Journal of Agrometeorology

ISSN : 0972-1665 (print), 2583-2980 (online)
Vol. No. 26 (4) : 501-504 (December - 2024)
https://doi.org/10.54386/jam.v26i4.2676
https://journal.agrimetassociation.org/index.php/jam

Short communication

Use of machine learning techniques in predicting inflow in Tarbela reservoir of Upper Indus Basin

SHABIR JAN*[1], UZAIR KHAN[1], ALAMGIR KHALIL[1], AMJAD ALI KHAN[1], HABIB AHMAD JAN[1] and IBAD ULLAH[2]

[1]University of Engineering and Technology, Peshawar 25120, Pakistan
[2]Institute of Chemical Sciences, University of Peshawar, Peshawar 25120, Pakistan
**Corresponding author's e-mail: engineershabirjan@gmail.com*

Tarbela dam is an important part of Pakistan's water infrastructure and is located on the Upper Indus River Basin, in the Haripur district of Khyber Pakhtunkhwa Province. It is also an essential part of the water resource management system in Pakistan. Tarbela dam was planned and constructed in 1976 to provide millions of people with potable water, hydroelectric power, regulation of floods, and advancement of agricultural development. Among its many applications are power generating, flood control, and irrigation (Khan, 2022). Tarbela dam experienced sudden and significant rises in 2010 and 2022 because there was insufficient future inflow information, leading to disastrous consequences for various aspects of life and infrastructure (Event Analysis, 2023). By optimizing how water is distributed, it ensures that different sectors get what they need without too much or too little. Predicting the inflow accurately helps the hydropower turbines work better, ensuring a steady and dependable electricity supply. This is essential for stability and economic progress (Tarbela Dam Project, 2023). Predicting how much water flows into a reservoir is essential for flood control.

Machine learning models are playing crucial role in making prediction in various fields like finance, agriculture and many more (Shrivastav and Kumar, 2021). In agriculture, different algorithms have been used to predict crop yield (Sridhara *et al.,* 2024; Sakthipriya and Thangavel, 2024). Various machine learning applications have been made to predict reservoir inflow accurately for managing water resources efficiently and making dam operations more effective and safer (Kumar *et al.,* 2023; Gupta and Kumar, 2022). In this article, an attempt has been made to predict the inflow of Tarbela reservoir of Pakistan using machine learning models.

Fig. 1 shows the location of Tarbela basin area with data stations. Forty-three years (1981 to 2023) data on precipitation, temperature and reservoir level and inflow of study area were collected from three main sources. Precipitation data was collected from the Climate Engine platform, accessible at (https://app.climateengine.org/climateEngine). This dataset provides high-resolution precipitation estimates derived from satellite observations. Temperature data was obtained from the NASA Power Project data access viewer, available at (https://power.larc.nasa.gov/data-access-viewer) & Reservoir level and inflow data was obtained from WAPDA (Water and Power Development Authority). Precipitation, temperature and reservoir level data was used as input feature and reservoir inflow was used as target variable (Gupta and Kumar, 2022).

Machine learning models

XGBoost

Extreme Gradient Boosting is an ensemble learning method for regression and classification that was derived from gradient boosting. This method generates a number of learners, finds the mistakes made by each student, and then brings those errors to correct. Resulting in a very precise model. XGBoost has revolutionized the field of machine learning due to its fantastic performance, scalability, flexibility, and interpretability (Rajesh *et al.,* 2022).

Random Forest

Random Forest is an ensemble machine learning technique that combines several decision trees in order to produce

Article info - DOI: *https://doi.org/10.54386/jam.v26i4.2676*
Received: 30 July 2024; Accepted: 25 August 2024; Published online : 01 December 2024

