International Journal of Advanced and Applied Sciences, 10(4) 2023, Pages: 114-120

Contents lists available at Science-Gate

International Journal of Advanced and Applied Sciences

Journal homepage: http://www.science-gate.com/IJAAS.html

Comparison of machine learning techniques for rainfall-runoff modeling in Punpun river basin, India

Shashi Shankar Ojha *, Vivekanand Singh, Thendiyath Roshni

Department of Civil Engineering, National Institute of Technology, Patna, Bihar 800005, India

ARTICLE INFO

Article history:
Received 22 June 2022
Received in revised form
13 October 2022
Accepted 18 January 2023

Keywords:
Machine learning
Random forest
Rainfall-runoff modeling

ABSTRACT

Machine learning (ML) models have emerged as potential methods for rainfall-runoff modeling in recent decades. The appeal of ML models for such applications is owing to their competitive performance when compared to alternative approaches, ease of application, and lack of rigorous distributional assumptions, among other attributes. Despite the promising results, most ML models for rainfall-runoff applications have been limited to areas where rainfall is the primary source of runoff. The potential of Random Forest (RF), a popular ML method, for rainfall-runoff prediction in the Punpun river basin, India, is investigated in this paper. The correlation coefficient (R), Root mean squared error (RMSE), Mean absolute error (MAE), and Nash–Sutcliffe efficiency (NSE) are four statistical metrics used to compare RF performance to that of alternative ML models. Model evaluation metrics indicate that RF outperforms all others. In the RF model, we got the best NSE score of 0.795. These findings offer new perspectives on how to apply RF-based rainfall-runoff modeling effectively.

1. Introduction

Precise rainfall-runoff modeling is a key factor for effective water resources management and planning. It has been scientifically shown that the rainfall-runoff modeling of the river system is a challenging task due to physical processes and natural changes associated with the river system (Mohammadi, 2021). Because of the complex relationship between rainfall and runoff, accurate runoff estimation is always a difficult problem for hydrologists. Changes in river runoff are influenced by a variety of meteorological factors, including evapotranspiration, solar radiation, wind speed, air temperature, and catchment-specific characteristics, such as topography, shape, slope, altitude, soil type, land cover, and soil moisture-holding capacity, among others. Data on all of these factors is normally necessary for the successful running of rainfall-runoff models, which is a complex and challenging task, especially in developing nations with sparsely gauged catchments.

Due to the relatively low input data requirements, black-box models are chosen over conceptual and physically based models for modeling rainfall-runoff processes when reliable data on the aforementioned meteorological and site-specific parameters are not available. For modeling such complex nonlinear and nonstationary processes, ML models, such as artificial neural network (ANN) based, fuzzy-based, and regression-based machine learning (ML) models have been applied successfully in recent years.

Over the past decade, ML techniques have gained immense popularity in hydrology research. ML techniques have been successfully implemented for various hydrological applications for example flood modeling (Mosavi et al., 2018; Janizadeh et al., 2019) drought assessment (Feng et al., 2019; Shamshirband et al., 2020; Rhee and Im, 2017), water demand studies (Villarin and Rodriguez-Galiano, 2019; Xenochristou et al., 2021) rainfall modeling (Cramer et al., 2017; Basha et al., 2020), runoff modeling (Kumar et al., 2019; Taşar et al., 2019). Some of the ML models specifically used for rainfall-runoff modeling include ANN (Sudheer et al., 2002; Srinivasulu and Jain, 2006), adaptive neuro-fuzzy inference system or adaptive network-based fuzzy inference system (ANFIS) (Nourani and Komasi, 2013; Talei et al., 2010), multivariate adaptive regression splines model (MARS) (Sharda et al., 2008) and M5 model tree (M5Tree) (Adnan et al., 2021; Nourani et al., 2019), support vector regression (SVR) (Hosseini and Mahjouri, 2016;

* Corresponding Author.
Email Address: shashio.pg15.ce@nitp.ac.in (S. S. Ojha)
https://doi.org/10.21833/ijaas.2023.04.014
Corresponding author's ORCID profile:
https://orcid.org/0000-0002-3289-1018

Sedighi et al., 2016). Extreme learning machine (ELM) (Roushangar et al., 2018), Group Method of Data Handling (GMDH) (Al-Juboori, 2022).

Among the decades-old regression-based ML models, the random forest (RF) regression model has been successfully used in many hydrological applications. This is a unique comparative study for the selected study area as we have compared six ML algorithms for rainfall-runoff modeling in the Punpun river basin of India.

