World Journal of Environmental Biosciences
Available Online at: **www.environmentaljournals.org**

Volume 12, Issue 4: 33-39

https://doi.org/10.51847/U72sgqfYRZ

ISSN 2277- 8047

Machine Learning Approaches for Prediction of Daily River Flow

Naser Shiri[1*], Sepideh Karimi[2], Jalal Shiri[2]

[1]Department of Civil Engineering, Sharif University of Technology, Tehran, Iran.
[2]Water Engineering and Science Research Institute, University of Tabriz, Tabriz, Iran.

ABSTRACT

River flow is an important parameter in hydrology, irrigation scheduling, groundwater pollution studies, and hydropower analysis. It depends on various climate and hydrologic factors, e.g. precipitation, temperature, river basin physiography, geological characteristics of the basin, etc. Although several factors may affect river flow quantity and quality during a certain period, it is difficult to account for all those variables in simulating/predicting river flow values due to the complex relations governing the hydrologic cycle in nature. Therefore, using simpler methods that can be used with fewer required input data would be necessary. A prediction task was implemented in the present study to obtain river flow values based on the previously recorded river flows using three machine learning approaches, namely, multi-variate adaptive regression spline (MARS), boosted regression tree (BT), and random forest (RF). Data from three stations in Iowa stat (U.S.A) covering daily records of five years were utilized for developing the ML models. Based on the results, all three applied models could simulate the river flow values well, when the time lags of two successive days were introduced to feed the model. An analysis was also made for detecting the variations of the applied statistical indicators per test stage of k-fold testing data assignment. This analysis showed obvious variations of indicators among the test stages, revealing the necessity of adopting k-fold testing in the studied region.

Keywords: *River flow, Random forest, MARS, Boosted regression tree*

Corresponding author: Naser Shiri
e-mail ✉ nasershiri7596@gmail.com
Received: 04 September 2023
Accepted: 11 December 2023

INTRODUCTION

Accurate prediction of streamflow records is very important in hydrology and water resources management and engineering, drought analysis (Edossa & Babel, 2011) ELGENDY (2022), environmental studies (Tennant, 1976), groundwater interactions (Gunduz & Aral, 2005) and river erosive capability (Kisi *et al*., 2012; Kaur *et al.,* 2023). There are some methods for the prediction of river flow values such as time series models (e.g. ARMA), empirical models, and physics-based models. Among the empirical group, machine learning (ML) approaches' applications have been viable due to their flexibility and capability to map nonlinear relations of the hydrologic cycle (Karimi *et al*., 2016; Çora & Çora, 2022). Substantial studies have used various ML techniques for river flow prediction worldwide including the use of neural networks, genetic programming, neuro-fuzzy systems, support vector machine, etc (e.g. Han *et al*., 2007; Latt & Wittenberg, 2014; Insom *et al.,* 2015; Maroufpoor *et al*., 2020; Shiri *et al*., 2021; Macharyulu *et al*., 2022; Shiri *et al*., 2022; Wegayehu & Muluneh, 2022; Bakhshi Ostadkalayeh, 2023; Terela & Strilets, 2023). Some others have tried to introduce coupled wavelet-ML techniques for flow predictions when data carry considerable noise and ML couldn't handle their simulations (e.g. Karimi *et al*., 2017; Dalkiliç & Hashimi, 2020; Jayavel & Sivagnanam, 2022; Yilmaz *et al*., 2022).

Despite the broad use of ML techniques in this context, there is still empty room to work on it and improve the simulation knowledge due to the different nature of the rivers and flow time series at various locations under different climatic/hydrologic conditions. Nevertheless, as the works deal with time series, information captured from each series is an essential factor in modeling success because the capacities of various models in handling different data sets vary among them. Further, tackling those series with a specified data management strategy is an important issue because some extreme events might fall within a certain part of the available events and affect the prediction accuracy in both model building (training) and validation (testing) phases. Karimi *et al*. (2016), and Mashhour *et al*. (2023) argued that using the k-fold testing mode for data assigning in the prediction of river flow values by ML might solve such difficulties considerably. Hence, a k-fold testing scenario for temporal assigning of ML techniques was adopted here to assess some methods for flow prediction.