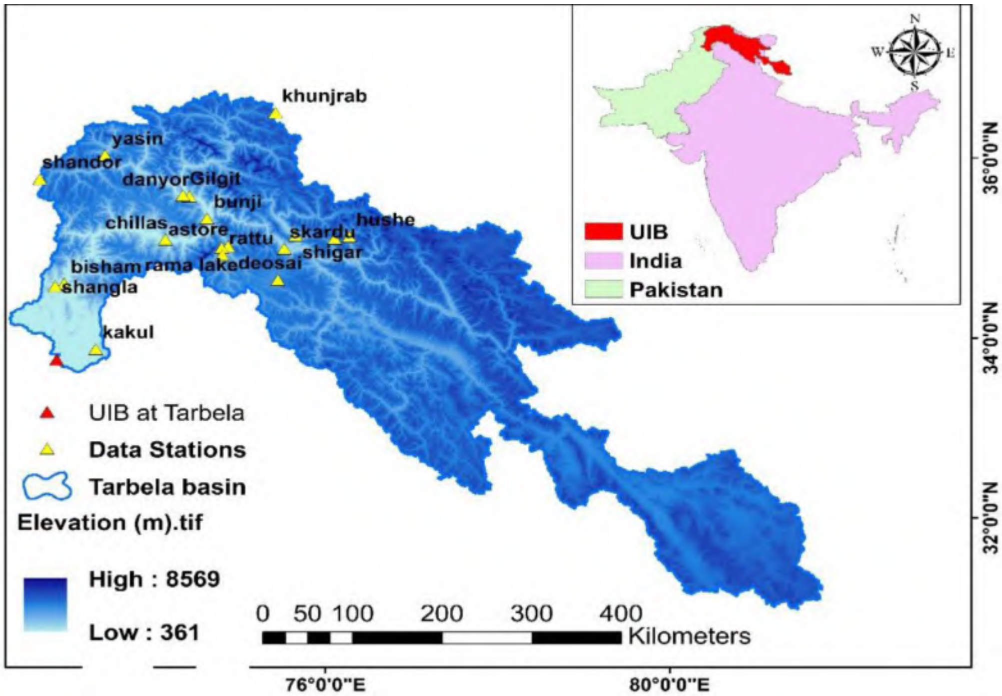

Fig. 1: Location of study area and data stations

a more reliable and accurate predictive model. Reduced overfitting and enhanced prediction performance were the main reasons for its introduction as an upgrade over single decision tree (Wang *et al.,* 2022)

Ensemble of XGBoost-Random Forest

The strengths of both algorithms are combined in an ensemble of Random Forest and XGBoost to enhance prediction performance. XGBoost is an efficient gradient boosting technique that builds decision trees one after the other with the goal of fixing each other's mistakes. & Random Forest, a bagging method, reduces over fitting and increases stability by building numerous decision trees independently and averaging their predictions. Combining these two techniques, the ensemble takes use of Random Forest's resilience and XGBoost's accuracy in simulating complex patterns, producing an accurate and broadly applicable model (Shrivastav and Kumar, 2021).

Forty-three years (1981-2023) data used for this research was split into training, testing & prediction. Data from 1981 to 2008 was used for training the models and from 2009 to 2015 was used for testing the accuracy of the model. Then, these trained and tested models were used for predicting inflow for eight years (2016 to 2023) and was compared with actual inflow to check the model's accuracy for future prediction. Mean bias error (MBE), root mean squared error (RMSE), mean absolute error (MAE) and R-squared (R^2) metrics were used to assess model's performance.

Model performance

Figs. 2, 3 and 4 visually demonstrate each model performance during training, testing and prediction periods, respectively. Models were evaluated using R^2, RMSE, MAE and MBE during training, testing and prediction period and are presented in Table 1. XGBoost trained to R^2 of 0.92, RMSE of 706.3, MAE of 434.5 and MBE of -4.5, suggested a good performance of the model while during testing period, the evaluation parameters R^2 (0.94), RMSE (668.2), MAE (419.2) and MBE (8.1) indicated better performance. During prediction period, however, these values were little less, though the performance was satisfactory with R^2 of 0.80. The Random Forest model showed a good predictive capability, with R^2 of 0.89, RMSE of 809.2, MAE of 469.5, and MBE of 8.9, during training phase and more or less similar during testing phase. The ensemble model (XGBoost-Random Forest) returned slightly lower results compared to the individual models during all the three phases of model's performance (Table 1). The performance analysis reveals that XGBoost and Random Forest models perform well during training and testing, with high accuracy and low errors. However, both models show a decline in prediction accuracy, with increased errors and a tendency to over predict inflows. The Ensemble model XGBoost-Random Forest underperforms in all phases compared to the individual models. From the graphs it is clear that XGBoost model has the capability of picking extreme inflows conditions while the other models are good in average inflow calculation. Overall, the results demonstrate the effectiveness of machine learning models for predicting reservoir inflows, with XGBoost standing out as the most precise and accurate model among those studied.