2. Study area and data collection

The Punpun river basin is located on the southern bank of the Ganga River, with latitudes of 24 11' and 25 25' N and longitudes of 84 9' and 85 20' E. (Fig. 1). The Punpun River is surrounded by the west by the Sone River and the east by the Kiul-Harohar-Falgu river system. The Punpun River, which originates in the Chhotanagpur hills of Jharkhand's Palamu district, is mostly a rain-fed river. It runs for 232 kilometers before joining the Ganga as a right bank tributary in the Fatuha region, 25 kilometers downstream of Patna. The Punpun River Basin, with its basin area of 8530 km2, has been chosen for this investigation. In the Punpun river basin, the average annual rainfall is between 960mm to 1020 mm.

Daily rainfall gridded data of 30 years i.e., from 1991 to 2020 have been obtained IMD website (www.imdpune.gov.in), for ten grid points of grid size 0.25×0.25, and monthly rainfall data has been calculated for the above grid points. The monthly discharge data of 30 years i.e., from 1991 to 2020 for the Sripalpur gauging site has been collected from Central Water Commission (CWC), Patna.

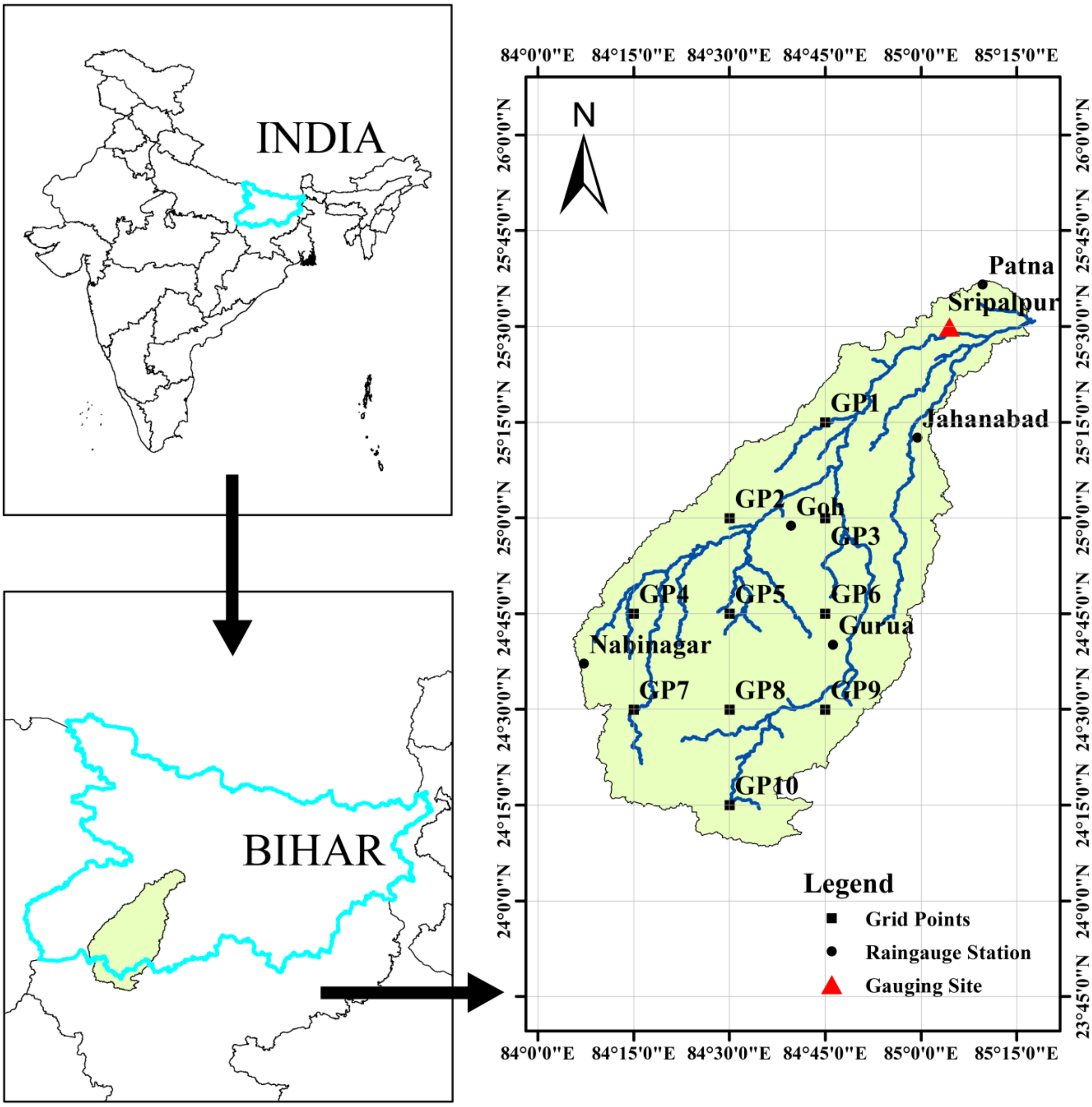


Fig. 1: Map showing the location of the Punpun river basin

3. Methodology

3.1. ML approaches for rainfall-runoff modeling

Different machine-learning model approaches were utilized to illustrate each model's accuracy to estimate the runoff.

