Appendix A:

Highlights:

- A combination approach has been proposed for modeling dye removal by electro-oxidation.

- Computational intelligence techniques were compared to propose the best techniques.

- A new IM has been utilized for generating more data from EDA in five dimension space.

- RO7 dye removal process has been effectively modeled and predicted by proposed approach.

- The interpretable relation between dye removal and pH, C, I, and Time has been presented.

- This approach has the potential to model various pollutants removal from wastewater.

Appendix B:

Table B.1
A brief review on recent years studies based on their advantage and disadvantage

Research	Study area	Subject	Algorithms	Advantages of study	Disadvantage of study
Song et al. (2020)	Wastewater treatment	Investigation N2O production mechanism in activated sludge tanks	RF	1. Using data mining algorithm	1. Not using hybrid algorithms 2. Not using different data mining methods 3. Not presenting interpretable relation 4. Lack of optimal treatment conditions prediction 5. Not using sufficient data
Ashrafi et al. (2020)	Wastewater treatment	Modeling sorption of Pb^{2+} ions from aqueous solutions onto modified walnut shell	ANN, MLR	1. Using data mining algorithm Presenting interpretable relation	1. Not using hybrid algorithms 2. Not using different data mining methods 3. Not presenting interpretable relation 4. Lack of optimal treatment conditions prediction 5. Not using sufficient data
Farzin et al. (2020)	Wastewater treatment	Electrochemical removal of Ciprofloxacin (CIP)	ANN, ANFIS, hybrid of LSSVM and FFA	1. Using different data mining algorithms 2. Using sufficient data	1. Not using of state of the art optimization algorithms 2. Not presenting interpretable relation 3. Lack of optimal treatment conditions prediction
Wu et al. (2021)	Wastewater treatment	Removal of arsenide from aqueous solutions using mesoporous CoFe2O4/ grapheme oxide nanocomposites	ANN, RBF-NN, and RF	1. Using different data mining algorithms 2. Prediction optimal treatment conditions	1. Not using of state of the art optimization algorithms 2. Not presenting interpretable relation 3. Not using sufficient data
Caglar Gencosman and Eker Sanli (2021)	Wastewater treatment	Polycyclic Aromatic Hydrocarbons (PAHs) removal from Wastewater	ANN, KNN, SVM, decision tree, RF	1. Using different data mining algorithms	1. Not using hybrid algorithms 2. Not presenting interpretable relation 3. Lack of optimal treatment conditions prediction 4. Not using sufficient data
K et al. (2021)	Wastewater treatment	Treatment of textile industry wastewater	ANN, and ANFIS	1. Using data mining algorithm	1. Not using different data mining 2. Not using hybrid algorithms 3. Not using sufficient data

Appendix C: Literature review about interpolation method

These interpolation methods were used in various fields such as air, water and soil for estimation of reference evapotranspiration in the hydrologic cycle (Hodam et al., 2017),the spatial distribution of soil organic carbon (Bhunia et al., 2018), air pollution mapping (Shukla et al., 2020) and investigation of ground-level $PM_{2.5}$ concentration (Zhang et al., 2021).

Appendix D:

