figures 8, 9 and 10, respectively. The comparison of the results for RF model shows that this model can successfully predict the adsorption amount. It can be observed that among the examined models, LS-SVM has led to a greater difference between the measured and the estimated adsorption values. In this sense, the performance of the SVM model has been better than the LS-SVM model and close to the RF model. Hence, the performance of the SVM model is better than that of LS-SVM and is closer to that of the RF model.

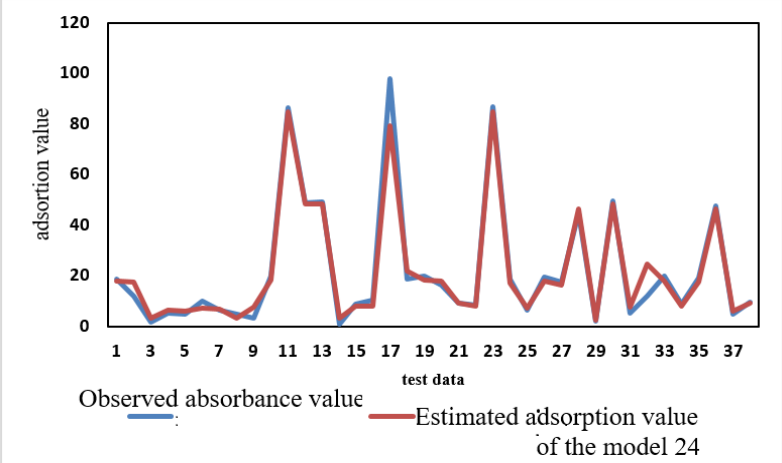

Figure 8- Comparison between the observed adsorption values and those estimated via RF model

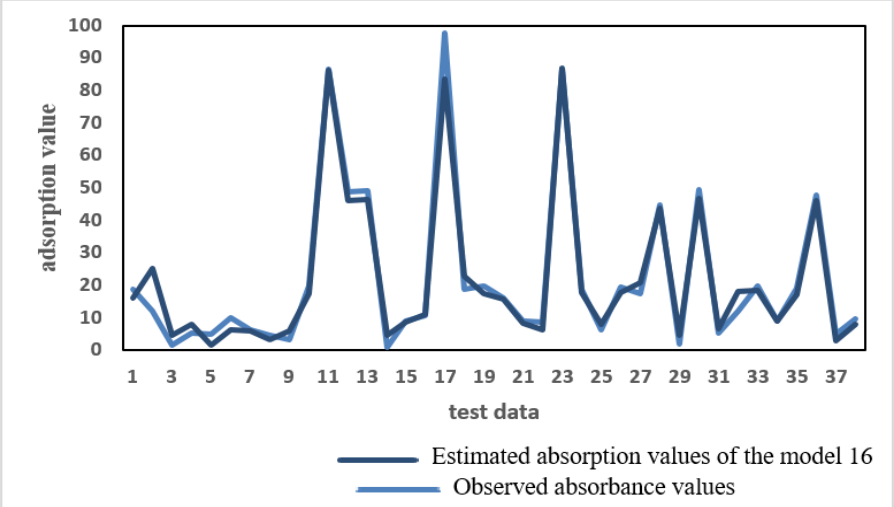

Figure 9- Comparison between the observed adsorption values and those estimated via SVM model