3.2. Multiple linear regression (MLR)

MLR is a popular ML technique for regression problems. One independent variable is present in basic linear regression, and the model must establish a linear relationship between it and the dependent variable. In MLR, on the other hand, the model must consider multiple independent variables to find a relationship. The MLR approach is employed when the response variable is affected by more than one predictor variable.

Furthermore, MLR is an extension of Simple Linear Regression in that it predicts the response variable using more than one predictor variable. The equation of MLR can be defined as follows (Niu et al., 2019).

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \dots\dots\dots + b_nx_n \tag{1}$$

where, b_0 is the intercept, b_1, b_2, b_3,.... b_n are coefficients or slopes of the independent variables x_1, x_2, x_3 ...,x_n and y is the dependent variable.

3.3. Least absolute shrinkage and selection operator (LASSO)

Tibshirani (1996) first presented the LASSO regression (LR) model. LASSO is a linear regression-based strategy that uses a shrinking process to confine the sum of the absolute values of the model parameters; as a result, the sum must be less than a predetermined number (Upper bound). The LR algorithm recommends simple, sparse models (models with fewer parameters), which are well-suited for models or data with high levels of multicollinearity or when we want to automate certain parts of model selection, such as variable selection or parameter elimination, using feature engineering. The LR algorithm employs the L1 regularization technique and is used when there are a large number of features because it automatically performs feature selection.

3.4. Ridge regression (RR)

In ML, RR is another type of regression procedure that is used when there is a high correlation between the independent variables or model parameters (Shariff and Duzan, 2018). The least-square estimates evaluate unbiased values as the correlation value increases. However, if the dataset's collinearity is very high, there may be some bias value. As a result, we include a bias matrix in the RR algorithm equation. It is a useful regression method in which the model is less susceptible to overfitting and thus works well even with small datasets.

3.5. Polynomial regression (PR)

PR is a type of regression analysis in which the relationship between the independent and dependent variables is represented by an nth-degree polynomial. The least-squares method is commonly used to fit PR models (Maulud and Abdulazeez, 2020) According to the Gauss-Markov Theorem, the least-square method minimizes the variance of the coefficients. PR is a subset of Linear Regression in which the data is fitted with a polynomial equation with a curvilinear relationship between the dependent and independent variables. The relationship between the independent and dependent variables in the data set does not have to be linear for PR to work. This is also one of the main differences between Linear and PR. PR is generally used when the linear regression model does not capture the points in the data and the linear regression fails in describing the best result clearly.

3.6. Support vector machine (SVR)

SVM was introduced by Vapnik (1995) based on supervised learning methods for classification and regression problems that analyze data and recognize patterns. The SVM learning system employs a hypothesis space of linear functions in a high dimensional feature space, which has been trained with an optimization theory learning algorithm that implements a statistical learning theory learning bias and structural risk minimization principle. In this present study, SVM is used for regression analysis therefore its SVR. SVR for regression is defined by the following equations.

Considering a set of training data $\{(x_1, y_1), \dots\dots, (x_n, y_n)\}$, where $x_i \in R^n$, $y_i \in R$, the decision function is represented by $f(x) = w^T\emptyset(x) + b$, with respect to $w \in R^n$, $b \in R$, where f denotes a nonlinear transformation from R^n to a high-dimensional space. The primal optimization problem is given by:

$$\text{Minimize } R_{reg}(f) = \frac{1}{2}\|w\|^2 + c\left(\sum_{i=1}^{n}|y - f(x)|_{\epsilon}\right) \tag{2}$$

3.7. RF

RF was introduced by Breiman (2001). RF is a supervised ML algorithm and also an ensemble ML technique for classification and regression that works by building a large number of decision trees. The principles behind ensemble learning techniques are based on the idea that it outperforms other ML algorithms in terms of accuracy because the combination of predictions outperforms any single constituent model. Distinct decision trees in RF tend to learn highly irregular patterns, i.e. to overfit their training data sets. The goal of RF is to minimize prediction variance by averaging multiple decision trees trained on different parts of the same training data set. RF modeling is suitable for simulating the nonlinear effect of variables. It can handle complex variable interactions and is unaffected by multicollinearity. The RF can evaluate the effects of all instructive variables at the same time and automatically ranks their importance in descending order. Because the generalization error in RF converges as the number of trees surges, the RF does not overfit the data.