MATERIALS AND METHODS

Study locations

The Des Moines River, which spans approximately 845 km from its farthest headwaters, is a tributary of the Mississippi River located in the upper Midwest region, U.S.A. It is the largest river passing through the Iowa state with a basin area of about 38.340 km^2. Beginning from southern Minnesota, the Des Moines River flows from northwest to southeast across the Iowa state, transitioning from the glaciated plains to the

World Journal of Environmental is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-Non Commercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms. (https://creativecommons.org/licenses/by-nc-sa/4.0/).

unglaciated hills near the capital city of Des Moines. The river then continues to flow in a southeastern direction and drains into the Mississippi River. Three stations, namely, Des Moines River at Humboldt (station 1), Des Moines River at Fort Dodge (station 2), and Des Moines River near Stratford (station 3) have been constructed along the north reach of the river for continuous monitoring of river flow characteristics **(Figure 1)**.

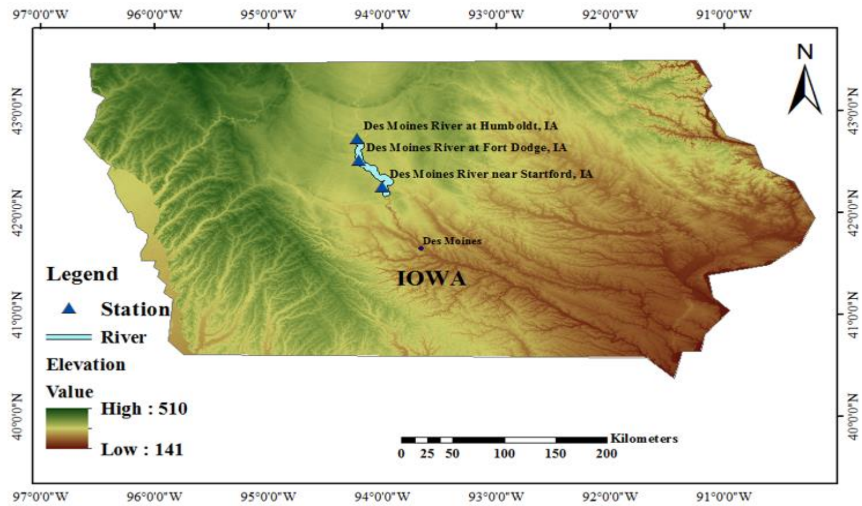


Figure 1. Geographical positions of the stations

Daily records of river flow consisting of 5-year patterns at all locations (2018 to 2023) were used, basic statistical characteristics for which have been listed in **Table 1**. The observed flow domains (differences between the maximum and minimum observed flow rate) for stations 1-3 were about, 16700, 30000, and 37000 ft^3/s, respectively. Among the three studied locations, station 3 presented the highest maximum river flow values during the study period. On the other hand, the standard deviation values of the stations were, respectively, 2048, 3603, and 4438 ft^3/s for the same sites. This clearly shows the higher variance of the flow rates in the third location, although the coefficient of variation suggests the highest dimensionless variance values for station 2. The same trend can be seen for the skewness coefficient values. So, it might be stated, based on these observations, that variations around the average flow rate values and the magnitude of discrepancy from normal distribution are higher for the second station. As a first hypothesis, this might affect the modeling performance in this location and make the simulation process difficult. However, this should be accepted or rejected after analysis of the obtained results in the next sections.

