DISCHARGE PREDICTION AT BAHADURABAD TRANSIT OF BRAHMAPUTRA-JAMUNA USING MACHINE LEARNING AND ASSESSMENT OF FLOODING

M. I. I. Rabbi[1*], M. M. H. Galib[1], M.A. Hasan[2], P. M. Toma[3], M. Ahamed[4]

[1] Undergraduate Student, Department of Civil Engineering, RUET, Bangladesh, (imran140105@gmail.com)*
[2] Postgraduate Student, Department of Geology, University of Dhaka, Bangladesh, (asifhasanimon96@gmail.com)
[3] Undergraduate Student, Department of Agriculture, Sher-e-Bangla Agricultural University, Dhaka, Bangladesh, (promamazumder31@gmail.com)
[4] Lecturer, Department of Civil Engineering, RUET, Bangladesh, (mirazahamed24@gmail.com)

Abstract

The river Brahmaputra and Jamuna have a significant contribution to water transportation, agriculture, and the livelihood of people living on the river banks. Ensuring the proper utilization of these channels, accurate water discharge prediction is a must since it can be of benefits for managing the river and allocating water resources. For hydrological prediction, time series model such as auto regressive moving average model and artificial neural network models are mostly used. In this study, three different machine learning algorithms, K-nearest neighbor, decision tree, and random forest regressor, along with different hyperparameters, are presented, which have been found more efficient among the other machine learning approaches. Daily water level and maximum velocity were used as explanatory variables, and water discharge was used as a response variable. Data from 2005 to 2013 was used for training the model, and 2014 to 2019 was used for evaluation. Among these three models, the k-nearest neighbor has performed exceptionally well. This model's R2 value and mean absolute percentage error are 0.9447 and 17.38, respectively. The obtained discharge rates are further compared with previously recorded discharge data before, during, and after major floods in those regions and which are found to have a linear relationship with river flooding.

Keywords: *Hydrological prediction; Machine learning; K-nearest neighbor; Water discharge; Brahmaputra-Jamuna.*

1. Introduction

The level of water plays a major role in the well-being and livelihoods of the population. For example, increases in water levels and discharge rate can influence physical processes, including the circulation of rivers, leading to changes in water mixing and resuspension of the bottom sediment. The prediction of the discharge rate is therefore increasingly necessary (Ali et al., 2013). For example, the Institute of Water and Flood Management (IWFM) recommends that further measures be put to develop approaches for water level control and

prediction. Change in the water level is dynamic hydrology because of its different regulated variables, including temperature and water sharing between the river and its watersheds (Mosavi et al., 2018). Some models must be chosen carefully to predict actual changes in the water level and discharge rate. Many conditions, such as influencing variables that influence the water level, take a significant amount of time and calibrate to ensure the forecast is correct (Sahoo et al., 1997). Since process-driven methodologies take too long, recent experiments have estimated water levels using an artificial neural network (ANN) based machine learning model (ML) (Corani et al., 2005). Several machine learning algorithms have been used in this paper to forecast water discharge rates (Benoudjit et al., 2019). Machine learning is often used nowadays to predict the level of water, discharge rate, and increasing accuracy; it is popular day by day. In the present report, three machine learning models – K-nearest neighbor, decision tree, and random forest regressor – were used using historical evidence. The average water level and discharge in Bahadurabad transits were observed in the years 2005 to 2019. And then, the ML model is compared to the ML model for predictive results. ML models are built by considering the influence of historical changes in water level and weather influences.

2. Study region

The Brahmaputra-Jamuna is one of Bangladesh's major rivers and the primary sources of water in the Northwest region of the country. According to the IPCC and Bangladesh Climate Change Cell Research, increased monsoon rainfall in the future will cause increased flood inundation. For this purpose, it is crucial to make accurate predictions of water level and discharge. In this study, the Bahadurabad discharge gauge station from the river Jamuna was selected. Daily Water level, discharge, and maximum velocity data from January 2005 to March 2019 of this station is collected from the Bangladesh Water Development Board (BWDB).