4. Model development

The selection of appropriate input variables in hydrological modeling studies would be critical in their applications. We developed hydrological model strategies that predict outputs based on past rainfall and runoff data. As a result, four input data combinations for runoff record periods were prepared as shown in Table 1.

4.1. Assessment of model performances

Many techniques are recommended for the assessment of model performance in the literature. However, we used 4 performance evaluation criteria used in this study are shown in Table 2.

In Table 2, Q_{obs} is observed runoff; Q_{for} is forecasted runoff; $\bar{Q}_{obs}$ is average observed; $\bar{Q}_{for}$ is average forecasted Q; N is number of data points (70% for training and 30% for testing of the data).

Table 1: Input data combination

Combination	Model input	Model output
Combination_1	rainfall(t-1), runoff(t-1), rainfall(t)	runoff(t)
Combination_2	rainfall(t-1), rainfall(t)	runoff(t)
Combination_3	rainfall(t-2), runoff(t-2), rainfall(t-1), runoff(t-1), rainfall(t)	runoff(t)
Combination_4	rainfall(t-2), runoff(t-2), rainfall(t-1), rainfall(t)	runoff(t)

t represents the current month; t-1 represents the one month back and so on

Table 2: Performance evaluation parameters

Statistical measures	Formula	Range	Optimum value
Coefficient of correlation (R)	$R=\frac{\sum_{i=1}^{N}(Q_{obs}-\bar{Q}_{obs})(Q_{for}-\bar{Q}_{for})}{\sqrt{\sum_{i=1}^{N}(Q_{obs}-\bar{Q}_{obs})^2(Q_{for}-\bar{Q}_{for})^2}}$	-1to1	0.8
Root mean square error (RMSE)	$RMSE=\sqrt{\frac{\sum_{i=1}^{N}(Q_{obs}-Q_{for})^2}{N}}$	The optimal value is 0	0
Nash-Sutcliffe efficiency (NSE)	$NSE=1-\left[\frac{\sum_{i=1}^{N}(Q_{obs}-Q_{for})^2}{\sum_{i=1}^{N}(Q_{obs}-\bar{Q}_{for})^2}\right]$	-∞ to 1	1
Mean absolute error (MAE)	$MAE=\frac{1}{N}\sum_{i=1}^{N}\lvert(Q_{obs}-Q_{for})\rvert$	The optimal value is 0	0

5. Results and discussion

This section represents and discusses the case study's obtained results using the proposed ml models. Table 1 shows the different combinations of rainfall and runoff models. The models were built using 70% of the dataset for training and the remaining 30% for testing. All developed models have undergone 10-fold cross-validation. Model evaluation metrics suggest RF performs among all. We obtained the best NSE scores in the RF model i.e., 0.795 for combination_1 which is followed by SVR (NSE=0.858) for the same combination. A comparative study of the best combination of each regression method has been shown in Fig. 2 for test data.

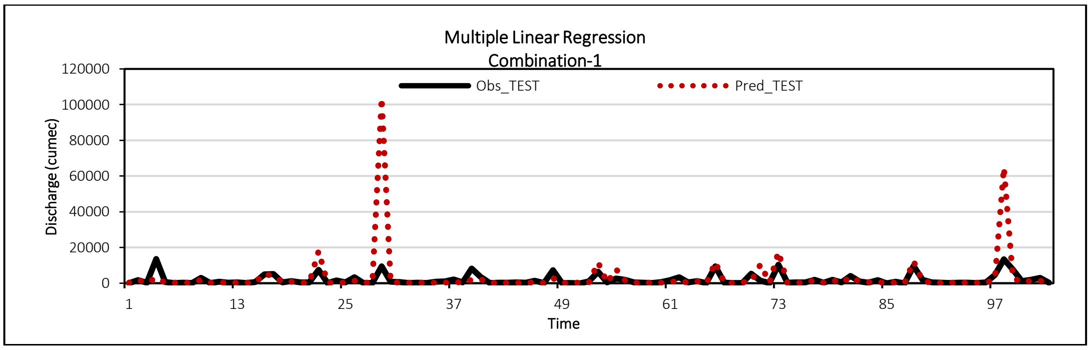
Table 3 shows that combination_1 and combination_2 models performed better than combination_3 and combination_4 in all the ML techniques except PR used in this study. In this study, the PR model developed is a second-degree polynomial model. Some authors give more priority to NSE and MAE matrices over R and RMSE for model selection because the former is less sensitive to extreme values. Further studies reported NSE is a normalized statistic that determines the relative magnitude of the residual variance compared to the measured data variance.

These results presented provide new insights for the effective application of RF-based rainfall-runoff modeling.

