Analyzing the temporal rainfall variability over hundred Indian cities through deep learning approach

Jagabandhu Panda¹, Sanjeev Singh¹, and Asmita Mukherjee¹
¹Department of Earth and Atmospheric Sciences, National Institute of Technology Rourkela, Odisha 769008, India
E-Mail: jagabandhu@gmail.com; pandaj@nitrkl.ac.in

ABSTRACT

Applications of rainfall analysis and forecasting, range from disaster management to agriculture, making it an essential component of modern-day research. With the increasing impacts of climate change, it is anticipated that in the near future, frequent and extreme rainfall events would trigger severe floods, landslides, etc. Therefore, it is extremely important to make a precise prediction so that the intensity of the impacts on life and property could be reduced. The dynamic nature of the Indian monsoon makes it volatile and difficult to predict. Initially, rainfall forecasting began with numerical weather prediction models, but in recent times, with the advancement of AI/ML applications in weather and climate science, it has become reasonably popular for such studies. Thus, the present work focuses on the use of climatological rainfall data sets for the analysis and prediction of monthly, seasonal, and annual rainfall patterns across India, by considering city-specific information. Deep learning (DL) approaches like Long Short Term Memory (LSTM), Bidirectional LSTM (BiLSTM), Gated Recurrent Unit (GRU), and Convolution 1D LSTM (Conv1DLSTM) are considered for long-term rainfall prediction over hundred selected smart cities of India based on their location. Performance indicators like root means square error (RMSE) are computed to test the model training accuracy. The initial results from the comparison of the considered DL models indicated that for univariate forecasting of accumulated monthly rainfall, BiLSTM performed better while for bivariate forecasting, GRU performed better than the others. Prior to the forecasting using DL models, city-based trend analysis of rainfall is performed using the Modified Mann-Kendall test. The current study would demonstrate the results obtained from the univariate and bivariate analysis and forecasting till 2031 by considering a city-based approach.

Keywords: Rainfall; ML; DL; LSTM; BiLSTM; GRU; Conv1DLSTM

Changing Dynamics of Arid Region and Impact on Weather and Climate over Indian Subcontinent
A study on rainfall variability over Indian cities through deep learning models: Climatological Analysis and Future Prediction

Presented

By

Prof. Jagabandhu Panda

#Department of Earth & Atmospheric Sciences
National Institute of Technology Rourkela, Odisha – 769008
Email: jagabandhu@gmail.com; pandaj@nitrkl.ac.in
Phase-I

45 selected Smart cities

Source:
Smart Cities Mission, Ministry of Housing and Urban Affairs

Data:
IMD gridded datasets used
Methodology

Raw Datasets of Daily rainfall, Daily Mean, Minimum and Maximum Temperature from CPC, IMD

Preprocessing and splitting data for training and testing

Deep Learning Models Adopted

LSTM  BiLSTM  GRU

Train the model using training dataset with batch size 64 and 128 for 500 epochs

Test the model using testing dataset and plot the predicted values along with true values.

Determine RMSE and MAE values

Compare RMSE and MAE values

Least RMSE and MAE determine the best model and configuration for prediction
REGION WISE COMPARISON (UNIVARIATE)

Region wise RMSE comparison (batch64)

Region wise RMSE comparison (batch128)

REGION WISE COMPARISON (MULTIVARIATE)

Region wise MAE comparison

Region wise RMSE comparison
CITY WISE COMPARISON (MULTIVARIATE)
GRU MAE comparison

GRU RMSE comparison

Error
Comparing all the results, it can be said that for univariate forecasting BiLSTM model has performed better than others on an average for all cities considered. However, GRU has been able to predict better for places of large range of rainfall variation. Hence, a combination of BiLSTM and GRU can be tried.

- The RMSE values range from 1.476 (Davangere) to 5.34 (Shillong).
- The MAE values range from 0.8232 (Jaipur) to 3.72 (Shillong).

For multivariate forecasting, LSTM model has performed better than others on an average for all cities considered.

- Here, RMSE values ranged from 1.376 (Tumukara) to 4.951 (Shillong). And MAE values ranged from 0.821 (Udaipur) to 3.057 (Shillong).
Study Area

Phase-II of the Study

Total Indian Cities/Towns Considered: 100 (as per the GOI’s smart city mission)
Data and Methodology