Table 1. Basic statistical characteristics of river flow time series

	station	Latitude	Longitude	Max(ft^3/s)	Min(ft^3/s)	Mean(ft^3/s)	SD(ft^3/s)	CV	Skew
station 1	Des Moines River at Humboldt, IA	42°43'10"	94°13'13"	16800	25.90	1481.98	2048.33	1.38	2.64
station 2	Des Moines River at Fort Dodge, IA	42°30'30"	94°12'12"	30900	57.70	2518.18	3603.10	1.43	3.03
station 3	Des Moines River near Stratford, IA	42°15'08.1"	93°59'50.88"	37100	85.70	3258.65	4438.28	1.36	2.74

Note: Max, min, mean, SD, CV, and Skew show the maximum, minimum, average, standard deviation, coefficient of variation, and skewness coefficient, respectively.

Applied models

Multi-variate adaptive regression alpine (MARS)

As a regression-based model, MARS (Friedman, 1991; Kryuchkova *et al*., 2022) technique applies the stepwise linear regression technique fundamentals. MARs have a high capability to identify and enhance the comprehension of complex interactions between the input and target parameters As a non-parametric model, it builds upon the linear regression model by incorporating flexibility. The general form of the MARS model Reads:

$$f(X)= \alpha_0+ \sum_{n=1}^{N} \alpha_0 h_n(X) \quad (1)$$

Where, the alpha (α) coefficients are computed by minimizing the residual errors and depend on weights (the variable importance) (Friedman & Roosen, 1995; Kisi & Parmar, 2016; Kisi *et al*., 2017; Shiri *et al*., 2020a).

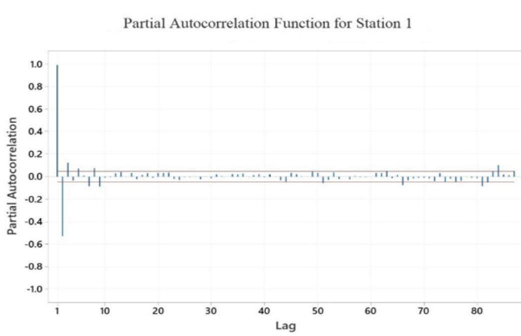
Random forest (RF)

Random Forest (RF) has been developed based on classification and regression trees (CART). The basic advantage of RF is in providing accurate predictions for high-dimensional input matrices without overfitting (Breiman, 2001; Domatskiy & Sivkova, 2022). RF starts with creating a tree through the random selection of a group of input variables to split on for each node. The best split is then computed based on training data. Second, the bagging procedure is used to resample the calibration data so that each time a new individual tree is grown (Breiman, 1996; Biau, 2012; Shiri *et al.*, 2020b; Osadchuk *et al.*, 2023). Different numbers of trees are normally evaluated to select the best RF method. Here, 150 trees (with eight cycles) were found to be optimal. The minimum child node size and the maximum number of levels were chosen as 5 and 10, respectively.

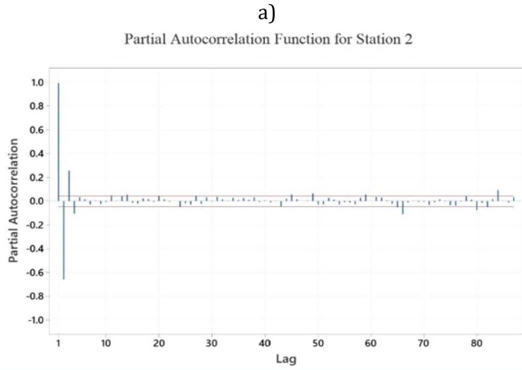
Boosted regression tree (BT)

BT incorporates a tree-based algorithm with boosting to enhance its regression-based machine learning strategy, which is an improvement over the traditional approach (Freidman *et al.*, 2000; Mekeres *et al.*, 2022). By fitting new trees to residual errors of existing trees, the boosting procedure enhances the model's accuracy. During each iteration, the existing tree remains unchanged and the optimal model is represented by a combination of linear trees (Elith *et al.*, 2008). The number of trees is automatically optimized via an internal cross-validation procedure. Control parameters, e.g. learning rate, tree complexity, and bag fraction were determined using a trial-and-error approach (França & Cabral 2015; Shiri *et al.*, 2020b; Al-Jaloud *et al.*, 2022). Various numbers of seeds for the random number generator were evaluated, and the optimal outputs were obtained for seed number = 1.