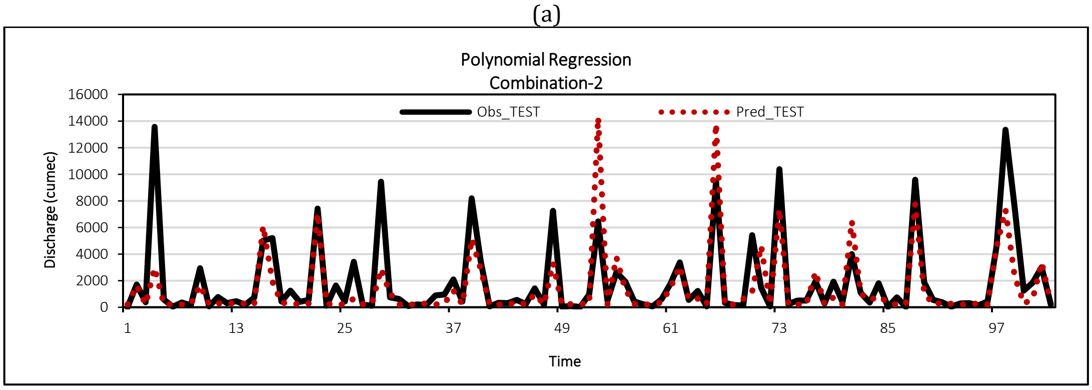
6. Conclusion

The research focused on the applicability of RF ML methods for rainfall-runoff modeling and compared the results to those of other ML algorithms. RF methods were implemented with several input scenarios, including previous rainfall and runoff data from the Punpun River basin in India.

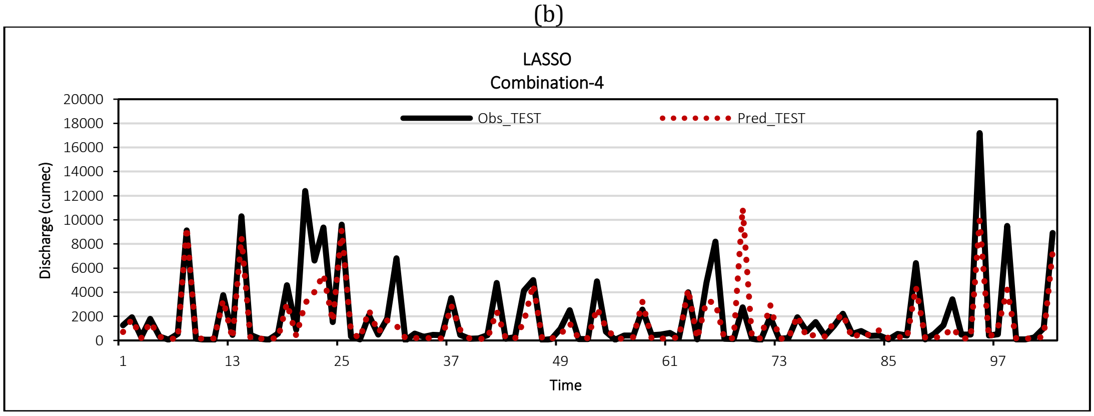
The RF model was developed using four different combinations of rainfall and runoff, and its results were compared to those of various machine-learning models. Given the Punpun river system's current and future vulnerabilities due to erratic rainfall, skilled rainfall-runoff modeling can have significant implications. RF and RF-based algorithms continue to gain popularity in hydrological studies due to their practical applications. Further study can be done using a coupled model and with combinations sorted by input selector techniques.


Multiple Linear Regression
Combination-1
Obs_TEST
Pred_TEST
Discharge (cumec)
Time

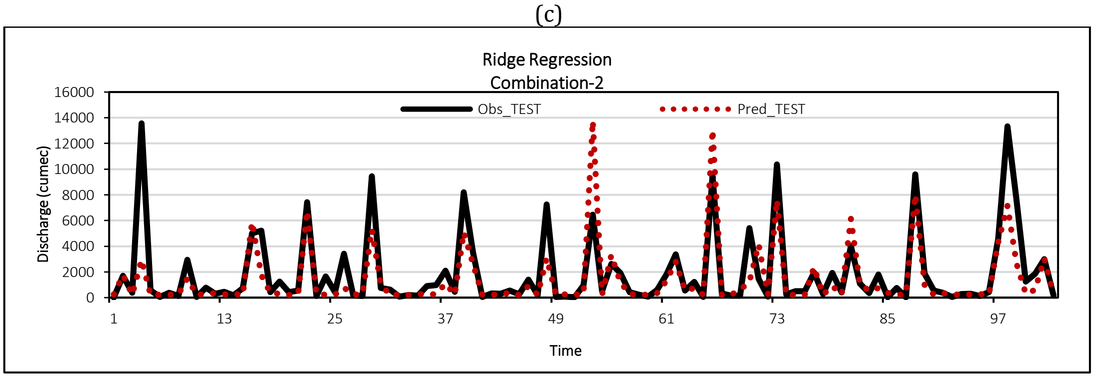
(a)