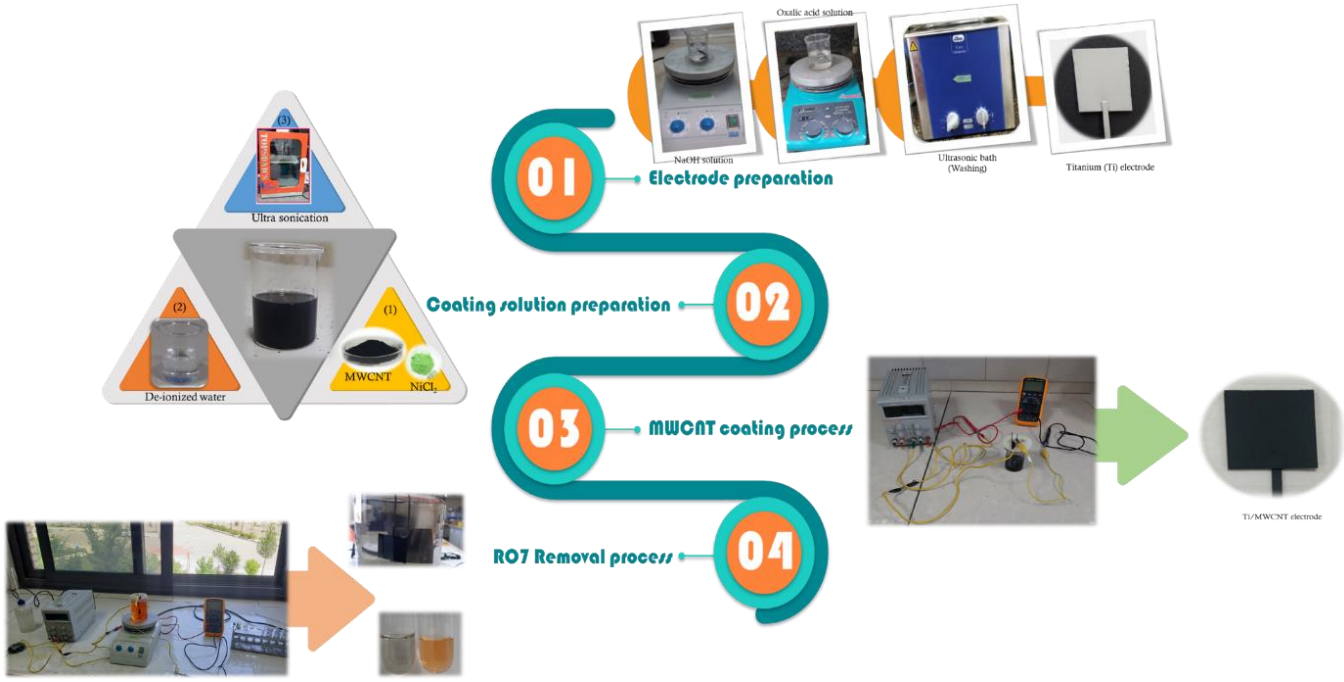


Figure D.1: Steps of preparing electrode, preparing coating solution, coating process, and removal process for RO7 electrochemical removal on Ti/MWCNT anode.

Table D.1
The original data and their statistical characteristics

Row	pH	C (g/L)	I (mA/cm^2)	Time (min)	DE (%)
1	7	1.5	3	10	44.61
2	9	2	2	15	19.54
3	5	1	4	5	36.73
4	7	1.5	3	0	0
5	5	1	4	15	76.84
6	7	1.5	3	10	42.9
7	5	2	4	15	81.12
8	7	1.5	1	10	15
9	9	1	2	5	12.58
10	7	1.5	5	10	50.5

11	7	1.5	3	10	44.96
12	7	1.5	3	10	39.31
13	5	2	2	5	14.08
14	5	1	2	5	9.75
15	7	1.5	3	10	46.92
16	9	1	4	5	11.82
17	5	1	2	15	58.09
18	3	1.5	3	10	59.11
19	7	1.5	3	10	46.4
20	7	0.5	3	10	22.77
21	7	2.5	3	10	33.91
22	9	2	4	5	28.26
23	7	1.5	3	10	45.07
24	9	1	2	15	38.14
25	7	1.5	3	20	64.56
26	9	1	4	15	36.86
27	9	2	2	5	9.34
28	9	2	4	15	32.46
29	5	2	2	15	53.72
30	11	1.5	3	10	10.56
31	5	2	4	5	57.13
Min	3.000	0.500	1.000	0.000	0.000
Max	11.000	2.500	5.000	20.000	81.120
Average	7.000	1.500	3.000	10.000	35.099
Standard deviation	1.960	0.490	0.980	4.899	22.354

Appendix E: Evaluation criteria

$$MAE = \frac{\sum_{i=1}^{n} \left| Y^{i}_{mod} - Y^{i}_{obs} \right|}{N} \tag{1}$$

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{\left(Y^{i}_{\mathrm{mod}} - Y^{i}_{Obs} \right)}{N}} \tag{2}$$

$$R = \left(\frac{\sum_{i=1}^{N} \left(Y^{i}_{\mathrm{mod}} - \bar{Y}_{\mathrm{mod}} \right) \left(Y^{i}_{obs} - \bar{Y}_{obs} \right)}{\sqrt{\sum_{i=1}^{N} \left(Y^{i}_{\mathrm{mod}} - \bar{Y}_{\mathrm{mod}} \right)^{2} \sum_{i=1}^{N} \left(Y^{i}_{obs} - \bar{Y}_{obs} \right)^{2}}} \right) \tag{3}$$

$$R^{2} = 1 - \frac{\sum_{i=1}^{N} \left(Y^{i}_{obs} - \bar{Y}_{\mathrm{mod}} \right)^{2}}{\sum_{i=1}^{N} \left(Y^{i}_{obs} - \bar{Y}_{obs} \right)^{2}} \tag{4}$$

In which, Y_{mod} is modeled output, Y_{obs} is observed or experimental output, $\bar{Y}_{\mathrm{mod}}$ is average of modeled output, $\bar{Y}_{obs}$ is observed output, and N is number of data.

Appendix F: Multivariate Linear Regression (MLR)

MLR is a multivariate data mining technique used to estimate the relationship between input variables (independent variable) and output variable (dependent variable). The main formula in MLR is written as follows:

$$Y_{\mathrm{mod}} = W_i X_i + b_i \tag{5}$$

In which, W_i, X_i, b_i and Y_{mod} are i$^{\mathrm{th}}$ weight, i$^{\mathrm{th}}$ input variables, random error, and modeled output, respectively. The W and b are obtained by least square method and minimizing the sum of square error between the observed and modeled output (Albergaria et al., 2014; Sousa et al., 2006).

Appendix G: Ridge Regression (RR)

RR is similar to MLR with this difference in that RR minimizes the weights of the input variables in addition to the sum of squares error. RR finds the W and b by minimizing follows function:

$$Min: \quad \sum_{i=1}^{N}\left(Y_{\mathrm{mod}} - W_i X_i - b_i\right)^2 + \gamma \sum_{i=1}^{n} W_i \tag{6}$$

Here, γ or gamma is the penalty coefficient, the first term denotes the sum of square error, and the second term is the sum of square weights. Considering these two terms leads to avoiding over fitting and increasing sensitivity to input values for training.