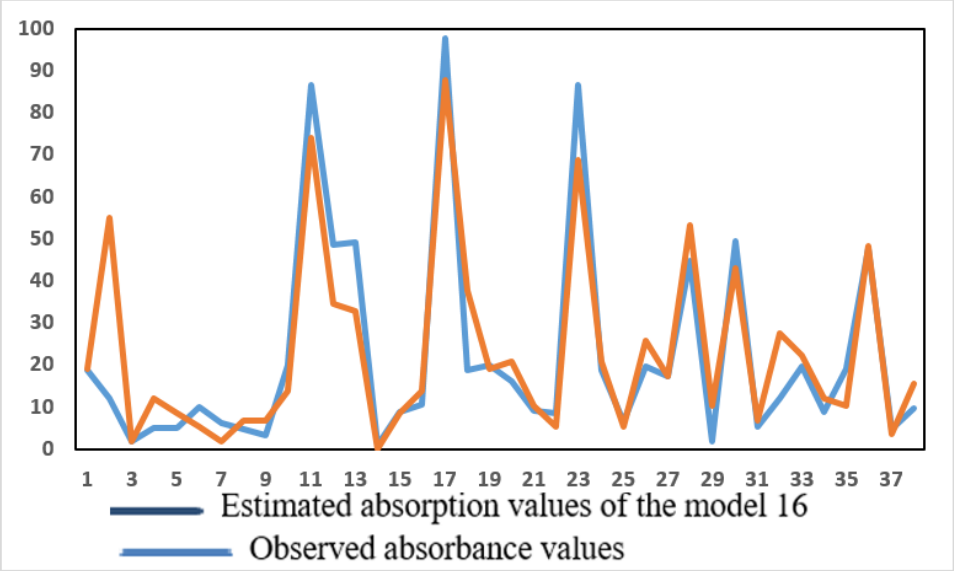

Figure 10- Comparison between the observed adsorption values and those estimated via the LS-SVM model

3.7. Results of uncertainty, reliability and resilience analyses

A three-dimensional comparison has been made for each model with respect to the observed data set (indicated by label on the horizontal axis) and shown in Figure 11 as a Taylor plot. The Taylor diagram is a single chart that summarizes several results evaluation indices. In this diagram, the predicting performance of the models is visually displayed on the polar plot by comparing with the actual values (Taylor, 2001). The reference point shows the observed values located on the horizontal axis (standard deviation). Also, the azimuth angle of the graph represents the correlation coefficient of the actual and predicted values. Moreover, the radial distance from the reference point describes the normalized standard deviation of the predicted values from the actual ones. Each point in this diagram shows the accuracy of the corresponding model, and the closer the model is to the reference point, the more accurate it is. Also, the minimum value of RMSD in the test phase has been obtained as 0.99, 0.99, 0.99, which is related to SVM, RF and LS-SVM model, respectively. As would be observed from Figure 11, SVM and RF models have met the highest correlation coefficient compared to LS-SVM model.

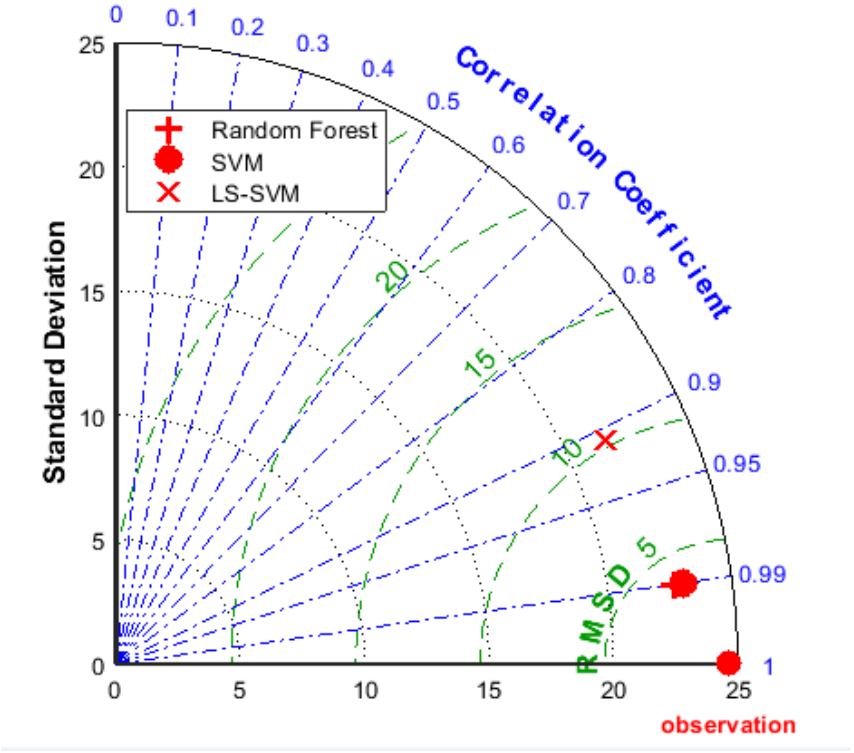

Figure 11-Taylor diagram, correlation coefficients of the examined models with respect to the actual values

