



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

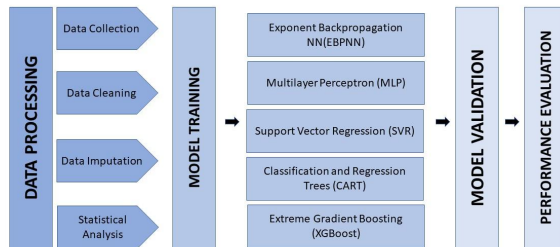
M. H. Haque¹, M. Sadia² & M. Mustaq³

¹ Department of Water Resources Engineering, ^{2,3} Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka

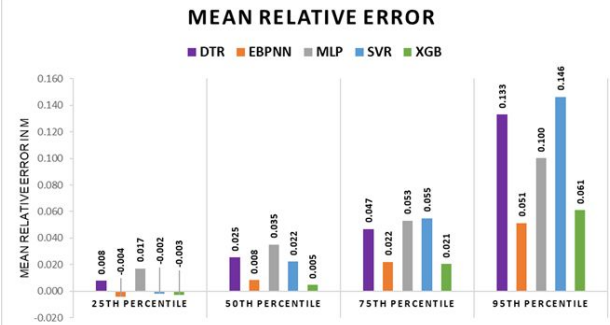
Background

- Each year, Bangladesh suffers from fluvial, coastal, urban and flash floods in different times of year. Physically based model for flood prediction requires intensive monitoring datasets for calibration, accurate soil properties information, and a heavy computational facility.
- Extreme meteorological events driven flooding can be difficult to accurately and quickly model using a large integrated process-based modeling system. Therefore, Machine Learning based data driven models can be a good alternative for better and timely prediction of such flood events.
- Five different machine learning techniques- EBPNN, MLP, SVR, DTR, and XGBoost were used to develop total 180 independent models to find the suitable one for the study area.
- Accuracy of these models was evaluated using different statistical indices - coefficient of determination, MSE, MAE, MRE and explained variance score.

Methodology



Performance Evaluation



Study Area and Data Used

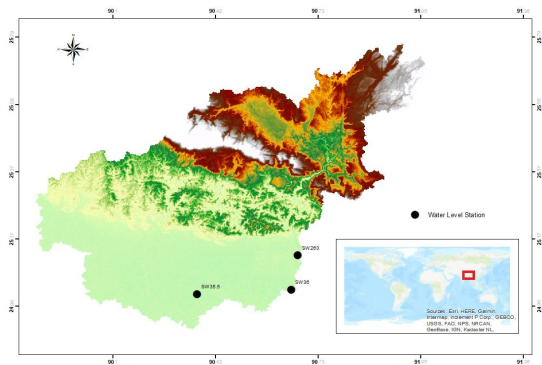


Figure 1: Study area is located at Someshwari-Kangsa sub-watershed of Bangladesh-India with a total area of 5772 Sq. KM. Water Level Station No. SW36 was used for model training, testing and validating. Water Level data were collected from Bangladesh Water Development Board.

Correlation Analysis

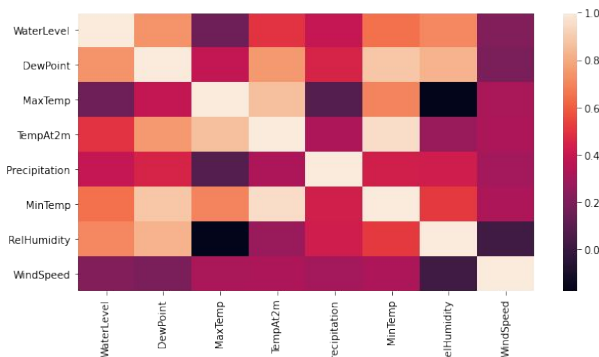


Figure 2: Correlation between different input datasets were analyzed which have been depicted through the above plot. High value of correlation indicates linear dependency between two features. High correlation with the output parameter was considered for training. Climate data were collected from NASA power.

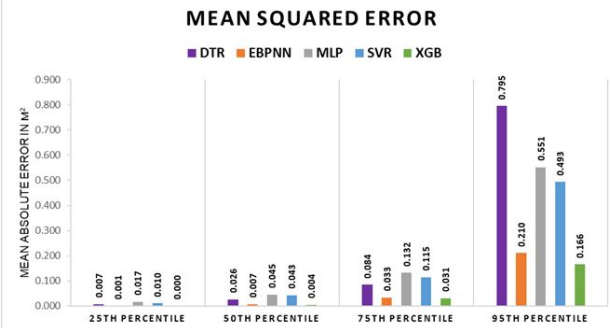


Figure 3: a) Depicts the Mean Relative Error (MRE) of the models developed using five different machine learning algorithms compared to the observed water level data in validation stage. Initial 25th percentile error data showed both the positive and negative bias but later models showed the positive bias only. b) Depicts the Mean Squared Error (MSE) of the models developed using five different machine learning algorithms in validation stage. Initial 50th percentile error data of the all models showed less than 0.01 m² error compared to the observed water level data.