Appendix H: Multiple Nonlinear Regression (MNLR)

MNLR is one of the nonlinear data mining techniques that has a simple structure that is used for solving nonlinear regression problems. The main relation of this model defined as follows (Kahraman and Şimşek, 2020):

$$Y_{\text{mod}} = b_0 + \sum_{i=1}^{n} W_i X_i + \sum_{i=1}^{n} W_{ij} X_i X_j + \sum_{i=1}^{n} W_{ii} X_i^2 \quad (7)$$

Here, Y_{mod} is modeled output, X_i is input, b_0 is bias, W_i is linear weight, W_{ij} is interactive weight, and W_{ii} is nonlinear weight. W_i and b_0 are computed by iterative reweighted least squares algorithm (Holland and Welsch, 1977).

Appendix I: Artificial Neural Network (ANN)

ANN is a simple model of the complex neural network of human brain. This algorithm uses several numbers of neurons in one, or multi its hidden layer, for processing inputs data, and creating a nonlinear relationship between inputs and output data. In this algorithm, the output of each layer is used as input to the next layer, and each link between neurons of one layer and other layers is equivalent to the one weight. In the process of learning ANN, the inputs to each neuron are multiple to corresponding weights, and a bias is added to its summation. Then this value is applied to a transfer function, and the output of transfer function is the output of the neuron. There are various algorithms for learning ANN, such as lbfgs, stochastic gradient descent (SGD), and stochastic gradient-based optimizer (Adam). Figure H1 illustrates the structure of two hidden layers of ANN. For more information about ANN please refers to (Deng et al., 2021).

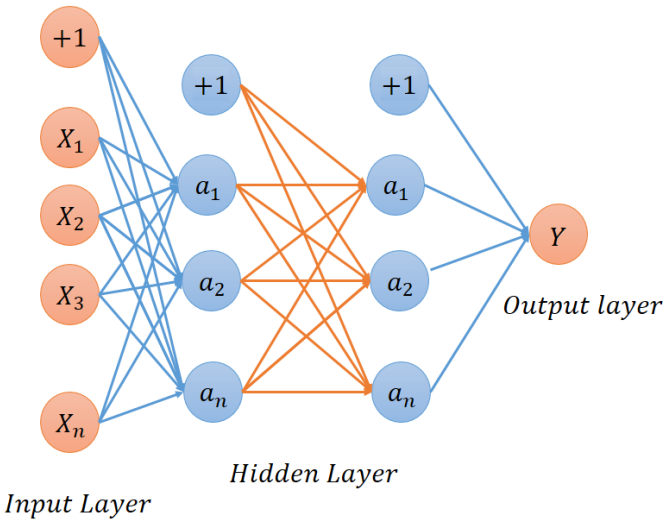


Fig. I1. The structure of ANN.

Appendix J: Classification and Regression Tree (CART)

CART is a data mining technique that is introduced by (Breiman et al., 1984). In this method, the original data is partitioned into subsamples using a binary recursive method. In recursive method first, some partitioning input variables and partitioning points are selected. Then, current data or parent node is partitioned into two inhomogeneous subsamples (children node). In each subsample, the modeled output is determined by least square method. The best partitioning input variable and partitioning point are determined based on the minimizing residual variance as follows (Breiman et al., 1984).

$$Var(v,s)=\min_{s}\left(\min_{s_1(v,s)}\left(\sum_{x_i\in s_1(v,s)}\left(y_{obs}-y_{\mathrm{mod}_1}\right)^2\right)+\min_{s_2(v,s)}\left(\sum_{x_i\in s_2(v,s)}\left(y_{obs}-y_{\mathrm{mod}_2}\right)^2\right)\right) \quad (8)$$

In which Var is input variables, s is partitioning points, x_i are inputs in child nodes, and y_{obs} is observation output. y_{mod_1}, y_{mod_2}, s_1, and s_2 are modeled output in first child node, modeled output in second child node, subsample for the first child, and subsample for the second child node, respectively. The growth of the tree (subsampling and fitting process) is down until the homogeneous divisions or terminal nodes. This method has different advantages such as insensitivity to the distribution of data, and efficiency in a wide range of data. For more details, see (Choubin et al., 2019). Figure I1 shows the scheme of the CART technique.

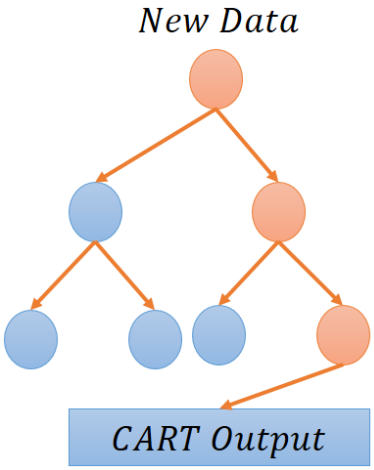


Fig. J. The scheme of CART.

Appendix K: K nearest neighbor (KNN)

KNN is a nonparametric data mining technique that used the similarity concept for estimating the nonlinear relationship between inputs and output. In this technique first, the k sample with more similarity with training data is selected. Generally, this similarity is considered based on the Euclidian distance. Then, the weighted average of the k nearest neighbors is calculated. This weight is determined based on the inverse of K nearest neighbor distance. Figure J1 presents the structure of KNN.