Polynomial Regression
Combination-2
Obs_TEST
Pred_TEST
Discharge (cumec)
Time

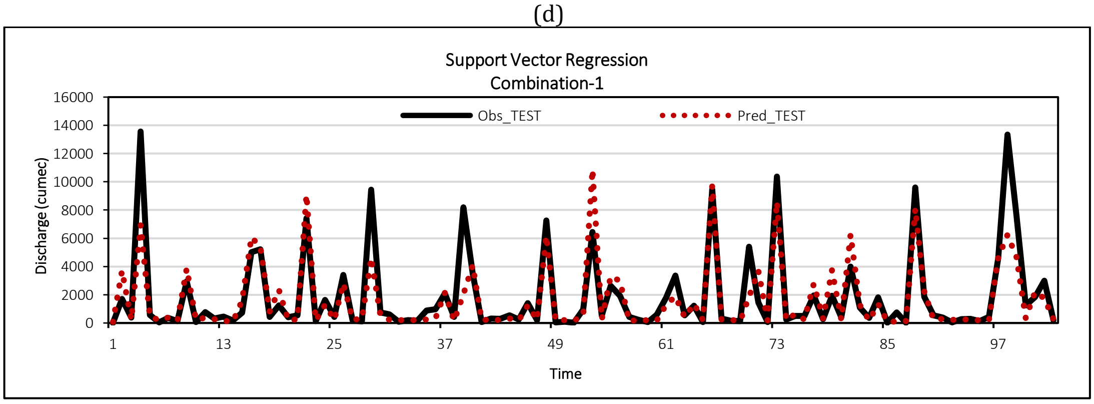
(b)


LASSO
Combination-4
Obs_TEST
Pred_TEST
Discharge (cumec)
Time

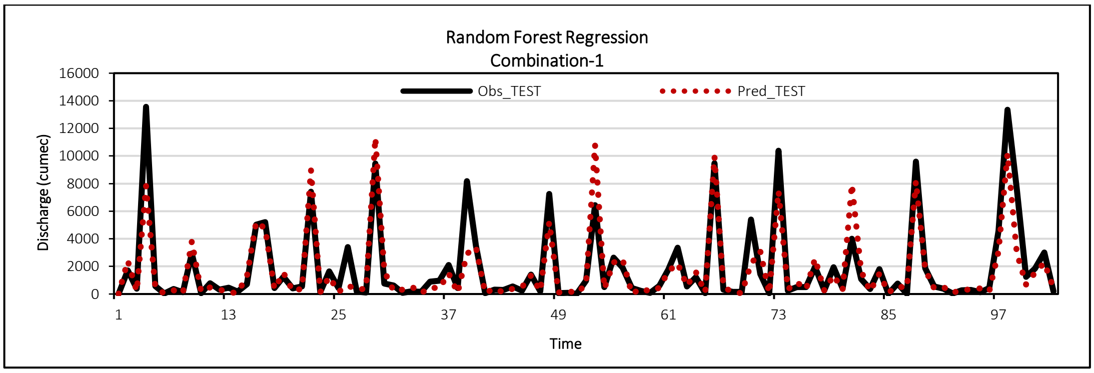
(c)


Ridge Regression
Combination-2
Obs_TEST
Pred_TEST
Discharge (cumec)
Time

(d)


Support Vector Regression
Combination-1
Obs_TEST
Pred_TEST
Discharge (cumec)
Time

(e)



(f)

Fig. 2: Graphical representation of best combination for different regression methods (a) MLR (b) PR (c) LR (d) RR (e) SVR, and (f) RF