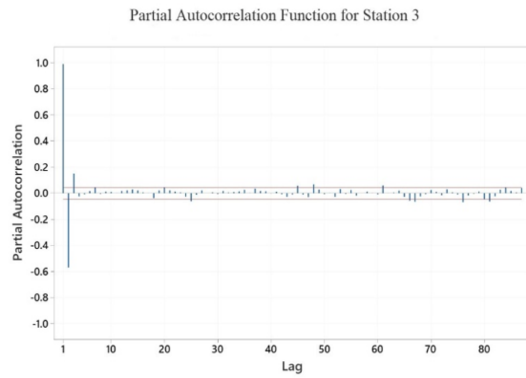
Study workflow description



a)



b)



c)

Figure 2. Partial autocorrelation function (PACF) of river flow records

The recorded river flow values were used as model inputs in this study. Hence, the time series of the recorded flow values were analyzed based on the partial autocorrelation function (PACF) to select the best time lag for feeding the models. This is a common approach in literature for identifying the suitable input parameters when time series should be simulated/predicted (e.g. Karimi *et al.*, 2016; Alexander *et al.*, 2018; Ahmed *et al.*, 2022). PACF diagrams of river flow at each location have been shown in **Figure 2**, from which, it can be observed that the first four lags presented significant values in terms of correlation among various time steps of flow records. Based on this observation, the flow values of 4 days were used as input of the applied models in a step-by-step mode, so that each time one parameter was included in the input set. **Table 2** summarizes the adopted input set for the applied models.

Table 2. The adopted input combinations

Model	Input variables
Mars 1, RF1, BT1	Q_{t-1}
Mars 2, RF2, BT2	Q_{t-1}, Q_{t-2}
Mars 3, RF3, BT3	Q_{t-1}, Q_{t-2}, Q_{t-3}
Mars 4, RF4, BT4	Q_{t-1}, Q_{t-2}, Q_{t-3}, Q_{t-4}

After constructing the input matrix, the assessing method of the models should be defined. As mentioned, the k-fold test approach was used here, where one part of available records (here, one year) was reserved each time as test patterns, and the models were trained using the remaining data. The process is then repeated for all reserved parts till all the parts can be incorporated into both the model training and testing stages. Considering that there are 5 years of daily data at each location, 5 training-testing processes were performed at each location for each model and a total of 45 processes were performed in the study.

Performance criteria

The variance accounted for (VF), the dimensionless RMSE (scatter index, SI), and the Nash and Sutcliffe coefficient (NS) were used as statistical measures of models' performance accuracy.

$$\mathrm{VAF} = \left[1 - \frac{\mathrm{Var}(Q_{io} - Q_{ie})}{\mathrm{Var}(Q_{io})}\right] \times 100 \quad (2)$$

$$\mathrm{SI} = \frac{\mathrm{RMSE}}{\overline{Q_o}} \quad (3)$$

$$\mathrm{NS} = 1 - \frac{\sum_{i=1}^{n}(Q_{io} - Q_{ie})^2}{\sum_{i=1}^{n}(Q_{io} - \overline{Q_o})^2} \quad (4)$$

Where Q_{io} and Q_{ie} define the recorded and predicted river flow, respectively. $\overline{Q_o}$ is the mean observed river flow values of n available pattern. A good model should show higher VAF and NS values and a lower SI magnitude. These measures were computed for each test year as well as for all the available patterns (complete patterns of 5 years).