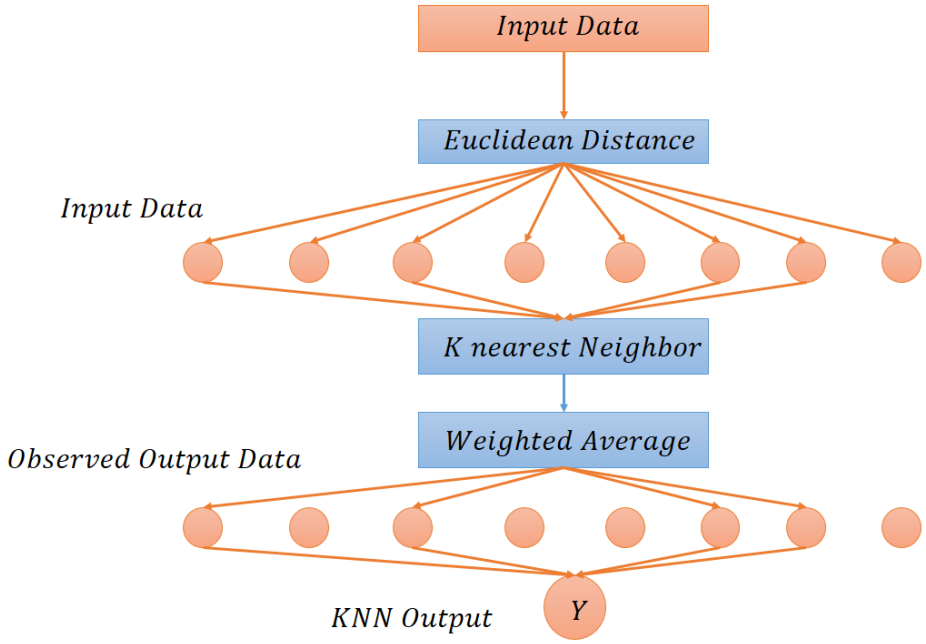


Fig. K. The KNN structure.

Appendix L: Random Forest (RF)

Generally, using one tree can be lead to overfitting, in this situation employing RF is useful. RF is a data mining technique that is proposed by (Breiman, 2001). This technique creates k random subsamples from the original data. Then, it fits a tree to each random subsample. Afterward, the average of all k tree responses is the final response of RF. The random subsampling in RF can lead to more accuracy, modeling nonlinearity problems, and as mentioned avoiding overfitting. The RF scheme is shown in Figure K1.

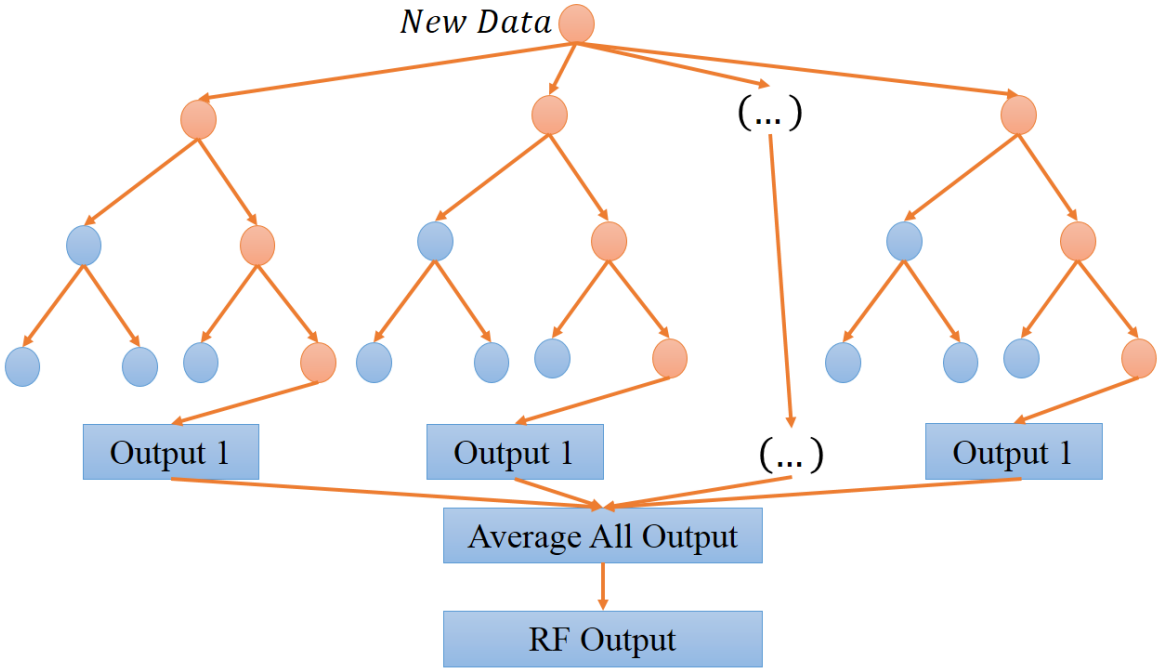


Fig. L. The RF scheme.

Appendix M: Sensitivity Analysis of Proposed Techniques

Appendix M.1 IDW and KRG Parameters

The main parameters of IDW and KRG are power and $tetha_0$, respectively. These parameters must be determined carefully, for yielding better results. Because the best interpolating method is used for generating new data, and the accuracy of the interpolating method impacts all other computation-based techniques. In order to perform sensitivity analysis for selecting the best IDW and KRG parameters, the power parameter is varied between 1 and 30 with unit intervals, as well as the $tetha_0$ parameter is considered between 0.001 to 0.1 with intervals of 0.001. The sensitivity analysis of IDW and KRG is shown in Figure L.1.1. The best value of the power parameter is equal to 28 with RMSE of 14.550 for IDW, and also the best value of tetha0 is 0.037 with RMSE of 9.179 for KRG. (see Figure L.1.1).