XGBoost stands out for its top performance, showing it can handle challenging inflow prediction tasks due to its advanced gradient boosting framework. The Random Forest model proved to make accurate predictions, showing its high accuracy in regression problems like inflow prediction. Surprisingly, when XGBoost and Random Forest were put together, they didn't perform better than when used separately. This shows that more complexity doesn't necessarily lead to improved results. Using artificial intelligence to

Table 1: Results of models' evaluation at different phases of computation

Model	Phase	R^2	RMSE	MAE	MBE
	Training	0.92	706.3	434.5	-4.5
XGBoost	Testing	0.94	668.2	419.2	8.1
	Prediction	0.80	1193.2	595.0	189.7
	Training	0.89	809.2	469.5	8.9
Random Forest	Testing	0.89	892.7	487.3	-0.15
	Prediction	0.79	1217.4	719.5	214.8
	Training	0.84	980.9	557.1	-44.4
Ensemble model (XGBoost-Random Forest)	Testing	0.81	1147.9	595.6	-73.0
	Prediction	0.77	1295.8	752.7	107.5

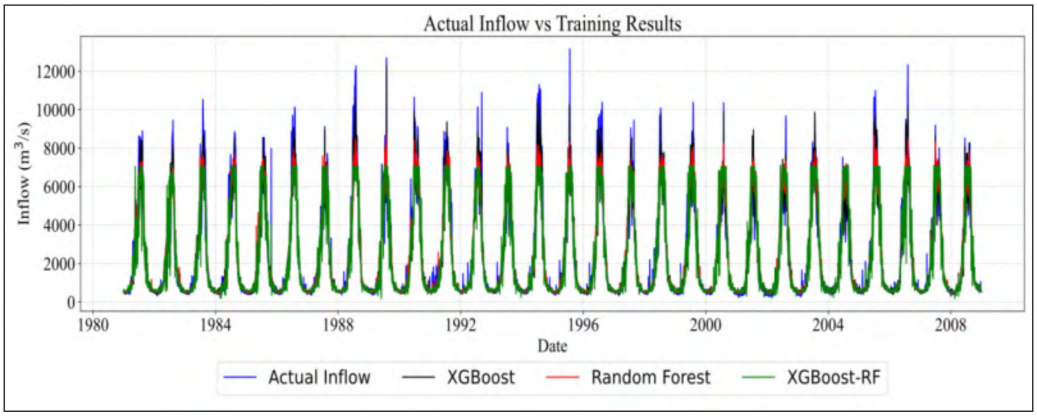

Fig. 2: Actual inflow and computed inflow by three models during Training period (1981-2008)

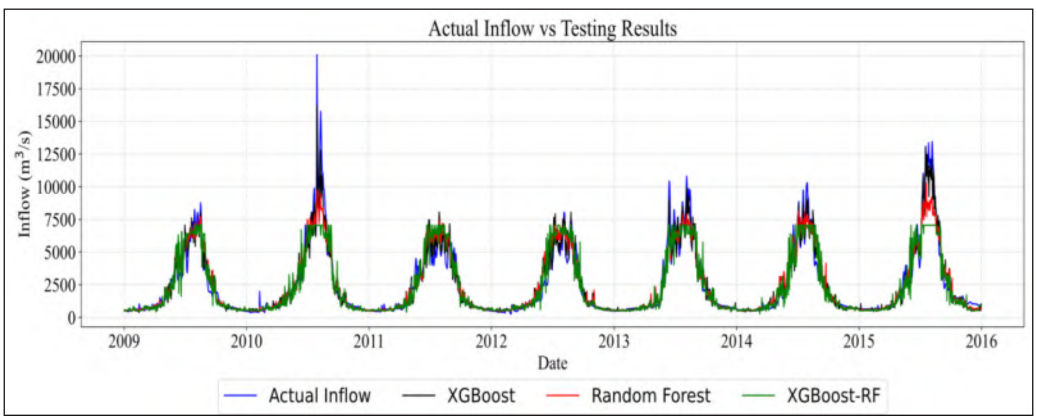

Fig. 3: Actual inflow and computed inflow by three models during testing period (2009-2016)