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

M. H. Haque¹, M. Sadia² & M. Mustaq³

¹Department of Water Resources Engineering, ^{2,3}Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology, Dhaka

[Continued]

Performance Evaluation

MEAN ABSOLUTE ERROR

■ DTR ■ EBPNN ■ MLP ■ SVR ■ XGB

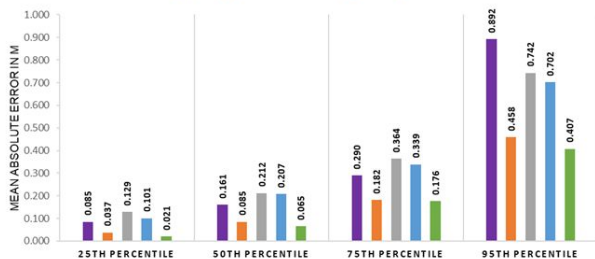


Figure 3: c) Depicts the Mean Absolute Error (MAE) of the models developed using five different machine learning algorithms compared to the observed water level data in validation stage. Initial 50th percentile error data showed that mean absolute error is less than 0.25m or 25 cm for all of the developed models. Model developed using XGBoost had mean absolute error of less than 0.40m or 40 cm till 95th percentile of total validation data.

Conclusion

- The objective of the study was to experiment different machine learning algorithms to develop reliable flood forecasting models and compare their performance to propose the most compatible machine learning based data-driven flood forecasting model for the study area.
- Total of 180 models were trained, tested, validated and compared using different error metrics. The comparison showed that the model developed using XGBoost algorithm outperforms all other models but slightly underperforms in long term flood forecasting.
- This study also concludes that models developed using EBPNN, SVR, and XGBoost performed with more than 90% accuracy at 3-day forecasting lead time.

Forecasting Horizon Analysis

Forecasting Horizon vs Accuracy

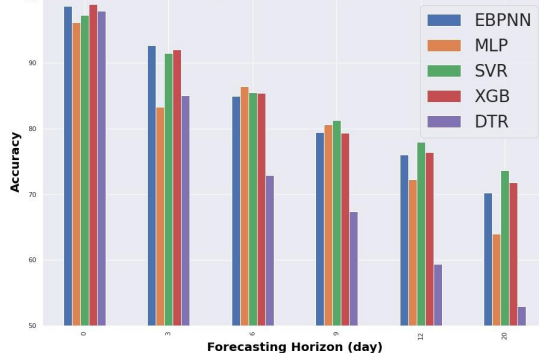
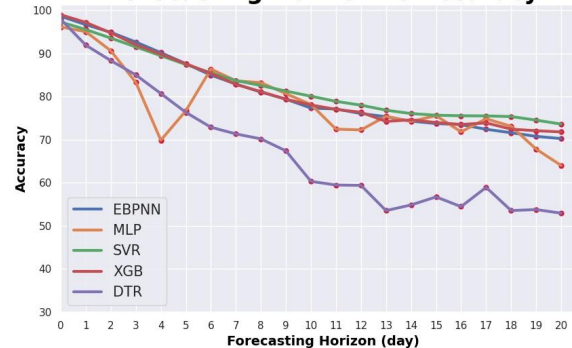


Figure 4: Among 180 independent models, best performed models of each algorithms were run for 0 to 20 Days (which has been termed as Forecasting Horizon). Modeled developed using XGBoost technique performed better than any other models for a 6 day lead time with accuracy greater than 85%. Then model developed using SVR technique performed better than any other models. Up to 3 day lead time, models developed using EBPNN, SVR, and XGBoost performed with more than 90% accuracy.

Forecasting Horizon vs Accuracy



Reference

- Adamowski, J.; Fung Chan, H.; Prasher, S.O.; Ozga-Zielinski, B.; Sliusarieva, A. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resour. Res.* 2012, 48
- Biswas R.K and Jayawardena A.W. (2014), Water Level Prediction By Artificial Neural Network In A Flashy Transboundary River Of Bangladesh, *Global NEST Journal*.
- Chowdhary, A. and Shrivastava, R.K. (2010). "River discharge prediction using artificial neural network", *Canadian Journal of Pure and Applied Sciences*, 4(3), 1275-1281.
- Dawson, C., Abrahart, R., Shamseldin, A., Wilby, R. and See, L. (2005). "Flood estimation at ungauged sites using artificial neural networks", *Journal of Hydrology*, 319, 391 – 409.
- Hsu, K., Gupta, H.V. and Sorooshian, S. (1995). "Artificial neural network modeling of the rainfall runoff process", *Water Resources Research*, 31, 2517-2530.