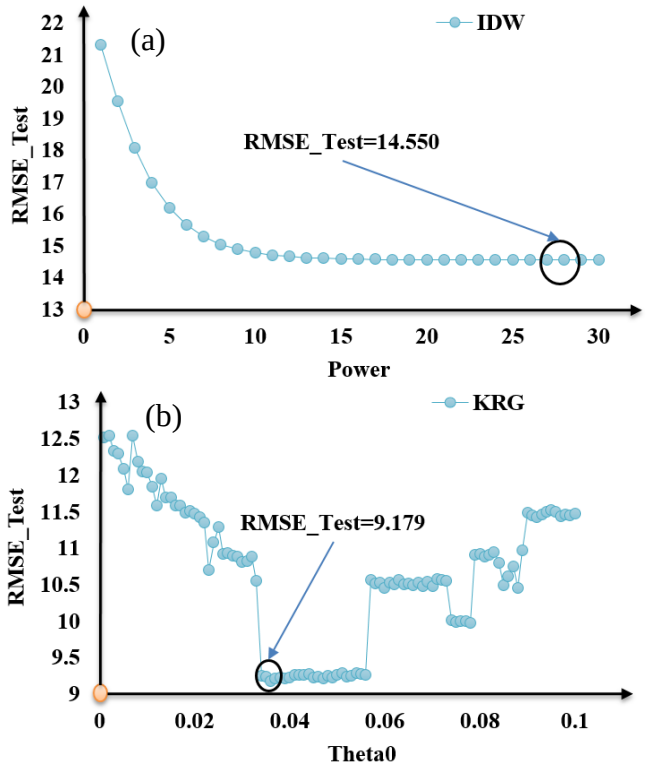


Fig. M.1.1 the sestivity analysis of interpolation methods (a) IDW and (b) KRG.

Appendix M.2: RR Parameters

The important parameter of RR technique is Gamma, this parameter for original and new data differs from 0.01 to 10 with the same intervals of 0.01. Figure L.2.1 shows the results of RR sensitivity analysis for original and new data. As can be seen in this Figure, the best value of Gamma for original data is equal to 0.02 with the RMSE value of 9.597. For new data, the minimum value of RMSE is equal to 10.347 and is related to 0.01 values of Gamma

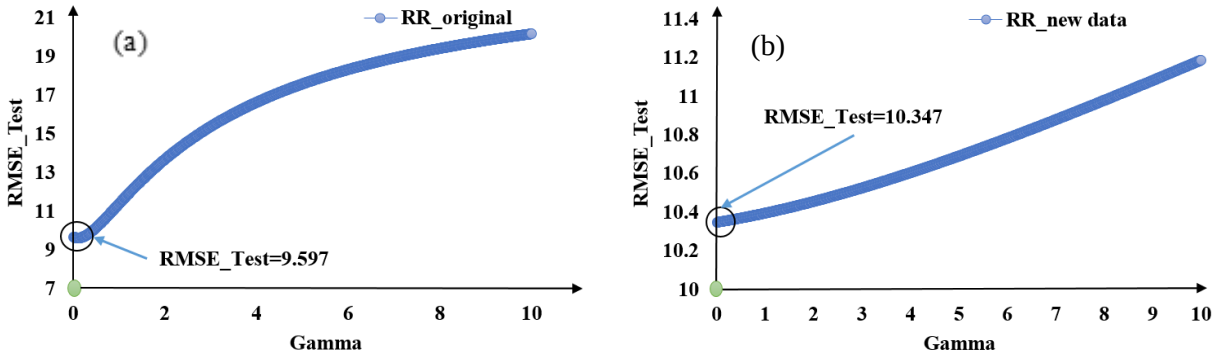
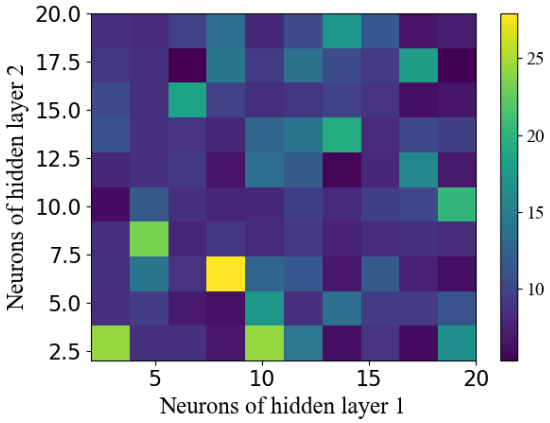


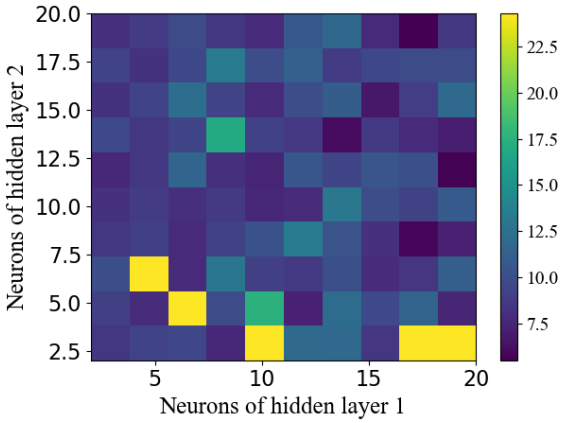
Fig. M.2.1. The sensitivity analysis of RR for (a) Original data and (b) New data.

