

EGU21-15294

<https://doi.org/10.5194/egusphere-egu21-15294>

EGU General Assembly 2021

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



## Development of Flood Forecasting System for Someshwari-Kangsa Sub-watershed of Bangladesh-India Using Different Machine Learning Techniques

**Md Hamidul Haque<sup>1</sup>**, Mushtari Sadia<sup>2</sup>, and Mashiat Mustaq<sup>2</sup>

<sup>1</sup>Department of Water Resources Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh (hhaque6944@gmail.com)

<sup>2</sup>Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka, Bangladesh (mushtarisadia98@gmail.com, mashiatmustaq98@gmail.com)

Floods are natural disasters caused mainly due to heavy or excessive rainfall. They induce massive economic losses in Bangladesh every year. Physically-based flood prediction models have been used over the years where simplified forms of physical laws are used to reduce calculations' complexity. It sometimes leads to oversimplification and inaccuracy in the prediction. Moreover, a physically-based model requires intensive monitoring datasets for calibration, accurate soil properties information, and a heavy computational facility, creating an impediment for quick, economical and precise short-term prediction. Researchers have tried different approaches like empirical data-driven models, especially machine learning-based models, to offer an alternative approach to the physically-based models but focused on developing only one machine learning (ML) technique at a time (i.e., ANN, MLP, etc.). There are many other techniques, algorithms, and models in machine learning (ML) technology that have the potential to be effective and efficient in flood forecasting. In this study, five different machine learning algorithms- exponent back propagation neural network (EBPNN), multilayer perceptron (MLP), support vector regression (SVR), DT Regression (DTR), and extreme gradient boosting (XGBoost) were used to develop total 180 independent models based on a different combination of time lags for input data and lead time in forecast. Models were developed for Someshwari-Kangsa sub-watershed of Bangladesh's North Central hydrological region with 5772 km<sup>2</sup> drainage area. It is also a data-scarce region with only three hydrological and hydro-meteorological stations for the whole sub-watershed. This region mostly suffers extreme meteorological events driven flooding. Therefore, satellite-based precipitation, temperature, relative humidity, wind speed data, and observed water level data from the Bangladesh Water Development Board (BWDB) were used as input and response variables.

For comparison, the accuracy of these models was evaluated using different statistical indices - coefficient of determination, mean square error (MSE), mean absolute error (MAE), mean relative error (MRE), explained variance score and normalized centred root mean square error (NCRMSE). Developed models were ranked based on the coefficient of determination ( $R^2$ ) value. All the models performed well with  $R^2$  being greater than 0.85 in most cases. Further analysis of the

model results showed that most of the models performed well for forecasting 24-hour lead time water level. Models developed using XGBoost algorithm outperformed other models in all metrics. Moreover, each of the algorithms' best-performed models was extended further up to 20 days lead time to generate forecasting horizon. Models demonstrated remarkable consistency in their performance with the coefficient of determination ( $R^2$ ) being greater than 0.70 at 20 days lead-time of forecasting horizon in most cases except the DTR-based model. For 10- and 5-days lead time of forecasting horizon, it was greater than 0.75 and 0.80 respectively, for all the model extended. This study concludes that the machine algorithm-based data-driven model can be a powerful tool for flood forecasting in data-scarce regions with excellent accuracy, quick building and running time, and economic feasibility.