Table 3: Shows the error statistics

		Training				Testing			
		R	RMSE(m^3/s)	NSE	MAE	R	RMSE(m^3/s)	NSE	MAE
MLR	Combination_1	0.69	4257.947	-0.752	1484.582	0.577	10828.6	-12.654	2449.41
	Combination_2	0.757	2803.36	0.241	1197.61	0.366	17978.11	-36.637	2656.296
	Combination_3	-0.021	6199.01	-2.702	2870.076	-0.021	5461.733	-2.474	2848.476
	Combination_4	-0.02	6206.52	-2.711	2871.31	-0.021	5429.685	-2.433	2837.801
PR	Combination_1	0.792	2409.49	0.439	1044.866	0.668	3278.537	-0.252	1349.06
	Combination_2	0.692	2906	0.184	1081.793	0.758	1989.975	0.539	951.241
	Combination_3	-0.008	5491.13	-1.905	2849.435	-0.004	4015.517	-0.878	2401.467
	Combination_4	0.008	5586.45	-2.007	2784.195	0	3807.61	-0.688	2357.897
LR	Combination_1	0.037	4983.86	-1.53	2807.163	0.1	4937.731	-1.457	2744.796
	Combination_2	0.04	4884.23	-1.43	2645.435	0.155	3716.545	-0.392	2175.494
	Combination_3	0.786	2561.59	0.332	943.648	0.776	2063.724	0.571	912.739
	Combination_4	0.751	2841	0.178	978.053	0.828	1869.023	0.648	845.011
RR	Combination_1	0.781	2451.11	0.42	1013.918	0.684	2925.639	0.003	1211.204
	Combination_2	0.704	2788.54	0.249	1058.059	0.786	1880.803	0.588	907.876
	Combination_3	0.001	5365.96	-1.774	2801.093	-0.017	3890.922	-0.763	2357.201
	Combination_4	0.01	5418.149	-1.828	2753.792	-0.006	3771.689	-0.657	2331.014
SVR	Combination_1	0.923	1258.79	0.847	612.156	0.858	1523.729	0.73	731.803
	Combination_2	0.876	1581.58	0.758	768.886	0.84	1627.469	0.692	783.317
	Combination_3	0.032	4247.81	-0.739	2583.252	-0.007	3818.672	-0.698	2396.982
	Combination_4	0.037	4268.52	-0.756	2591.426	-0.033	3947.262	-0.814	2471.34
RF	Combination_1	0.978	774.798	0.942	339.094	0.894	1326.147	0.795	663.822
	Combination_2	0.964	926.955	0.917	405.262	0.835	1648.898	0.683	825.167
	Combination_3	0.009	4250.63	-0.741	2620.312	-0.033	3901.616	-0.773	2460.745
	Combination_4	0.019	4192.31	-0.693	2583.135	-0.025	3757.663	-0.644	2395.793

Compliance with ethical standards

Conflict of interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

Adnan RM, Petroselli A, Heddam S, Santos CAG, and Kisi O (2021). Comparison of different methodologies for rainfall–runoff modeling: Machine learning vs conceptual approach. Natural Hazards, 105(3): 2987-3011. https://doi.org/10.1007/s11069-020-04438-2

Al-Juboori AM (2022). Solving complex rainfall-runoff processes in semi-arid regions using hybrid heuristic model. Water Resources Management, 36(2): 717-728. https://doi.org/10.1007/s11269-021-03053-5

Basha CZ, Bhavana N, Bhavya P, and Sowmya V (2020). Rainfall prediction using machine learning and deep learning techniques. In the International Conference on Electronics and Sustainable Communication Systems (ICESC), IEEE, Coimbatore, India: 92-97. https://doi.org/10.1109/ICESC48915.2020.9155896

Breiman L (2001). Random forests. Machine Learning, 45(1): 5-32. https://doi.org/10.1023/A:1010933404324

Cramer S, Kampouridis M, Freitas AA, and Alexandridis AK (2017). An extensive evaluation of seven machine learning methods for rainfall prediction in weather derivatives. Expert Systems with Applications, 85: 169-181. https://doi.org/10.1016/j.eswa.2017.05.029

Feng P, Wang B, Li Liu D, and Yu Q (2019). Machine learning-based integration of remotely-sensed drought factors can improve the estimation of agricultural drought in South-Eastern Australia. Agricultural Systems, 173: 303-316. https://doi.org/10.1016/j.agsy.2019.03.015

Hosseini SM and Mahjouri N (2016). Integrating support vector regression and a geomorphologic artificial neural network for daily rainfall-runoff modeling. Applied Soft Computing, 38: 329-345. https://doi.org/10.1016/j.asoc.2015.09.049

Janizadeh S, Avand M, Jaafari A, Phong TV, Bayat M, Ahmadisharaf E, and Lee S (2019). Prediction success of machine learning methods for flash flood susceptibility mapping in the Tafresh watershed, Iran. Sustainability, 11(19): 5426. https://doi.org/10.3390/su11195426

Kumar A, Kumar P, and Singh VK (2019). Evaluating different machine learning models for runoff and suspended sediment simulation. Water Resources Management, 33(3): 1217-1231. https://doi.org/10.1007/s11269-018-2178-z

Maulud D and Abdulazeez AM (2020). A review on linear regression comprehensive in machine learning. Journal of Applied Science and Technology Trends, 1(4): 140-147. https://doi.org/10.38094/jastt1457

Mohammadi B (2021). A review on the applications of machine learning for runoff modeling. Sustainable Water Resources Management, 7: 98. https://doi.org/10.1007/s40899-021-00584-y