RESULTS AND DISCUSSION

Error statistics of the applied ML models for all studied stations are given in **Table 3**. Since the models were assessed by k-fold testing, two sets of statistics were computed, namely, global and individual. The individual indicators were computed for each test stage, while the global indicators were obtained by averaging the individual measures at each station. In this section, the global measures were analyzed first, and then a breakdown of indicators will be presented and discussed per test stage. The global indicators presented there showed that, in general, except for the first input combination, the rest of the combinations had similar performance accuracy in terms of the indicators in all stations. This shows although four-time lags have a significant correlation in time series (according to PAF), including more inputs beyond the second lag has negligible impact on modeling performance improvement (only VAF values differ) (Gupta *et al.*, 2023). Comparing the applied models, the MARS presented lower SI and higher NS and VF values than the RF and BT in all locations and inputs, although differences are low in this case, too. So, based on the global indicators of each input combination, it might be stated that all three models have had similar performance, and the second input combination, relying on two days river flow values would be a suitable choice for the prediction task in the studied stations. A further global performance comparison was made between the adopted input combinations as can be seen in **Table 4**. The indicators presented in **Table 4** have been obtained by averaging the global values of each indicator for all stations. Comparing the values in this table revealed that increasing the input variables beyond the second time lag has a negligible impact on modeling performance and the second input combination can be used as the optimum input set for the prediction task in this case. The reason for the selection of this combination is to use relatively fewer input variables for modeling, which reduces the model size and computational cost (Guzek *et al.*, 2023).
Analyzing the indicator's temporal variations at the studied locations (not presented here), revealed that obvious changes in the SI, NS, and VAF values occur per test stage when each time one part of data (here, one year) was used for testing the developed models.

Table 3. Error statistics of the RF, MARS, and BT models in the studied stations

			MARS	RF	BT
Station 1	Qt-1	SI	0.17	0.26	0.17
		NS	0.98	0.97	0.97
		VAF	98.45	96.55	97.40
	Qt-1, Qt-2	SI	0.13	0.24	0.22
		NS	0.99	0.97	0.97
		VAF	99.07	97.03	97.49
	Qt-1, Qt-2, Qt-3	SI	0.13	0.25	0.22
		NS	0.99	0.97	0.97
		VAF	99.17	96.63	97.48
	Qt-1, Qt-2, Qt-3, Qt-4	SI	0.12	0.23	0.22
		NS	0.99	0.97	0.97
		VAF	99.19	97.18	97.48
Station 2	Qt-1	SI	0.18	0.26	0.23
		NS	0.99	0.97	0.98
		VAF	98.46	96.74	97.49
	Qt-1, Qt-2	SI	0.13	0.26	0.23
		NS	0.99	0.97	0.98
		VAF	99.22	96.58	97.52
	Qt-1, Qt-2, Qt-3	SI	0.12	0.29	0.22
		NS	0.99	0.96	0.98
		VAF	99.29	95.91	97.54
	Qt-1, Qt-2, Qt-3, Qt-4	SI	0.12	0.28	0.22
		NS	0.99	0.96	0.98
		VAF	99.32	96.08	97.54
Station 3	Qt-1	SI	0.21	0.28	0.24
		NS	0.98	0.96	0.97

		VAF	97.73	95.85	96.87
	Qt-1, Qt-2	SI	0.16	0.29	0.24
		NS	0.99	0.96	0.97
		VAF	98.62	95.35	96.91
	Qt-1, Qt-2, Qt-3	SI	0.16	0.32	0.24
		NS	0.99	0.95	0.97
		VAF	98.64	94.34	96.91
	Qt-1, Qt-2, Qt-3, Qt-4	SI	0.16	0.32	0.24
		NS	0.99	0.95	0.97
		VAF	98.66	94.64	96.91

Table 4. Global performance indicators of the input combinations

Models	**MARS**			**RF**			**BT**		
	SI	NS	VAF	SI	NS	VAF	SI	NS	VAF
Input configuration I	0.44	0.98	98.21	0.49	0.97	96.38	0.46	0.97	97.25
Input configuration II	0.42	0.99	98.97	0.49	0.97	96.32	0.47	0.98	97.30
Input configuration III	0.41	0.99	99.04	0.50	0.96	95.63	0.47	0.98	97.31
Input configuration IV	0.41	0.99	99.06	0.49	0.96	95.97	0.47	0.98	97.31