Appendix M.3: ANN Parameters

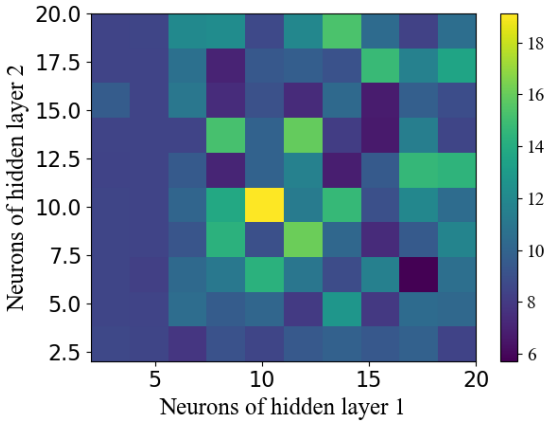
In this section, based on the sensitivity analysis, the best learning algorithm of ANN, and the best structure of ANN is selected for original and new data In Figure L.3.1, the sensitivity of RMSE in the test period with respect to the number of neurons in the first and second hidden layers is investigated by considering lbfgs, sgd, and adam learning algorithm which the neuron of hidden layers differs from 1 to 20. It is obvious from this Figure that the best value of RMSE is equal to 5.370 for original data, and it is related to lbfgs learning algorithm with 6 and 18 neurons for the first and second layers. Furthermore, the minimum RMSE for new data is equal to 1.399, for lbfgs learning algorithm with 20 and 16 neurons in the first and second layers, respectively.



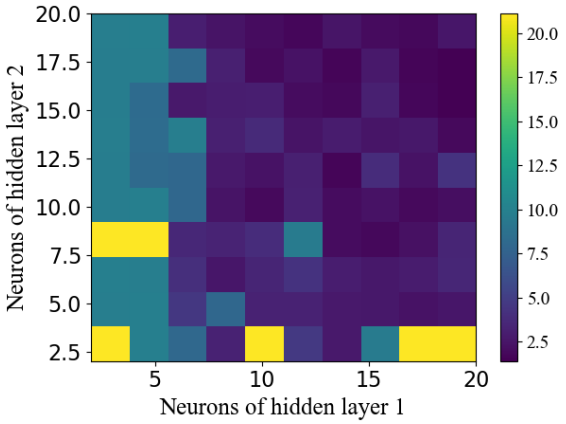
(a)



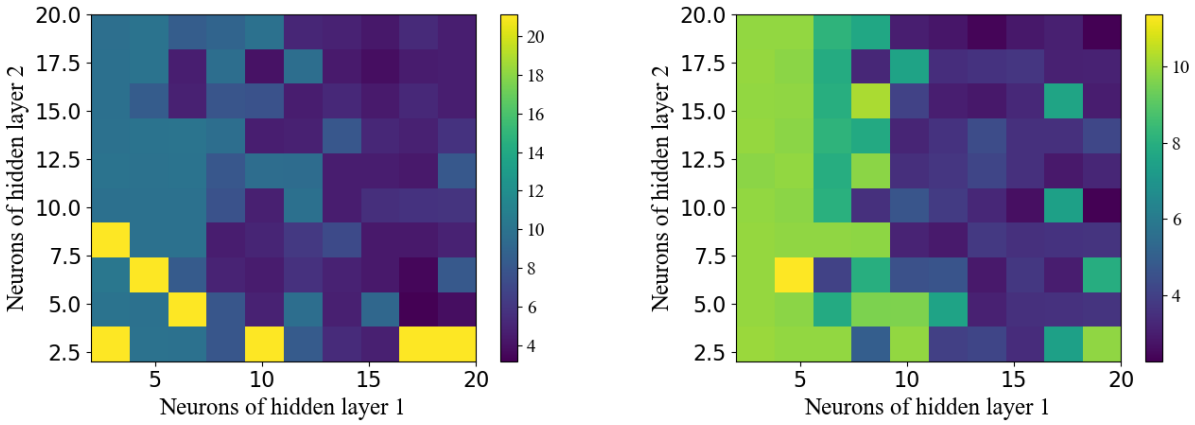
(b)



(c)



(d)



(e) (f)

Fig. M.3.1. The senstivity analysis of ANN(RMSE for (a) Original_lbfgs, (b) Original_sgd, (c) Original_adam, (d) New_lbfgs , (e) New_sgd and (f) New_adam).

Appendix M.4: KNN Parameters

The important parameter for running KNN is the number of neighbors. This parameter is varied from 1 to 10 and 1 to 45 with unit intervals for original data and new data, respectively. Figure I.4.1 illustrates the sensitivity of RMSE in the test period respect to the number of neighbors. It can be seen that the minimum value of RMSE for original data is related to neighbor number 6 and equals to 14.362. While, for new data, the minimum value of RMSE is equal to 5.049 which obtained by neighbor number 6.