Biochar can reduce the availability of nutrients except for metals, leading to calcium, phosphorus and nitrogen deficiency in plants or even reduced competition with cations for metal uptake. Moreover, biochar and vermicompost not only improve plant growth and yield, but also reduce yield through the negative impact of agriculture on water quality (Ebrahimi et al., 2021).

Although biochar compost did not improve copper amendment in temperate vineyard soils, it provided ecosystem services by increasing microbial abundance and functional activity (McLay et al. 2015). Careful selection of plants, soil modification, and cultivation conditions. These are the main factors that determine the success of phytoremediation.

Chen et al (2023) investigated the positive effects of biochar application on Rhizophagus irregularis, rice seedlings, and phosphorus cycling in paddy soil. This study investigated the effects of biochar and Rhizophagus irregularis inoculation on the organic and inorganic P contents and phosphatase activity of paddy soil, rice seedling growth, and AMF colonization ability. Compared with the control, biochar enhanced the percentage spore germination at day 7, hyphal length, most probable number, and mycorrhizal colonization rate of R. irregularis by, on average, 32%, 662%, 70%, and 28%, respectively. Biochar and R. irregularis altered soil P cycling and transformed soil P availability. Biochar and R. irregularis, either individually or in combination, increased the soil AP content by 2%–48%. Rice seedlings treated with biochar and R. irregularis produced greater biomass and showed improved root morphology and higher nutrient uptake compared with those of the control. The results suggest that the combined application of biochar and R. irregularis is beneficial for rice cultivation in paddy soils with high total P but low AP content.

Li et al(2023) investigated biochar preparation and evaluation of its effect in the composting mechanism: A review. This article provides an overview of biochar application for organic waste co-composting and its biochemical transformation mechanism. As a composting amendment, biochar works in the adsorption of nutrients, the retention of oxygen and water, and the promotion of electron transfer. These functions serve the microorganisms (physical support of the niche) and determine changes in community structure beyond the succession of composing primary microorganisms. Biochar mediates resistance genes, mobile gene elements, and biochemical metabolic activities of organic matter degradation. The participation of biochar enriched the α-diversity of microbial communities at all stages of composting, and ultimately reflects the high γ-diversity. Finally, easy and convincing biochar preparation methods and characteristics need to be explored; in turn, the mechanism of biochar on composting microbes at the microscopic level can be studied in depth.

4.Conclusion

According to the obtained results, RF and SVM models were found to have similar performance in terms of higher correlation (R^2=0.98) between the observed and the estimated adsorption values. Among the studied models, LS-SVM has met a lower correlation (R^2=0.82) between the adsorption values. The minimum values of RMSD in the testing phase were obtained as 0.99, 0.99 and 0.99, which corresponded to SVM, RF and LS-SVM models, respectively. Further to these, SVM and RF models provided the highest correlation coefficient compared to the LSSVM one. The results of the Taylor diagram indicated that RF and SVM had a higher correlation between the observed and the estimated data. Therefore, based on the obtained results, it can be concluded that RF model is more consistent with the observed data and predicts the nitrate variation well.

Acknowledgment

Gonbad Kavous University financially supports this research with registration number 269/6. We thank the university research department for providing suitable conditions for conducting this research.

Statements and Declarations

The authors declare that no funds, grants, or other support were received during the preparation of this

manuscript.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Funding

It was not funding

Author Contributions

Masumeh Farasati**:** Methodology, writing – original draft, Perform Experiment.

Seyed Morteza Seyedian**:** review & editing.