CONCLUSION

Given the rising applications of machine learning (ML) techniques in various disciplines including hydrology and water resources management, prediction of river flow values has been considered as one of the important tasks in this context. So, a simulation study was performed in this research to investigate the capabilities of three ML techniques, namely, MARS, RF, and BT using data from three gauging stations at Des Moines River, Iowa, U.S.A. As this river is the largest river passing the Iowa state, accurate predictions of river flow time series are very important. These stations have been located upstream of the river. The applied models were constructed using four input combinations defined based on temporal correlations among the time series patterns. Total available patterns belonged to five years of daily records of river flow. The models were trained and tested based on adopting a temporal k-fold testing strategy. The obtained results showed that including two sets of river flow variables (records of two successive days) would be enough for accurate prediction of the river flow at all three locations. MARS, RF, and BT showed similar performance, although the error statistics fluctuated for different input combinations/stations, monotonously. Although the fundamentals of these three models are different, similar performance accuracy may belong, despite their higher capacity to map nonlinear complex systems, to the natural characteristics of river flow time series that made it easy to handle the prediction task based on chronologic information on data.

ACKNOWLEDGMENTS: None

CONFLICT OF INTEREST: None

FINANCIAL SUPPORT: None

ETHICS STATEMENT: None

REFERENCES

Ahmed, A. A. B., Alruwaili, M. N., Alanazi, J. F., Alanazi, D. F., & Alanazi, A. S. (2022). Awareness of diabetic patients regarding diabetes complications in Saudi Arabia: Systematic review. *Pharmacophore, 13*(5), 58-63. doi:10.51847/UQQWem7NOV

Alexander, A. A., Thampi, A. G., & Chithra, N. R. (2018). Development of hybrid wavelet-ANN model for hourly flood stage forecasting. *Hydrology Research, 49*(3), 658-669.

Al-Jaloud, M. M., Al-Osaidi, K. S., Al-Anzi, S. S., Al-Jalban, H. A., Al-Shahrani, F. M., Al-Omari, S. D., Al-Shahrani, A. S., Al-Enezi, M. M., Almusjan, S. A., & Al-Faridi, S. A. (2022). Effect of various distraction techniques on pain and anxiety of pediatric dental patients: A systematic review. *Pharmacophore, 13*(5), 105-111. doi:10.51847/vTfTdjBaws

Bakhshi Ostadkalayeh, F., Moradi, S., Asadi, A., Moghaddam Nia, A., & Taheri, S. (2023). Performance improvement of LSTM-based deep learning model for streamflow forecasting using Kalman filtering. *Water Resources Management*, 37, 3111-3127.

Biau, G. (2012). Analysis of a random forests model. *The Journal of Machine Learning Research*, *13*, 1063-1095.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*, 123-140.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5-32.

Çora, H., & Çora, A. N. (2022). An international relations study: Turks in the Western world's history perspective and Ataturk's approach. *Journal of Organizational Behavior Research, 7*(1), 96-107. doi:10.51847/eegcptya4E

Dalkiliç, H. Y., & Hashimi, S. A. (2020). Prediction of daily streamflow using artificial neural networks (ANNs), wavelet neural networks (WNNs), and adaptive neuro-fuzzy inference system (ANFIS) models. *Water Supply*, *20*(4), 1396-1408.

Domatskiy, V. N., & Sivkova, E. I. (2022). Opisthorchiasis - Is an urgent medical and social problem in Russia. *Journal of Biochemical Technology, 13*(4), 20-29. doi:10.51847/lXm5zetax6

Edossa, D. C., & Babel, M. S. (2011). Application of ANN-based streamflow forecasting model for agricultural water management in the Awash River Basin, Ethiopia. *Water Resources Management*, *25*, 1759-1773.