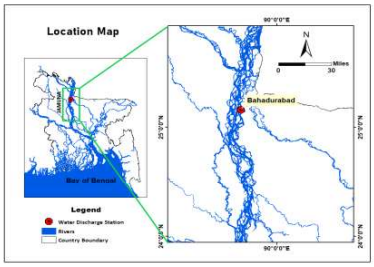


Figure 1: Location of the Study Area.

3. Methodology

3.1 Data preprocessing

After removing the outliers, the entire dataset was split into two categories. Data from January 2005 to September 2013 was used as a train set to train the models. And the rest of the data was used as the test set to check the accuracy and performance of the models. Afterwards both the sets were again split into explanatory and response variables. The water level and maximum velocity were considered explanatory variables, while the discharge was regarded as the response variable.

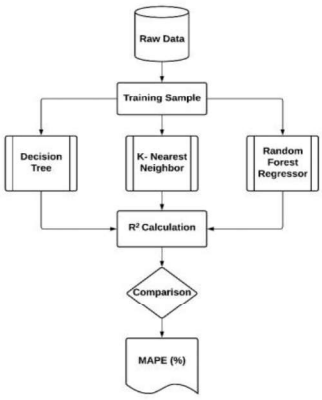


Figure 1 : Flow Chart of the Process

3.2 K Nearest Neighbor

Many researchers have found that when it comes to traffic flow prediction, the non-parametric model performs relatively better than the parametric models. K nearest Neighbor is one of the non-parametric models used in both regression and classification analysis. K nearest neighbor uses 'Feature Similarity' for prediction (Zhao et al., 2019). When introduced with a new data point from the test set, it calculates the distances between the latest data point and the training data point. For distance measurement, Euclidean distance has been used as the distance metric. Let dist p, q be the distance between two feature vectors and with dimension. Then the equation of Euclidean distance is

$$\mathbf{dist^2_{p,\,q} = (p_1 - q_1)^2 + (p_2 - q_2)^2 + (p_3 - q_3)^2 + \ldots\ldots + (p_n - q_n)^2}$$

3.3 Decision Tree Regressor

Decision Tree is another popular machine learning approach that can be used for both classification and regression analysis. Using the train set, this model forms a tree consists of several branches and leaves. These branches can also be called Decision nodes, and the leaves can be called terminal nodes. While constructing the branches, there's a decision by which the explanatory variables will be split into the following branches (Noor et al., 2017). And when the variables come across all the branches and the decision nodes, it reaches the terminal node where the response variable of the corresponding explanatory variable will be calculated. By our train set, a tree and a model have been built where the decision nodes are split according to the water level and maximum velocity, which were taken as explanatory variables in our study. For improving the model accuracy, a hyper-parameter named 'Maximum Depth' was tuned.

3.4 Random Forest Regressor

Random Forest regressor is a supervised machine learning algorithm that uses an ensemble learning method for prediction. When a model combines the prediction results from several machine learning algorithms instead of doing it alone, it is called the ensemble technique (Lukman et al., 2016). A decision tree works fine with a specific set of data. But if there is a slight variation in the test set, the result differs much in this algorithm. Here comes the

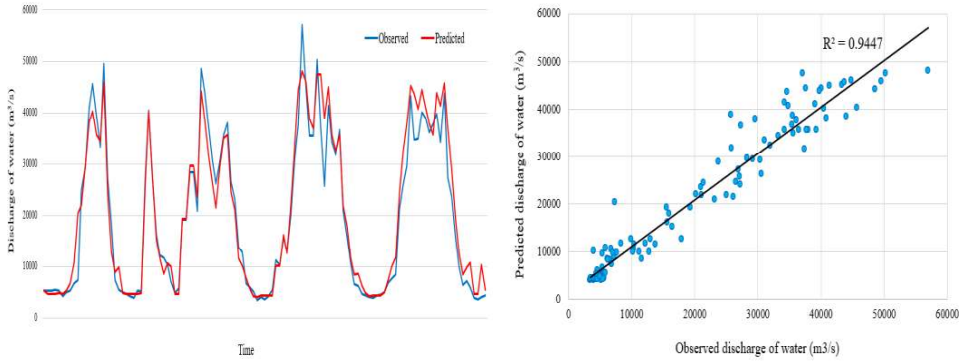
bagging ensemble technique, random forest regressor, to overcome this situation and prevent overfitting. Random forest regressor is nothing but a set of decision trees that run in parallel, and after the completion, the results are aggregated with the mean value of the results achieved from the decision tree models.