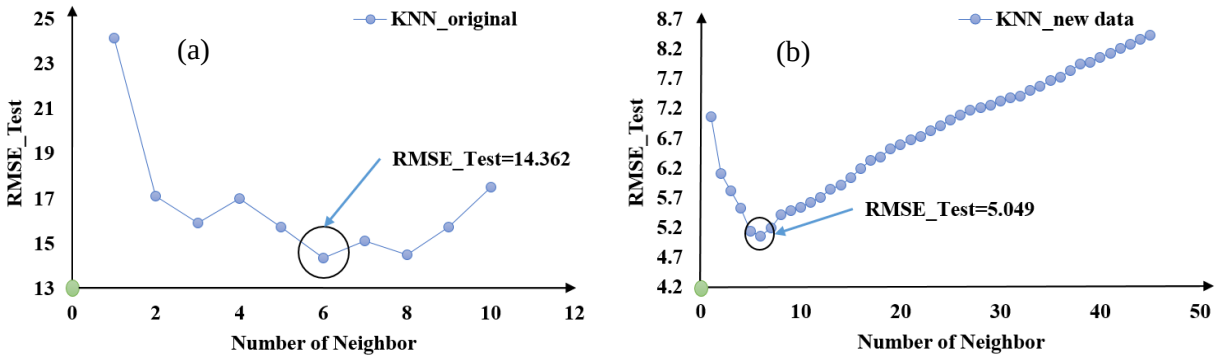
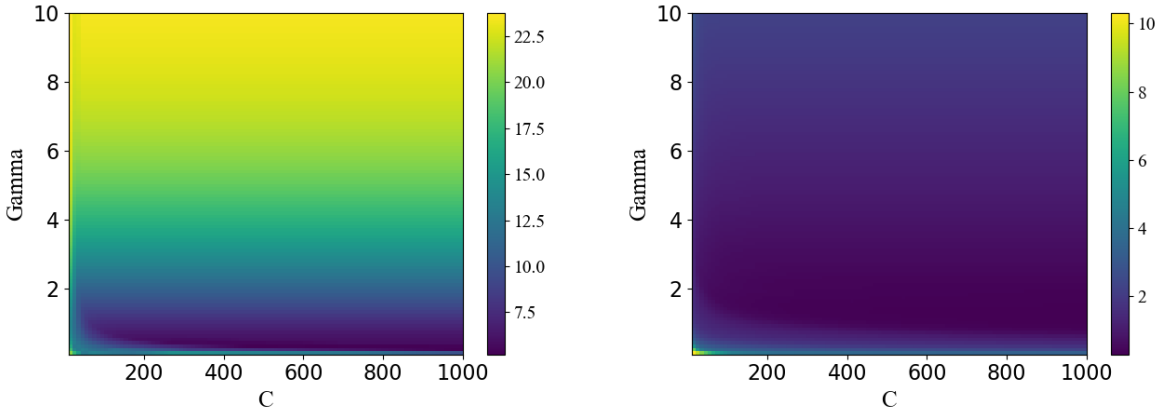


Fig. M.4.1. The sensitivity analysis of KNN for (a) Original data and (b) New data.

Appendix M.5: SVM Parameters

SVM technique has two parameters include C and Gamma for running. For sensitivity analysis, these two parameters are differed from from 0.1 to 10, and from 10 to 1000 with intervals of 0.1 and one, respectively. For this purpose, the SVM is run 10000 times and obtained 10000 RMSE values that the difference of these values with respect to C and Gamma is shown in Figure L.5 as color bar diagrams.

As seen in this Figure, by increasing the distance from the best C and Gamma (1000, and 0.2 with best RMSE_Test = 5.207), the RMSE is increased for original data. Also, the best values of C, Gamma, and correspond RMSE for new data are equal to 1000, 1.2, and 0.297, respectively, which by increasing distance from this point the RMSE is increased.



(a) (b)

Fig. M.5.1. The sensitivity analysis of SVM (RMSE for (a) Original data and (b) New data).

Appendix M.6: RF Parameters

The most important parameter of RF is the number of trees. The best value of this parameter is determined by considering different values of 50 to 1000 with intervals of 50. Figure I.6.1 shows the sensitivity analysis of this parameter for original and new data. According to this figure, the minimum values of RMSE for original and new data are 17.724, and 5.084, respectively. Thus, for the original dataset and the new dataset, the best number of trees is equal to 250, and 900, respectively.

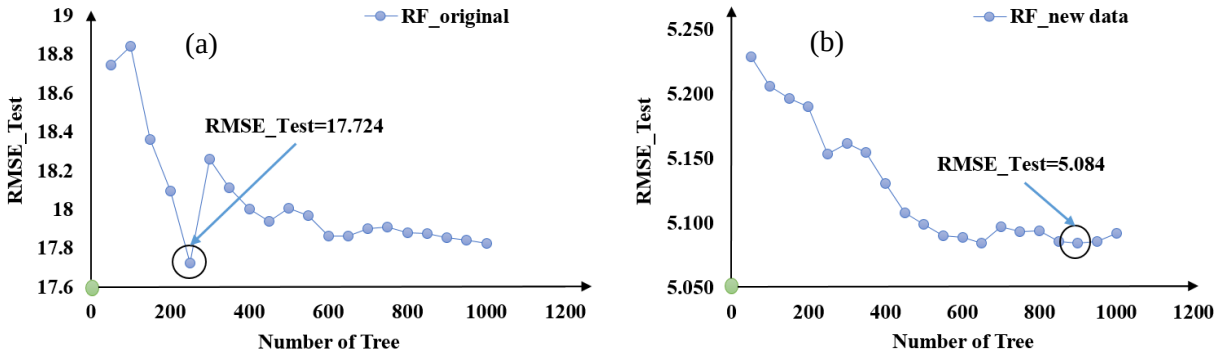


Fig. M.6.1 The sensitivity analysis of RF for (a) Original data and (b) New data.