Elgendy, T. Y. A. A. A. (2022). Proposed model for selection of the internal auditor using analytical network process: Case study. *Journal of Organizational Behavior Research, 7*(1), 138-155. doi:10.51847/LNRB0ZkBTU

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802-813.

França, S., & Cabral, H. N. (2015). Predicting fish species richness in estuaries: Which modelling technique to use? *Environmental Modelling & Software*, *66*, 17-26.

Freidman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting. *Annals of Statistics, 28*(2), 337-407.

Friedman, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, *19*(1), 1-67.

Friedman, J. H., & Roosen, C. B. (1995). An introduction to multivariate adaptive regression splines. *Statistical Methods in Medical Research*, *4*(3), 197-217.

Gunduz, O., & Aral, M. M. (2005). River networks and groundwater flow: A simultaneous solution of a coupled system. *Journal of Hydrology*, *301*(1-4), 216-234.

Gupta, R. S., Roy, S., Bose, R., & Mandal, P. (2023). Impact of Ayush alternative medicine interventions in the COVID-19 pandemic. *Archives of Pharmacy Practice, 14*(2), 37-43. doi:10.51847/QdA0hbAnDG

Guzek, K., Stelmach, A., Rożnowska, A., Najbar, I., Cichocki, Ł., & Sadakierska-Chudy, A. (2023). A preliminary study of genetic polymorphisms potentially related to the adverse effects of aripiprazole. *Archives of Pharmacy Practice*, *14*(2), 13-18. doi:10.51847/bNMXd353X2

Han, D., Chan, L., & Zhu, N. (2007). Flood forecasting using support vector machines. *Journal of Hydroinformatics*, *9*(4), 267-276.

Insom, P., Cao, C., Boonsrimuang, P., Liu, D., Saokarn, A., Yomwan, P., & Xu, Y. (2015). A support vector machine-based particle filter method for improved flooding classification. *IEEE Geoscience and Remote Sensing Letters*, *12*(9), 1943-1947.

Jayavel, K., & Sivagnanam, S. (2022). The current scenario regarding the narrative advancement of oral cancer. *Clinical Cancer Investigation Journal, 11*(2), 7-13. doi:10.51847/fEhFVfwAsL

Karimi, S., Shiri, J., Kisi, O., & Shiri, A. A. (2016). Short-term and long-term streamflow prediction by using wavelet–gene expression programming approach. *ISH Journal of Hydraulic Engineering*, *22*(2), 148-162.

Karimi, S., Shiri, J., Kisi, O., & Xu, T. (2017). Forecasting daily streamflow values: Assessing heuristic models. *Hydrology Research*, *49*(3), 658-669.

Kaur, H., Mishra, D., Roychoudhury, A., Sharma, M. C., Bhalla, A. S., Mridha, A. R., Kakkar, A., Yadav, R., Kala, S., & Mishra, S. (2023). Giant cells lesions of oral and maxillofacial region – A proposed diagnostic algorithm. *Clinical Cancer Investigation Journal, 12*(1), 17-26. doi:10.51847/jT6kJbFKDg

Kisi, O., & Parmar, K. S. (2016). Application of least square support vector machine and multivariate adaptive regression spline models in long term prediction of river water pollution. *Journal of Hydrology*, *534*, 104-112.

Kisi, O., Dailr, A. H., Cimen, M., & Shiri, J. (2012). Suspended sediment modeling using genetic programming and soft computing techniques. *Journal of Hydrology*, *450*, 48-58.

Kisi, O., Parmar, K. S., Soni, K., & Demir, V. (2017). Modeling of air pollutants using least square support vector regression, multivariate adaptive regression spline, and M5 model tree models. *Air Quality, Atmosphere & Health*, *10*, 873-883.