4. Results:

There is a reasonably good agreement between predicted and observed water levels with correlation coefficient values (R^2) above 0.87 in all three models. The K-Nearest Neighbour seems to produce the highest R2 values (>0.9) with mean absolute percentage errors (MAPE) varying from 17.38 up to 27.33 % (Table 1), which indicates a more accurate prediction. Besides, from the figure 3(a), it is observed that there is less discrepancy between observed and predicted curve in comparison with others two models. In contrast, the other two models produced almost similar correlation coefficient values. However, their high MAPE (>30%) makes them less suitable for forecasting.

Table 1: Comparison of three different models along with different hyper-parameters (Parameter set: water level + velocity).

Model	Hyper parameter		R^2 vlue	MAPE (%)
	Name	Value		
K-Nearest Neighbor	n	5	0.9121	27.33
		10	0.9265	24.55
		15	0.932	20.46
		20	0.9393	18.1
		25	0.9447	17.38
Decision Tree	Maximum Depth	4	0.8813	35.13
		5	0.8833	30.02
		8	0.8719	35.44
Random Forest	Number of Estimator	3	0.8941	30.39
		7	0.8945	32.87
		10	0.8869	32.32



(a) (b)

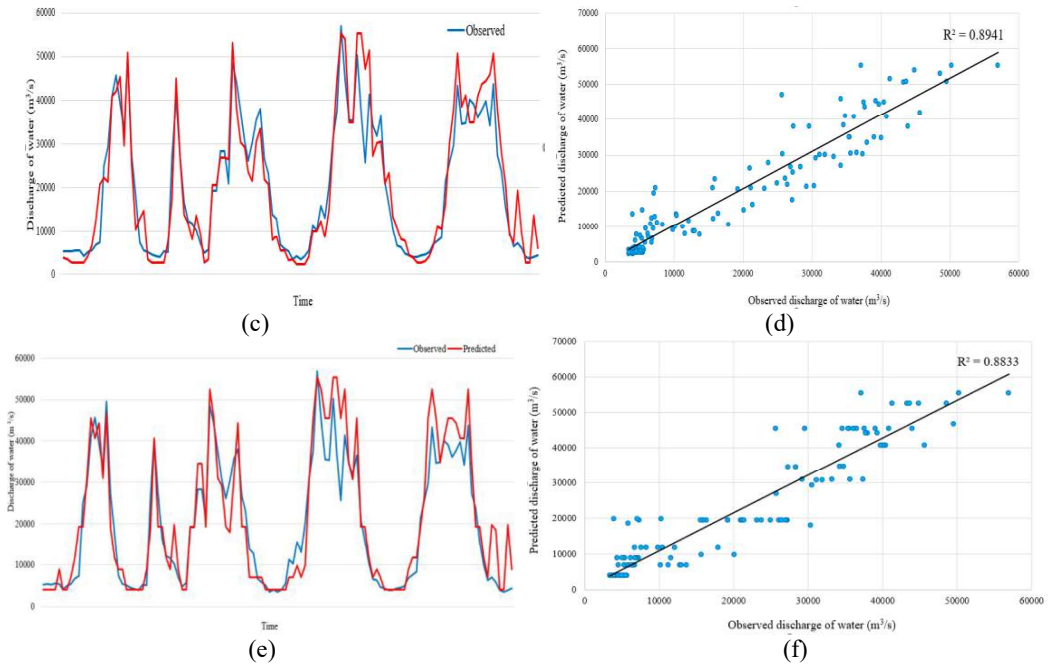


Figure 3: Water discharge prediction using- K- Nearest Neighbor model – (a) discharge of water varying with time, (b) predicted discharge of water variation with respect to observed discharge of water; Random Forest model - (c) discharge of water varying with time, (d) predicted discharge of water variation with respect to observed discharge of water; Decision Tree Model - (e) discharge of water varying with time, (f) predicted discharge of water variation with respect to observed discharge of water.

