EGU25-1959
EGU General Assembly 2025

A Comparative Analysis of Data-Driven Machine Learning Models for Rainfall Forecasting in Bangladesh

Mir Mahmid Sarker[1], Arish Morshed Zobeyer[2], Tasnuva Rouf[3], and S M Mahbubur Rahman[4]
[1]Bangladesh University of Engineering and Technology (BUET) , Dhaka, Bangladesh (1916007@wre.buet.ac.bd)
[2]Institute of Water Modeling (IWM), Dhaka, Bangladesh (arishzobeyer@gmail.com)
[3]National Oceanic and Atmospheric Administration (NOAA), Maryland, United States (tasnuva.rouf@noaa.gov)
[4]Institute of Water Modelling (IWM), Dhaka, Bangladesh (smr@iwmbd.org)

Accurate rainfall forecasting is crucial for effective urban planning and disaster management in Dhaka, the capital of Bangladesh, a city highly vulnerable to urban flooding and extreme weather events. Traditional forecasting methods often struggle to capture the region's complex rainfall patterns, resulting in inaccurate rainfall forecasts. This study evaluates the performance of two traditional machine learning algorithms, Random Forest Regression and Multi-layer Perceptron (MLP), alongside one deep learning algorithm, the Long Short-Term Memory (LSTM) network. These models are trained and tested to forecast rainfall over 1 to 5-day lead times, emphasizing their ability to handle temporal dependencies in time series data. Atmospheric and hydrologic variables, including temperature, surface pressure, evaporation, solar surface radiation, total column rainwater, large-scale precipitation, and total cloud cover, from the ECMWF (European Centre for Medium-Range Weather Forecasts) Reanalysis v5 (ERA5) dataset, were used as model inputs. Model forecasts were validated against ERA5 rainfall data and compared with the forecasts from the Global Forecast System (GFS) model. Results indicate that the Random Forest model outperforms all others, achieving an RMSE of 6.11 mm and Pearson's correlation coefficient (R) of 0.74 for a 1-day lead time. The LSTM model achieved an RMSE of 7.46 mm, while the MLP performed less effectively than both RF and LSTM, with an RMSE of 7.61 mm. In comparison, the GFS forecasts displayed an RMSE of 9.16 mm. The RF model outperformed the other models at all lead times; however, its accuracy decreased as the lead time increased. This study highlights the potential of machine learning to improve short to medium range rainfall forecasts, contributing to timely decision-making for urban resilience and resource management.