Seyed Javad Sajadi: Data analysis

References

Acikalin, K., Karaca, F., Bolat, E. 2012. Pyrolysis of pistachio shell: Effects of pyrolysis conditions and analysis of products. Fuel 95: 169-177.

Breiman, L .1984. Classification and regression trees CA. Wadsworth International Groups.

Chan, K.Y., Xu, Z. 2009. Biochar: Nutrient properties and their enhancement. PP. 67-84. In: Biochar for Environmental Management Science and Technology, Earthscan Publishing, UK.

Chen, Y., Wen, Z., Meng, J., Liu, Z., Wei, J., Liu, X., Ge, Z., Dai, W., Lin, L., Chen, W. The positive effects of biochar application on Rhizophagus irregularis, rice seedlings, and phosphorus cycling in paddy soil. Journal of Pedosphere.

Ebrahimi, M., Souri, M., Mousavi, A., Sahebani, N. 2021. Biochar and vermicompost improve growth and physiological traits of eggplant (Solanum melongena L.) under deficit irrigation. Chemical and Biological Technologies in Agriculture 8(1):19.

Farasati, M., BoroomandNasab, S., abedi kopaei, J., Jafarzadeh, N., Moazed, H., Seydian, M.2018. Investigating the effect of reed nanostructure and sugarcane pulp on nitrate removal from polluted water, Irrigation and Drainage Ph.D thesie, Shahid Chamran University of Ahvaz.

Farasati, M., Seyedian, M., Daab, K. 2021. Evaporation modeling from the free surface of water using support vector machine and least squares support vector machine. Iranian Journal of Irrigation and Water Engineering, 11(43).

Guo, X., Liu, H., Wu, S. 2019. Humic substances developed during organic waste composting: Formation mechanisms, structural properties, and agronomic functions. Science of the Total Environment. 662:501-510.

Iran Standard and Industrial Research Institute. 1388. Drinking water - characteristic of physical and spiritual theory, fifth revision.

Lian, F., Huang, F., Chen, W., Xing, B., Zhu, L. 2011. Sorption of apolar and polar organi contaminants by waste tire rubber and its chars in single- and bi-solute systems. Environ. Pollution 159: 850-857.

Li, Y., Kumar, M., Sindhu, R., Binod, P., Zhang, Z., Taherzadeh, J. 2023. Biochar preparation and evaluation of its effect in composting mechanism: A review. Bioresource Technology. 384, 129329.

Manju, G.N., Raji, C., Anirudhan, T.S. 1998. Evaluation of coconut husk carbon for the removal of arsenic from water. Water Res. 32 (10): 3062–3070.

Mc lay, A. 2001. Predicting groundwater nitrat concent-rations in region of mixed agricultura landuse. Env-iron Pollut 115: 191-204.

Pourmohammad, P., Farasati, M., Farhadi, B., Pirsaheb, M., 2018. Cadmium Removal and Recovery from Aqueous Solution Using Conocarpus Nanostructure. Journal of Water and Wastewater, 29. (a).

Sayadi, M., Farasati, M., Gharehmahmudlo, M., Rostamicherati, F. 2018. Nitrate removal using modified plant-based adsorbent palonia.

Sayadi, M., Farasati, M., Gharehmahmudlo, M., Rostami cherati, F. 2020. Removal of Nitrate, Ammonium and Phosphate from Water Using Conocarpus and Paulownia Plant Biochar. Iran. J. Chem. Chem. Eng.

Simkovic, I. Laszlo, J.A., Thompson, A.R. 1996. Preparation of a weakly basic ion exchanger by crosslynking starch with epichlorohydrin in the presence ofNH4OH. Carbohydrate Polymer 30: 25–30.

Taylor, K. 2001. Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research Atmospheres 106(D7):7183-7192.

Vapnik, V.N. 1998. Statistical Learning Theory. Wiley. New York.

Yao, Y., Gao, B., Chen, J., Yang, L. 2013. Engineered biochar reclaiming phosphate from aqueous solutions: mechanisms and potential application as slow-release fertilizer. Environmental Science andTechnology. 47: 8700-8708.