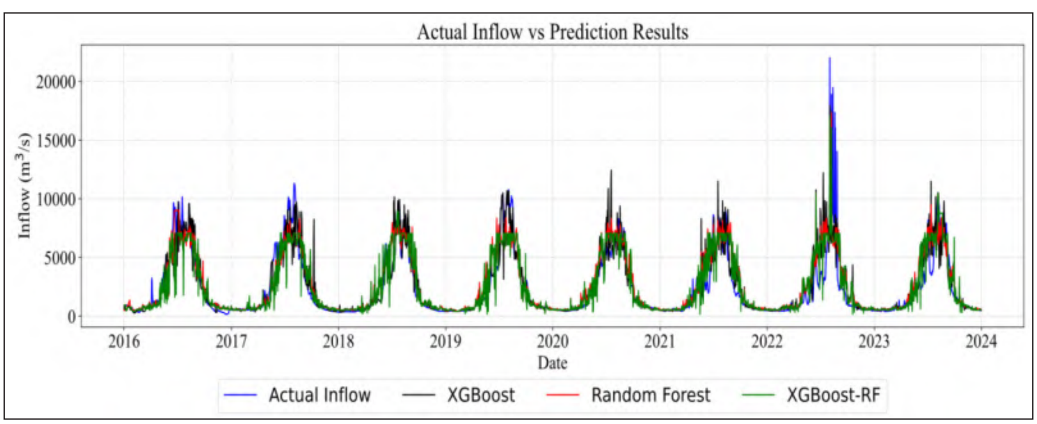

Fig. 4: Actual inflow and predicted inflow by three models during 2016-2023

predict water inflow at Tarbela Dam is a big step forward in managing water resources. This research shows that machine learning models are effective, with XGBoost being the most accurate.

ACKNOWLEDGMENTS

The authors are thankful to Water and Power Development Authority for providing reservoir data.

Fundings: The research received no specific funds from any agency in the public, commercial or not-for-profit sectors

Conflict of Interest: The Authors declare that there is no conflict of interest

Data availability statement: Data will be made available upon request to corresponding author.

Authors contributions: **S. Jan:** Conceptualization, model development, methodology & final review; **U. Khan**: Data collection; **A. Khalil:** Supervision; **A. A. Khan**: Statistical analysis; **H. A. Jan**: Literature review, and **I. Ullah:** Visualizations.

Disclaimer: The contents, pinions, and views expressed in the research communication published in the Journal of Agrometeorology are the views of the authors and do not necessarily reflect the views of the organizations they belong to.

Publisher's Note: The periodical remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

Event Analysis (2023). Floods in Swat Valley, Pakistan - Pakistan | ReliefWeb. (n.d.). Retrieved 25 May 2024, from https://reliefweb.int/report/pakistan/event-analysis-2022-floods-swat-valley-pakistan

Gupta, A. and Kumar, A. (2022). Two-step daily reservoir inflow prediction using ARIMA-machine learning and ensemble models. *J. Hydro-Environ. Res., 45*, 39-52. https://doi.org/10.1016/J.JHER.2022.10.002

Khan, F. (2022). Water availability and response of Tarbela Reservoir under the changing climate in the Upper Indus Basin, Pakistan. *Sci. Reps.,* 12(1), 15865. https://doi.org/10.1038/s41598-022-20159-x

Kumar, V., Kedam, N., Sharma, K. V., Mehta, D. J. and Caloiero, T. (2023). Advanced Machine Learning Techniques to Improve Hydrological Prediction: A Comparative Analysis of Streamflow Prediction Models. *Water (Switzerland)*, 15(14). https://doi.org/10.3390/w15142572

Rajesh, M., Indranil, P. and Rehana, S. (2022). Reservoir Inflow Forecasting Based On Gradient Boosting Regressor Model - A Case Study Of Bhadra Reservoir, India. *18th Annu. Meet. Asia Oceania Geosci. Soc.,* 64-66. https://doi.org/10.1142/9789811260100_0022

Sakthipriya, D. and Thangavel, C. (2024). Comparison of machine learning classification algorithms based on weather variables and seed characteristics for the selection of paddy seed. *J. Agrometeorol.,* 26(2): 209-214. https://doi.org/10.54386/jam.v26i2.2553

Shrivastav, L. K. and Kumar, R. (2021). An Ensemble of Random Forest Gradient Boosting Machine and Deep Learning Methods for Stock Price Prediction. *J. Inf. Technol. Res.,* 15(1), 1-19. https://doi.org/10.4018/JITR.2022010102

Sridhara, S., Soumya B. R. and Kashyap, G. R. (2024). Multistage sugarcane yield prediction using machine learning algorithms. *J. Agrometeorol.,* 26(1): 37-44. https://doi.org/10.54386/jam.v26i1.2411

Tarbela Dam Project (2023). Tarbela Dam Project, Haripur District, Pakistan. (n.d.). Retrieved 25 May 2024, from https://www.water-technology.net/projects/tarbela-dam-project/

Wang, M., Feng, D., Li, D. and Wang, J. (2022). Reservoir Parameter Prediction Based on the Neural Random Forest Model. *Front Earth Sci.,* 10. https://doi.org/10.3389/feart.2022.888933