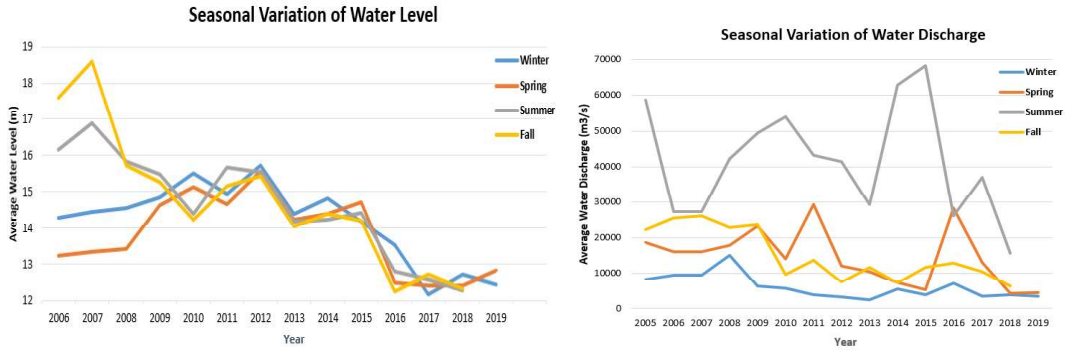


Figure 4: (a) Seasonal variation of water level; (b) Seasonal variation of water discharge.

The predicted water levels are compared with the available observations, and later the setup is validated using the seasonal average discharge data available from the past 15 years. The results show that water level varies seasonally with river discharge (Figure 3). The mean water level varies seasonally between 14 m (dry season) and 16 m (wet season). The strong monsoonal river discharge during the end of the summer season results in higher water levels. In particular, it can significantly rise up to 18 m at high discharge (Figure 4).

5. Discussions & Conclusions

The present state of ML modeling for discharge prediction is very young and in the early stage of development. The results presented in this study are important for understanding, modeling and managing complex river systems like the Ganges-Brahmaputra-Meghna. The established model setups can be further applied to investigate flood risk assessment.

References

Sahoo, A., Samantaray, S. & Ghose, D.K. Prediction of Flood in Barak River using Hybrid Machine Learning Approaches: A Case Study. *J Geol Soc India* **97,** 186–198 (2021). https://doi.org/10.1007/s12594-021-1650-1.

Mosavi A, Ozturk P, Chau K-w. Flood Prediction Using Machine Learning Models: Literature Review. *Water*. 2018; 10(11):1536. https://doi.org/10.3390/w10111536.

G. Corani and G. Guariso, "Coupling fuzzy modeling and neural networks for river flood prediction," *IEEE Trans. on Systems Man and Cybernetics*, vol.35, no.3, pp.382-390, Aug.2005.

D.P. Lettenmaier and E.F. Wood, 1993, Hydrological Forecasting, Chapter 26 in *Handbook of Hydrology*. (D. Maidment, ed.), McGraw-Hill. D.P.

Benoudjit, A., & Guida, R. (2019). A Novel Fully Automated Mapping of the Flood Extent on SAR Images Using a Supervised Classifier. *Remote Sensing*, 11(7), 779.

Zhao, G., Pang, B., Xu, Z., Peng, D., & Xu, L. (2019). Assessment of urban flood susceptibility using semi-supervised machine learning model. *Science of The Total Environment,* 659, 940-949.

H. M. Noor, D. Ndzi, G. Yang and N. Z. M. Safar, "Rainfall-based river flow prediction using NARX in Malaysia," 2017 *IEEE 13th International Colloquium on Signal Processing & its Applications (CSPA)*, Batu Ferringhi, 2017, pp. 67-72.

Q. A. Lukman, F. A. Ruslan and R. Adnan, "5 Hours ahead of time flood water level prediction modelling using NNARX technique: Case study terengganu," 2016 *7th IEEE Control and System Graduate Research Colloquium (ICSGRC),* Shah Alam, 2016, pp. 104-108.

Ali, M., Qamar, A. M., & Ali, B. (2013). Data analysis, discharge classifications, and predictions of hydrological parameters for the management of Rawal Dam in Pakistan. In 2013 *12th International Conference on Machine Learning and Applications.1*, pp. 382-385. IEEE.