References

Albergaria, J.T., Martins, F.G., Alvim-Ferraz, M.C.M., Delerue-Matos, C., 2014. Multiple Linear Regression and Artificial Neural Networks to Predict Time and Efficiency of Soil Vapor Extraction. Water, Air, Soil Pollut. 225, 2058. https://doi.org/10.1007/s11270-014-2058-y

Ashrafi, M., Borzuie, H., Bagherian, G., Chamjangali, M.A., Nikoofard, H., 2020. Artificial neural network and multiple linear regression for modeling sorption of Pb 2+ ions from aqueous solutions onto modified walnut shell. Sep. Sci. Technol. 55, 222–233. https://doi.org/10.1080/01496395.2019.1577437

Bhunia, G.S., Shit, P.K., Maiti, R., 2018. Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC). J. Saudi Soc. Agric. Sci. 17, 114–126.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Breiman, L., Friedman, J., Stone, C.J., Olshen, R.A., 1984. Classification and regression trees. CRC Press.

Caglar Gencosman, B., Eker Sanli, G., 2021. Prediction of Polycyclic Aromatic Hydrocarbons (PAHs) Removal from Wastewater Treatment Sludge Using Machine Learning Methods. Water, Air, Soil Pollut. 232, 87. https://doi.org/10.1007/s11270-021-05049-8

Choubin, B., Moradi, E., Golshan, M., Adamowski, J., Sajedi-Hosseini, F., Mosavi, A., 2019. An ensemble prediction of flood susceptibility using multivariate discriminant analysis, classification and regression trees, and support vector machines. Sci. Total Environ. 651, 2087–2096. https://doi.org/10.1016/j.scitotenv.2018.10.064

Deng, Y., Zhou, X., Shen, J., Xiao, G., Hong, H., Lin, H., Wu, F., Liao, B.-Q., 2021. New methods based on back propagation (BP) and radial basis function (RBF) artificial neural networks (ANNs) for predicting the occurrence of haloketones in tap water. Sci. Total Environ. 772, 145534. https://doi.org/10.1016/j.scitotenv.2021.145534

Farzin, S., Chianeh, F.N., Anaraki, M.V., Mahmoudian, F., 2020. Introducing a framework for modeling of drug electrochemical removal from wastewater based on data mining algorithms, scatter interpolation method, and multi criteria decision analysis (DID). J. Clean. Prod. 266, 122075. https://doi.org/10.1016/j.jclepro.2020.122075

Hodam, S., Sarkar, S., Marak, A.G.R., Bandyopadhyay, A., Bhadra, A., 2017. Spatial interpolation of reference evapotranspiration in India: comparison of IDW and Kriging Methods. J. Inst. Eng. Ser. A 98, 511–524.

Holland, P.W., Welsch, R.E., 1977. Robust regression using iteratively reweighted least-squares. Commun. Stat. - Theory Methods 6, 813–827. https://doi.org/10.1080/03610927708827533

K, A., Mungray, A., Agarwal, S., Ali, J., Chandra Garg, M., 2021. Performance optimisation of forward-osmosis membrane system using machine learning for the treatment of textile industry wastewater. J. Clean. Prod. 289, 125690. https://doi.org/10.1016/j.jclepro.2020.125690

Kahraman, Ö., Şimşek, İ., 2020. Color removal from denim production facility wastewater by electrochemical treatment process and optimization with regression method. J. Clean. Prod. 267, 122168. https://doi.org/10.1016/j.jclepro.2020.122168

Song, M.J., Choi, S., Bae, W. Bin, Lee, J., Han, H., Kim, D.D., Kwon, M., Myung, J., Kim, Y.M., Yoon, S., 2020. Identification of primary effecters of N2O emissions from full-scale biological nitrogen removal systems using random forest approach. Water Res. 184, 116144. https://doi.org/10.1016/j.watres.2020.116144

Sousa, S.I.V., Martins, F.G., Pereira, M.C., Alvim-Ferraz, M.C.M., 2006. Prediction of ozone concentrations in Oporto city with statistical approaches. Chemosphere 64, 1141–1149. https://doi.org/10.1016/j.chemosphere.2005.11.051

Wu, X., Hu, J., Wang, X., Xin, L., Li, C., Wei, X., 2021. Effective removal of arsenide from aqueous solutions using mesoporous CoFe2O4/graphene oxide nanocomposites assisted by artificial intelligence. Carbon Lett. https://doi.org/10.1007/s42823-021-00239-0

Zhang, H., Zhan, Y., Li, J., Chao, C.-Y., Liu, Q., Wang, C., Jia, S., Ma, L., Biswas, P., 2021. Using Kriging incorporated with wind direction to investigate ground-level PM2.5 concentration. Sci. Total Environ. 751, 141813. https://doi.org/10.1016/j.scitotenv.2020.141813