Kryuchkova, A. V., Tunguzbieva, R. U., Tokaeva, K. S., Isaev, A. A., Elmaeva, L. R., & Mikhailenko, V. V. (2022). Collagen hydrolysates in the prevention and treatment of arthritis. *Journal of Biochemical Technology, 13*(4), 54-59. doi:10.51847/3SWKk0nMnP

Latt, Z. Z., & Wittenberg, H. (2014). Improving flood forecasting in a developing country: A comparative study of stepwise multiple linear regression and artificial neural network. *Water Resources Management*, *28*, 2109-2128.

Macharyulu, I. S., Satapathy, D. P., Sahoo, A., Samantaray, S., Mohanta, N. R., & Ray, A. (2022). Performance evaluation of MLP and cnn models for flood prediction. In *Intelligent System Design: Proceedings of INDIA 2022* (pp. 273-281). Singapore: Springer Nature Singapore.

Maroufpoor, S., Jalali, M., Nikmehr, S., Shiri, N., Shiri, J., & Maroufpoor, E. (2020). Modeling groundwater quality by using hybrid intelligent and geostatistical methods. *Environmental Science and Pollution Research*, *27*(2), 28183-28197.

Mashhour, K., Saad, E., Abdelghany, H., & Hashem, W. (2023). 3D-CRT versus SIB IMRT acute toxicity outcomes in preoperative concurrent chemo-radiotherapy for locally advanced rectal cancer. *Clinical Cancer Investigation Journal, 12*(1), 36-42. doi:10.51847/uBAn5N4CCd

Mekeres, G. M., Buhaş, C. L., Bulzan, M., Marian, P., & Hozan, C. T. (2022). Objective criteria in evaluating the consequences of the posttraumatic scars. *Pharmacophore, 13*(1), 56-61. doi:10.51847/nSmkjXUdzR

Osadchuk, M. A., Osadchuk, A. M., Vasilieva, I. N., & Trushin, M. V. (2023). The state biology museum named after kliment arkadyevich timiryazev as a scientific and educational center. *Journal of Biochemical Technology, 14*(1), 7-12. doi:10.51847/OLKERwxo55

Shiri, J., Kisi, O., Yoon, H., Kazemi, M. H., Shiri, N., Poorrajabali, M., & Karimi, S. (2020b). Prediction of groundwater level variations in coastal aquifers with tide and rainfall effects using heuristic data driven models. *ISH Journal of Hydraulic Engineering*, *28*(sup1), 188-198.

Shiri, N., Shiri, J., Kazemi, M. H., & Xu, T. (2022). Estimation of CO2 flux components over northern hemisphere forest ecosystems by using random forest method through temporal and spatial data scanning procedures. *Environmental Science and Pollution Research*, 29(11), 16123–16137.

Shiri, N., Shiri, J., Nourani, V., & Karimi, S. (2020a). Coupling wavelet transform with multivariate adaptive regression spline for simulating suspended sediment load:

Independent testing approach. *ISH Journal of Hydraulic Engineering*, *28*(sup1), 356-365.

Shiri, N., Shiri, J., Yaseen, Z. M., Kim, S., Chung, I. M., Nourani, V., & Zounemat-Kermani, M. (2020). Development of artificial intelligence models for well groundwater quality simulation: Different modeling scenarios. *Plos one*, *16*(5), e0251510.

Tennant, D. L. (1976). Instream flow regimens for fish, wildlife, recreation and related environmental resources. *Fisheries*, *1*(4), 6-10.

Terela, H., & Strilets, V. (2023). Supervision and control over compliance with labor law in Ukraine during the martial law. *Journal of Organizational Behavior Research, 8*(1), 39-51. doi:10.51847/mEsB03yCM9

Wegayehu, E. B., & Muluneh, F. B. (2022). Short-term daily univariate streamflow forecasting using deep learning models. *Advances in Meteorology*, *2022*. doi:10.1155/2022/1860460

Yilmaz, M., Tosunoğlu, F., Kaplan, N. H., Üneş, F., & Hanay, Y. S. (2022). Predicting monthly streamflow using artificial neural networks and wavelet neural networks models. *Modeling Earth Systems and Environment*, *8*(4), 5547-5563.