



# Rainfall Prediction in Bangladesh

Mohaiminul Islam (1905018)

Tanveer Rahman (1905025)



# Problem Definition

- We have various weather related information collected from 35 stations of our country. The information ranges from 1948-2022.
- From this dataset, we need to use various machine learning algorithms to predict the rainfall and analyze the predictions.
- Comparing to other countries, less work is done in our country because of the unavailability of relevant datasets.
- Few previous works :
  - [Data-driven multi-step prediction and analysis of monthly rainfall using explainable deep learning](#)
  - [Precipitation Forecasting in Northern Bangladesh Using a Hybrid Machine Learning Model](#)
  - [Standardization Of Rainfall Prediction In Bangladesh Using Machine Learning Approach](#)



# Dataset Preparation

- [65 years weather dataset](#) from Kaggle was taken as the basis. This dataset ranges from 1948 to 2013.
- The recent data from 2014 to 2022 was taken from [Bangladesh Agriculture Research Council](#).
- Combining this two we get a dataset of shape (24868, 10) .
- The 10 features are:
  - Station, Year, Month, Max Temp, Min Temp,
  - Rainfall(mm), Humidity (%), Wind Speed (m/s), Cloud Coverage (Octs)
  - Sunshine (hrs)
- The Station column contains the data of 35 stations in total.



## Proposed Solution

We plan to develop a hybrid deep learning model that combines **LSTM** with an **attention** mechanism, seamlessly integrated with **XGBoost** for enhanced performance. Where :

- LSTM will capture temporal dependencies.
- Attention mechanism will allow to improve its ability to handle long-range dependencies.
- The XGBoost layer will further ensure accurate and robust predictions.



# Performance metrics

- Mean Squared Error (MSE)

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

- **R-squared Test:** Proportion of variance in the dependent variable explained by independent variable. Higher R-square means better fit.
- **Mean Absolute Error (MAE):** The average of the absolute differences between predicted and actual values