Mosavi A, Ozturk P, and Chau KW (2018). Flood prediction using machine learning models: Literature review. Water, 10(11): 1536. https://doi.org/10.3390/w10111536

Niu WJ, Feng ZK, Feng BF, Min YW, Cheng CT, and Zhou JZ (2019). Comparison of multiple linear regression, artificial neural network, extreme learning machine, and support vector machine in deriving operation rule of hydropower reservoir. Water, 11(1): 88. https://doi.org/10.3390/w11010088

Nourani V and Komasi M (2013). A geomorphology-based ANFIS model for multi-station modeling of rainfall–runoff process. Journal of Hydrology, 490: 41-55. https://doi.org/10.1016/j.jhydrol.2013.03.024

Nourani V, Davanlou Tajbakhsh A, Molajou A, and Gokcekus H (2019). Hybrid wavelet-M5 model tree for rainfall-runoff modeling. Journal of Hydrologic Engineering, 24(5): 04019012. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001777

Rhee J and Im J (2017). Meteorological drought forecasting for ungauged areas based on machine learning: Using long-range climate forecast and remote sensing data. Agricultural and Forest Meteorology, 237: 105-122. https://doi.org/10.1016/j.agrformet.2017.02.011

Roushangar K, Alizadeh F, and Nourani V (2018). Improving capability of conceptual modeling of watershed rainfall–runoff using hybrid wavelet-extreme learning machine approach. Journal of Hydroinformatics, 20(1): 69-87. https://doi.org/10.2166/hydro.2017.011

Sedighi F, Vafakhah M, and Javadi MR (2016). Rainfall–runoff modeling using support vector machine in snow-affected watershed. Arabian Journal for Science and Engineering, 41(10): 4065-4076. https://doi.org/10.1007/s13369-016-2095-5

Shamshirband S, Hashemi S, Salimi H, Samadianfard S, Asadi E, Shadkani S, and Chau KW (2020). Predicting standardized streamflow index for hydrological drought using machine learning models. Engineering Applications of Computational Fluid Mechanics, 14(1): 339-350. https://doi.org/10.1080/19942060.2020.1715844

Sharda VN, Prasher SO, Patel RM, Ojasvi PR, and Prakash C (2008). Performance of multivariate adaptive regression splines (MARS) in predicting runoff in mid-Himalayan micro-watersheds with limited data/Performances de régressions par splines multiples et adaptives (MARS) pour la prévision d'écoulement au sein de micro-bassins versants Himalayens d'altitudes intermédiaires avec peu de données. Hydrological Sciences Journal, 53(6): 1165-1175. https://doi.org/10.1623/hysj.53.6.1165

Shariff NSM and Duzan HMB (2018). An application of proposed ridge regression methods to real data problem. International Journal of Engineering and Technology, 7(4.30): 106-108. https://doi.org/10.14419/ijet.v7i4.30.22061

Srinivasulu S and Jain A (2006). A comparative analysis of training methods for artificial neural network rainfall–runoff models. Applied Soft Computing, 6(3): 295-306. https://doi.org/10.1016/j.asoc.2005.02.002

Sudheer KP, Gosain AK, and Ramasastri KS (2002). A data-driven algorithm for constructing artificial neural network rainfall-runoff models. Hydrological Processes, 16(6): 1325-1330. https://doi.org/10.1002/hyp.554

Talei A, Chua LHC, and Wong TS (2010). Evaluation of rainfall and discharge inputs used by adaptive network-based fuzzy inference systems (ANFIS) in rainfall–runoff modeling. Journal of Hydrology, 391(3-4): 248-262. https://doi.org/10.1016/j.jhydrol.2010.07.023

Taşar B, Unes F, and Varcin H (2019). Prediction of the rainfall–runoff relationship using neurofuzzy and support vector machines. In the 2019 Air and Water–Components of the Environment Conference Proceedings, Wessex Institute, Cluj-Napoca, Romania: 237-246. https://doi.org/10.24193/AWC2019_24

Tibshirani R (1996). Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society: Series B (Methodological), 58(1): 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Vapnik V (1999). The nature of statistical learning theory. Springer Science and Business Media, Berlin, Germany. https://doi.org/10.1007/978-1-4757-3264-1_8

Villarin MC and Rodriguez-Galiano VF (2019). Machine learning for modeling water demand. Journal of Water Resources Planning and Management, 145(5): 04019017. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001067

Xenochristou M, Hutton C, Hofman J, and Kapelan Z (2021). Short-term forecasting of household water demand in the UK using an interpretable machine learning approach. Journal of Water Resources Planning and Management, 147(4): 04021004. https://doi.org/10.1061/(ASCE)WR.1943-5452.0001325