Data Acquisition

Data Preprocessing

Train Test Split

Selection of Parameters

Model Training

Calculate RMSE and MAE for all models

Compare RMSE of all Models

Least RMSE value determine the model for future prediction

Non-parametric test (Seasonal Mannkendall test)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Spatial resolution</th>
<th>Temporal Resolution</th>
<th>Unit</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Precipitation</td>
<td>$0.5 \times 0.5$ degree</td>
<td>Monthly</td>
<td>Monthly Accumulated in mm</td>
<td>CRU (<a href="https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.07/cruts.2304141047.v4.07/pre/">https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.07/cruts.2304141047.v4.07/pre/</a>)</td>
</tr>
<tr>
<td>Temperature</td>
<td>$0.5 \times 0.5$ degree</td>
<td>Monthly</td>
<td>Monthly Mean in degree Celsius</td>
<td>CRU (<a href="https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.07/cruts.2304141047.v4.07/tmp/">https://crudata.uea.ac.uk/cru/data/hrg/cru_ts_4.07/cruts.2304141047.v4.07/tmp/</a>)</td>
</tr>
</tbody>
</table>

- **Data considered for the duration**: 1901-2021
- **Non parametric MK test is adopted for analyzing rainfall trends**
- **DL Models used in Phase-II**: LSTM, BiLSTM, GRU, Conv1dLSTM
Segregation based on intensity of rainfall

RMSE for high intensity rainfall cities 10 cities

RMSE for very high intensity rainfall cities 5 cities

RMSE for low intensity rainfall cities 46 cities

RMSE for moderate intensity rainfall cities 39 cities

LSTM_RMSE, GRU_RMSE, BiLSTM_RMSE, Conv_RMSE
Overall comparison in the univariate approach for intensity-based segregation
Forecasts

a. Low intensity
b. Moderate intensity
c. High intensity
d. Very intensity
Segregation based on homogeneous rainfall regions

RMSE for Central_NE cities 17 cities

RMSE for North West cities 19 cities
Overall comparison in the univariate approach for segregation based on homogeneous rainfall regions
Forecasts

a. Central North East
b. North West
c. North East
d. West Central
Forecasts

- **e) Peninsular Region**: Graph showing the 10-year prediction of rainfall for the Peninsular Region, with actuals and forecasted data plotted over time from 1900 to 2020.

- **f) Hilly Region**: Graph showing the 10-year prediction of rainfall for the Hilly Region, with actuals and forecasted data plotted over time from 1900 to 2020.
Multivariate (Bivariate) Rainfall Analysis and Forecasting

✓ Segregation based-on intensity

✓ Segregation based-on rainfall homogeneous regions
✓ Analysis was done similar to that of the univariate approach

✓ Forecasting was done till 2031 over all cities separately

✓ Trend analysis performed over all individual cities in both univariate and multi-variate approach by adopting MK trend test (a modified version considered)

✓ The trend test was done for both analysis (1901-2021) and forecasting (till 2031) too
For univariate analysis, when the cities were segregated on the basis of intensity of rainfall, Conv1DLSTM performed better except for the case of the cities with high intensity rainfall where BiLSTM was performing well.

Decreasing trend was found in cities with the moderate, high and very high intensity rainfall which is expected to continue and remain decreasing.

For most of the cities showing either increasing or no trend, they may change to either no trend or increasing trend with most showing a tendency to change to increasing trend in future.

For univariate analysis, when the cities were segregated on the basis of homogenous rainfall regions, Conv1DLSTM performed better, except for the cities located in CNE where BiLSTM performed well. Decreasing trend can be seen in the NEI cities which will remain same in future. And most of the cities showed either increasing trend in future.
• For multivariate analysis, cities segregated on the basis of rainfall intensity, GRU performed well in all the cases.
• Decreasing trend was found in the cities having moderate, high and very high intensity rainfall which will continue to remain same.

• For the multivariate analysis, when the cities were segregated on the basis of homogenous rainfall regions GRU performed better except for cities in peninsular regions where BiLSTM performed well.
• Decreasing trend can be seen in NEI cities which will remain same in future.
• The errors in the case of multivariate models is very less compared to univariate models except for few scenarios.
For rainfall analysis and forecasting, one can safely adopt Conv1DLSTM in the univariate approach for most of the Indian cities, while GRU can be used in the multi-variate approach.

Deep learning models have nothing to do with actual dynamics of the rainfall. But it depends on the correlation of the input parameters. In fact, its prediction is based on the sequence of the input parameters. The model learns the variation of the rainfall with respect to variation of the input parameters from the training set and it predicts the rainfall on the basis of training. So, in this work, these techniques are explored.

AI/ML framework may provide some additional understanding to what we can get with the usual way of analysing the available data sets. For instance, such a framework can always be helpful in the data based future prediction.....

In order to identify the primary issues encountered and to develop an effective ML model, researchers need to have a decent understanding of the primary objective concerning climate studies besides the relevant AI/ML framework.

AI/ML may be used within many components of NWP workflow which would make the complicated flow easier.
Thank You.

Email for queries and correspondence: jagabandhu@gmail.com;
pandaj@nitrkl.ac.in